Learning the structure of correlated synaptic subgroups using stable and competitive spike-timing-dependent plasticity

H. Meffin,[1,*] J. Besson,[1,2] A. N. Burkitt,[1,3] and D. B. Grayden[1,3]
[1]*The Bionic Ear Institute, 384–388 Albert Street, East Melbourne, Vic 3002, Australia*
[2]*School of Computer and Communication Sciences and Brain-Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
[3]*Department of Electrical & Electronic Engineering, The University of Melbourne, Vic 3010, Australia*
(Received 3 August 2005; revised manuscript received 12 January 2006; published 10 April 2006)

Synaptic plasticity must be both competitive and stable if ongoing learning of the structure of neural inputs is to occur. In this paper, a wide class of spike-timing-dependent plasticity (STDP) models is identified that have both of these desirable properties in the case in which the input consists of subgroups of synapses that are correlated within the subgroup through the occurrence of simultaneous input spikes. The process of synaptic structure formation is studied, illustrating one particular class of these models. When the learning rate is small, multiple alternative synaptic structures are possible given the same inputs, with the outcome depending on the initial weight configuration. For large learning rates, the synaptic structure does not stabilize, resulting in neurons without consistent response properties. For learning rates in between, a unique and stable synaptic structure typically forms. When this synaptic structure exhibits a bimodal distribution, the neuron will respond selectively to one or more of the subgroups. The robustness with which this selectivity develops during learning is largely determined by the ratio of the subgroup correlation strength to the number of subgroups. The fraction of potentiated subgroups is primarily determined by the balance between potentiation and depression.

DOI: 10.1103/PhysRevE.73.041911 PACS number(s): 87.19.La, 87.18.Sn, 87.10.+e

I. INTRODUCTION

Learning in the brain is believed to involve the modification of the synaptic connections between neurons so as to either increase or decrease their strength in response to stimuli. Two crucial aspects of this plasticity are competition and stability. Competition occurs when the presence of a stimulus results in some synapses being strengthened and others weakened. It is important because it allows a neuron to learn to respond selectively to only some stimuli by promoting the formation of a stimulus-related synaptic structure from an initially random configuration. However, such competition requires an instability in the dynamics of the synaptic weights that must be reconciled with the need for those weights to remain bounded and ultimately reach a stable configuration.

One form of synaptic plasticity, called spike-timing-dependent plasticity (STDP), depends on the relative timing of the pre- and post-synaptic action potentials (spikes) [1–5]. A causal pre-before-post ordering results in synaptic strengthening (potentiation), whereas a post-before-pre ordering results in synaptic weakening (depression). A number of theoretical studies have investigated the issues of competition and/or stability in the context of STDP [6–10]. In the most common model, potentiation and depression are independent of synaptic strength [1,8,11–15]. This typically results in strong competition and globally unstable dynamics so that hard upper and lower limits must be enforced to contain the weights. While it is clear that natural limits exist in the form of limited resources for the upper bound and zero synaptic current for the lower bound, it is also possible that the weights have a "soft" bound that arises from the synaptic dynamics. Competition is so prevalent in the model with weight-independent potentiation and depression that synaptic structure will emerge even when there is no structure inherent in the inputs [12]. On the other hand, models in which the dependence of depression and/or potentiation is linear in the weight typically result in distributions that are stable but also unimodal due to the lack of competition [16,17]. Recently, Gütig *et al.* [18] introduced a model with soft limits that continuously interpolates between the additive and multiplicative models and permits both competition and stability, which typically results in a bimodal distribution of weights. This shows that by introducing weight dependence into the learning dynamics, it is possible to bound the weights in a graded fashion while retaining competition. This mechanism for achieving competition and stability is plausible since STDP has been observed experimentally to have a dependence upon the weight [3]. Present experimental data are too sparse and noisy to determine the analytical form of weight dependence for the depression. Thus, as a guide to future experiments, it seems useful to determine the essential features of weight dependence in STDP that allow for both competition and stability in synaptic dynamics.

In this paper, we introduce general criteria for the weight dependence of potentiation and depression that are necessary for a STDP model to exhibit both competition and stability. This is done for the case in which the input consists of subgroups of synapses that are correlated within the subgroup through the occurrence of simultaneous input spikes. Under these circumstances, symmetry breaking may occur, in which case all the synapses of one (or more) subgroup(s) become

*Electronic address: meffin@zi.biologie.uni–muenchen.de, hmeffin@yahoo.com

potentiated and the synapses of the remaining subgroups become depressed. This simple synaptic structure results in the neuron responding selectively with an elevated firing rate to those subgroups with potentiated inputs. We describe how the number of potentiated subgroups depends on the parameters describing both the inputs and the plasticity. We also describe when and how robustly such synaptic structure forms. All equations are given in terms of a general weight dependence and illustrated with a case in which potentiation is constant and depression has a cubic dependence on the weights. This particular case typifies synaptic learning rules that are both stable and competitive.

II. MODEL

A. Plasticity

We consider a single neuron receiving input on multiple synapses. The strength of a synapse, i, is characterized by its weight, w_i (see the following subsection for its precise role). The spike-timing-dependent plasticity (STDP) is implemented by making the following changes to the synaptic weight, w_i, when a synaptic input on the ith fiber arrives at time t_{in} and an output spike is generated at time t_{out} (giving a difference of $\delta t = t_{\text{out}} - t_{\text{in}}$):

$$\delta w = \begin{cases} \eta f_-(w_i)\mathcal{T}(\delta t) & \text{if } \delta t < 0, \\ \eta f_+(w_i)\mathcal{T}(\delta t) & \text{if } \delta t \geqslant 0, \end{cases} \tag{1}$$

where η is the learning rate, f_+ and f_- are functions describing the weight dependence of the plasticity, and the time window function $\mathcal{T}(\delta t)$ is defined as

$$\mathcal{T}(\delta t) = \begin{cases} -e^{\delta t/\tau_-} & \text{if } \delta t < 0, \\ e^{-\delta t/\tau_+} & \text{if } \delta t \geqslant 0, \end{cases} \tag{2}$$

and τ_+ and τ_- are, respectively, the potentiation and depression time constants. Bi and Poo [3] measured values of $\tau_+ = 17 \pm 9$ ms and $\tau_- = 34 \pm 13$ ms in cultured hippocampal neurons. We further considered the case in which the time extent of the input/output interactions is restricted so that each output interacts only with its nearest synaptic inputs (in time). Consequently, each output spike in this model contributes a potentiation component that arises by interaction with the most recent input spike and a depression component by interaction with the first subsequent input spike (i.e., if additional input spikes fall within the STDP time window of an output spike, these interactions are neglected) [16,17]. The input-output interactions included in this model are illustrated in Fig. 1,which shows that output spikes interact only with the two input spikes that form the temporal limits of the input interspike interval (ISI) in which they are generated. The reasons for choosing this type of input-output restriction are that it may be more biologically realistic [16,17,19] and that it allows the calculation of the diffusion function $B(w, \bar{w})$ as explained in Appendix A.

B. Inputs

We consider a neuron with N excitatory synaptic inputs. The firing times of any given input are described by a homogeneous Poisson process of rate λ_{in}, the same for all inputs. Inputs are partitioned into M subgroups, such that the spike times of inputs within a subgroup are correlated, but the spike times of inputs from different subgroups are independent. The within-group correlations are introduced so that, for any input, a given portion of its spikes occurs at the same time as some other spikes within its subgroup, while the remainder occur at independent times. This is achieved by selecting spikes for each input from two reference Poisson spike trains each of rate λ_{in} [18]. The first train is for the correlated events and is the same for all inputs in the subgroup, but with independent trains for each subgroup. The spikes are selected for the input from this train with probability $\sqrt{c}$, independent of the other neurons in its subgroup. Thus only a portion of all the neurons in a subgroup participate in each correlated event. The second is the input's own independent train. The spikes are selected from this train with probability $1-\sqrt{c}$. The correlation coefficient is then given by [18]

$$c = \frac{\text{cov}(X_i(t), X_j(t))}{\sqrt{\text{var}(X_i(t))\text{var}(X_j(t))}}, \tag{3}$$

where the random variable $X_i(t)$ is 1 if there is an input spike at synapse i at time t and zero otherwise, and i and j are synapses from the same subgroup.

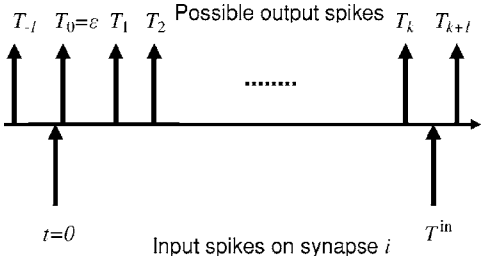

FIG. 1. Schematic representing a typical interspike interval for a synapse i. The input spikes define the interval's beginning, at $t=0$, and end, at $t=T^{\text{in}}$, as shown by the two vertical arrows below the time line. Output spikes, as shown by vertical arrows above the time line, may occur at times $T_0=\epsilon$, due the input(s) at time $t=0$, or at other times T_n, $n \in \mathbb{Z}\ n \neq 0$, depending on the arrival of inputs on other synapses (Sec. II B) and the probability of a consequent output spike (Sec. II C). Contributions to plasticity changes via Eq. (1) occur only for interactions involving each output spike and the two input spikes that define the interspike interval to which that output belongs. Thus the output spikes at $t=T_0, \ldots, T_k$ interact only with the input spikes at $t=0$ and T^{in}, while the output spikes at T_{-1} and T_{k+1} interact with the previous and subsequent input intervals, respectively.

C. Output

The spiking activity of the output neuron model is an inhomogeneous Poisson process with the instantaneous rate function [18]

$$\lambda_{\text{out}}(\{w_j\},t) = \frac{1}{N}\sum_{j=1}^{N} w_j(t) S_j^{\text{in}}(t-\epsilon), \tag{4}$$

where the sum is over all synapses j with input spike trains $S_j^{\text{in}}(t)=\Sigma_k \delta(t-t_j^k)$ for spike times t_j^k and the parameter ϵ denotes a small constant delay in the output. This delay is much smaller than the STDP time constant τ_+ (or τ_-) and so we assume that $e^{\epsilon/\tau_{+,-}} \approx 1$ throughout this work. The effect of this prescription is that whenever a spike event arrives on a subset $\mathcal{I}$ of synapses, an output spike occurs with probability $\frac{1}{N}\Sigma_{j\in\mathcal{I}} w_j$ at a time ϵ later.

