

# Spike-Time-Dependent Plasticity and Heterosynaptic Competition Organize Networks to Produce Long Scale-Free Sequences of Neural Activity

Ila R. Fiete,<sup>1,\*</sup> Walter Senn,<sup>2</sup> Claude Z.H. Wang,<sup>3</sup> and Richard H.R. Hahnloser<sup>3</sup>

<sup>1</sup>Center for Learning and Memory, University of Texas at Austin, Austin, TX 78712, USA

<sup>2</sup>Department of Physiology, University of Bern, CH-3012 Bern, Switzerland

<sup>3</sup>Institute for Neuroinformatics, ETH and the University of Zurich, CH-8057 Zurich, Switzerland

\*Correspondence: [fiete@mail.clm.utexas.edu](mailto:fiete@mail.clm.utexas.edu)

DOI 10.1016/j.neuron.2010.02.003

## SUMMARY

Sequential neural activity patterns are as ubiquitous as the outputs they drive, which include motor gestures and sequential cognitive processes. Neural sequences are long, compared to the activation durations of participating neurons, and sequence coding is sparse. Numerous studies demonstrate that spike-time-dependent plasticity (STDP), the primary known mechanism for temporal order learning in neurons, cannot organize networks to generate long sequences, raising the question of how such networks are formed. We show that heterosynaptic competition within single neurons, when combined with STDP, organizes networks to generate long unary activity sequences even without sequential training inputs. The network produces a diversity of sequences with a power law length distribution and exponent  $-1$ , independent of cellular time constants. We show evidence for a similar distribution of sequence lengths in the recorded premotor song activity of songbirds. These results suggest that neural sequences may be shaped by synaptic constraints and network circuitry rather than cellular time constants.

## INTRODUCTION

Reaching for or throwing objects, walking, and vocalizing are a few of the ways vertebrates interact with the world. Vertebrates also plan, visualize, or review action and event sequences. Underlying the time-varying patterns of muscle activation or sequential cognitive processing are sequences of neural activity. Such sequences are found in various parts of the brain, including the cortex (Schwartz and Moran, 1999; Andersen et al., 2004; Pulvermüller and Shtyrov, 2009; Luczak et al., 2007; Buonomano, 2003; Ikegaya et al., 2004; Tang et al., 2008), hippocampus (Nádasdy et al., 1999; Louie and Wilson, 2001; Pastalkova et al., 2008; Davidson et al., 2009), basal ganglia (Barnes et al., 2005), and the songbird HVC (Hahnloser et al., 2002; Kozhevni-

kov and Fee, 2007), under various behavioral states. The ubiquity of repeating sequential neural patterns across species, task and nontask conditions, and even in vitro suggests that the mechanisms for creating sequence-producing circuits may be quite general and robust. Yet little is known, from experiment or theory, about what these mechanisms might be. In this work, we investigate plasticity rules that could sculpt sequence-producing neural circuits out of initially disordered networks.

What are some of the properties of sequential neural activity patterns? Sequences are frequently much longer than the membrane and synaptic time-constants of individual neurons. The coding of sequences is sparse. For instance, individual premotor neurons in motor cortex are active in only small portions of a figure-eight arm tracing trajectory in monkeys (Schwartz and Moran, 1999). Similarly, hippocampal place cells fire at one or a few locations of a long track while the animal runs or as it rehearses its possible forward trajectories at a decision point (Pastalkova et al., 2008) or as it replays in sleep its place cell activation sequence (Louie and Wilson, 2001). Zebra finches produce song motifs lasting up to 1 s, while individual neurons in the high-level premotor center are each active for only single bursts of about 6 ms duration (Hahnloser et al., 2002) over the full song sequence. In other words, the high-level coding of sequential activity in the brain is sparse, with single neurons firing for small portions of the entire sequence.

Many sequential behaviors are also “modular,” composed of gestures or shorter sequences that can be flexibly arranged and combined. The underlying neural codes are also found to be modular, sometimes even when the behavior itself is not obviously so. For example, although the song of a zebra finch consists of a largely stereotyped single sequence of syllables, the neural drive underlying the song appears to consist of a concatenation of a disjoint set of separate subsequences of neural activity (Tanji, 2001; Glaze and Troyer, 2006; Wang et al., 2008; Davidson et al., 2009).

Several network-level models seek to explain the propagation of sequential neural activity. A number of such models can be grouped into the category of “synaptic chain” networks (Amari, 1972; Kleinfeld and Sompolinsky, 1988; Abeles, 1991; Drew and Abbott, 2003; Li and Greenside, 2006; Jin et al., 2007). In synaptic chain networks, the connectivity matrix is asymmetric or directional, with one group of neurons connecting to the next, and so on. Activity in the network flows in the direction of

the underlying connections. Such an architecture is consistent with the dynamics of sequence generation in the premotor nucleus HVC of songbirds (discussed in Fiete and Seung, 2008; Weber and Hahnloser, 2007). Synaptic chain models represent a first step toward understanding neural sequence generation, but the requisite network connectivity is hand designed and hard wired. There is relatively little experimental or theoretical understanding of how initially unstructured networks may organize into synaptic chain configurations. Similarly, it remains unknown how multiple neural subsequences of varying length (Tanji, 2001; Wang et al., 2008; Davidson et al., 2009) are formed.

Spike-time-dependent plasticity (STDP) rules demonstrably allow networks to perform next-step recall of sequentially presented inputs: STDP rules translate repeated sequential activations of pairs of neurons into a permanent increase in the synaptic strength from the first onto the second neuron (ordered by time of activation), while weakening the reverse connection (Bi and Poo, 1998, 2001; Abbott and Nelson, 2000), making STDP a natural candidate for explaining synaptic chain formation.

But simulation studies make it clear that STDP rules with bounds on individual synaptic strengths are largely unsuccessful at producing networks that autonomously generate long or sparse neural activity sequences (Aviel et al., 2003; Levy et al., 2001; Suri and Sejnowski, 2002; Rao and Sejnowski, 2003; Nowotny et al., 2003). This is because STDP tends to enhance population synchrony (temporal bunching) and concentrate activity in a few winning neurons (spatial bunching): the forward synapse between a pair of neurons firing in close succession will be strengthened, thereby further decreasing the lag between their firing times and thus promoting their synchrony (Buonomano, 2005). A neuron that fires frequently early on will have its inputs strengthened and will also tend to successfully drive its outputs, quickly becoming a hub that drives simultaneous activity in a large fraction of the network. These results illustrate the difficulty encountered in explaining how various brain areas could organize to generate sequential patterns of neural activity.

One method for forming long sequences using STDP is to consider a network of intrinsically bursting neurons and sequentially grow a chain by restricting synaptic plasticity to just the few neurons at the end of the growing chain (Jun and Jin, 2007). However, this approach does not allow for the simultaneous formation of multiple chains and requires a separate scheme for producing a range of chain lengths.

A cellular property that has not been linked with sequence-producing networks or sequence learning is heterosynaptic competition. Heterosynaptic competition for synapse growth or total synaptic strength has been documented at both pre- and postsynaptic neurons. For example, postsynaptic neurons balance activity-dependent potentiation of an input synapse by inducing heterosynaptic depression among other input synapses, conserving the total synaptic weight onto the neuron (Royer and Paré, 2003). Similarly, the dependence of long-term potentiation on the synthesis of new proteins provides neurons with the ability to constrain the strengthening and weakening of outgoing synapses on the full-cell level (Huber et al., 2000; Fonseca et al., 2004, 2006).

We show that when STDP is combined with heterosynaptic competition for scarce synapse-building resources on the level

of individual neurons, initially random neural networks robustly self-organize to form multiple synaptic chains of different lengths. If inputs to the network are sequential and dense, the combined plasticity rules drive the network to rapidly learn unary versions of the input sequence. Surprisingly, a network with these plasticity rules self-organizes to produce long unary chains of activity even if the training inputs are temporally random, with no sequential structure.

For concreteness, we identify our model network with the songbird premotor area HVC. The reasons for this choice are that, first, HVC appears to originate sequential activity, rather than inheriting it as sequential input from an upstream area (Nottebohm et al., 1976, 1982; Bottjer et al., 1984; Hahnloser et al., 2002; Fee et al., 2004; Long and Fee, 2008); second, the constituent neuron types and their activity patterns during song are well-characterized (Mooney, 2000; Hahnloser et al., 2002; Mooney and Prather, 2005; Kozhevnikov and Fee, 2007); and third, HVC is thought to possess an underlying synaptic chain structure (arguments in Fiete and Seung, 2008; Seung, 2009). We demonstrate that the lengths of the chains formed by learning obey a power law that resembles the distribution of HVC chain lengths, as inferred from electrical stimulation experiments in songbirds. The model, because of its genericness, could be applied to other areas where sequences are known to originate and where the underlying network architecture is that of a synaptic chain. In these cases, it would lead to similar predictions on the distribution of chain lengths and on the elements required for chain formation.

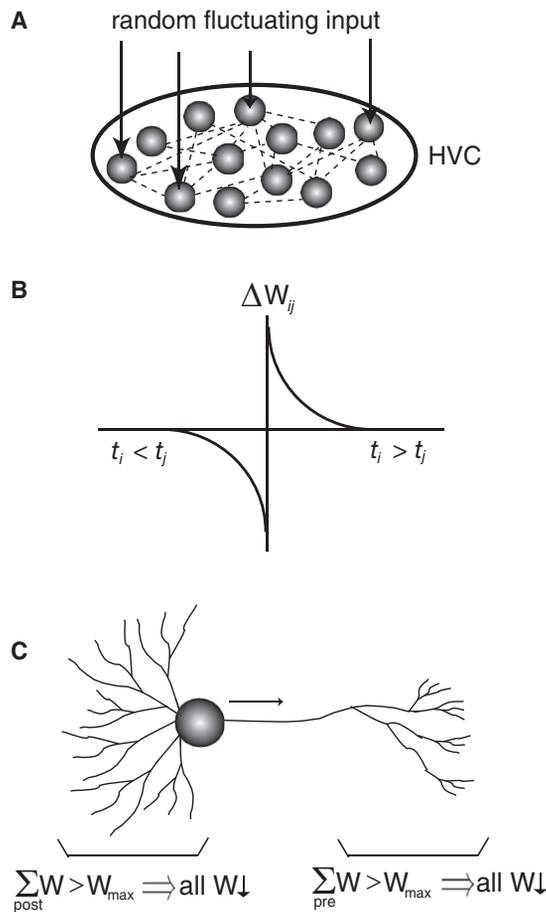
## RESULTS

### The Model

The songbird HVC consists of three cell populations.  $HVC_{RA}$  neurons display unary activity sequences, send recurrent collaterals within HVC, and project downstream to the next nucleus (RA) in the motor pathway. Inhibitory interneurons fire tonically throughout the song motif and project within HVC.  $HVC_X$  cells send outputs to a distinct anterior forebrain pathway that is not necessary for song production in adults.

