Characteristics of sequential activity in networks with temporally asymmetric Hebbian learning

Maxwell Gillett\textsuperscript{a,b}, Ulises Pereira\textsuperscript{a}, and Nicolas Brunel\textsuperscript{a,b,c,d}

\textsuperscript{a}Department of Neurobiology, The University of Chicago; \textsuperscript{b}Department of Neurobiology, Duke University; \textsuperscript{c}Department of Statistics, The University of Chicago; \textsuperscript{d}Department of Physics, Duke University

Sequential activity has been observed in multiple neuronal circuits across species, neural structures, and behaviors. It has been hypothesized that sequences could arise from unsupervised learning processes. However, it is still unclear whether biologically plausible synaptic plasticity rules can organize neuronal activity to form sequences whose statistics match experimental observations. Here we investigate temporally asymmetric Hebbian rules in sparsely connected recurrent rate networks, and develop a theory of the transient sequential activity observed after learning. These rules transform a sequence of random input patterns into synaptic weight updates. After learning, recalled sequential activity is reflected in the transient correlation of network activity with each of the stored input patterns. Using mean-field theory, we derive a low-dimensional description of the network dynamics and compute the storage capacity of these networks. Multiple temporal characteristics of the recalled sequential activity are consistent with experimental observations. We find that the degree of sparseness of the recalled sequences can be controlled by non-linearities in the learning rule. Furthermore, sequences maintain robust decoding, but display highly labile dynamics, when synaptic connectivity is continuously modified due to noise or storage of other patterns, similar to recent observations in hippocampus and parietal cortex. Finally, we demonstrate that our results also hold in recurrent networks of spiking neurons with separate excitatory and inhibitory populations.

Sequential activity has been reported across a wide range of neural systems and behavioral contexts, where it plays a critical role in temporal information encoding. Sequences can encode choice-selective information (1), the timing of motor actions (2), planned or recalled trajectories through the environment (3), and elapsed time (4–6). This diversity in function is also reflected at the level of neuronal activity. Sequences occur at varying timescales, from those lasting tens of milliseconds during hippocampal sharp-wave ripples, to those spanning several seconds in the striatum (7, 8). Sequential activity also varies in temporal sparsity. In Nucleus HVC of the zebra finch, highly-precise sequential activity is present during song production, where many neurons fire only a single short burst during a syllable (9). In primate motor cortex, single neurons are typically active throughout a whole reach movement, but with heterogeneous and rich dynamics (10).

Numerous models have explored how networks with specific synaptic connectivity can generate sequential activity (11–20). These models span a wide range of single neuron models, from binary to spiking, and a wide range of synaptic connectivities. A large class of models employs temporally asymmetric Hebbian (TAH) learning rules to generate a synaptic connectivity necessary for sequence retrieval. In these models, a sequence of random input patterns are presented to the network, and a Hebbian learning rule transforms the resulting patterns of activity into synaptic weight updates. In networks of binary neurons, TAH learning together with synchronous update dynamics lead to sequence storage and retrieval (21). With asynchronous dynamics, additional mechanisms such as delays are needed to generate robust sequence retrieval (13).

Another longstanding and influential class of model has centered on the ‘synfire chain’ architecture (22) in networks of spiking neurons, in which discrete pools of neurons are connected in a chain-like, feedforward fashion, and a barrage of synchronous activity in the first pool propagates down connected pools, separated in time by synaptic delays (22, 23). Theoretical studies have shown that synfire chains can self-organize from initially unstructured connectivity through an appropriate choice of plasticity rule and stimulus (16, 19). Other studies have shown that these chains produce sequential activity when embedded into recurrent networks if appropriate recurrent and feedback connections are introduced among pools (24–26).