In this model, the post-synaptic excitation of the neuron is mediated by a δ function [see Eq. (4)]. A related assumption is that correlations in the inputs occur through the simultaneous arrival of spikes at different synapses. These two assumptions mean that correlations are always "instantaneous" in the model and have no finite duration. In real neurons this is never the case, so the model used here should be considered an approximation to the case in which the time constants governing synaptic correlations and post-synaptic potentials are short compared to the plasticity time constants $\tau_\pm$ and the neuron is acting primarily as a temporal coincidence detector. This may apply in some cortical areas, especially given the much smaller membrane time constant reported *in vivo* compared to the passive value obtained from *in vitro* studies [20,21].

III. RESULTS

The evolution of the synaptic weights w_i can be described by a Langevin equation [22,23]

$$\frac{dw_i(t)}{dt} = A(w_i,\{w_j\}) + C(w_i,\{w_j\})\xi(t), \tag{5}$$

where $A(w_i,\{w_j\})$ is the mean drift and $C(w_i,\{w_j\})$ describes the magnitude of the stochastic fluctuations about this mean as characterized by Gaussian white noise $\xi(t)$ with zero mean and δ function autocorrelation, $\langle \xi(t)\xi(t')\rangle = \delta(t-t')$. The drift function may be calculated according to [22,23]

$$A(w_i,\{w_j\}) = \int (d\Delta w_i)\Delta w_i Q(\Delta w_i | w_i,\{w_j\}), \tag{6}$$

where $Q(\Delta w_i | w_i,\{w_j\})$ is the conditional probability of weight i changing by an amount Δw_i as the result of STDP, independent of STDP weight changes occurring at other times and given the current weights w_i and $\{w_j\}$. As shown in Appendix A, this leads to

$$A(w_i,\{w_j\}) = \frac{\eta\lambda_{\text{in}}}{N}\Bigg\{[\Gamma_+ f_+(w_i) - \Gamma_- f_-(w_i)]\sum_{j=1}^{N} w_j + f_+(w_i) \times (1-\Gamma_+)c\sum_{j\in\mathcal{G}_i} w_j + (1-\Gamma_+)(1-c)f_+(w_i)w_i\Bigg\}, \tag{7}$$

where $\Gamma_\pm = \lambda_{\text{in}}/(\lambda_{\text{in}} + 1/\tau_\pm)$ and $\mathcal{G}_i$ is the subgroup to which synapse i belongs. The first term in Eq. (7) is the contribution from output spikes that are independent of synapse i's input spikes. It will typically be negative in the neighborhood of a fixed point of the dynamics and thus represents the competition between synapses. The second and third terms are the contributions from output spikes that are causally related to synapse i's input spikes and can be thought of as a "spike-triggering effect." They provide positive feedback to the synapse weight and thus promote symmetry breaking among the population of synapses. The first of these is the contribution from correlations within the synapses' subgroup, while the second is the contribution from all the input spikes of synapse i which are independent of the others in its subgroup. Since N is typically very large in the cortex ($N\sim 10^4$ synapses per neuron) [24], this latter term is very small and can be neglected provided $cN \gg (1-c)M$. (In the case of additive potentiation and depression, this term can lead to symmetry breaking even though it is very small if there is no structure inherent in the input; this is not a problem for the choice of f_+ and f_- used in this work.) The first of these terms represents symmetry breaking between subgroups and is typically the dominant symmetry breaking term.

A. Small learning rates

In the limit of a very small learning rate, the fluctuations in Eq. (5) can be ignored and the equilibrium weights, w_i^*, are well approximated by the zeros of the drift functions,

$$A(w_i^*,\bar{w}) = \eta\lambda_{\text{in}}\Bigg\{[\Gamma_+ f_+(w_i^*) - \Gamma_- f_-(w_i^*)]\bar{w} + \frac{(1-\Gamma_+)c}{M} f_+(w_i^*)\hat{w}_i\Bigg\} = 0, \tag{8}$$

where the mean weight over a subgroup is $\hat{w}_i = M/N\Sigma_{j\in\mathcal{G}_i} w_j$ and the global mean weight is $\bar{w} = 1/N\Sigma_{j=1}^{N} w_j$. We will require that any symmetry breaking occur between subgroups, not within them, a situation that leads to the partitioning of subgroups into those that are potentiated and those that are depressed. In this case, the subgroup weight distribution is unimodal and the subgroup mean is well approximated by its modal value: $\hat{w}_i \approx w_i^*$. Then for every synapse, i, the global mean weight $\bar{w}$, which is independent of i, can be expressed as a function of the equilibrium weight

$$\bar{w} = \frac{cw_i^*}{MF(w_i^*)} \quad \text{for } i = 1,\ldots,N, \tag{9}$$

where $F(w) = [\Gamma_- f_-(w) - \Gamma_+ f_+(w)]/[(1-\Gamma_+)f_+(w)]$. For a symmetry breaking solution, we require that Eq. (9) has multiple (nonidentical) stable solutions for some value of $\bar{w}$ so that $w_i^* \neq w_j^*$ for some choices of $i \neq j$. We will also require that solutions are positive and that the homogeneous solution, $w_i^* \cong w_0 \approx \bar{w}\ \forall i$, exists and is unique. Uniqueness is enough to guarantee that symmetry breaking only occurs between subgroups (see Appendix A). (Conversely, if the homogeneous solution is not unique, examples can be found in which symmetry breaking occurs within a subgroup.) Given the preceding requirements, $F(w)$ must have the following

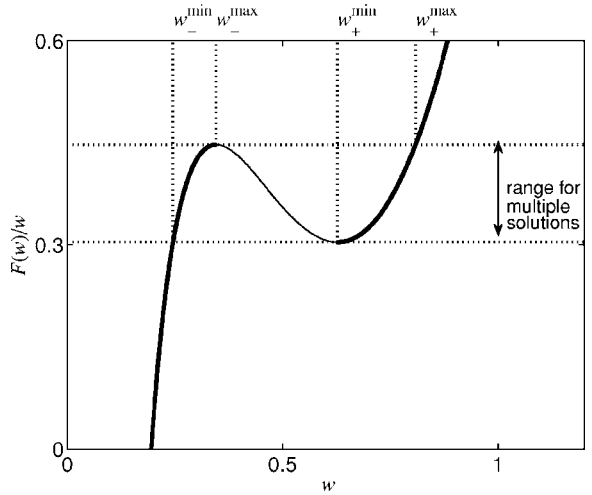

FIG. 2. The function $F(w)/w$ appearing in Eq. (9) is plotted to show the range over which there are multiple stable solutions to this equation as required for symmetry breaking. Stable portions of the curve have negative slope (bold line) while unstable portions have positive slope (faint line). Thus stable bimodal solutions have their depressed synapses between $w_-^{\min}$ and $w_-^{\max}$, while their potentiated synapses are between $w_+^{\min}$ and $w_+^{\max}$. The choice of $f_+(w)$ and $f_-(w)$ for $F(w)$ is the cubic set given in Eq. (10) with parameters $a=1.0$, $b=0.92$, $m=3.0$, $q=0.5$, $\tau_+=17$ ms, $\tau_-=34$ ms, and $\lambda_{\text{in}}=100$ Hz.

properties for smooth functions $f_+(w)$ and $f_-(w)$ on the domain, $\mathcal{D}$, in which w is defined (see Appendix B).

The conditions are as follows:

(i) $F(w)$ always has a non-negative slope, i.e., $F'(w)\geqslant 0\ \forall w\in\mathcal{D}$.

(ii) $F(0)$ is negative, i.e., $F(0)<0$.

(iii) $F(w)$ has an inflection point with positive third derivative, i.e., $\exists w_f\in\mathcal{D}: F''(w_f)=0$ and $F'''(w_f)>0$.

A simple example of a pair of functions (f_+,f_-) satisfying these criteria is

$$f_+(w)=a,$$

$$f_-(w)=b+m(w-q)^3 \tag{10}$$

for m, $q>0$ and $\Gamma_+a-\Gamma_-(b-mq^3)<0$. A plot of the function $F(w)/w$ in Fig. 2 illustrates the range of values of $c/(M\bar{w})$ for which there are multiple solutions to Eq. (9), for a particular choice of the parameters a, b, m, q, and λ_{in}. We shall typically use functions of the sort described in Eq. (10) to give numerical examples throughout, but all equations and arguments will be given in terms of the general weight dependence. The parametrization in Eq. (10) is by no means unique, it simply represents one realization that encapsulates conditions (i)–(iii) above. Another pair of functions that satisfies these conditions, but applies to weights that are explicitly restricted to lie in the interval $w\in[0,1]$, is $f_+(w)=(1-w)^\mu$ and $f_-(w)=\alpha w^\mu$ for $\mu\in[0,1]$, as proposed by Gütig *et al.* [18]. All learning rules satisfying the criteria (i)–(iii) behave in a way that is qualitatively similar to those with the explicit cubic dependence for depression given in Eq. (10). The weight dependence given in Eq. (10) is consistent with experimental data which, while noisy, show a monotone increasing dependence on weight for depression and weak or inverse dependence on weight for potentiation [3]. (Note that this description of the absolute weight changes is equivalent to the data given in Bi and Poo [3], which showed that the relative amount of depression is roughly independent of the weight and the relative amount of potentiation is roughly inversely related to the weight. See also [16].)

We shall assume that for any value of the mean weight $\bar{w}$, there are at most two modal solutions to Eq. (9) denoted by w_+ (for the potentiated weight) and w_- (for the depressed weight) (when the solution is unimodal we shall write $w_+=w_-=w_0$). This is always the case for the choice of $(f_+(w), f_-(w))$ in Eq. (10). This allows us to conceptualize the solution of the M equations in Eq. (9) as a two-dimensional problem. A first equation is obtained by eliminating $\bar{w}$,

$$w_+F(w_-)=w_-F(w_+), \tag{11}$$

which is independent of the input parameters c and M. A second equation is found by making the small learning rate approximation $\bar{w}\approx rw_++(1-r)w_-$, where r is the proportion of potentiated synapses. r is an important measure of the emergent synaptic structure since rM gives the number of subgroups that the neuron responds to with an elevated spike probability. The approximation gives

$$w_-=-\frac{rw_+}{1-r}+\frac{cw_+}{(1-r)MF(w_+)},\quad r=0,\frac{1}{M},\frac{2}{M},\ldots,\frac{M-1}{M}. \tag{12}$$

For each partition, $(r,1-r)$, of the M subgroups into potentiated and depressed synapses, there is a valid pair of Eqs. (11) and (12) which must be simultaneously solved to find the solution(s). Consequently, there may be more than one stable solution whenever symmetry breaking occurs as specified in conditions (i)–(iii) for f_+ and f_-.