Our simple network model consists of excitatory neurons with modifiable recurrent synapses (Figure 1A). These represent the  $HVC_{RA}$  neurons. The model includes an inhibitory unit that sums the activity of all excitatory neurons and in turn provides equal global inhibition to all of them. This global inhibitory unit represents the pool of inhibitory interneurons in HVC. We do not include  $HVC_X$  neurons in our model.

The excitatory neurons receive external inputs with temporally random activations (no sequential structure or temporal correlations), except where specifically noted. Initially, the recurrent weights between excitatory neurons are all assumed to be small and random. All weights between the excitatory neurons undergo STDP with an antisymmetric learning window, schematized in Figure 1B. Crucially for the success of sequence formation, in addition to STDP we impose a nonlinear competition across synapses at each neuron, by imposing heterosynaptic long-term depression (hLTD) when the weights at a neuron hit a limit (Figure 1C). The rule is summarized by the summed-weight limit rule.



**Figure 1. Schematic of Model Network and Synaptic Plasticity Hypothesis**

(A) A naive network has initially weak but plastic weights (dashed lines) between all pairs of excitatory neurons. These neurons receive random drive from higher-level areas. The excitatory neurons represent RA-projecting HVC neurons in the song system of songbirds. Global inhibition, proportional to the summed activation of the excitatory neurons, is also present (inhibitory interneurons not shown). The higher-level drive to HVC may arrive from the nucleus interface of the nidopallium (Nif) or the thalamic nucleus uvaformis (Uva).

(B) Neural activity leads to long-term strengthening and weakening of the plastic weights through an antisymmetric spike-time-dependent plasticity (STDP) rule, as depicted here in continuous time. Discrete-time simulations use a discrete-time version of this rule.

(C) Synapses at each neuron are subject to a “soft” limit on their total strength: when the summed weight of synapses into (or out of) a neuron exceeds a limit, all the incoming (or outgoing) synapses to that neuron undergo a slight heterosynaptic long-term depression (hLTD).

### Summed-Weight Limit Rule

If an outgoing (incoming) synapse at neuron  $i$  undergoes long-term potentiation (LTP) and the sum of all outgoing (incoming) synaptic weights at the neuron exceeds a limit  $W_{\text{max}}$ , then all outgoing (incoming) weights at that neuron are slightly reduced.

This competitive rule enforces a “soft bound” on the total outgoing and total incoming weights at any neuron. The bound is called soft because the rule does not explicitly force the total weight of a neuron’s synapses to stay below the specified limit

$W_{\text{max}}$ . Instead, it penalizes all weights by causing an equal amount of hLTD in all synapses when such a bound is crossed. Individual synaptic weights are nonnegative and allowed to vary within the interval  $[0, W_{\text{max}}]$ , where  $w_{\text{max}}$  is a hard bound (smaller than or equal to  $W_{\text{max}}$ ) on the maximum strength of each synapse. More detail is supplied in [Experimental Procedures](#).

### Sequence Formation in a Simple Neuron Network

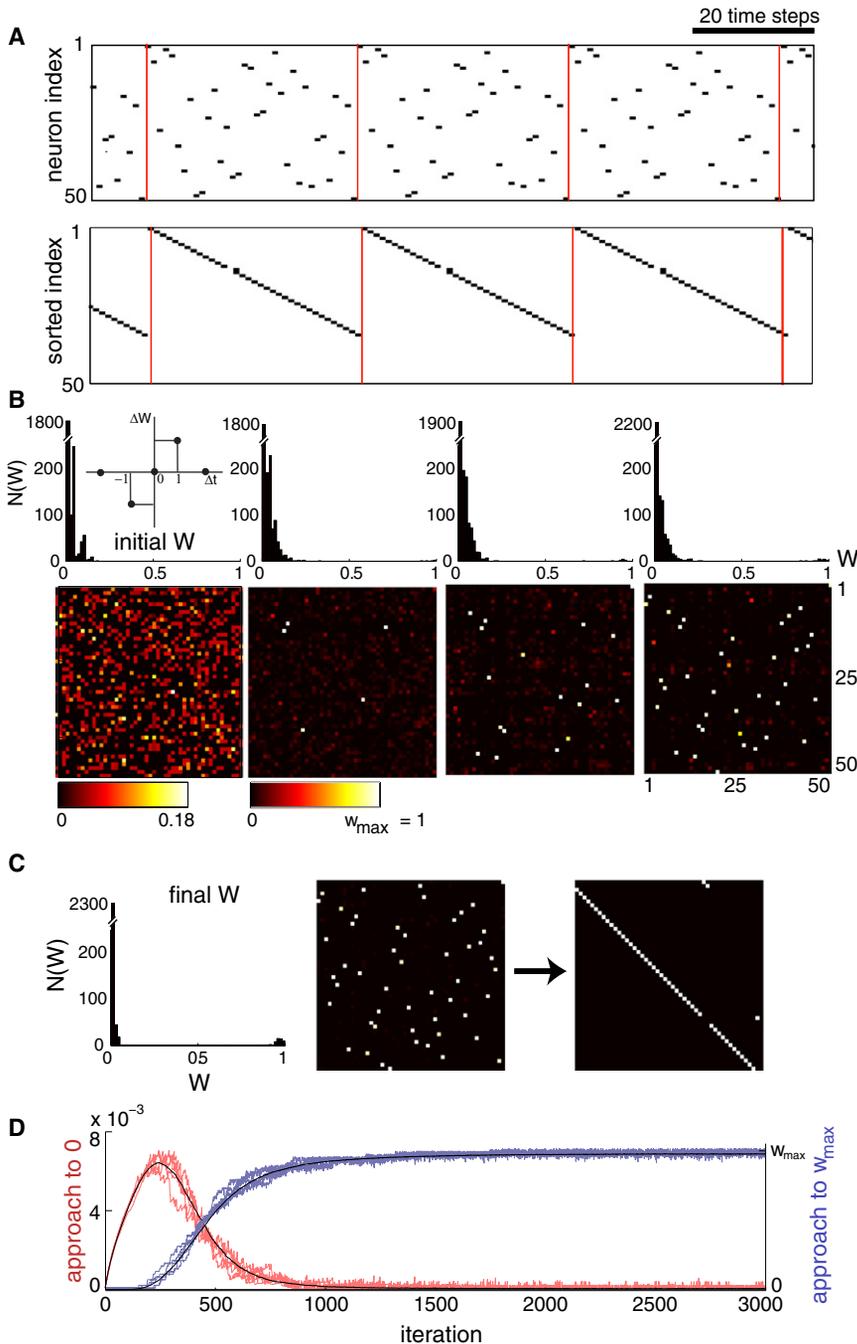
We first consider the case of binary neurons simulated in discrete time, where the states 1 or 0 mean the neuron is bursting or quiescent, and where one simulated time-step corresponds to a duration of 6 ms (the duration of a burst in HVC<sub>RA</sub> neurons; see [Experimental Procedures](#) for details) (Figure 2). The maximum allowed strength of individual synapses is equal to the maximum allowed summed strength,  $w_{\text{max}} = W_{\text{max}}$ . Neurons are driven by random external inputs that are uncorrelated across neurons and time. Under the learning rule, the weights in the connectivity matrix evolve from small random initial values and robustly converge to a steady state (Figure 2B). The steady state of the weight matrix is such that after removal of the ongoing random external input and ignition of neural activity by an input barrage to a random subset of neurons, stereotyped and nondecaying neural activity sequences are observed. In the particular run illustrated in Figure 2, the network can support the propagation of two distinct sequences, only one of which is shown (Figure 2A). The activity sequence is self-propagating: the HVC network requires no external inputs to sustain activity. Each HVC<sub>RA</sub> neuron participates in exactly one chain or sequence, and within that sequence is active exactly once. Thus, activity in this network is unary.

That the resulting neural activity sequences must be unary could have been inferred directly from the converged connectivity matrix. The matrix evolves from a random initial state into a matrix that has exactly one nonzero element per row and per column (Figure 2B, left). The value of the nonzero element in all cases is  $w_{\text{max}} = W_{\text{max}}$ . Such matrices are called permutation matrices and have very special properties: a permutation matrix, applied to a vector (e.g., the state vector of neural activities) simply rearranges the vector components. If the vector consists of one active neuron, the permutation matrix shifts the register so that the next neuron in line according to the permutation ordering becomes active; in the next time-step, the matrix again shifts the index of the active neuron, and so on. Thus, the resulting neural activity is chain-like, with one active neuron per time-step. By rearranging the rows and columns of the weight matrix, we can see that the final network topology is a set of disjoint synaptic chains of different lengths (Figure 2C, right). The time course of learning is shown in Figure 2D.

These results are fairly generic: they do not require parameter values to be fine-tuned (Figure S1 and Matlab code in the [Supplemental Information](#)).

### Playback of Formed Sequences

To induce activity playback in the formed network, as in Figure 2A, the network is given a random barrage of input. If global inhibition in the network is weak, multiple chains may be simultaneously activated. But if global inhibition is sufficiently strong, only one chain will remain activated. The activated



**Figure 2. Evolution of Network Connectivity to Produce Long Chains**

(A) A 50 neuron network has organized to produce neural activity sequences of length  $\sim 35$ . Red lines are visual guides highlighting the period of the activity loop.

(B) Evolution of the network connectivity matrix during learning. (Left) An initially random matrix with dense but weak connectivity evolves to produce a few strong synapses. The inset depicts the STDP window, of two burst widths or time-steps, used in these simulations. (Right) A single winner weight emerges per row and column.

(C) The converged weight matrix is actually a permutation matrix (shuffled identity matrix): re-sorting the matrix makes the chain structure of the connectivity apparent (here there are two chains).

(D) Learning curves. The distance to  $w_{max}$  of weights that eventually reach  $w_{max}$  decreases, and all weights converge to their steady-state values within 1000 iterations (blue curves for several simulation runs, scale on right; each curve is the average of all weights that end up close to  $w_{max}$  from one run; black is the average over all runs). The distance to 0 of all weights that eventually go to 0 initially grows due to random strengthening of weights by STDP, but then steadily shrinks because of synaptic competition, and converges to zero (red curves for several simulation runs, scale on left; black is the average).

approaches or exceeds  $N/2$ , half as long as the longest possible chain that the network could produce by stringing all neurons end-to-end into a single chain (Figure 2). The average sequence length is comparable to the size of the network,  $\langle L \rangle \sim N$ . Since no time-constant in the simulation exceeds one time-step, the average chain is far longer than any neural or synaptic time-constant. Thus, the formation of long chains is an emergent property of the network-level learning dynamics and does not require long cellular time-constants.

Given that the lengths of the formed chains are not determined by intrinsic time-constants, what governs their distribution?

sequence repeats in a loop, unless another input induces switching to a different chain.