Despite decades of research on this topic, the relationship between network parameters and experimentally observed features of sequential activity are still poorly understood. In particular, few models account for the temporal statistics of sequential activity during retrieval and across multiple recording sessions. In many sequences, a greater proportion of neurons encode for earlier times, and the tuning width of recruited neurons increases in time (5, 27, 28). Neural sequences are also not static, but change over the timescales of days, even when controlling for the same environment and task constraints (29, 30). TAH learning provides an appealing framework for how sequences of neuronal activity are embedded in the connectivity matrix. While TAH learning has been extensively explored in binary neuron network models, it remains an open question as to whether these learning rules can produce sequential activity in more realistic network models, and also reproduce the observed temporal statistics of this activity. It is also unclear whether a theory can be developed to understand the resulting activity, as has been done with networks of binary neurons. Rate networks provide a useful framework for investigating these questions, as they balance analytical tractability and biological realism.

We start by constructing a theory of transient activity that can be used to predict the sequence capacity of these networks, and focus on two features of sequences that have been described in the experimental literature: temporal sparsity, and the selectivity of single unit activity. We find that several temporal characteristics of sequential activity are reproduced by the network model, including the temporal distribution of firing rate peak widths and times. Introduction of a nonlinearity to the learning rule allows for the generation of temporally sparse sequences. We also find that storing multiple sequences yields selective single unit activity resembling that found during two-alternative forced choice tasks (1). Consistent with experiment, we find that the population activity of these sequences changes over time in the face of ongoing synaptic changes, while decoding performance is preserved (29, 30).
nally, we demonstrate that our findings also hold in networks of excitatory and inhibitory leaky-integrate and fire neurons, where learning takes place only in excitatory-to-excitatory connections.

Results
We have studied the ability of large networks of sparsely-connected rate units to learn sequential activity from random, unstructured input, using analytical and numerical methods. We focus initially on a network model composed of $N$ neurons, each described by a firing rate $r_i$ ($i = 1, \ldots, N$), that evolves according to standard rate equations (31)

$$
\tau \frac{dr_i}{dt} = -r_i + \phi(h_i), \quad h_i = \sum_{j \neq i} J_{ij} r_j
$$

where $\tau$ is the time constant of rate dynamics, $h_i$ is the total synaptic input to neuron $i$, and $J_{ij}$ is the connectivity matrix. $\phi$ is a sigmoidal neuronal transfer function,

$$
\phi(h) = \frac{r_{\text{max}}}{2} \left( 1 + \text{erf} \left( \frac{h - \theta}{\sqrt{2} \sigma} \right) \right)
$$

which is described by three parameters: $r_{\text{max}}$, the maximal firing rate (set to 1 in most of the following), $\theta$, the input at which the firing rate is half its maximal value, and $\sigma$, a parameter which is inversely proportional to the slope (gain) of the transfer function. Note that when $\sigma \to 0$ the transfer function becomes a step function: $\phi(h) = 1$ for $h > \theta$, 0 otherwise. Note also that the specific shape of the transfer function (Eq 2) was chosen for the sake of analytical tractability, but our results are robust to other choices of transfer function (see Fig. 7).

Learning rule. We assume that the synaptic connectivity matrix is the result of a learning process through which a set of sequences of random inputs presented to the network has been stored using a temporally asymmetric Hebbian (TAH) learning rule. Specifically, we consider $S$ sequences of $P$ random independent and identically distributed (i.i.d.) Gaussian input patterns $\xi_i^s$ at neuron $i$, where $s = \{1, \ldots, S\}$ and $\mu = \{1, \ldots, P\}$. For each successive pair of presented inputs in a sequence, the strength $J_{ij}$ of a synaptic connection from neuron $j$ to neuron $i$ is modified according to a temporally asymmetric Hebbian learning rule. This rule is simply the product of transformed pre-synaptic activity at a given time, and post-synaptic activity at the next time step in the sequence, $\Delta J_{ij} \propto f(\xi_i^{s+1})g(\xi_j^s)$. The transformations $f$ and $g$ are monotonically increasing functions of their argument that describe how post- and presynaptic inputs affect synaptic strength, respectively. After presentation of all $S$ sequences, the connectivity matrix is given by

$$
J_{ij} = A \frac{c_{ij}}{K} \sum_{s=1}^{S} \sum_{\mu=1}^{P-1} f(\xi_i^{s+1})g(\xi_j^s)\quad[3]
$$

where $c_{ij}$ is a matrix of i.i.d. Bernoulli random variable: $c_{ij} = 1, 0$ with probabilities $c, 1-c$. $c_{ij}$ describes the ‘structural’ connectivity matrix (i.e whether a connection from neuron $j$ to neuron $i$ exists or not), and $K = Nc$ represents the average in-degree of a neuron. The parameter $A$ controls the strength of the recurrent connectivity.