This multiplicity of solutions is illustrated in Fig. 3 for the case of $M=3$ subgroups and a correlation of $c=0.5$ and for the same choice of parameters as used in Fig. 2. It shows the solutions to the simultaneous pairs of Eqs. (11) and (12) as points of intersection in the (w_+,w_-) plane. There are two curves satisfying Eq. (11). The first is the diagonal line corresponding to homogeneous, unimodal solutions and the second is the diagonally symmetric ovoid shape corresponding to symmetry breaking, bimodal solutions. This situation is typical of functions f_+ and f_- that permit symmetry breaking as specified in conditions (i)–(iii). Also seen are two curves corresponding to the solutions of Eq. (12) for $r=1/3$ and $2/3$ (online in blue and green, respectively). Locally stable solutions correspond to points of intersection on the bold portions of the curves, while unstable solutions correspond to the points of intersection on the faint portions. In this example, there are two distinct stable solutions: one bimodal solution with $r=1/3$ and $(w_+,w_-)\approx(0.74,0.23)$, and a second, unimodal solution with $(w_+,w_-)\approx(0.30,0.30)$. In general, it is enough to restrict attention to the part of the graph on or below the diagonal since a solution for a particular

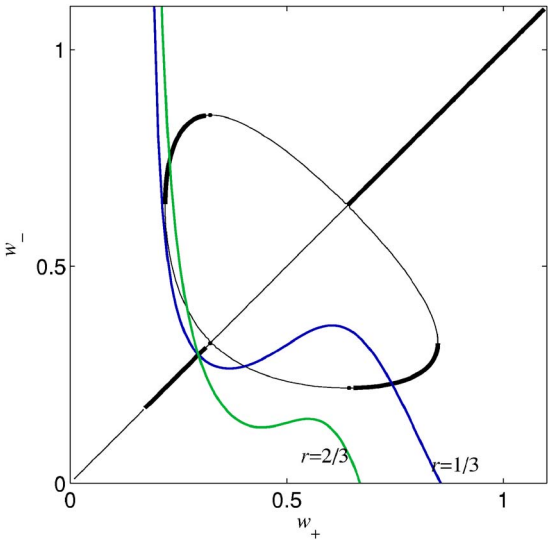

FIG. 3. (Color online) The solutions to the pairs of Eqs. (11) and (12) as points of intersection of curves in the (w_+,w_-) plane for $M=3$ subgroups and a correlation $c=0.5$ (other parameters as in Fig. 2). The diagonal line and the ovoid curve correspond to solutions to Eq. (11), while the two other curves correspond to the solutions to Eq. (12) for $r=1/3$ (blue) and $r=2/3$ (green) as marked. Portions of the curves corresponding to stable solutions for Eq. (11) are shown in bold.

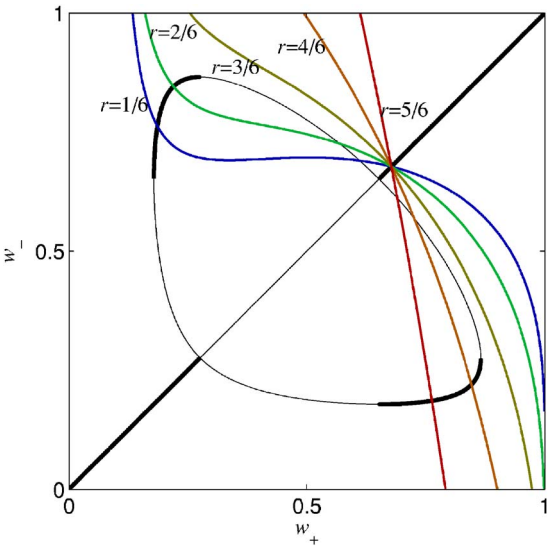

FIG. 4. (Color online) The solutions to the pairs of Eqs. (11) and (12) similar to Fig. 3 but using the plasticity rule of Gütig *et al.* [18]. Note that multiple stable solutions for different values of r also occur using this rule. Parameters are $M=6$, $c=1$, $\alpha=0.875$, $\mu=0.03$, and $\lambda_{\text{in}}=100$ Hz.

value of r in this part will have a symmetric solution with $1-r$ in the upper part, and we are interested in solutions with $w_+ \geqslant w_-$ (by definition). This can be seen here with the bimodal $r=1/3$ and $r=2/3$ solutions. The $r=0$ curve from Eq. (12) is redundant and is omitted from Fig. 3. This is because it always gives just the unimodal solution that can otherwise be found as the common point of intersection of curves with $0<r<1$ and the diagonal line. This is seen in Fig. 3 for the $r=1/3$ and $r=2/3$ curves, and the diagonal, which all intersect at the unimodal solution $(w_+,w_-)\approx(0.30,0.30)$. For unimodal solutions, we will adopt the convention that $r=0$ implies that all synapses are depressed, i.e., $w_i^*<q\ \forall i$, while $r=1$ implies that all synapses are potentiated i.e. $w_i^*>q\ \forall i$. Thus the unimodal solution in this example has a ratio of $r=0$. Multiple stable solutions are commonplace for functions f_+ and f_- satisfying conditions (i)–(iii). Figure 4 provides a second example, but using the plasticity rule of Gütig *et al.* [18].

With different choices of parameters, other solution sets arise. Focusing on the plasticity parameters a, b, m, and q, we note from Eq. (8) that a can be set to 1 by absorbing it as an overall factor into the learning rate η, and that q determines an arbitrary scale for the weights, which we will set to $q=0.5$ throughout. The effect of the two remaining parameters, b and m, on the solution set can be shown in a bifurcation diagram, as illustrated in Fig. 5(a), again for input parameters $M=3$, $c=0.5$, and $\lambda_{\text{in}}=100$ Hz. First, the straight lines forming the upper and lower borders in the figure correspond to the constraints that $F(0)<0$ and $f_-(0)>0$, respectively. Next, the number of solutions for any choice of (b,m) is shown in gray scale, and ranges from one (lightest) to three (darkest). The lines delineating the borders between these regions are marked by color (online) and text indicating the ratio $r=0,1/3,2/3$, or 1 of the solution that changes existence at the border. For example, the blue ("$\frac{1}{3}$") line on the right of the diagram marks the transition from the existence of the $r=1/3$ stable solution, on the left of the line, to its nonexistence, on the right of the line. The transition is accompanied by the appropriate change in gray tone. A second (blue, "$\frac{1}{3}$") $r=1/3$ line can be seen in the left of the diagram, marking the opposite transition (so that the $r=1/3$ stable solution exists between these two lines). The same situation occurs for the $r=2/3$ solution (green, "$\frac{2}{3}$"). The $r=0$ and $r=1$ (unimodal) solutions are slightly different, in that they each have only one line, and the $r=0$ solution is stable to the right of its line (white, "0"), while the $r=1$ solution is stable to left of its line (black, "1"). In between these lines there is no stable unimodal solution—the opposite pattern to the bimodal r values.

Representative plots, similar to Fig. 3, illustrating the solution sets for the various qualitatively distinct regions of Fig. 5(a), are shown in Figs. 5(b)–5(g), following a trajectory of increasing b for fixed $m=3.0$ [see labels (b)–(g) in Fig. 5(a)]. b is a measure of the strength of depressing contributions to the weight relative to potentiation, and this is reflected in the change in the solution sets observed as b increases. For the weakest level of depression, $b=0.83$, only the $r=1$ solution exists, corresponding to all synapses being potentiated [Fig. 5(b)]. When $b=0.875$, the bimodal $r=2/3$ solution is added to the unimodal $r=1$ solution, leading to a bistable system [Fig. 5(c)]. The system becomes tristable with a further increase in b to 0.885, with the $r=1/3$, $2/3$, and 1 solutions all being stable [Fig. 5(d)]. The $r=1$ solution loses stability when $b=0.9$, leaving the two symmetry breaking solutions $r=1/3$ and $2/3$ [Fig. 5(e)]. Another increase in b to 0.92 results in the $r=0$ depressed, unimodal solution gaining stability and disappearance of the $r=2/3$ bimodal

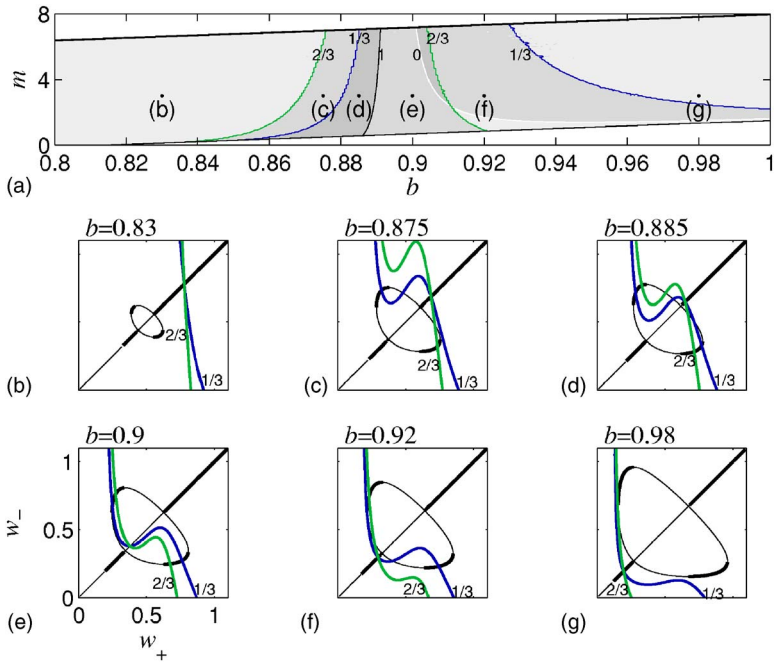

FIG. 5. (Color online) Role of the parameters b and m (that determine the relative strength of synaptic depression compared to potentiation [see Eq. (10)]) as shown by a (b,m) bifurcation diagram, (a), and six explicit solution sets in the (w_+,w_-) plane (as in Fig. 3) for particular choices of b (as marked) and $m=3.0$; (b)–(g) (other parameters as in Fig. 3). In the bifurcation diagram, (a), the number of distinct solutions in a region of (b,m) parameter space is given by the degree of gray shade, with the lightest corresponding to one solution and the darkest corresponding to three solutions. Borders between these regions are delineated by lines that are marked by the value of the ratio $r=0$ (white), 1/3 (blue), 2/3 (green), or 1 (black) of the solution that loses stability or existence at this border (thus the solution is stable and exists on the darker side of the line). The schema for (b)–(g) are the same as for Fig. 3.

distribution [Fig. 5(f)]. Finally by $b=0.98$ the remaining $r=1/3$ bimodal distribution has also disappeared leaving only the $r=0$ solution with all synapses depressed [Fig. 5(g)]. This example shows that the relative strength of depression, b, compared to potentiation is an important parameter in determining the number of subgroups that a neuron learns to respond selectively to, as measured by r. However, since there are often multiple solutions with different values of r, there can be ambiguity about the value of r that will emerge from the learning process for these low rates of learning. These observations apply in general to STDP rules that are both stable and competitive as specified in conditions (i)–(iii) at the beginning of this subsection.