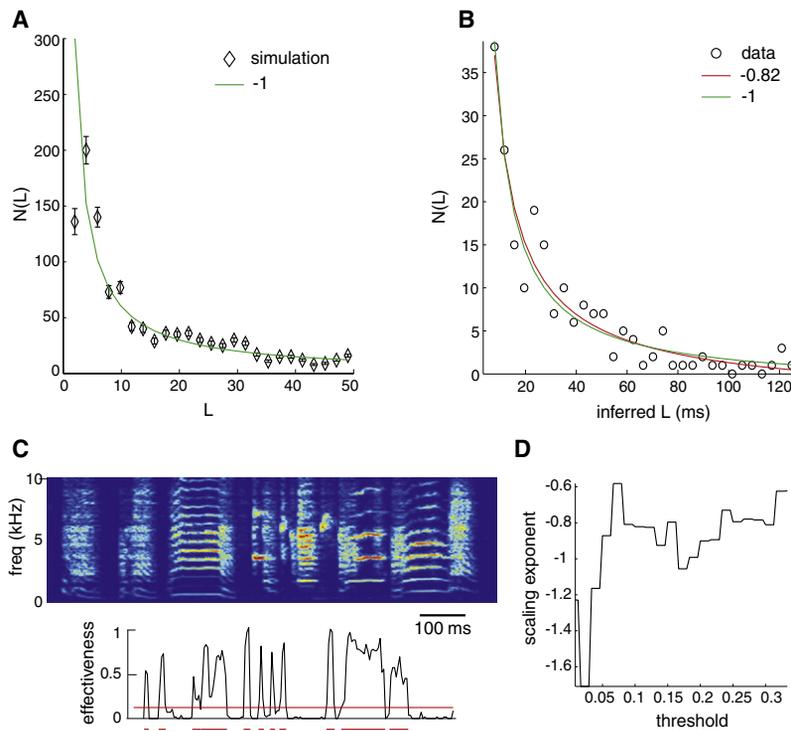
The control of sequence ordering (activation order of different chains) during playback would require external executive inputs (Hahnloser and Fee, 2007), which in turn may be triggered by auditory feedback or internal timing cues.

### The Distribution of Chain Lengths Is Scale Free

In different runs, the network weights converge to different permutation matrices. Typically, each permutation matrix contains multiple sequences, including a sequence whose length

approaches or exceeds  $N/2$ , half as long as the longest possible chain that the network could produce by stringing all neurons end-to-end into a single chain (Figure 2). The average sequence length is comparable to the size of the network,  $\langle L \rangle \sim N$ . Since no time-constant in the simulation exceeds one time-step, the average chain is far longer than any neural or synaptic time-constant. Thus, the formation of long chains is an emergent property of the network-level learning dynamics and does not require long cellular time-constants.

Given that the lengths of the formed chains are not determined by intrinsic time-constants, what governs their distribution? If we assume that the learning process randomly generates, with equal probability, any permutation matrix from the set of all possible permutation matrices of  $N$  elements, we can calculate the probability  $P(L)$  of finding a chain of length  $L$ , and find:  $P(L) = c/L$  for  $L \leq L_{max} = N$ , and  $P(L) = 0$  otherwise (see Supplemental Information).  $c$  is a normalization constant.



**Figure 3. Distribution of Formed Sequence Lengths: Model and Zebra Finch Data**

(A) The probability of a sequence of length  $L$ . Black diamonds: simulation data from 300 trials with  $n = 50$  neurons. Green curve: theoretical prediction of  $1/L$ . Binomial error bars as shown.

(B) Distribution of inferred HVC<sub>RA</sub> chain length from HVC stimulation experiments (right and left HVC stimulation, effectiveness threshold = 12%). The best power-law fit has exponent  $-0.82$  (red curve). A fit with scaling exponent  $-1$  is shown for comparison (green curve).

(C) Example HVC stimulation and how the chain lengths are inferred. The spectrogram of a zebra finch song motif is shown on top (blue to red: from low to high sound amplitudes). The effectiveness of electrical stimulation (0.2 ms, 500  $\mu$ A biphasic current pulse) is highly variable as a function of stimulation time in ongoing song (bottom, black line). HVC<sub>RA</sub> chain lengths (horizontal red rasters, bottom) are inferred from periods of suprathreshold effectiveness (threshold of 12%, red horizontal line).

(D) Scaling exponent of the optimal fit as a function of the effectiveness threshold. For a large range of thresholds, the scaling exponent is in the vicinity of  $-1$ .

$\sim 69\%$ . The probability that a run will produce a chain longer than  $0.6N$  is just over 50%. The  $c/L$  distribution of chain lengths is called “scale free” because it is a power law, which, unlike exponential distributions with a characteristic decay constant, has no inherent scale.

### The Distribution of Premotor Sequences in Zebra Finches

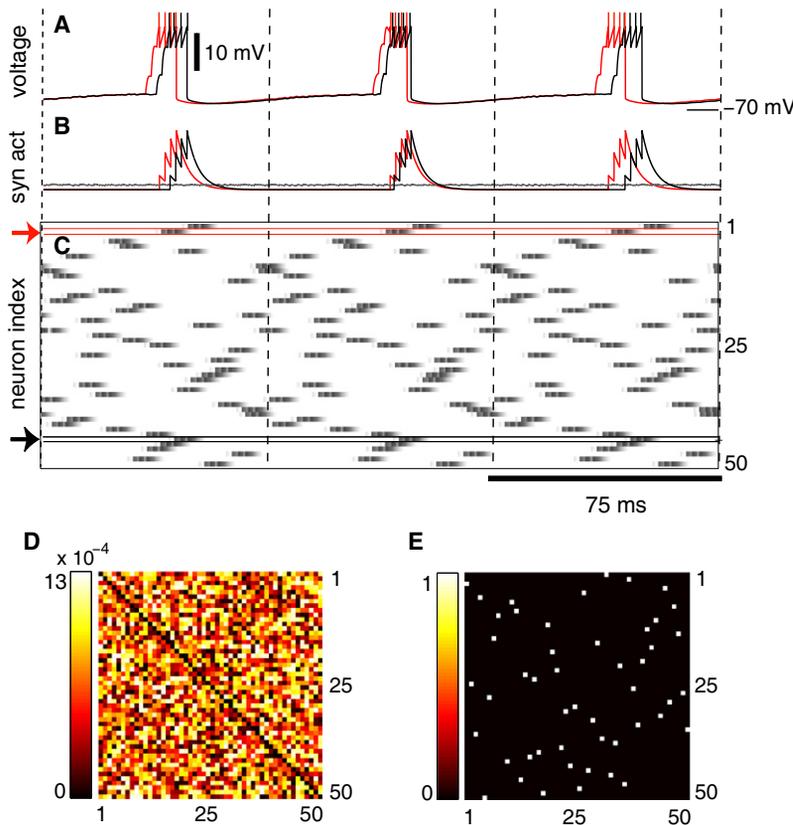
We now compare the chain-length distribution in our model with the distribution of sequence elements in zebra finch song nuclei, as inferred from electrical stimulation. It has long been known that brief electrical perturbation of HVC can disrupt ongoing songs, with effects ranging from brief syllable distortions to halting of the song. How effective a particular stimulation is can be quantified by the fraction of time bins following stimulation in which song amplitudes deviate significantly from unstimulated catch trials. For fixed-amplitude HVC stimulation, the effectiveness with which songs are distorted was found to vary dramatically over the time course of a song motif, with highly effective and completely ineffective stimulation times following each other in short intervals (Figure 3) (Wang et al., 2008). Furthermore, stimulation effectiveness tends to be complementary across the two cerebral hemispheres, such that a stimulation effective at disrupting song over some time interval in one hemisphere tends to be ineffective in the other hemisphere over the same interval.

These data imply that there are blocks of time when one hemisphere is involved in sequence production (effective interval) and that the hemispheres take turns in driving different parts of the song. We interpret one effective interval in one hemisphere as the playing out of one synaptic chain, with several distinct chains formed by the HVC<sub>RA</sub> neuron populations in each hemisphere.

We analyze the stimulation data from  $n = 19$  birds subjected to random HVC stimulation during singing, as in Wang et al. (2008). We infer HVC<sub>RA</sub> chain lengths from the time intervals between consecutive threshold crossings of stimulation effectiveness curves (first crossing from below and second crossing from above). The inferred distribution of chain lengths is reasonably well fit by a power law (Figure 3B). For a wide range of effectiveness thresholds, we find that the scaling exponent of the best fit is close to  $-1$  (Figure 3D). Hence, we find support for a  $1/L$  chain-length distribution in experimental data.

### Numerical Experiments in Networks of Conductance-Based Spiking Neurons

Our results so far rest on networks of highly simplified binary neurons lacking any temporal dynamics and with time discretized into 6 ms chunks that contain the entire burst of a single neuron. Different neurons firing single bursts in song can have either perfect or zero activity correlation: partial correlations are impossible. The STDP time window was also narrow, spanning only two burst durations (Figure 2B). To test whether sequence formation also holds with more realistic continuous-time neuron models and wider STDP windows, we performed numerical experiments in which we applied the learning rule to a network of spiking neurons (Figure 4). The neurons were intrinsic bursters with a membrane time constant of 25 ms and interacted through conductances. Plasticity was governed by an antisymmetric STDP function (Figure 4D; see Experimental Procedures) of 20 ms width. With time discretized more finely than burst lengths, neurons that each fire a single burst have graded activity overlap durations, potentially complicating the chaining problem. We assume no synaptic delays or slow rise times that could help to stabilize asynchronous (sequential) neural activity patterns.



**Figure 4. Numerical Experiment with Conductance-Based Spiking Neurons**

Like the binary neuron discrete-time networks, a network of conductance-based bursting neurons in continuous time organizes under STDP and the summed-weight limit rule and random feedforward input, to produce long unary sequences. The vertical dashed black lines (in A and B) are a guide, highlighting the periodic repetition of the neural activity sequence. The voltage (A) and synaptic activation (B) traces of two sample integrate-and-burst HVC neurons in the network (black and red) during a 75 ms interval. (C) Synaptic activations of all 50 neurons during the sequence, and the initial (D) and converged (E) connectivity matrices for the network.

gation of *wide* unary chains of neural activity. First, the maximum summed synaptic weight into (out of) each neuron should exceed the maximum weight of individual synapses:  $W_{max} = mw_{max}$ , with  $m > 1$ . Second, groups of neurons should receive correlated external input. The first condition allows single neurons to send and receive more than one strong synapse. The second condition, which corresponds to each external neuron driving more than one HVC neuron, allows neurons to become recurrently coupled into groups, which then connect with each other to form a wide synaptic chain.

Despite these multiple steps toward biological realism, connectivity still evolves into a permutation matrix that supports the propagation of long, unary activity chains (Figure 4). Sequence formation is more robust when STDP updates represent fractional changes in synaptic strength (multiplicative STDP) rather than absolute changes: multiplicative STDP, combined with a uniform subtractive hLTD across synapses in response to the summed-weight limit, allows large weights to grow sufficiently rapidly and intensifies the competition across weights. Aside from this general scaling prescription, our networks did not require tuning of neural thresholds or levels of global inhibition or fine-tuning of other neural or network parameters. Hence, unary sequences can robustly form in networks of conductance-based neurons, if the total synapse strength at each neuron is limited and the limits are enforced through cross-synaptic competition.