Note that this learning rule makes a number of simplifications, for the sake of analytical tractability: (i) Plasticity is assumed to depend only on external inputs to the network, and recurrent synaptic inputs are neglected during the learning phase (ii) Inputs are assumed to be discrete in time, and the learning rule is assumed to associate a presynaptic input at a given time, with postsynaptic input at the next time step. In other words, the temporal structure in the input should be matched with the pre-post delay maximizing synaptic potentiation. We note that this temporal asymmetry is consistent with the spike-timing dependent plasticity (STDP), but that widely different time scales have been reported, from tens of milliseconds (32), to seconds (33). We come back to these assumptions in the Discussion.

Retrieval of stored sequences. We first ask the question of whether such a network can recall the stored sequences, and characterize the phenomenology of the retrieved sequences. We start by considering a network that stores a single sequence of $P$ i.i.d. Gaussian patterns, $\xi_i^s \sim \mathcal{N}(0, 1)$ for all $i = 1, \ldots, N$, $\mu = 1, \ldots, P$, using a bilinear learning rule ($f(x) = x$ and $g(x) = x$). Without loss of generality we choose $A = 1$, as this parameter can be absorbed into the transfer function by defining new parameters $\theta^* = \frac{\theta}{\sigma}$ and $\sigma^* = \frac{\sigma}{\sqrt{2}}$. The Gaussian
assumption can be justified for a network in which neurons receive a large number of weakly correlated inputs, since the sum of such inputs is expected to be close to Gaussian due to the central limit theorem. The bilinearity assumption is made for the sake of analytical tractability, and will be relaxed below.

In Fig. 1a, we show the retrieval of a single sequence of sixteen input patterns from the perspective of a single unit. The filled black circles correspond to the sequence of firing rates driven by input patterns for this particular neuron (i.e. $\phi(\xi^\mu)$, for $\mu = 1, \ldots, 16$). The solid red line displays the unit’s activity following initialization to the first pattern in that sequence, after learning has taken place. Note that for this learning rule, the network dynamics do not reproduce exactly the stored input sequence, but rather produce a dynamical trajectory that is correlated with that input sequence. In Fig. 1a, we have specifically selected a unit whose activity transits close to the stored patterns during recall. In Fig. 1b, the activity of several random units is displayed during sequence retrieval. As each unit experiences an uncorrelated random sequence of inputs during learning, activity is highly heterogeneous, and units differ in the degree to which they are correlated with input patterns (Supplementary Figure 1). Furthermore, retrieval is highly robust to noise - a random perturbation of the initial conditions of magnitude 75% of the pattern itself leads to sequence retrieval, as shown in Fig. 2a (compare the two solid lines, corresponding to unperturbed and perturbed initial conditions, respectively - see also Supp. Info for more details about noise robustness).

A natural quantity to measure sequence retrieval at the population level is the overlap (dot product) of the instantaneous firing rates with a given stored pattern in the sequence: $m_l(t) = \frac{1}{N} \sum_{i=1}^{N} r_i(t) \xi^\mu_i$. Another quantity is the (Pearson) correlation between network activity and the stored pattern, which for the bilinear rule corresponds to the overlap divided by the standard deviation of the firing rates across the population. In the general case, we consider the correlation between the transformed presynaptic input $g_i(\xi^\mu)$ and network activity $r(t)$ (see Supp. Info). Overlaps and correlations are shown in Fig. 1c and d, respectively, showing that all patterns in the sequence are successively retrieved, with an approximately constant peak overlap ($\sim 0.1$) and correlation ($\sim 0.4$). Note that the maximum achievable overlap and correlation for the bilinear rule are 0.388 and 0.825 respectively, following retrieval from initialization to the first pattern (see Supp. Info).