B. Intermediate and large learning rates

The existence of multiple, stable solutions described above holds only for sufficiently small learning rates, η. As η increases, so too do the noisy fluctuations of weights about their modal values. For intermediate values of η, this can lead to trajectories in weight space that stray outside of the basin of attraction of a fixed point and into the basin of an adjacent fixed point, wherein it may be captured. As η increases, this occurs with increasing frequency so that such solutions become essentially unstable on any reasonable time scale over which learning occurs. At high values of η, the fluctuations become so large that transitions back and forth between the fixed points can occur so frequently that no synapse remains permanently potentiated or depressed. These arguments apply in general to functions f_+ and f_- that exhibit multiple stable solutions at low learning rates as is typical whenever they satisfy conditions (i)–(iii) given in Sec. III A.

This effect of learning rate is illustrated in Fig. 6 for the bistable system shown in Fig. 3, with $M=3$ and stable solutions corresponding to ratios of $r=0$ and $r=1/3$. Figure 6(a) shows the results of simulations in which all $N=120$ weights (i.e., 40 in each subgroup) were initially at the $r=0$ fixed point $[(w_+,w_-)\approx(0.30.0.30)]$. The plot shows the probability density, $P(w)$, combined across all 120 synapses of the neuron in gray scale as a function of η between 3×10^{-4} and 3×10^{-2}. For the smallest values of η, the $r=0$ unimodal solution is stable as indicated by the single dark streak

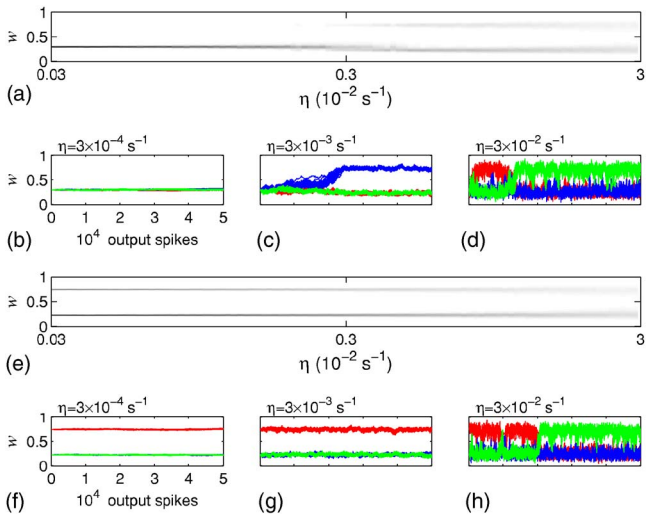

FIG. 6. (Color online) The effect of the learning rate, η, on the stability of the two solutions found to be stable for small η in Fig. 3 (with $M=3$ subgroups). Simulations were run in which all the weights were initially set according to either (i) the unimodal solution [(a)–(d)] or (ii) the bimodal solution [(e)–(g)]. (a) and (e) The probability density, $P(w)$, combined across all synapses of the neuron is given in gray scale as a function of η between 3×10^{-4} and 3×10^{-2} for the unimodal and bimodal initial conditions, respectively. (b)–(d) and (f)–(h) show the time evolution of the weights of the three synaptic subgroups (in different shades/colors) for small, intermediate, and large values of the learning rate ($\eta=3\times10^{-4}$, 3×10^{-3}, or 3×10^{-2} s^{-1}, respectively).

around $w=0.30$ at these small η values (i.e., there are no streaks for any other values of w for small η). This can also be seen in Fig. 6(b), which shows the evolution of all 120 weights when $\eta=3\times10^{-4}$ as a function of time, with weights belonging to different subgroups shown in different shades (colors, online). It is evident that all weights begin and remain around the $r=0$ solution of $w_0\approx0.30$. However, for intermediate and large values of η, this solution becomes unstable as seen by gradual fading out of the initial $w\approx0.30$ dark streak in Fig. 6(a) and the simultaneous emergence of the two streaks at $w\approx0.74$ and 0.23 corresponding to the $r=1/3$ bimodal solution. This can also be seen in Fig. 6(c), which is similar to Fig. 6(b) but with $\eta=3.0\times10^{-3}$, and shows that although all the weights began at $w=0.30$, the weights of one subgroup rapidly became potentiated while the others were slightly depressed as expected for the $r=1/3$ solution. For this intermediate value of η, once a particular subgroup becomes potentiated and the others are depressed, they remain so. This is not the case for large values of η. The fluctuations become so large that subgroups can often change whether they are potentiated or depressed, as seen in Fig. 6(d), which shows a transposition in which two subgroups simultaneously swap their potentiated and depressed status. In contrast, if the weights are initiated to conform with the $r=1/3$ solution, they always remain faithful to this solution as seen in Figs. 6(e)–6(h) [although subgroups can switch back and forth between modes in Fig. 6(h)].

It is useful to have analytical techniques for investigating the formation of synaptic structure when η is not small, since numerical simulations are computationally intensive and (possibly prohibitively) time consuming. An analysis is possible in the large η limit using a one-dimensional Fokker-Planck equation. It requires that r may be treated as a continuous parameter, which is a reasonable approximation in the large η limit in which weights constantly make transitions back and forth between modes. The approximation is best when the correlations in weight changes between subgroups are weak, which is generally the case but improves as M becomes larger. As usual, we assume that the subgroup mean can be approximated by the weight of any one of its members. The Fokker-Planck equation is given by [22,23]

$$\frac{\partial P(w,\bar{w})}{\partial t}=-\frac{\partial}{\partial w}[A(w,\bar{w})P(w,\bar{w})]+\frac{1}{2}\frac{\partial^2}{\partial w^2}[B(w,\bar{w})P(w,\bar{w})], \tag{13}$$

where $A(w,\bar{w})$ is given by Eq. (8) and

$$B(w,\bar{w})=\int(d\Delta w)\Delta w^2 Q(\Delta w|w,\bar{w}) \tag{14}$$

$$=\eta^2\lambda_i\left\{\left[1+2\Gamma_{1,0}\left(\bar{w}-\frac{c}{M}w\right)\right]\left[\frac{c}{M}w+\Gamma_{2,0}\left(\bar{w}-\frac{c}{M}w\right)\right]f_+^2(w)-2\left[1+(\Gamma_{0,0}+\Gamma_{1,1})\left(\bar{w}-\frac{c}{M}w\right)\right]\right.$$
$$\left.\times\left[\Gamma_{0,1}\frac{c}{M}w+\Gamma_{1,0}\Gamma_{0,1}\left(\bar{w}-\frac{c}{M}w\right)\right]f_+(w)f_-(w)+\left[1+2\Gamma_{0,1}\left(\bar{w}-\frac{c}{M}w\right)\right]\left[\Gamma_{0,2}\frac{c}{M}w+\Gamma_{0,2}\left(\bar{w}-\frac{c}{M}w\right)\right]f_-^2(w)\right\}, \tag{15}$$

where $\Gamma_{\alpha,\beta}=\lambda_{\rm in}/(\lambda_{\rm in}+\alpha/\tau_++\beta/\tau_-)$. The derivation for $B(w,\bar{w})$ is given in Appendix A. With the free boundary conditions that apply here, the stationary solution is given by

$$P(w,\bar{w})=\frac{\mathcal{N}}{B(w,\bar{w})}\exp[\Psi(w,\bar{w})], \tag{16}$$

where $\mathcal{N}$ is the normalization factor and

$$\Psi(w,\bar{w})=\int^w dw'2A(w',\bar{w})/B(w',\bar{w}). \tag{17}$$

The mean weight, $\bar{w}$, can be calculated self-consistently by solving the equation

$$\bar{w}=\int_0^\infty dw'w'P(w',\bar{w}). \tag{18}$$

In practice, this is computationally expensive since three numerical integrals must be performed for each evaluation of the right-hand side of Eq. (18), so it is faster to approximate $P(w,\bar{w})$ as the sum of Gaussian distributions centered on the modes w_+ and w_- as described in Appendix C. This is often highly accurate and at minimum gives a good initial approximation from which to find the self-consistent $\bar{w}$ from the full theory. We emphasize that this approach applies to any choices of the functions $f_+(w)$ and $f_-(w)$ describing the weight dependence of STDP.