### Formation of Wide Chains

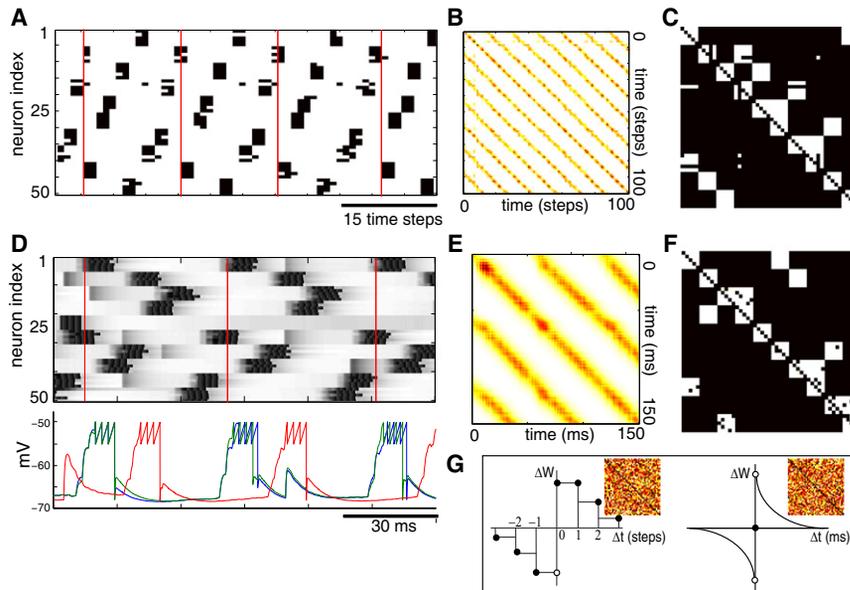
For expositional convenience, we illustrated the formation of long chains of width one. However, chains of width one are not robust, in the sense that deletion of a neuron or failures of neural or synaptic activation would break the chain of activity. This property is biologically implausible, and a width one chain is inconsistent with estimates of several ( $\sim 200$  out of  $\sim 2 \times 10^4$ ) coactive HVC<sub>RA</sub> neurons at any time in song (Hahnloser et al., 2002).

If we generalize two ingredients of the model above, a network with initially random lateral connectivity will spontaneously organize into a synaptic chain organization that supports the propa-

An *input group* is a set of all HVC neurons that receive a common external input. Input groups can be overlapping: if two external neurons drive one HVC neuron, it will belong to two groups. We first assume disjoint input groups. (Below, we consider overlapping input groups.) Each external input projects to  $k$  HVC neurons, and each HVC neuron receives a single external input. Such organization could result if the external inputs to HVC have spatially segregated arborization patterns or if external inputs compete to drive their HVC targets, with a single winner per HVC neuron (e.g., Sanes and Lichtman, 1999; Hashimoto et al., 2009).

Wide chains endow noise tolerance, so we modeled the dynamics as stochastic: individual neurons and synapses independently and probabilistically respond given above-threshold input or presynaptic firing, respectively. The external inputs are temporally random with no sequential structure. With sufficiently large  $m$ , a network with weak, random, all-to-all connectivity and disjoint input groups organizes into a structure that generates unary sequences of essentially uniform width  $q \leq k$  (i.e.,  $q \leq k$  neurons are active per time). We illustrate this result in networks of binary neurons (Figures 5A–5C and 5G) and conductance-coupled integrate-and-burst neurons (Figures 5D–5G). The chains are long in the sense that the longest of the formed chains scales like  $\sim N/q$ .

The resulting weight matrices look superficially quite different from the permutation matrices formed for chains of width one (data not shown). Each row and column contains multiple nonzero entries, meaning that each neuron receives input from



**Figure 5. Formation of Wide Chains**

The total summed-weight limit is larger than the single-synapse maximum ( $W_{max} = mW_{max}$ , with  $m > 1$ ). Groups of  $k$  HVC neurons receive correlated external input (see text).

(A) Binary neuron networks stably generate wide activity chains after learning is complete:  $k = 5$ , and  $k$  neurons are active per time-step, excepting occasional failures to fire due to variability in neural and synaptic responses. The sequence (with  $n = 50$  neurons) is 20 steps long, with individual neurons active only for two consecutive time-steps.

(B) The temporal cross-correlation of network activity illustrates sequence stability: the sequence replays without fading.

(C) The resulting weight matrix has multiple nonzero entries per row and column, in blocks at the diagonal and at off-diagonal locations. A diagonal block represents strong connectivity between neurons within a formed group, causing coordinated firing even if the external input is not tightly coordinated. The off-diagonal blocks form a permutation matrix, if each block is viewed as a single element.

(D) Integrate-and-burst conductance-coupled neuron networks produce wide activity chains after learning: intensity plot of the voltages of all 50 neurons (top). Black lines are spikes. Below: subthreshold voltages of neurons 4, 5 (blue and green, group 1), and 6 (red, group 2).

(E) Temporal cross-correlation of network response.

(F) The final weight matrix: there is symmetric connectivity within groups (diagonal blocks), and permutation matrix connectivity between groups (off-diagonal blocks), demonstrating that the network has formed synaptic chains.

(G) The STDP windows used for the binary (left) and integrate-and-burst (right) simulations, and in insets, the initial weights for the two networks. Open (closed) circles: the open (closed) interval. These simulations include unreliable (probabilistic) responses (see [Experimental Procedures](#)).

and sends output to multiple neurons. But sorting the matrix reveals an orderly underlying structure ([Figures 5C and 5F](#)). If coactive neurons during playback are assigned consecutive indices, the weights form diagonal blocks, together with blocks of off-diagonal nonzero entries ([Figure 5B](#)) (center and right). Each diagonal block represents a highly interconnected *formed group*. When the input groups are disjoint, the formed groups are essentially the same as the input groups. When input groups are overlapping, the formed groups may be quite different from the input groups (below). The off-diagonal blocks form a block permutation matrix, corresponding to projections from one formed group to the next. Thus, the learning rule results in a synaptic chain organization, from an initially random network topology.

If the formed groups contain  $q$  neurons each, then each neuron typically has  $2q - 1$  incoming and as many outgoing connections:  $q$  inputs from another group and  $q - 1$  inputs from within the group. As mentioned, the formed groups may be smaller than the input groups ( $q \leq k$ ), if  $m < 2k - 1$  and hLTD is strong (data not shown). In this case, there are  $k - q$  unused neurons per input group during playback. The advantage is that  $m$  and  $k$  need not be exactly tuned relative to each other (e.g., synaptic chains form even if  $m \neq 2k - 1$ ).

The probability of finding a chain of length  $L \leq L_{max}$  is  $c/L$ , as before, but with  $L_{max} = N/q$ , and  $c$  defined accordingly (not shown). The  $q$ -wide activity chains are robust to neuron or synapse deletion when  $q$  is large. For example, if the summed weight  $W_{max}$  is  $p$  times larger than necessary to produce activity in a postsynaptic neuron, then  $q/p$  synapses from one group to

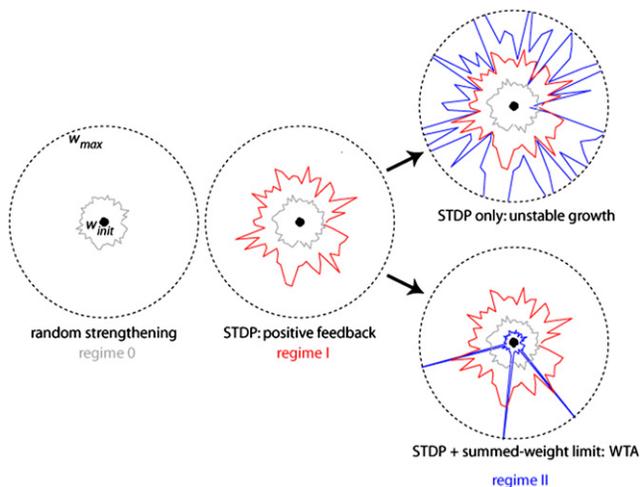
the next can be severed, without failure of sequential activity propagation. For similar reasons, during learning and in playback mode the network can tolerate unfaithful neural and synaptic responses.

In the brain, as in our model, neurons from the same group need not be spatially localized, so a wide-chain network need not display spatial clusters of coactive neurons. Neurons of a group may be spatially distributed, and only distinguishable as members of a group by their correlated activation times and through the shared synaptic inputs from within the group and from the external inputs.

### Random Input Groupings

In our model, it is possible to observe stable chain formation even if the input groups are not disjoint. We let each external input neuron randomly select HVC neurons as targets, without excluding HVC neurons already selected by different external inputs. A total of  $N/k$  (or the nearest rounded-down integer) external input neurons innervated  $k$  HVC neurons each, in this random manner. Thus, individual HVC neurons frequently belonged to multiple input groups. Despite the overlapping input group structure, if  $m$  is large enough, the network connectivity organizes into largely disjoint formed groups that are connected sequentially ([Figure S2](#)).

Sequence playback is unary: each neuron is active at only one part of the sequence. The learning process orthogonalizes the formed groups, as well as stringing formed groups into chains. This is visible in the approximate block structure of the resulting weight matrix, in contrast with the more overlapping structure of



**Figure 6. How Unary Sequences Form with STDP and a Limit on the Summed Synaptic Weights at Each Neuron**

All outgoing weights from neuron  $i$  are arranged radially (schematic). Initially (left), all weights are small and do not contribute to neural firing; weight changes are due to externally driven random activity (regime 0). In regime I (center, red), the weights contribute to lateral excitation within the network, even though the summed weight has not exceeded  $W_{max}$  (so the summed-weight limit has no impact) and no individual weights have saturated at  $w_{max}$ . STDP strengthens strong weights. In regime II (right, blue) the summed weight has reached  $W_{max}$ . STDP continues to amplify large weights. With STDP alone (top), many weights may reach  $w_{max}$ , contributing to runaway network activity (cf. Figure S3). But hLTD combined with STDP produces a winner-take-all competition in the weights, so only a few weights can win and the rest are driven to zero (illustrated are three winning weights, which would result for  $m = 3$  wide chains).

the input groups (Figure S2). Yet despite the difference between formed groups and input groups, some of the input group structure is preserved in the formed groups: two neurons from the same formed group are more likely to receive the same external inputs than are two neurons from different formed groups. The key qualitative difference compared to the disjoint input group case is that the formed groups are of varying size, causing the activity chain to vary in width along the sequence.

Random input groups result in an inefficient use of neurons, even though chain formation remains possible. If  $N$  neurons are randomly selected for external drive (with replacement) from a bag of neurons labeled 1 to  $N$ , then typically more than a third ( $1/e$ ) of the neurons will never be selected. These neurons will not receive external drive or be included in any chain. If more than  $N$  neurons are selected, a larger fraction of neurons receives some external input, but the overlap in input group membership grows, causing problems: the network either fails to form chains or forms only short chains (data not shown). If fewer than  $N$  elements are selected, the overlap between groups shrinks, and chain formation becomes more robust (data not shown); in the limit where the selected number is far smaller than the total number of HVC neurons, the random input groups will be statistically disjoint, and all previous results from disjoint groups hold. The cost, however, is  $>1/e$  unused neurons.