**Temporal characteristics of the retrieved sequence.** We next consider the temporal characteristics of retrieved sequences. The speed of retrieval of a sequence is defined as the inverse of the interval between the peaks in the overlaps with successive patterns in the sequence. We find that the network retrieves the sequence at a speed of approximately one pattern retrieved per time constant $\tau$. For instance, in Fig. 1, in which the time constant of the units is 10ms, the whole sequence of 16 patterns is retrieved in approximately 150ms (the first pattern was retrieved at time zero). Empirically, we find that the time it takes to retrieve the whole sequence is approximately invariant to the number of sequences stored and the parameters of the rate transfer function, and is well predicted by $t_{\text{retrieval}} = \tau(P - 1)$ (Supplementary Figure 3).

While the sequence is retrieved at an approximately constant speed, several other features of retrieval depend strongly on the time within the sequence. In particular, the peak overlap magnitudes decay as a function of time within the sequence, but the width increases as $\tau \sqrt{t}$, where $l > 1$ denotes the index of the overlap (see Fig. 2b and Supp. Info). Similarly, neurons at the beginning of a sequence are sharply tuned, and those toward the end have more broad profiles (Fig. 2a). This is consistent with experimental findings in rat mPFC and CA3, and in monkey IPFC (5, 27, 28). Another commonly reported experimental observation involves the temporal ordering of these peaks. Across these same areas and species, a larger number of neurons appear to encode for earlier parts of the sequence than for later segments, resulting in sequences that are not uniformly represented in time by neuronal activity (5, 27, 28). A similar pattern of over-representation is found in the model. The cumulative density of firing rate peak times deviates from a uniform distribution, while the density of the overlap peak times are roughly uniform in time (Fig. 2c).

**A low-dimensional description of dynamics.** To better understand the properties of these networks, we employed mean-field theory to analyze how the overlaps in a typical network realization evolve in time. To this end, we defined several ‘order parameters’. The first set of order parameters measure the typical overlap of network activity in time with the $\mu$-th pattern: $q_\mu(t) = \mathbb{E}(\xi^\mu r(t))$, where the average is taken over the statistics of the patterns. Note that these are distinct from the previously defined overlaps $m$, which are empirically measured values for a particular realization of sequential activity patterns. However, for the large, sparse networks considered here, these quantities describe any typical realization of random input patterns and connectivity (see Supp. Info). We also define the average squared rate $M(t) = \mathbb{E}(\xi^2 r(t))$, the average two-point rate correlation function $C(u, t) = \mathbb{E}(r(u) r(t))$, and the network memory load as the number of patterns stored per average synaptic in-degree, i.e. $\alpha = \frac{\mathbb{E}(m \mu)}{\mu}$.

Framing the network in these terms yields a low-dimensional description in which the time-dependent evolution of the overlaps can be written as an effective delay line system (see Supp. Info). In this formulation, the activation level of each overlap is driven by a gain-modulated version of the previous one in the sequence, where the gain depends on the sequential load, rate variability, and norm of the overlaps

$$
\tau \frac{d q_\mu}{dt} = -q_\mu + q_{\mu - 1} G \left( \sum_{\nu=1}^{P-1} q^2_\nu(t) + \alpha M \right), \quad \mu = 2, \ldots, P, \ [4]
$$

and the gain function $G$ is given by

$$
G(x) \equiv \frac{1}{\sqrt{2\pi(\sigma^2 + x)}} \exp \left( -\frac{q^2}{2(\sigma^2 + x)} \right). \ [5]
$$

The dynamics of $M$ and $C$ are given by coupled integrodifferential equations (see Supp. Info). With a constant gain $G$ below one, recall of a sequence would decay to zero, as each overlap would become increasingly less effective at driving the next one in the sequence. Conversely, a constant gain above one would eventually result in runaway growth (see Supp. Info). In Eq. (4), the gain changes in time due to its dependence on the norm of $q$ and $M$. During successful retrieval, the gain transiently rises to a value that is slightly larger than one, allowing for overlap peaks to maintain an approximately constant value (Supplementary Figure 6b). We
find that there are two different regimes for successful retrieval that depend on the shape of $G$ (and therefore the transfer function parameters): one in which sequences of arbitrarily small initial overlap are retrieved, and another in which a small but finite initial overlap is required (Supplementary Figure 6a, c).