Typically Eq. (18) yields a unique solution for $\bar{w}$, indicating that the multistable solutions encountered for small values of η are absent in the large η limit, in agreement with simulations. It is also interesting to know the fraction of potentiated synapses, r. This can be calculated by substituting the self-consistent value for $\bar{w}$ into Eq. (16) and integrating, $r=\int_q^\infty P(w,\bar{w})$, or by using a Gaussian approximation as given in Appendix C [Eq. (C3)]. In Fig. 7(a), r has been calculated in the latter fashion and shown in gray scale in the (b,m) plane for the $M=3$, $c=0.5$ example that was considered in Fig. 5(a) in the small η limit. In contrast to the small η limit shown in Fig. 5(a), for large η there is only one solution for each (b,m) pair in Fig. 7(a) as indicated by the

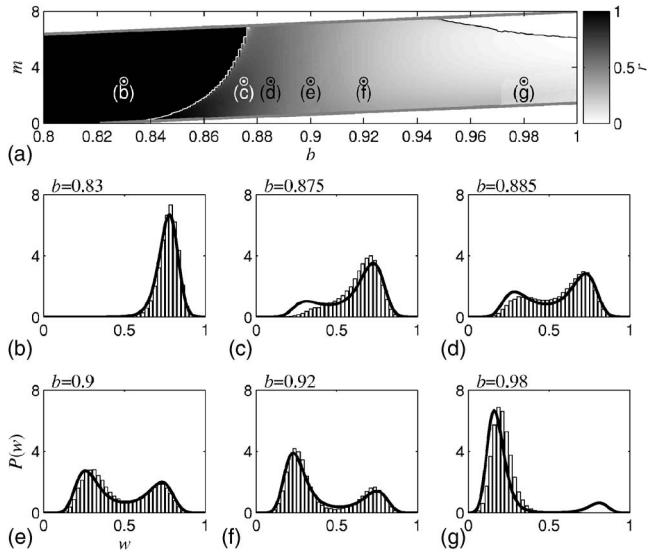

FIG. 7. Predictions of the large learning rate theory for the same parameters as used in Fig. 5 (but with a large value of $\eta=3\times10^{-2}$ s^{-1} instead of the small η limit of Fig. 5). (a) A bifurcation diagram showing the ratio of potentiated synapses, r, in gray scale (colorbar) as a function of (b,m). (b)–(g) Comparison of the distributions $P(w,\bar{w})$ predicted by the theory [Eq. (16), solid lines] to those obtained from simulations (histograms) for six values of b (as marked) and $m=3.0$. These are the same six points used in Figs. 5(b)–5(g), and they are also marked on part (a) of the present figure.

unique gray tone specifying the value of r (see adjacent color bar). Two solid lines mark the boundaries between qualitatively distinct regions of the (b,m) plane. On the left of Fig. 7(a) a white line indicates the transition from the $r=1$ unimodal solution (pure black shading) to the bimodal solution with $r<1$. Choices of (b,m) in the left half of the plot typically result in the majority of synapses being potentiated. As the synaptic depression parameter b increases, so do the fraction of depressed synapses as indicated by continuously diminishing values of r (lighter shading). For sufficiently strong depression, b, all synapses become depressed, resulting once more in a unimodal solution, but with $r=0$ (pure white shading). Part of this boundary can be seen as the black line in the top right corner of the plot.

There is generally fair agreement between the distribution predicted from theory [Eq. (16)] and that obtained by simulation. Comparative examples of the full distribution, $P(w,\bar{w})$, are shown in Figs. 7(b)–7(g) for the six points (b,m) marked in Fig. 7(a) [and used in Figs. 5(b)–5(g) for the small η limit]. The histograms give the results of simulations, while the lines give the results of the theory. The transition from the unimodal $r=1$, through the bimodal $0<r<1$, to the unimodal $r=0$ solutions is apparent in these figures as the degree of synaptic depression, b, increases.

For intermediate values of η lying between the small and large η limit, the corresponding limiting theories described above provide indications about the number and properties of the solution. The value of r for the synaptic distribution in the intermediate η regime typically appears to be the nearest consistent rational approximation to r as predicted by the large η theory. For example, if a value of $r=0.7$ results from the large η theory in a case with $M=4$ subgroups, then the nearest consistent rational approximation is $r=3/4=0.75$. However, this is an observation rather than a consequence of the theory. Furthermore, the theory does not indicate which values of η can be considered intermediate in this sense. A theory that applies in this regime may be possible by resorting to a higher-dimensional Fokker-Planck equation (equal to the number of subgroups), but since the derivation appears formidable and the resulting equation is unlikely to be either analytically solvable or computationally tractable (for M much greater than 3), this approach is not pursued here.

C. Effect of input parameters

Thus far we have considered the effect of the learning parameters b, m, and η while keeping c, M, and λ_{in} fixed. The effect of these input parameters on learning is now described.

1. Effects when $c \propto M$

Inspecting the expressions for $A(w_i,\bar{w})$ [Eq. (8)] and $B(w_i,\bar{w})$ [Eq. (15)], it is seen that the parameters M and c appear only as a ratio, suggesting that input with a different number of subgroups, M, but the same ratio will exhibit the same behavior. While this is true in the large η limit, it does not hold in the small and intermediate η regimes. This is because of the implicit dependence on M in Eq. (12), through the allowable values of $r=0,1/M,2/M,\ldots,(M-1)/M$. This is illustrated in Fig. 8 for the case $c/M=1/6$, of which our standard $M=3$, $c=1/2$ case is an example. The plot shows bifurcation diagrams in the (b,m) plane and small η limit for the six possible values of $M=1$, 2, 3, 4, 5, and 6 (corresponding to $c=1/6$, 1/3, 1/2, 2/3, 5/6, and 1). Since the correlation $c\leqslant 1$, then M can be no larger than 6 since $c/M=1/6$. In Fig. 8, the diagram with $M=3$ and $c=1/2$ is identical to Fig. 5(a) and readers may refer to the explanation of this latter figure to help interpret the schema for the present figure. From the series of diagrams, it is clear that, although there are similarities between them, they are all distinct in detail. There is an increase in complexity of the diagrams as M increases due to the greater number of allowable r values, from $r\in\{0,1\}$ in the first diagram to $r\in\{0,1/6,\ldots,5/6,1\}$ in the final diagram. The following points about each plot should be noted. (i) For any permissable ratio $r=0,1/M,2/M,\ldots,(M-1)/M,1$, there is a region of the (b,m) plane within which a stable solution with this ratio exists, termed the "stable-solution-region." (ii) Such regions show a progression from right to left across the plane as r increases in the following manner. For any two permissable ratios, the greater one will have at least part of its stable-solution region to the left of the other's. Conversely, the lesser ratio will have at least part of its stable-solution region to the right of the other's. (iii) For bimodal ratios $(0<r<1)$, there is always a partial overlap between their stable-solution regions. Further, there is always a region of the (b,m) plane (albeit possibly a very small one) in which stable solutions exist for all permissable bimodal ratios $(0<r<1)$ and one unimodal ratio $(r=0$ or 1). (iv) The

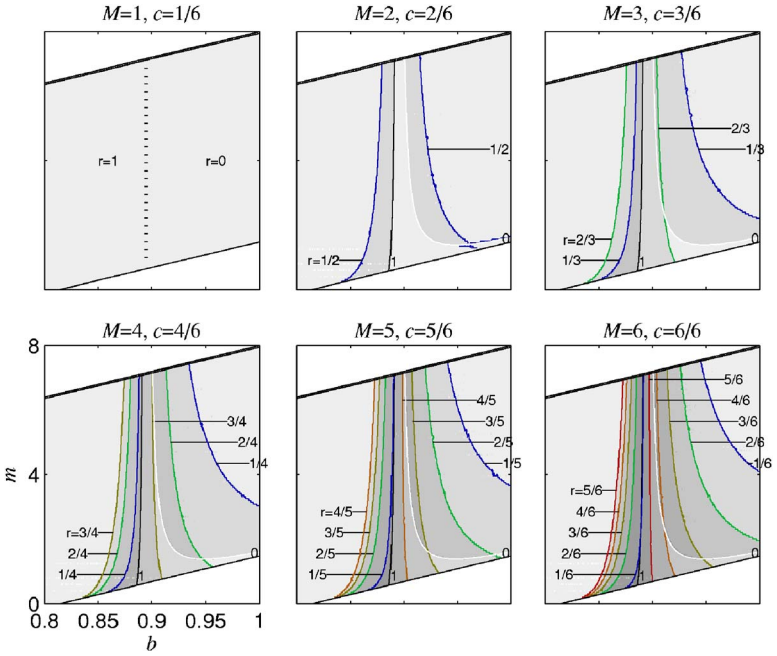

FIG. 8. (Color online) Bifurcation diagrams in the (b,m) plane in the small η limit for $M=1$ through $M=6$ subgroups and with the value of the correlation c chosen in each case so that the ratio $c/M=1/6$ is constant throughout. In each bifurcation diagram, the number of distinct solutions in a region of (b,m) parameter space is given by the degree of gray shade, with the lightest corresponding to one solution and the darkest (seen only in the $M=6$ diagram) corresponding to six solutions. Borders between these regions are delineated by lines (color online) that are marked by the value of the ratio r of the solution that loses stability or existence at this border (so the solution is stable and exists on the darker side of the line). For the $M=1$, $c=1/6$ diagram no bifurcations take place, but there is a transition from the unimodal $r=0$ distribution to the unimodal $r=1$ distribution as the value of $\bar{w}$ passes through q from below. This transition has been marked with a dotted line. A constant ratio, c/M, leaves the drift $A(w_i,\bar{w})$ [Eq. (8)] and diffusion $B(w_i,\bar{w})$ [Eq. (15)] invariant under changes in M and c. Nevertheless, it is clear from the plots that the bifurcation diagrams change as M and c change, despite the constant ratio, c/M (see text for explanation).

stable-solution region for a given ratio, r, is the same regardless of the value of M (i.e., regardless of which plot it appears in). For example, the $r=0$ and 1 regions are the same in every plot, as are the $r=1/3$ and $2/3$ regions in the $M=3$ and 6 plots.

In contrast to the above variation with M for small η, when η is large the unique, stable synaptic distribution is the same for a given ratio, c/M, independent of the particular value of M. Thus the example given in Fig. 7 for $M=3$ and $c=1/2$ remains unaltered for any of the other five possible values of M that can conform to this ratio $c/M=1/6$ (i.e., $M=1$, 2, 4, 5, and 6).