These results suggest two possibilities. First, the brain may construct nonrandom disjoint input groups, either by physical

or chemical topography in the external inputs to HVC or by competitive target innervation (Sanes and Lichtman, 1999; Hashimoto et al., 2009). Second, innervation may be sparse and random, producing disjoint groups and low neural usage, with elimination or subsequent use of the unused neurons. Synaptic maturation for neurons already in chains and continued plasticity for unused neurons could allow neural recruitment for repair or novel sequence acquisition, without disrupting existing sequences.

### Mechanism of Chain Formation

How do the two processes, STDP and a competitive constraint on the summed weight of synapses per neuron, interact to produce synaptic chains, when STDP alone fails (Figure S3)?

Again, we first focus on the  $m = 1$  case. (The wide-chain case is similar, and we address it next.) The learning rule converts an initially random matrix into a (block) permutation matrix in which all rows and columns each contain exactly one nonzero element (or block of elements in the case of wide chains, after removing the diagonal blocks). From this matrix structure, we can infer that the rule effectively drives a simultaneous winner-takes-all (WTA) competition across synapses within each row and within each column of the initial weight matrix.

Consider a single neuron  $i$ . Initially (Figure 6, regime 0) the weights  $W_{ji}$  from neuron  $i$  to all other neurons  $j$  in the network are close to zero, far from any weight bounds and too small to drive neural activity. Activity is driven by random external inputs, and STDP produces a statistically uniform strengthening of all weights, with small differences reflecting random activation histories. Next, when weights are large enough to contribute to neural activation (Figure 6, regime I) but are still below any individual or summed limits, then STDP acts as a positive feedback process: larger weights increase the chance of post-synaptic spiking in response to presynaptic activity and are further strengthened by STDP. Thus, in regime I, STDP amplifies the small randomly induced weight differentials from regime 0.

In regime II (Figure 6), individual weights  $W_{ji}$  are still typically smaller than  $w_{max}$ , but the summed-weight limit of  $W_{max}$  has been hit. This limit drives uniform hLTD in all outgoing synapses from  $i$  whenever any outgoing synapse undergoes LTP. Uniform subtraction more strongly penalizes smaller weights (as a percentage of their values), while STDP tends to selectively strengthen strong weights, amplifying weight discrepancies. These are the conditions of WTA dynamics, across all the outgoing synapses at neuron  $i$ . The same is true at each neuron and across incoming synapses.

The  $1/L$  likelihood of finding a chain of length  $L$  (Figure 3A) corresponds to the expected distribution of chain lengths if the final weight matrix were assumed drawn at random from the set of all possible permutation matrices. This suggests that, remarkably, the WTA dynamics drives *without bias* the formation of any possible permutation matrix (e.g., not favoring permutation matrices with specific chain lengths). In other words, chain formation does not take into account the lengths of existing chains when adding elements to them: the choice of a winner among eligible elements in an empty row is random.

For wide chains,  $W_{max} = mw_{max}$ , so  $m$  incoming (outgoing) synapses per neuron can reach  $w_{max}$ , while pushing the remaining synapses to 0 (group WTA). Thus, there are  $m$  nonzero entries per row and column of the weight matrix. The question is how these multiple nonzero entries are coordinated to be in block form. In other words, neurons form into recurrent groups, and essentially every neuron in one group projects to every neuron in the next group. How is the synaptic fan-in and fan-out coordinated across neurons in a group? It results from the correlated activity of neurons within each group. The within-group correlations are initially due to correlated external inputs, but are greatly enhanced by the formation of recurrent within-group connections through STDP. Now suppose neuron a1 in group A projects to neuron b1 in group B, contributing to its firing. But if a1 is active, then all neurons in A are likely active. Thus, if a1 tends to fire before b1, most neurons in A will tend to fire before b1, causing synapses from all A to b1 to be strengthened. Similarly, if b1 fired, all neurons in B are likely firing. Thus, if a1 contributes to activity in b1, it also does so for all B. This causes synapses from a1 to all B to be strengthened.

#### Network Formation with Sequential Training Inputs

If the inputs to the network are spatiotemporally patterned, in the form of a sequence of length  $T$ , the network connectivity still evolves to support the propagation of neural activity sequences (Figure S4). In fact, learning is extremely rapid and can be complete in as few as 30 to 40 presentations of the input sequence. This may help explain observations that some songbirds can acquire an internal song template after hearing their tutor song only a few times.

The learning rule converts a dense (non-unary) input sequence into a unary sequence: the active neurons at any part of the formed sequence are typically a subset of the neurons driven at that point by the training sequence. Different runs produce a formed unary sequence that is a different sparse version of the same dense training sequence (Figures S4C and S4D).

Finally, the formed chain has the same length (number of steps) as the input sequence, if the input sequence is shorter than the maximum length the network can support (the number of neurons in the network divided by formed chain width). Thus, when the input is sequential, it can alter or override the scale-free distribution of chain lengths found when the input is random.

#### Varying the Neural Constraint on Synapses: Weight-Growth Limit

To illustrate that the outcome can depend strongly on the form of the trigger for hLTD, we briefly consider a different neuron-level constraint on synapses. Suppose limits are placed not on the total resources for synaptic building, as above, but on their rates of production. We call this constraint the “weight-growth limit.” As before, synapses undergo STDP. Unlike the summed-weight limit rule, hLTD is not triggered when the summed weight exceeds a threshold. Rather, whenever an incoming (outgoing) synapse undergoes LTP, all other incoming (outgoing) synapses at that neuron undergo slight hLTD (see Experimental Procedures).

This rule also generates synaptic chain connectivity to support the stable propagation of long neural activity sequences

(Figure 7A). However, each network produces a single sequence or chain (Figures 7A and 7B). The chain is unary but wide, with multiple coactive neurons per time. Sequences appear in  $<1/10^{\text{th}}$  the time taken by the summed-weight limit, with temporally random input.

There are other important differences. The weight-growth limit with random inputs generates sequences only if neurons have a slow adaptation time-constant (Experimental Procedures). However, the weight-growth limit readily produces wide chains without the help of correlated input groups. Unlike the summed-weight rule, the weight matrix with the weight-growth limit rule never converges to a steady state while the network continues to receive full-amplitude external random inputs and the learning rate remains nonzero. The weights continue to evolve, with formed chains morphing into different ones. To obtain stable sequences, the external inputs to the network must be annealed to zero (Experimental Procedures).

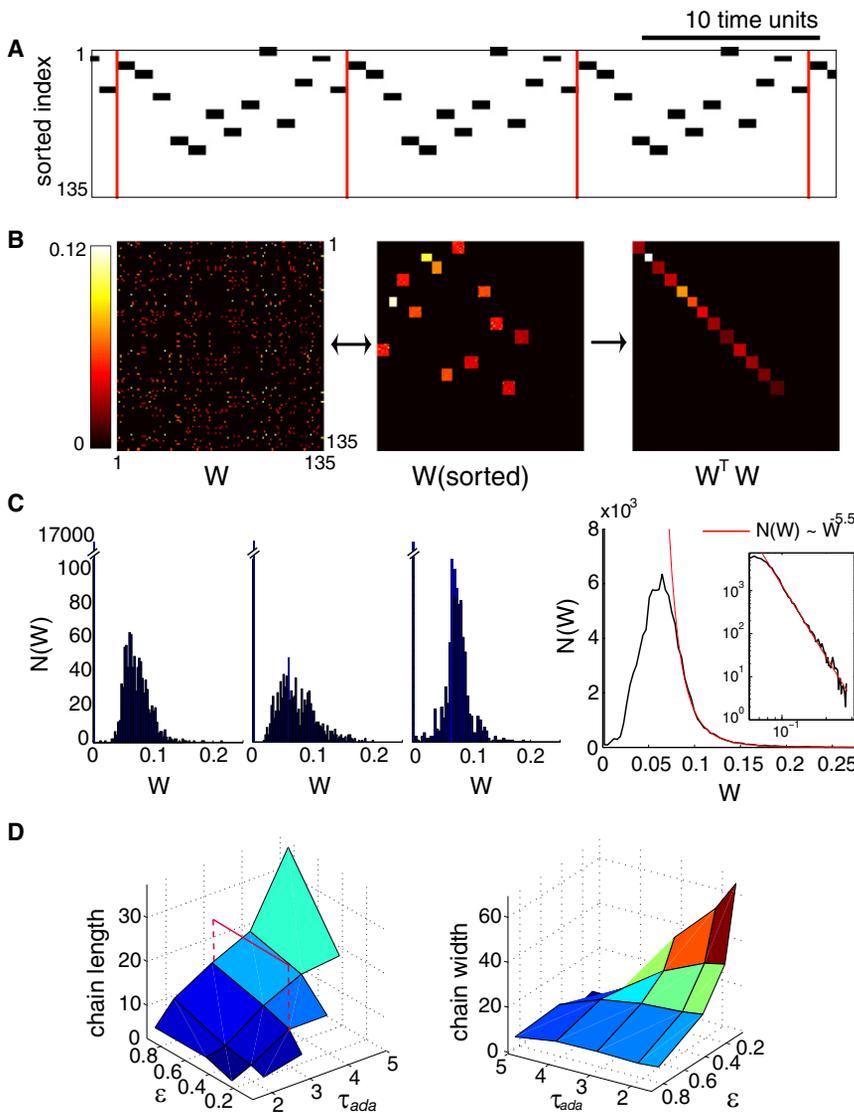
The resulting weight matrix looks different from that obtained with the summed-weight limit (Figure 7B). Each row and column of the weight matrix contains several nonzero entries of varying magnitudes. Yet the underlying structure is familiar: the sorted matrix reveals recurrently connected neuron groups (Figure 7B, center, right), with groups connected in a permutation matrix structure. However, the blocks are of different sizes, meaning that the neural activity chains are not of constant width. Unlike in the summed-weight rule, where synapse strengths become essentially binary (Figures 2C, 4E, 5C, and 5F), the distribution of nonzero weights is wide, with a power-law (heavy) tail (Figure 7C).

In multiple runs with random inputs and fixed parameters, the formed chain length is exactly the same and is governed by the adaptation time-constant (Figure 7D).

The network can be trained to produce sequences longer than the adaptation time-constant if the input is sequential and longer than the adaptation time (data not shown). Thus, the cellular time-constant governs chain length if the inputs are random, but plays little role in restricting chain length if the inputs are sequential. This ability to produce a long sequence in response to an equally long sequential training input is shared by the summed-weight limit rule, as described above (and seen in Figure S4). However, it is an especially important feature for the weight-growth rule, for which the maximum chain length is otherwise strictly limited by the cellular adaptation time.