2. Effect of c/M

As the ratio c/M decreases, the region of (b,m)-parameter space capable of supporting bimodal distributions diminishes, so that for small values of c/M both b and m must be very finely tuned in order for synaptic structure to emerge from the learning process. This region can be defined from Eq. (9) by noting that $M\bar{w}/c$ must lie between the local minima and the local maxima of the function $w/F(w)$ (shown in Fig. 2). These two local stationary points of $w/F(w)$ define the minimum and maximum values of w_- and w_+ that are consistent with bimodal distributions, denoted $w_-^{\min}, w_-^{\max}, w_+^{\min}$, and $w_+^{\max}$, respectively (see Fig. 2). Then, as shown in the Appendix B, the following constraints are necessary for a bimodal distribution:

$$F(w_-^{\min}) \leqslant \frac{c}{M} \leqslant F(w_+^{\max}). \tag{19}$$

Equality in these constraints requires the unimodal limits as $r\to 0$ for the lower constraint and $r\to 1$ for the upper constraint. Thus constraints are outer bounds that are never met in practice for any permissible value of $r=1/M,2/M,\ldots,(M-1)/M$. Constraints that are sufficient for a bimodal distribution can be obtained from the condition that the unimodal solution is unstable (see Appendix B),

$$F(w_-^{\max}) \leqslant \frac{c}{M} \leqslant F(w_+^{\min}). \tag{20}$$

These constraints assume that the conditions (i)–(iii) in Sec. III A on $F(w)$ apply. Consequently, they are stronger than these conditions and they also constrain c/M. These constraints are plotted in Fig. 9in the (b,m) plane for six values of $c/M=1/2$, 1/4, 1/8, 1/16, 1/32, and 0. Note that c/M is bounded above by 1 and that the maximum value consistent with a bimodal distribution is $c/M=1/2$. Apart from the first and last plot, each plot has four curves in addition to the two straight parallel lines depicting the non-negativity constants $F(0)<0$ and $f_-(0)>0$. The region of the (b,m) plane that

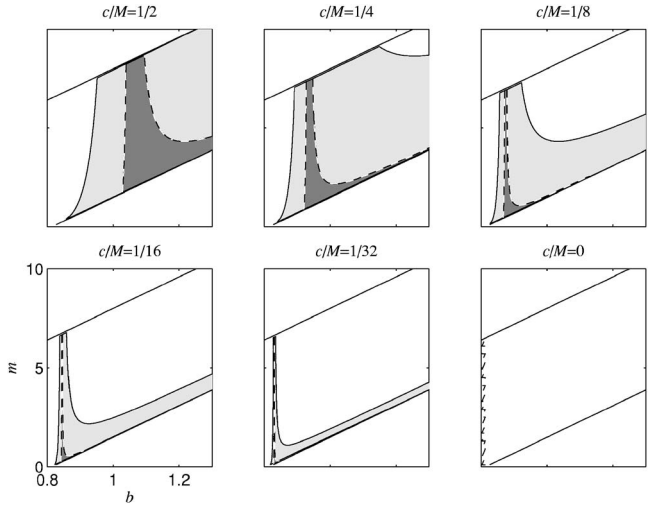

FIG. 9. The necessary and sufficient constraints for the existence of a bimodal solution [as given by Eqs. (19) and (20), respectively] are plotted in the (b,m) plane for decreasing choices of the ratio c/M. The region corresponding to sufficiency is shown in dark shade while the region corresponding to necessity is shown in light and dark shade. In the first plot, where $c/M=1/2$, the curve corresponding to one of the necessity constraints lies to the right of the visible portion of the (b,m) plane. In the final plot, where $c/M=0$, these regions vanish so that there are no values of (b,m) consistent with a bimodal distribution. λ_{in}, a, and q are as in previous figures.

satisfies the necessity constraints [Eq. (19), light and dark shade] lies between the left- and rightmost curves (solid lines). The region that satisfies the sufficiency constraints [Eq. (20), dark shade] lies between the innermost curves (dashed lines). In the final diagram in Fig. 9, where $c/M=0$ (i.e., $M=\infty$ or $c=0$), these regions vanish so that there are no values of (b,m) consistent with a bimodal distribution. For finite values of c/M, these regions appear and grow as c/M increases. [In the first plot, where $c/M=1/2$, the curve corresponding to one of the necessity constraints lies to the right of the visible portion of the (b,m) plane.]

3. Effect of the input rate

The effect of the input rate, λ_{in}, is minor for the range of neural firing rates observed in the cortex (1–100 Hz) given the experimental values reported for $\tau_+=17$ ms and $\tau_-=34$ ms [3].

The main effect of λ_{in} appears through the Γ symbols in Eq. (8) for the drift and Eq. (15) for the diffusion. It is useful to consider the low and high λ_{in} limits. When $\lambda_{in}\tau_\pm \ll 1$, we have $\Gamma_\pm \to \lambda_{in}\tau_\pm$ and we recover the model without any input restrictions for input-output interactions in STDP. This model behaves in a qualitatively similar fashion to the input restricted model examined in this paper with moderate λ_{in}. The drift function becomes

$$A(w,\bar{w}) = \eta\lambda_{in}\left\{[\lambda_{in}\tau_+ f_+(w) - \lambda_{in}\tau_- f_-(w)]\bar{w} + \frac{c}{M}f_+(w)w\right\}. \tag{21}$$

There will be a greater tendency for the term $(c/M)f_+(w)w$ to break symmetry because of its relatively large magnitude compared to the other term when $\lambda_{in}\tau_\pm \ll 1$. For the values of $\tau_\pm$ measured in the cortex, this limit would only be encountered when neurons are firing at their background rate of a few hertz.

In the opposite limit, in which $\lambda_{in}\tau_\pm \gg 1$, $\Gamma_\pm \to 1$. In this case, the symmetry breaking term $(1-\Gamma_+)(c/M)f_+(w)w$ vanishes and

$$A(w,\bar{w}) = \eta\lambda_{in}[f_+(w) - f_-(w)]\bar{w}. \tag{22}$$

Consequently, there can be no bimodal distributions. The stationary distribution has all the weights at exactly one value w_0 given by $f_+(w_0)=f_-(w_0)$, since the diffusion

$$B(w,\bar{w}) = \eta^2\lambda_{in}[f_+(w) - f_-(w)]^2 \times \left[1 + 2\left(\bar{w} - \frac{c}{M}w\right)\right]\bar{w} \tag{23}$$

also vanishes at this point. Fortunately, this undesirable behavior only occurs at firing rates of more than several hundred hertz, which is outside the normal operating range found in the cortex.

For neural firing rates above background rates but still observed in the cortex (20–100 Hz), there are considerable and overlapping regions of (b,m) parameter space that support bimodal distributions. The detailed variation in these parameter regions as λ_{in} varies within the cortical operational range is complicated and is not investigated any further here.

IV. DISCUSSION AND CONCLUSION

In this paper, we have described the conditions that permit the formation of synaptic structure through stable and competitive STDP when inputs consist of correlated synaptic subgroups. This has been done in general by specifying the conditions (i)–(iii) in Sec. III A that the potentiation and depression functions, f_+ and f_-, must satisfy for the emergence of bimodal distributions to be possible during learning. It has also been done in detail by analyzing the role of both input and plasticity parameters in determining the structure of the synaptic distributions that can emerge from learning, for the particular choice of f_+ and f_- given in Eq. (10). We will now summarize the main findings and discuss them with reference to previous work on STDP.

The key aspect of the STDP models considered here is that they are both competitive and stable. This may be an important consequence of the weight dependence in STDP that is observed experimentally [3]. Much previous work on STDP has considered rules that are either competitive (e.g., the rule with "additive" potentiation and depression) [1,8,11–15] or stable (e.g., the rule with "additive" potentiation and "multiplicative" depression) [16,17]. The model proposed by Gütig *et al.* [18] interpolated continuously between the "additive" and "multiplicative" models and was the first to exhibit both competition and stability. In conditions (i)–(iii) in Sec. III A we have given general conditions for the kind of models that exhibit both these properties.

One new, and for us surprising, result is the existence of multiple stable synaptic distributions when the learning rate is sufficiently small. In this case, the particular distribution

that emerges depends on the initial condition of the weights and the particular stochastic realization of the individual weight changes. These alternative possible distributions can be very different from each other, for example a unimodal distribution with the weights of all subgroups depressed ($r=0$) versus a bimodal distribution with the weights of half the subgroups potentiated and the remaining half depressed ($r=1/2$). In general the number of alternative distributions can be very numerous but will be less than the number of subgroups M. The existence of multiple stable distributions in the small η limit is common to all STDP models that are both competitive and stable (at least when the inputs consist of correlated subgroups to a Poisson linear neuron as considered here). This is due to the multiple values of r giving rise to different versions of Eq. (12). For instance, multiplicity also occurs for the model of Gütig *et al.* [18]. When the model is competitive, but not stable, as in the "additive" model, a related multiplicity of synaptic structure emerges. For example, in the "additive" model, weights evolve along the eigenvector of $\partial A(w_i)/\partial w_j$ whose eigenvalue has the greatest real part. It turns out that this eigenvalue has very high multiplicity and thus the eigenspace through which the weights evolve has a correspondingly high dimension giving a large number of possible weight configurations. In contrast, for STDP models that are stable, but not competitive, there is usually only one stable distribution, which is unimodal [16].

In contrast to the multiplicity of solutions and values of r typically seen at low learning rates, at higher rates there is usually a unique value of r. For very high rates, the synaptic structure is not stable because large stochastic fluctuations in the weights cause potentiated and depressed subgroups to spontaneously flip places. This results in neurons that do not have consistent response properties over time. For intermediate learning rates, a unique and stable synaptic structure typically emerges. In this case, the value of r is analogous to the sparseness of coding in a purely feed forward network in which all the neurons receive the same correlated subgroup input as specified here. The primary determinant of r is the strength of depression compared to potentiation, as parametrized through b and m (see, for example, Fig. 7). Empirically the value of r for synaptic distributions in the intermediate η regime typically appears to be the nearest consistent rational approximation to r as predicted by the large η theory.

Finally the role of the input parameters M, c, and λ_{in} was also considered (in Sec. III C). An important parameter is c/M: the subgroup correlation divided by the number of subgroups. This quantity plays an important role in determining the robustness with which a subgroup structure forms on the synapses during learning. When c/M is very small, the learning parameters b and m must be very finely tuned in order for bimodal distributions to form (see, e.g., Fig. 9 with $c/M=1/32$). This is intuitively correct since it corresponds to a situation of very weak correlation or a very large number of subgroups (or both).

ACKNOWLEDGMENTS

The authors thank Michael Eager for a critical reading of the manuscript. This work was funded by the Australian Research Council (ARC Discovery Project No. DP0453205) and the Bionic Ear Institute. A.N.B. acknowledges funding from the LEW Carty Foundation and the Jack Brockhoff Foundation. D.B.G. acknowledges funding from the Victorian Lions Foundation.

APPENDIX A: DERIVATION OF THE DRIFT AND DIFFUSION FUNCTIONS

The definitions of $A(w_i,\bar{w})$ [Eq. (6)] and $B(w_i,\bar{w})$ [Eq. (14)] require that we identify small weight changes, Δw_i, independent of other independent weight changes. These are not congruent with the individual contributions, δw_i, described by Eq. (1), because these latter contributions will be correlated whenever they share the same input times, t_{in}, despite having independent output times, t_{out}. However, for the input restricted model considered here, the sum total of all individual contributions during a single input interspike interval represents an independent contribution, Δw_i, since no input-output interactions occur outside this interval. Possible contributions during a typical interval are illustrated in Fig. 1. The first possible contribution is at the start of the interval at time $T_0=\epsilon\approx 0$, and occurs if the input from synapse i (and possibly inputs from other synapses in its subgroup) produces an output spike. All further inputs during this interval occurring at consecutive times $T_n>0$ cannot involve synapse i until this synapse again has an input at time T^{in} which terminates the interval. Thus

$$\Delta w_i = \eta[f_+(w_i) - f_-(w_i)e^{-T^{\text{in}}/\tau_-}]\chi(S_0) + \eta\sum_{n=1}^{\infty}[f_+(w_i)e^{-T_n/\tau_+} - f_-(w_i)e^{-(T^{\text{in}}-T_n)/\tau_-}]\chi(S_n)H(T^{\text{in}}-T_n). \tag{A1}$$

Here the Heaviside function $H(T^{\text{in}}-T_n)$ ensures that the contribution is counted if and only if the time T_n falls inside the input interspike interval, while the decision function $\chi(S_n)$ ensures that the contribution is counted if and only if the event S_n="The input spike(s) at time T_n caused an output spike" is true,

$$\chi(S) = \begin{cases} 1 & \text{if } S \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{A2}$$

These independent contributions must be integrated in Eqs. (6) and (14) to find $A(w_i,\bar{w})$ and $B(w_i,\bar{w})$, respectively, according to the conditional probability density $Q(\Delta w_i|w_i,\bar{w})$. This density is defined by (i) the arrival of inputs in Poisson spike trains as described in Sec. II B and (ii) the consequent probability that they will produce an output spike, as given by Eq. (4). We address each of these factors in turn. First, the input spike trains to the neuron can be partitioned into the events in which synapse i has an input and those in which it does not. Since all input spike times are described by homogeneous Poisson processes, these two disjoint classes of input events are given by independent Poisson processes with rates λ_i for those in which i participates, and λ_o for those in which it does not. For inputs consisting of M correlated subgroups, recall that events arrive at synapse i according to independent and correlated Poisson input streams. Let I_j and

C_j denote independent and correlated events on synapse j, respectively. Let s_j be the event in which there is an input spike on synapse j, and $\bar{s}_j$ be the event that there is not. Define $\lambda(E)$ to be the rate at which a given event E occurs and $P(E)$ be the probability of that event. Then according to the definitions (Sec. II B) we have $\lambda(I_j)=\lambda(C_j)=\lambda_{\text{in}}$, $P(s_j|I_j)=1-\sqrt{c}$, and $P(s_j|C_j)=\sqrt{c}$ for all synapses j. Now λ_i contains contributions (in the order in which they appear below) from its own independent events and from correlated events of its own subgroup, with the condition that synapse i participates in both event types. Similarly λ_{o} contains contributions from synapse i's independent events and from correlated events of i's subgroup, provided that synapse i does not participate in either event type, as well as contributions from independent events of other synapses and correlated events of other subgroups. Thus

$$\lambda_i = \lambda(I_i)P(s_i|I_i) + \lambda(C_i)P(s_i|C_i) = \lambda_{\text{in}}(1-\sqrt{c}) + \lambda_{\text{in}}\sqrt{c} = \lambda_{\text{in}}, \tag{A3}$$

$$\begin{aligned}\lambda_{\text{o}} &= \lambda(I_i)P(\bar{s}_i|I_i) + \lambda(C_i)P(\bar{s}_i|C_i) + \sum_{j\neq i}\lambda(I_j)P(\bar{s}_i|I_j) \\ &\quad + \sum_{\mathcal{G}_k\neq\mathcal{G}_i}\lambda(C_k)P(\bar{s}_i|C_k) \\ &= \lambda_{\text{in}}\sqrt{c} + \lambda_{\text{in}}(1-\sqrt{c}) + (N-1)\lambda_{\text{in}} + (M-1)\lambda_{\text{in}} \\ &= (M+N-1)\lambda_{\text{in}}.\end{aligned} \tag{A4}$$

In the calculation of λ_{o}, note that an independent event in which the synapses cannot participate has zero probability of causing an output spike. This is taken into account in the second part of the calculation, in which the consequent probability of an output spike is also considered on the basis of whether synapse i participated or not. The probability of an output spike given an input event in which synapse i participated, $P_i(S)$, contains the same contributions from both correlated and independent events listed above for λ_i and is given by

$$P(S|s_i) = P(S|I_i,s_i)P(I_i|s_i) + P(S|C_i,s_i)P(C_i|s_i). \tag{A5}$$

Now denote a partition of subgroup $\mathcal{G}_i$ into those synapses that participate and those that do not by $\{\rho,\bar{\rho}\}$. Then

$$\begin{aligned}P(S|C_i,s_i) &= \sum_{\{\rho,\bar{\rho}\}} P(s_\rho,\bar{s}_{\bar{\rho}}|C_i,s_i)\sum_{k\in\rho}\frac{w_k}{N} \\ &= \sum_{j\in\mathcal{G}_i}\frac{w_j}{N}\sum_{\{\rho,\bar{\rho}\}:j\in\rho}P(s_\rho,\bar{s}_{\bar{\rho}}|C_i,s_i) = \sum_{j\in\mathcal{G}_i}P(s_j|C_i,s_i)\frac{w_j}{N} \\ &= \sqrt{c}\sum_{j\in\mathcal{G}_i,j\neq i}\frac{w_j}{N} + \frac{w_i}{N} = \frac{\sqrt{c}}{M}\hat{w}_i + \frac{1-\sqrt{c}}{N}w_i.\end{aligned} \tag{A6}$$

Then returning to Eq. (A5) we obtain

$$\begin{aligned}P_i(S) &= \frac{w_i}{N}(1-\sqrt{c}) + \left(\frac{\sqrt{c}}{M}\hat{w}_i + \frac{1-\sqrt{c}}{N}w_i\right)\sqrt{c} \\ &= \frac{c}{M}\hat{w}_i + \frac{1-c}{N}w_i.\end{aligned} \tag{A7}$$

The corresponding probability given that i did not participate contains the same contributions as previously described for λ_{o}. It is

$$\begin{aligned}P_{\text{o}}(S) &= P(S|\bar{s}_i) = P(S|I_i,\bar{s}_i)P(I_i|s_i) + P(S|C_i,\bar{s}_i)P(C_i|s_i) \\ &\quad + \sum_{j\neq i}P(S|I_j)P(I_j) + \sum_{\mathcal{G}_k\neq\mathcal{G}_i}P(S|C_k)P(C_k) \\ &= 0\left(\frac{\sqrt{c}}{M+N-1}\right) + \left(\frac{\sqrt{c}\hat{w}_i}{M} - \frac{\sqrt{c}w_i}{N}\right)\left(\frac{1-\sqrt{c}}{M+N-1}\right) \\ &\quad + \sum_{j\neq i}\left(\frac{(1-\sqrt{c})w_j}{N}\right)\left(\frac{1}{M+N-1}\right) + \sum_{\mathcal{G}_k\neq\mathcal{G}_i}\left(\frac{\sqrt{c}\hat{w}_k}{M}\right) \\ &\quad \times\left(\frac{1}{M+N-1}\right) = \frac{1}{M+N-1}\left(\bar{w} - \frac{c}{M}\hat{w}_i - \frac{1-c}{N}w_i\right).\end{aligned} \tag{A8}$$

In the above, several of the probabilities have been derived in a fashion similar to that of Eq. (A6). The (normalized) conditional probability density may now be expressed as

$$\begin{aligned}&\int (d\Delta w_i)Q(\Delta w_i|w_i,\bar{w}) \\ &\quad = \lambda_i\int_0^\infty dT^{\text{in}}p_{\lambda_i}(T^{\text{in}})\sigma_0\int_0^\infty dT_1 p_{\lambda_{\text{o}}}(T_1)\sigma_1 \\ &\qquad \times\int_{T_1}^\infty dT_2 p_{\lambda_{\text{o}}}(T_2-T_1)\sigma_2\cdots \\ &\qquad \times\int_{T_{k-1}}^\infty dT_k p_{\lambda_{\text{o}}}(T_k-T_{k-1})\sigma_k\cdots,\end{aligned} \tag{A9}$$

where $p_\lambda(T)\equiv\lambda\exp(-\lambda T)$ describes an exponential distribution, and the bivariate distributions for the existence of the output spike at T_0 or T_n for $n\in\mathbb{N}$ are given by

$$\sigma_0 \equiv P_i(S_0)\chi(S_0) + [1-P_i(S_0)]\chi(\bar{S}_0), \tag{A10}$$

$$\sigma_n \equiv P_{\text{o}}(S_n)\chi(S_n) + [1-P_{\text{o}}(S_n)]\chi(\bar{S}_n), \tag{A11}$$

respectively [the terms with $\chi(\bar{S}_n)$ vanish when the integral is performed since $\delta w_i=0$ in this case]. The first integral in Eq. (A9) describes the distribution of synapses i's input interval, while the remaining integrals describe the distribution of consecutive input events occurring after the input spike on synapse i at $t=0$ (see Fig. 1). In the calculation of the change of weight, Eq. (A1), the Heaviside step function cuts off the T_n integral at T^{in}. Consequently, in any explicit calculation the higher T_k integrals give a multiplicative factor of 1 and all the T_k integrals for $k\leqslant n$ are cut off at T^{in}.