#### DISCUSSION

Models suggest that the propagation of long neural activity sequences requires very special network connectivities (Amari, 1972; Abeles, 1991; Drew and Abbott, 2003; Li and Greenside, 2006; Jin et al., 2007). We have demonstrated that plausible constraints on synapse strength combine with STDP to spontaneously produce such special connectivity matrices, even without temporally patterned training inputs. Our results demonstrate that competitive constraints on synapses at the level of individual neurons can perform the global computation involved in evenly distributing activity across the network and across time so that the code is sequential and unary. The network produces



**Figure 7. Sensitivity to the Functional Form of Weight Limits: A Limit on Weight Growth Produces Long Wide Chains If a Long Cellular Time Constant Is Present**

(A) A 135 neuron network has self-organized to produce a 13 step long neural activity sequence, with seven to eight neurons active per time for a participation of 100 neurons in the sequence. Red lines highlight the periodicity of the replaying activity loop.

(B) The converged weights appear disordered (left). On rearrangement, the matrix is clearly of block form (center) and the blocks form a permutation matrix over the active neurons (right).

(C) (Left) Three examples of the distribution of synaptic strengths for an annealed network that produces stable sequences. (Right) The cumulative histogram pooling together many separate runs has a long power-law tail.

(D) (Left) The length of the formed sequence depends weakly on the strength of weight subtraction  $\epsilon$ , and strongly on the cellular time-constant of adaptation  $\tau_{ada}$ . (Right) The width of the formed sequence (number of coactive neurons) depends strongly on  $\epsilon$  and more weakly on  $\tau_{ada}$ .

**Sensitivity to the Form of the Learning Rule**

Varying the trigger for hLTD can profoundly affect network connectivity and the resulting activity sequences. In the summed-weight limit rule, cellular time-constants proved to be unimportant, multiple chains formed simultaneously, synaptic weights were relatively strong, and the distribution of chain lengths was scale free. In the weight-growth limit rule, the network formed a single chain, synaptic strengths were widely distributed, and a slow cellular

time-constant was essential (if the training input was nonsequential). Because the requirements and results of these forms of heterosynaptic competition differ greatly, it is possible to distinguish which form, if either, exists in a sequence-forming area. Below, we suggest some tests of our model.

**Insensitivity to Details of Neural and Synaptic Dynamics**

The main ingredients used in our model are intrinsic neural bursting, STDP, hLTD triggered by a constraint on summed weights at each neuron, and some group structure in the initial network whereby neurons within a group received similar external input. These ingredients produce synaptic chain networks, even when the dynamics of the neurons and synapses are varied. In this sense, the detailed dynamics of neurons and synapses do not matter for chain formation.

time-constant was essential (if the training input was nonsequential). Because the requirements and results of these forms of heterosynaptic competition differ greatly, it is possible to distinguish which form, if either, exists in a sequence-forming area. Below, we suggest some tests of our model.

**Tests of the Model**

The most basic tests of the model involve verification of its hypotheses: STDP, heterosynaptic competition at each neuron, and correlated external inputs to sets of HVC neurons. STDP, though found across the brain and across species, has not yet been documented in juvenile songbird HVC. The heterosynaptic competition hypothesis, motivated by numerous experiments (discussed below), can be tested in sequence-producing brain areas using similar experimental protocols.

Our predictions of multiple neural subsequences and a power-law length distribution are consistent with the zebra finch data (Wang et al., 2008) analyzed here. However, effective

stimulation duration is an indirect measure of neural sequence length. More direct approaches, based on neural recording and perturbation or behavioral song analysis (Glaze and Troyer, 2006), should help reveal the neural underpinnings of behavioral sequences.

Neurons within a formed group are predicted to receive similar external inputs with higher probability than neurons from separate groups, even if playback is no longer dependent on external input. In songbirds, Uva or Nlf input to coactive HVC neurons should be more correlated than the input to non-coactive HVC neurons.

The main (summed-weight limit) rule does not depend on slow neural adaptation or other slow time-constants. In accord, the autocorrelation traces of individual HVC<sub>RA</sub> neurons in sleeping zebra finches are flat (Hahnloser and Fee, 2006) with no refractoriness. If true in the awake bird, such results would preclude the weight-growth limit rule, which depends on firing adaptation, from explaining sequence formation in zebra finches (or would at least necessitate sequential training input).

Our results suggest that sequential training input may not be necessary for sequence formation. In the juvenile bird, HVC may still organize into synaptic chains if the neurons are randomly but strongly stimulated while all sequential auditory input is blocked. Sequence formation may be slow and result in multiple shorter sequences. With sequential training input, learning could take <50 presentations, with formed chain length given by the length of the input sequence.

### Experimental Evidence of Heterosynaptic Plasticity

Reports of heterosynaptic plasticity abound. Some show that homosynaptic LTP lowers the LTP threshold for synaptic neighbors (Harvey and Svoboda, 2007). Other reports of heterosynaptic plasticity demonstrate hLTD following homosynaptic LTP, closer to the rule hypothesized here for sequence formation. For instance, homosynaptic LTP in the lateral perforant path to the dentate gyrus causes hLTD in the medial perforant synapses to the same neuron (Abraham et al., 1985; Doyère et al., 1997). Heterosynaptic effects are specific to the induction or maintenance of long-term but not short-term potentiation (Abraham et al., 1985). Two studies, one physiological and the other structural, lend more specific support to our hypothesis. Postsynaptic neurons balance activity-dependent LTP of an input synapse by inducing heterosynaptic depression among other input synapses, conserving the total synaptic weight onto the neuron (Royer and Paré, 2003). Similarly, the total synaptic area in postsynaptic hippocampal neurons remains constant after LTP induction, even though the distribution of synapse sizes changes significantly (Bourne and Harris, 2010).

LTP depends on protein synthesis (through transcription and translation) at the cell body (Frey and Morris, 1997). Curtailing synthesis drives competition for these proteins across synapses in a single neuron for the maintenance (but not induction) of LTP (Fonseca et al., 2004, 2006). LTD relies on protein degradation, but also on protein synthesis through translation (Huber et al., 2000; Malenka and Bear, 2004). These studies suggest that transcription is a limiting step, and scarce posttranscriptional proteins may drive heterosynaptic competition.

### Playback

Bengalese finch songs show repeated playback of individual syllables, with stochastic transitions to other syllables, which may be consistent with the existence of loops in the underlying circuit. Even in zebra finch song, which does not usually contain repeated elements, when a syllable is repeated it is stuttered: the repeats happen in quick succession. These observations hint that in some songbirds, song may consist of several loops, with each loop representing a song element (e.g., a syllable) that tends to repeat unless there is an active command (e.g., from higher areas Uva or Nlf) to switch between loops. In non-loopy zebra finch songs, additional plasticity mechanisms may cut each loop into an open sequence, then paste them into a longer open sequence. In this case, sequential playback of the full motif would require less executive control from higher areas.

Sequence playback speed may differ from the speed of a training sequence (Davidson et al., 2009). A training sequence running at any speed consistent with mechanisms for STDP may successfully drive chain formation, but the speed of sequence playback will be generic, largely determined by the strength of the chaining weights, the threshold for spike generation, and the rise-time of synaptic currents.

### Sequence Generation in Mammals

The mammalian hippocampus appears to contain synaptic chains, given the widespread observations of robustly propagating neural activity sequences in behaving, resting, and sleeping animals (Nádasdy et al., 1999; Louie and Wilson, 2001; Pastalkova et al., 2008; Davidson et al., 2009). Muscle trajectories are driven by neural sequences in motor and premotor cortex (Schwartz and Moran, 1999; Andersen et al., 2004). Microstimulation in cortical slices evokes reliable sequential neural activity (Buonomano, 2003). Further, spontaneous in vitro and in vivo activity in primary sensory cortex appears sequential (Pulvermüller and Shtyrov, 2009; Luczak et al., 2007; Ikegaya et al., 2004), suggesting underlying synaptic chains (but some of these spontaneous activity data are also consistent with random activation of uncoupled neurons [Mokeichev et al., 2007; Roxin et al., 2008]). The apparent ubiquity of sequential neural activity suggests the mechanism for chain formation is robust and generic.

The robustness with which sequence-generating circuits arise in our model in the presence of an appropriate form of synaptic competition and STDP and little network infrastructure beyond correlated external inputs to sets of neurons suggests that the learning rule is generic enough to be a candidate model of sequence formation in diverse areas where synaptic chains are found.

### Comparison with Other Work

Buonomano (2005) suggested an alternative approach to sequence learning based on an antiassociative rule and driven by the difference between the time-averaged activity of each neuron and a specified target value for the same. During training, a group of initiator neurons is repeatedly stimulated. The antiassociative rule builds a chain beginning at the initiator group, by minimizing the mismatch between actual and target neural activity. Yet specifying a target firing rate may not always be

the appropriate prescription for learning. For example, if a bird sings the same song frequently at dawn, and less in the afternoon, the neural firing rates will differ greatly at those times, even though the actual and target songs are the same. By contrast, our framework imposes and requires no specific constraints on neural activity. Instead, constraints are given on synaptic weights, which govern the spreading of activity over the network by restricting synaptic fan-out.

A closer approach to ours (Jun and Jin, 2007) involves STDP and synaptic competition that is implemented discretely: once an outgoing synapse from a neuron at the leading edge of a forming chain crosses a threshold strength, it is allowed to grow, while the remaining outgoing synapses are pruned. As we have seen from the variant form of the neural limit on weight growth, subtle differences in synaptic competition produce notable differences in the results. Like Buonomano (2005), chain development in Jun and Jin (2007) is purely serial, starting from an initiator group and resulting in the development of a single chain in the network. This is in contrast to the parallel formation of multiple sequences of different lengths produced by our approach.

### Future Work

We have illustrated two neural limits on plasticity, which lead to very different formed sequences. Combining both limits may lead to interesting hybrid results. For instance, if the relative dominance of the rules is controlled by a tunable parameter, that parameter could determine whether single or multiple chains form (as in zebra finches or canaries and swamp sparrows, respectively). Intermediate parameter values might produce new regimes in the distribution of sequence lengths.

## EXPERIMENTAL PROCEDURES

### Zebra Finch Data Analysis

Stimulation data were analyzed as in Wang et al. (2008). Briefly, for each discretized stimulation time, we tested whether the sound amplitudes in 3.9 ms bins after stimulation were different from amplitudes in matched time bins during catch trials using the Kolmogorov-Smirnov (KS) test ( $p = 0.01$ ). For each set, we quantified the stimulation effect by the fraction of time bins in which significant differences were detected. The effectiveness or late-effect curve of Figure 3 is based on bins ranging from 78 to 312 ms after stimulation (bins 21 to 80).

### Binary Neuron Network Dynamics

Neurons are active ( $x_i = 1$ ) or inactive ( $x_i = 0$ ). Time is discretized in units of a burst duration, so one time step is equivalent to  $T_{burst}$ . The activity of the  $i$ th HVC neuron is given by

$$x_i(t) = \Theta(I_i^E(t-1) + I_i^{glob}(t-1) + I_i^{ada}(t-1)) \quad (1)$$

where  $\Theta$  is the Heaviside (step) function,  $I_i^E = \sum_j W_{ij}x_j + W_o b_i$  is the summed excitatory drive to the neuron, with  $W_{ij}$  the strength of the connection from neuron  $j$  to  $i$ ,  $W_o$  is the strength of the feedforward input (equal for all neurons), and  $b_i \in \{0, 1\}$  is the external input. Global inhibition is given by  $I_i^{glob} = -\beta \sum_j x_j$ , and adaptation is modeled as a threshold dependent on past activity,  $I_i^{ada} = -\alpha y_i$ , where  $y_i$  is a linearly low-pass-filtered version of  $x_i$ , with time-constant  $\tau_{ada}$ .  $\alpha$  is the adaptation strength. External inputs  $b_i(t)$  are assigned i.i.d. for each neuron in each time-step, with probability  $p(b_i(t) = 1) = p_{in}$  (except for wide-chain formation in the summed-weight rule, for which the procedure for generating spatially correlated  $b_i(t)$  is described below). In Figure 4, we also allow for stochastic neural and synaptic responses. Details in “Parameters and Initial Conditions,” below.

### Conductance-Based Neuron Network Dynamics

The membrane potentials of the conductance-based model HVC neurons are governed by “leaky integrate-and-burst” (LIB) dynamics. The subthreshold evolution of voltages is the same as leaky integrate-and-fire (LIF) neurons:

$$C_m \frac{dV_i}{dt} = -g^L(V_i - V_L) - g^E(V_i - V_E) - g^I(V_i - V_I) \quad (2)$$

The threshold condition is modified so that when  $V_i(t) = V_\theta$ , then disregarding the future inputs the neuron will receive between  $t$  and  $t + T_{burst}$ , the neural voltage segment  $V_i(t, t + T_{burst})$  is assigned to be the fixed sequence  $V_{burst}$ , a pre-determined train of four spikes spaced evenly over the duration  $T_{burst}$ , ending in a repolarization of the neuron to  $V_i(t + T_{burst}) = V_{reset}$ . At  $t + T_{burst}$ , the subthreshold leaky voltage dynamics of Equation 2 take over again. The synaptic activation  $s_i(t)$  is updated in the ordinary way:  $s_i(t)$  is incremented by 1 in response to each spike in neuron  $i$  and decays between spikes according to

$$\frac{ds_i}{dt} = -\frac{s_i}{\tau_s} \quad (3)$$

The excitatory conductance of the  $i$ th neuron is  $g_i^E = \sum_j W_{ij}s_j + W_o b_i$ , where  $b_i(t)$  is the external input. In each time-step, the external input to each neuron is active with probability  $r_{in} dt$ , where  $r_{in}$  is the mean firing rate of the external input in Hz, and  $dt$  is the duration of the time-step in seconds. The inhibitory conductance  $g_i^I = g_i^{glob} + g_i^{ada}$  is made up of a global inhibition term,  $g_i^{glob} = (A_{ij}/N) \sum_j s_j$ , and where applicable, a neural adaptation term, modeled by  $g_i^{ada} = A_{ia} s_i^{ada}(t)$ .  $N$  is the number of neurons in the network, and  $s_i^{ada}$  satisfies the same dynamical equation as  $s_i$ , but with time-constant  $\tau_{ada}$ . In Figure 4, we also allow for stochastic synaptic responses (see “Parameters and Initial Conditions”).

### Learning

In all cases, synaptic weights change under the STDP rule:

$$\Delta_{ij}^{STDP}(t) = \left( \frac{W_{ij}}{W_{max}} + 0.001 \right) \times \left[ x_i(t)K(0)x_j(t) + \sum_{\tau=0}^t x_i(t)K(\tau)x_j(t-\tau) - x_i(t-\tau)K(\tau)x_j(t) \right] \quad (4)$$

For binary neurons:  $x_i(t) = 1$  represents a burst in neuron  $i$  at time  $t$ . For LIB neurons:  $x_i(t) = 1$  if neuron  $i$  fired a spike at time  $t$ , and is zero otherwise.

For binary neurons (Figures 2, 5A–5C, S2, and S3),  $t$  takes values of 0, 1, 2, ..., and  $\tau_{STDP}$  is in units of burst durations. For Figure 2, the STDP window is short (one burst duration):  $K(t) = 1$  for  $t = 1$  and 0 otherwise. The window can be widened without qualitative effects. For wide-chain binary neuron simulations and all LIB simulations (Figures 4, 5, and S2), the STDP kernel is  $K(t) = \exp(-t/\tau_{STDP})$  for  $t > 0$ , and 0 otherwise. For binary neuron wide-chain simulations, the convention is  $K(0) = 1$ . (For robust formation of wide chains, it is important to allow coactive neurons to become strongly connected with each other. The convention  $K(0) = 1$  allows the weights between coactive neurons to grow. If  $K(0) = 0$ , then direct weights between coactive neurons would never undergo STDP because the one time-step occupies the entire burst duration.) For LIB neurons, the convention is  $K(0) = 0$ , producing an anti-symmetric STDP window. (The burst in a LIB neuron simulation occupies many time-steps. Thus, neurons with overlapping bursts overlap at nonzero time-lags and can undergo STDP even if  $K(0) = 0$ .)

The total weight update is given by STDP as above, and competitive hLTD triggered by a threshold on summed weight (or weight growth):

$$W_{ij}(t) = W_{ij}(t-1) + \eta \Delta_{ij}^{STDP}(t) - \epsilon \eta \theta_{in}(t) - \epsilon \eta \theta_{out}(t) \quad (5)$$

with clipping to keep each weight within  $[0, W_{max}]$ .  $\theta_{in}$  represents the competitive hLTD triggered at all outgoing synapses from neuron  $i$ , and  $\theta_{out}$  represents the same for all incoming synapses into neuron  $i$ . For the summed-weight limit,  $\theta_{in} = \max(0, \sum_k (W_{ik} + \Delta_{ik}^{STDP}) - W_{max})$ , and  $\theta_{out} = \max(0, \sum_k (W_{ki} + \Delta_{ki}^{STDP}) - W_{max})$ . For the weight-growth limit,  $\theta_{in} = \sum_k W_{ik} \Theta(\Delta_{ik}^{STDP})$ , and  $\theta_{out} = \sum_k W_{ki} \Theta(\Delta_{ki}^{STDP})$  where  $\Theta$  is the Heaviside (step) function. The diagonal (self-interaction) weights are fixed at zero,  $W_{ii} = 0$ .  $\eta$  and  $\epsilon$  set the learning step size and the strength of the heterosynaptic constraint.

**Parameters and Initial Conditions**

Initially in all simulations,  $W_{ij} = w_{max}/N$  or  $W_{ij}$  is random in the interval  $[0, w_{max}/N]$ , for all  $i \neq j$ .  $N$  is the network size.  $W_{ij} = 0$  for all  $i$ . For binary neurons,  $W_o = 1$  (sufficiently large to activate the postsynaptic neuron).

**Summed-Weight Limit, Binary Neurons**

$\alpha = 0$  (i.e., no adaptation; thus, *tada* is irrelevant),  $\beta = 0.25$ ,  $p_{in} = 2/N$ ,  $N = 50$ ,  $\eta = 0.025$ ,  $\epsilon = 0.125$ . For  $m = 1$  (width-1 chains),  $W_{max} = 1$ ,  $w_{max} = 1$ . Parameters do not need to be finely tuned (see Figure S1). Matlab code for this case (Figure 2) is posted online.

For Figures 5A–5C (wide chains), the  $N = 50$  neuron network is divided into ten nonoverlapping groups of five neurons each, and  $p_{in} = 2.5/N$  (here  $p_{in}$  refers to the i.i.d. probability that the external input to any input group is activated). Given superthreshold input, individual neurons within the group respond i.i.d. with a burst with probability 0.95, or fail to respond with probability 0.05. (Because each response represents a burst of three to six spikes, the 0.05 probability failure to respond represents a failure of every spike in the burst given that the input is superthreshold.) There are no temporal correlations in the inputs within or across groups.  $\eta = 0.001$ ,  $\epsilon = 0.05$ ,  $W_{max} = 1.8$ ,  $b = 0.15$ ,  $\tau_{STDP} = 2$ , and  $w_{max} = W_{max}/m$ , with  $m = 7$  or  $m = 9$ , which produced qualitatively similar results but different chain widths. Whenever neuron  $i$  spiked, the  $ji$  synapse was activated with probability 0.9, i.i.d. for each target  $j$ . (Because the  $ji$  model synapse represents the total connection between the  $ij$  pair of neurons, it may represent multiple individual synapses. Furthermore, failure of synaptic transmission represents failure of synaptic transmission over every spike within the burst.)

**Weight-Growth Limit, Binary Neurons**

$\alpha = 1$ ,  $\tau_{ada} = 4$  time-steps,  $N = 135$ ,  $p_{in} = 1/N$ , and  $\beta = 0$  (i.e., global inhibition is not necessary; however, using  $\beta \neq 0$  does not qualitatively change the results). Learning:  $\eta = 0.0125$ ,  $\epsilon = 2$ ,  $w_{max} = 2.5$ . Over the first 2000 time-steps (burst durations), the input probability grows from 0 to  $p_{in}$  to avoid synchronously activating (and then rendering adapted) the population. After 10,000 time-steps,  $p_{in}$  is decreased to zero with a time-constant of 4000 time-steps.

**LIB Neurons**

$dt = 0.02$  ms,  $C_m = 1$   $\mu$ F/cm<sup>2</sup>,  $V_L = -60$  mV,  $V_E = 0$  mV,  $V_I = -70$  mV,  $g_L = 0.4$  mS/cm<sup>2</sup>,  $W_o = 0.5$  mS/cm<sup>2</sup>,  $V_\theta = -50$  mV,  $V_{reset} = -55$  mV,  $T_{burst} = 6$  ms,  $\tau_S = 4$  ms, and  $r_{in} = 4$  Hz.

**Summed-Weight Limit, LIB Neurons**

In Figure 4,  $N = 50$ ,  $w_{max} = 0.14$ ,  $W_{max} = w_{max}$  ( $m = 1$ ),  $\eta = 0.002$ ,  $\epsilon = 72.5$ ,  $A_g = 0.4$  mS/cm<sup>2</sup>,  $A_a = 0.9$  mS/cm<sup>2</sup>,  $\tau_{STDP} = 20$  ms, and *tada* = 15 ms. The poisson rate of input neurons is 2 Hz. For wide chains (Figures 5D–5F), we used disjoint input groupings of size  $k = 5$ , with  $w_{max} = W_{max}/9$  ( $m = 9$ ),  $A_g = 0.2$  mS/cm<sup>2</sup>,  $W_{max} = 0.26$ ,  $\eta = 0.0001$ ,  $\epsilon = 30$ , and a poisson rate of 10 Hz for the input neurons. Whenever HVC or input neuron  $i$  spiked, the  $ji$  synapse for each target  $j$  was activated i.i.d. with probability 0.9. Other parameters were as in the  $m = 1$  case.