Using Eqs. (A1) and (A9), $A(w_i,\bar{w})$ and $B(w_i,\bar{w})$ can be evaluated from their definitions Eqs. (6) and (14), respec-

tively. The nested integrals are performed using Laplace transforms as given in [17]. The results are

$$A(w_i,\bar{w}) = \eta\lambda_i\{[P_i(S)+\Gamma^{\mathrm{o}}_{1,0}P_{\mathrm{o}}(S)]f_+(w_i) - [\Gamma^{i}_{0,1}P_i(S) + \Gamma^{\mathrm{o}}_{0,1}P_{\mathrm{o}}(S)]f_-(w_i)\}, \tag{A12}$$

$$B(w_i,\bar{w}) = \eta^2\lambda_i\{[1+2\Gamma^{\mathrm{o}}_{1,0}P_{\mathrm{o}}(S)][P_i(S)+\Gamma^{\mathrm{o}}_{2,0}P_{\mathrm{o}}(S)]f_+^2(w_i) - 2[1+(\Gamma^{\mathrm{o}}_{0,0}+\Gamma^{\mathrm{o}}_{1,1})P_{\mathrm{o}}(S)][\Gamma^{i}_{0,1}P_i(S) + \Gamma^{i}_{1,0}\Gamma^{\mathrm{o}}_{0,1}P_{\mathrm{o}}(S)]f_+(w_i)f_-(w_i) + [1+2\Gamma^{\mathrm{o}}_{0,1}P_{\mathrm{o}}(S)] \times[\Gamma^{i}_{0,2}P_i(S)+\Gamma^{\mathrm{o}}_{0,2}P_{\mathrm{o}}(S)]f_-^2(w_i)\}, \tag{A13}$$

where $\Gamma^{u}_{\alpha,\beta}=\lambda_u/(\lambda_i+\alpha/\tau_+ +\beta/\tau_-)$.

APPENDIX B: CONDITIONS FOR THE EXISTENCE OF BIMODAL DISTRIBUTIONS

The necessary conditions [(i)–(iii)] in Sec. III A for functions F associated with bimodal distributions can be derived as follows. (i) $F'(w)\geqslant 0\ \forall w\in\mathcal{D}$. $F(w)$ must be monotone for all positive w, otherwise the homogeneous solution given by $F(w_0)=c/M$ will have more than one solution for some choice of c/M, which contradicts our assumptions. Further, $F'(w)$ must be positive for $w>0$, since if it were negative there would only be one solution to Eq. (9) for any choice of $c/M\bar{w}$, which is also a contradiction. Notice also that the monotonicity of $F(w)$ is enough to ensure that symmetry breaking never occurs within subgroups. This follows from a rearrangement of Eq. (9) to give $F(w_i^*)=(c\hat{w}_i)/(M\bar{w})$, where we have retained the original average weight over the subgroup $\hat{w}_i$as it appears in Eq. (8). Using the fact that the right-hand side of this expression is the same for all members of a subgroup and that $F(w)$ is monotone, we find that $w_j^*=w_i^*$ for any synapses j and i that are members of the same subgroup. This in turn implies the absence of symmetry breaking within subgroups. (ii) $F(0)<0$, since otherwise there would be no positive homogeneous solution to $F(w_0)=c/M$ for sufficiently small choices of c/M. (iii) $\exists w_f\in\mathcal{D}: F''(w_f)=0$ and $F'''(w_f)>0$. Assume two stable fixed points w_- and w_+ and one unstable fixed point w_0 such that $w_-<w_0<w_+$. Notice that the drift may be written as $A(w,\bar{w})=\eta\lambda_{\mathrm{in}}f_+(w)\bar{w}(1-\Gamma_+)[-F(w)+cw/M\bar{w}]$. A stable fixed point w^* requires that $A'(w^*)<0$, while $A'(w^*)>0$ implies an unstable fixed point. Thus we must have $F'(w_\pm)>c/M\bar{w}$ and $F'(w_0)<c/M\bar{w}$. Since $F(w)$ is smooth [recall $f_+(w)$ and $f_-(w)$ are smooth and positive], this implies that $F'(w)$ has a local minimum between w_- and w_+, i.e., $\exists w_f>0$ such that $F''(w_f)=0$ and $F'''(w_f)>0$as required.

The inequalities giving necessary [Eq. (19)] and sufficient [Eq. (20)] conditions for the existence of bimodal solutions in terms of the model parameters can be found as follows. The existence of bimodal solutions implies that Eq. (9) has multiple nonidentical solutions, which we observe from Fig. 2 requires that $M\bar{w}/c$ lie between the local minima and the local maxima of $w/F(w)$ [the existence of which are guaranteed by the conditions on $F(w)$]. Thus

$$\frac{w_-^{\max}}{F(w_-^{\max})} < \frac{M\bar{w}}{c} < \frac{w_+^{\min}}{F(w_+^{\min})}, \tag{B1}$$

where $w_-^{\max}$ and $w_+^{\min}$ are defined in Fig. 2. By the definitions of $w_-^{\min}$ and $w_+^{\max}$ (see Fig. 2), we also have

$$\frac{w_+^{\max}}{F(w_+^{\max})} < \frac{M\bar{w}}{c} < \frac{w_-^{\min}}{F(w_-^{\min})}. \tag{B2}$$

The necessary bounds [Eq. (19)] follow since we must always have $w_-^{\min}<\bar{w}<w_+^{\max}$ for bimodal solutions. The sufficient bounds [Eq. (20)] arise because the homogeneous solution is always unstable whenever $w_-^{\max}<w_0\approx\bar{w}<w_+^{\min}$ in Eq. (B1) (implying the existence of a stable bimodal solution). This follows because $w/F(w)$ is positive and has positive slope in this region, which implies $F'(w)<F(w)/w$ for $w>0$. Thus at the fixed point, where $F(w^*)/w^*=c/M\bar{w}$, we have $A'(w^*)=\eta\lambda_{\mathrm{in}}f_+(w)\bar{w}(1-\Gamma_+)[-F'(w^*)+c/M\bar{w}]>0$, which implies instability.

APPENDIX C: GAUSSIAN APPROXIMATION

To approximate $P(w,\bar{w})$ [Eq. (16)] as the sum of Gaussian distributions centered on the modes w_+ and w_-, one expands $\Psi(w,\bar{w})$ [Eq. (17)] as a Taylor polynomial around each mode to second order. This gives

$$P(w,\bar{w}) \approx \mathcal{N}\sum_{\phi=+,-}\sqrt{\frac{-\exp\{2\Psi(w_\phi,\bar{w})\}}{A'(w_\phi,\bar{w})B(w_\phi,\bar{w})}} \times\left[\frac{1}{\sqrt{2\pi\sigma_\phi^2}}\exp\left\{-\frac{(w-w_\phi)^2}{2\sigma_\phi^2}\right\}\right], \tag{C1}$$

where the variance about a mode, $\phi\in\{+,-\}$, is $\sigma_\phi^2\approx -B(w_\phi,\bar{w})/2A'(w_\phi,\bar{w})$. Using this approximation, Eq. (18) becomes

$$\bar{w} = \left[\sum_\phi w_\phi\sqrt{-\frac{\exp\{2\Psi(w_\phi,\bar{w})\}}{A'(w_\phi,\bar{w})B(w_\phi,\bar{w})}}\right] \times\left[\sum_\phi\sqrt{-\frac{\exp\{2\Psi(w_\phi,\bar{w})\}}{A'(w_\phi,\bar{w})B(w_\phi,\bar{w})}}\right]^{-1}, \tag{C2}$$

where the modes w_ϕ should be thought of as implicit functions of $\bar{w}$ which can be found as the real positive roots of $A(w,\bar{w})$ as given by Eq. (8). After solving Eq. (C2) for $\bar{w}$, it can be substituted into Eq. (16) to obtain the distribution if required, or used as an initial approximation to solve the full theory [Eq. (18)]. One can also estimate the ratio of potentiated synapses, r, from the Gaussian approximation as

$$r = \left[\sqrt{-\frac{\exp\{2\Psi(w_+,\bar{w})\}}{A'(w_+,\bar{w})B(w_+,\bar{w})}}\right] \times\left[\sum_\phi\sqrt{-\frac{\exp\{2\Psi(w_\phi,\bar{w})\}}{A'(w_\phi,\bar{w})B(w_\phi,\bar{w})}}\right]^{-1}. \tag{C3}$$

[1] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, Nature (London) **383**, 76 (1996).
[2] H. Markram, L. Lübke, M. Frotscher, and B. Sakmann, Science **275**, 213 (1997).
[3] G.-Q. Bi and M.-M. Poo, J. Neurosci. **18**, 10464 (1998).
[4] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M.-M. Poo, Nature (London) **395**, 37 (1998).
[5] D. Debanne, B. H. Gähwiler, and S. M. Thompson, J. Physiol. (London) **507**, 237 (1998).
[6] K. D. Miller and D. J. C. MacKay, Neural Comput. **6**, 100 (1994).
[7] K. D. Miller, Neuron **17**, 371 (1996).
[8] S. Song, K. D. Miller, and L. F. Abbott, Nat. Neurosci. **3**, 919 (2000).
[9] L. F. Abbott and S. B. Nelson, Nat. Neurosci. **3**, 1179 (2000).
[10] R. P. N. Rao and T. J. Sejnowski, Neural Comput. **13**, 2221 (2001).
[11] L. F. Abbott and K. I. Blum, Cereb. Cortex **6**, 406 (1996).
[12] R. Kempter, W. Gerstner, and J. L. van Hemmen, Phys. Rev. E **59**, 4498 (1999).
[13] R. Kempter, W. Gerstner, and J. L. van Hemmen, Neural Comput. **13**, 2709 (2001).
[14] P. D. Roberts, J. Comput. Neurosci. **7**, 235 (1999).
[15] J. L. van Hemmen, in *Handbook of Biological Physics (Vol. 4): Neuro-Informatics and Neural Modelling*, edited by F. Moss and S. Gielen (Elsevier, Amsterdam, 2001), pp.771–823.
[16] M. C. W. van Rossum, G. Q. Bi, and G. G. Turrigiano, J. Neurosci. **20**, 8812 (2000).
[17] A. N. Burkitt, H. Meffin, and D. B. Grayden, Neural Comput. **16**, 885 (2004).
[18] R. Gütig, R. Aharonov, S. Rotter, and H. Sompolinsky, J. Neurosci. **23**, 3697 (2003).
[19] P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson, Neuron **32**, 1149 (2001).
[20] D. Paré and A. Destexhe, J. Neurophysiol. **79**, 1450 (1998).
[21] A. Destexhe and D. Paré, J. Neurophysiol. **81**, 1531 (1999).
[22] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[23] H. Risken, *The Fokker-Planck Equation*, 3rd ed. (Springer, Berlin, 1996).
[24] M. Abeles, *Local Cortical Circuits: An Electrophysiological Study* (Springer, Berlin, 1982).