**Weight-Growth Limit, LIB Neurons**

Parameters are the same as in the  $m = 1$  summed-weight limit case above, except that  $g_L = 0.1$  mS/cm<sup>2</sup>,  $N = 80$ ,  $W_o = 0.5$ ,  $w_{max} = 3$ ,  $\eta = 0.038$ ,  $\epsilon = 4.8/N$ ,  $A_g = 0$  mS/cm<sup>2</sup>,  $A_a = 0.5$  mS/cm<sup>2</sup>,  $\tau_{ada} = 20$  ms. The synaptic activation has an exponential rise-time of 1 ms and decay-time of 4 ms. The input is annealed away starting at 3 s. Annealing was done by exponentially decaying the input firing rate, with a time-constant of 6 s.

**Parameter Tuning**

It is important (summed-weight rule) to start with small weights  $W_{ij}$ :  $W_{ij} \ll w_{max}$  for each  $ij$ , and  $\sum_i W_{ij} \ll w_{max}$  and  $\sum_j W_{ij} \ll w_{max}$ . Parameters must be consistent with the existence of a self-propagating activity solution:  $w_{max}$  has to be large enough so that if, in the desired configuration, most of the presynaptic inputs to a neuron are simultaneously active, the postsynaptic neuron, driven by the summed weights  $W_{max}$ , will fire. The learning rate  $\eta$  must be large enough that upward random fluctuations in the weight matrix in regime I (see Results) can contribute significantly to the probability of activation of the postsynaptic neuron. Yet  $\eta$  must be small enough for learning to be stable. To allow for the growth of some weights to reach  $w_{max}$ , competitive hLTD per synapse should be smaller than the largest STDP-driven LTP. This involved tuning  $\epsilon$  once  $\eta$  was fixed (however,  $\epsilon$  did not need to be finely tuned, and its value could vary widely without qualitatively affecting the results—

Figure S1). These steps brought the dynamics into the right regime for learning. In this regime, the parameters did not require fine-tuning and could be perturbed without qualitative effects. Finally, we verified for the summed-weight limit rule that the results are unaffected if the dynamics of weight subtraction are slow, i.e., if subtraction does not follow instantaneously after  $w_{max}$  is reached through an LTP event, but is instead implemented in batch mode after a number of LTP events, or is delayed continuously by a low-pass filter, with actual hLTD or subtraction performed on a slower timescale.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes a calculation (of cycle lengths expected from the group of permutation matrices), Matlab code, and four figures and can be found with this article online at doi:10.1016/j.neuron.2010.02.003.

**ACKNOWLEDGMENTS**

We are grateful to the referees for their helpful comments. This work was partially supported by NSF-PHY 99-07949 to the Kavli Institute for Theoretical Physics, the Broad Fellows program at Caltech, and the Swiss National Science Foundation. I.R.F. is a Sloan Foundation Fellow.

Accepted: January 27, 2010

Published: February 24, 2010

**REFERENCES**

- Abbott, L.F., and Nelson, S.B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* 3 (Suppl), 1178–1183.
- Abeles, M. (1991). *Corticonics* (London: Cambridge University Press).
- Abraham, W.C., Bliss, T.V., and Goddard, G.V. (1985). Heterosynaptic changes accompany long-term but not short-term potentiation of the perforant path in the anaesthetized rat. *J. Physiol.* 363, 335–349.
- Amari, S.-I. (1972). Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* C-21, 1197–1206.
- Andersen, R.A., Burdick, J.W., Musallam, S., Pesaran, B., and Cham, J.G. (2004). Cognitive neural prosthetics. *Trends Cogn. Sci.* 8, 486–493.
- Aviel, Y., Mehring, C., Abeles, M., and Horn, D. (2003). On embedding synfire chains in a balanced network. *Neural Comput.* 15, 1321–1340.
- Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437, 1158–1161.
- Bi, G.Q., and Poo, M.M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472.
- Bi, G., and Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* 24, 139–166.
- Bottjer, S.W., Miesner, E.A., and Arnold, A.P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* 224, 901–903.
- Boume, J., and Harris, K. (2010). Coordination of size and number of excitatory and inhibitory synapses results in a balanced structural plasticity along mature hippocampal ca1 dendrites during LTP. *Hippocampus*, in press. Published online January 25, 2010. 10.1002/hipo.20768.
- Buonomano, D.V. (2003). Timing of neural responses in cortical organotypic slices. *Proc. Natl. Acad. Sci. USA* 100, 4897–4902.
- Buonomano, D.V. (2005). A learning rule for the emergence of stable dynamics and timing in recurrent networks. *J. Neurophysiol.* 94, 2275–2283.
- Davidson, T.J., Kloosterman, F., and Wilson, M.A. (2009). Hippocampal replay of extended experience. *Neuron* 63, 497–507.
- Doyère, V., Srebro, B., and Laroche, S. (1997). Heterosynaptic LTD and depotentiation in the medial perforant path of the dentate gyrus in the freely moving rat. *J. Neurophysiol.* 77, 571–578.

- Drew, P.J., and Abbott, L.F. (2003). Model of song selectivity and sequence generation in area HVC of the songbird. *J. Neurophysiol.* **89**, 2697–2706.
- Fee, M.S., Kozhevnikov, A.A., and Hahnloser, R.H.R. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N Y Acad. Sci.* **1016**, 153–170.
- Fiete, I.R., and Seung, H.S. (2008). Neural network models of birdsong production, learning, and coding. In *The New Encyclopedia of Neuroscience*, L. Squire, T. Albright, F. Bloom, F. Gage, and N. Spitzer, eds. (New York: Elsevier).
- Fonseca, R., Nägerl, U.V., Morris, R.G.M., and Bonhoeffer, T. (2004). Competing for memory: hippocampal LTP under regimes of reduced protein synthesis. *Neuron* **44**, 1011–1020.
- Fonseca, R., Nägerl, U.V., and Bonhoeffer, T. (2006). Neuronal activity determines the protein synthesis dependence of long-term potentiation. *Nat. Neurosci.* **9**, 478–480.
- Frey, U., and Morris, R.G. (1997). Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536.
- Glaze, C.M., and Troyer, T.W. (2006). Temporal structure in zebra finch song: implications for motor coding. *J. Neurosci.* **26**, 991–1005.
- Hahnloser, R., and Fee, M. (2006). Sleep-related spike bursts in hvc are driven by the nucleus interface of the nidopallium. *J. Neurophysiol.* **97**, 423–435.
- Hahnloser, R.H.R., and Fee, M.S. (2007). Sleep-related spike bursts in HVC are driven by the nucleus interface of the nidopallium. *J. Neurophysiol.* **97**, 423–435.
- Hahnloser, R.H., Kozhevnikov, A.A., and Fee, M.S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70.
- Harvey, C.D., and Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200.
- Hashimoto, K., Ichikawa, R., Kitamura, K., Watanabe, M., and Kano, M. (2009). Translocation of a “winner” climbing fiber to the Purkinje cell dendrite and subsequent elimination of “losers” from the soma in developing cerebellum. *Neuron* **63**, 106–118.
- Huber, K.M., Kayser, M.S., and Bear, M.F. (2000). Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288**, 1254–1257.
- Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., and Yuste, R. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Science* **304**, 559–564.
- Jin, D., Ramazanoglu, F., and Seung, H. (2007). Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area hvc. *J. Comput. Neurosci.* **23**, 283–299.
- Jun, J.K., and Jin, D.Z. (2007). Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS ONE* **2**, e723.
- Kleinfeld, D., and Sompolinsky, H. (1988). Associative neural network model for the generation of temporal patterns. Theory and application to central pattern generators. *Biophys. J.* **54**, 1039–1051.
- Kozhevnikov, A.A., and Fee, M.S. (2007). Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97**, 4271–4283.
- Levy, N., Horn, D., Meilijson, I., and Ruppin, E. (2001). Distributed synchrony in a cell assembly of spiking neurons. *Neural Netw.* **14**, 815–824.
- Li, M., and Greenside, H. (2006). Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus hvc. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **74**, 011918.
- Long, M.A., and Fee, M.S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194.
- Louie, K., and Wilson, M.A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* **29**, 145–156.
- Luczak, A., Barthó, P., Marguet, S.L., Buzsáki, G., and Harris, K.D. (2007). Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl. Acad. Sci. USA* **104**, 347–352.
- Malenka, R.C., and Bear, M.F. (2004). LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21.
- Mokeichev, A., Okun, M., Barak, O., Katz, Y., Ben-Shahar, O., and Lampl, I. (2007). Stochastic emergence of repeating cortical motifs in spontaneous membrane potential fluctuations in vivo. *Neuron* **53**, 413–425.
- Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J. Neurosci.* **20**, 5420–5436.
- Mooney, R., and Prather, J.F. (2005). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci.* **25**, 1952–1964.
- Nádasy, Z., Hirase, H., Czurkó, A., Csicsvari, J., and Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *J. Neurosci.* **19**, 9497–9507.
- Nottebohm, F., Stokes, T.M., and Leonard, C.M. (1976). Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* **165**, 457–486.
- Nottebohm, F., Kelley, D.B., and Paton, J.A. (1982). Connections of vocal control nuclei in the canary telencephalon. *J. Comp. Neurol.* **207**, 344–357.
- Nowotny, T., Rabinovich, M.I., and Abarbanel, H.D. (2003). Spatial representation of temporal information through spike-timing-dependent plasticity. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **68**, 011908.
- Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327.
- Pulvermüller, F., and Shtyrov, Y. (2009). Spatiotemporal signatures of large-scale synfire chains for speech processing as revealed by MEG. *Cereb. Cortex* **19**, 79–88.
- Rao, R., and Sejnowski, T. (2003). Self-organizing neural systems based on predictive learning. *Philos. Transact. Ser. A Math. Phys. Eng. Sci.* **361**, 1149–1175.
- Roxin, A., Hakim, V., and Brunel, N. (2008). The statistics of repeating patterns of cortical activity can be reproduced by a model network of stochastic binary neurons. *J. Neurosci.* **28**, 10734–10745.
- Royer, S., and Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* **422**, 518–522.
- Sanes, J.R., and Lichtman, J.W. (1999). Development of the vertebrate neuromuscular junction. *Annu. Rev. Neurosci.* **22**, 389–442.
- Schwartz, A.B., and Moran, D.W. (1999). Motor cortical activity during drawing movements: population representation during lemniscate tracing. *J. Neurophysiol.* **82**, 2705–2718.
- Seung, H.S. (2009). Reading the book of memory: sparse sampling versus dense mapping of connectomes. *Neuron* **62**, 17–29.
- Suri, R.E., and Sejnowski, T.J. (2002). Spike propagation synchronized by temporally asymmetric Hebbian learning. *Biol. Cybern.* **87**, 440–445.
- Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J.L., Patel, H., Prieto, A., Petrusca, D., Grivich, M.I., Sher, A., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J. Neurosci.* **28**, 505–518.
- Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annu. Rev. Neurosci.* **24**, 631–651.
- Wang, C.Z.H., Herbst, J.A., Keller, G.B., and Hahnloser, R.H.R. (2008). Rapid interhemispheric switching during vocal production in a songbird. *PLoS Biol.* **6**, e250.
- Weber, A.P., and Hahnloser, R.H.R. (2007). Spike correlations in a songbird agree with a simple markov population model. *PLoS Comput. Biol.* **3**, e249.