The predictions from mean-field theory agree closely with the full network simulations. Fig. 3a shows the dynamics of a network in which two sequences of identical length are stored. At time zero, we initialize to the first pattern of the first sequence (corresponding overlaps shown in red). At 250 ms, we present an input lasting 10 ms corresponding to the first pattern of the second (blue) sequence. The solid lines show the overlap of the full network activity with each of the stored input patterns. The predicted overlaps from simulating mean-field equations are shown in dashed lines. The average squared rate is also predicted well by the theory, as is the two-point rate correlation (Supplementary Figure 5).

Storage capacity of the network. We next asked how the properties of retrieved sequences depend on the memory load $\alpha$, and what is the maximal storage capacity of the network, defined as the largest value of $\alpha$ for which sequences can be retrieved successfully. Figure 3b shows the peak correlation value that the network attained with the final pattern in a sequence as a function of the sequence load, obtained using the mean-field equations. It shows that the correlation with the last pattern decreases with $\alpha$ until it reaches, for long enough sequences, a value close to zero at $\alpha \sim 0.5$. We calculated analytically this sequence capacity $\alpha_c$, and found that it is implicitly determined by solving the relation $G(\alpha_c, M_c) = 1$ for $\alpha_c$, where $M_c$ is the average squared firing rate at capacity (see Supp. Info). At capacity, for a fixed number of incoming connections $K$, the network can store one long patterned sequence of length $K\alpha_c$, or any number of sequences that collectively sum to this length.

In Fig. 3c, the analytically computed capacity curve is shown as a function of the inverse gain parameter of the rate transfer function $\sigma$, and for three values of $\theta$ (solid lines). Below a specified capacity curve, a sequence can be retrieved, and above it decays to zero. A small inverse gain corresponds to a transfer function with a very steep slope, whereas a large inverse gain produces a more shallow slope. The parameter $\theta$ determines the threshold of input required to drive a unit to half its maximum firing rate $r_{\text{max}}$, where a larger $\theta$ corresponds to a higher threshold. For Gaussian patterns, we find that capacity is maximal for transfer functions with high thresholds, and that above a critical $\theta$, capacity drops abruptly to zero (Supplementary Figure 6a, 7). The storage capacity obtained from simulating the mean-field equations (symbols) agrees well with the analytical result (solid lines).

**Nonlinear learning rules produce sparse sequences.** We have explored so far the storage of Gaussian sequences, and found that the stored sequences can be robustly retrieved. Activity at the population level is heterogeneous in time (Fig. 1b), with many units maintaining high levels of activity throughout recall. However, many neural systems exhibit more temporally sparse sequential activity (9). While the average activity level of the network can be controlled to some degree by the neuronal threshold $\theta$, retrieved sequences are non-sparse even for the largest values of $\theta$ for which the capacity is non-zero. Altering the statistics of the stored patterns will in turn change the statistics of neuronal activity during retrieval, which could lead to more sparse activity.

We therefore revisited Eq. 3, and explored nonlinear pre and post-synaptic transformations $(g(x)$ and $f(x)$, respectively). We focus on a simplified functional form, similar to one shown to fit learning rules inferred from in vivo cortical data (34), and shown to maximize storage capacity in networks storing fixed point attractors (35). For each synaptic terminal (pre and post), we apply a transformation that binarizes the activity patterns $\xi$ into high and low values according to a threshold (Fig. 4a):

$$f(x) = \begin{cases} q_f & \text{if } x > x_f \\ q_f - 1 & \text{if } x \leq x_f \end{cases} \quad [6]$$

$$g(x) = \begin{cases} q_g & \text{if } x > x_g \\ q_g - 1 & \text{if } x \leq x_g \end{cases} \quad [7]$$

In the following we choose $x_f = x_g$, and set $q_x$ such that $\int Dzg(z) = 0$. This fixes the mean connection strength to

---

Fig. 2. Temporal characteristics of a retrieved sequence. a. Distribution of single neuron peak widths, defined as continuous firing intervals occurring one standard deviation above the single neuron time-averaged firing rate. Black error bars denote mean and standard deviation of widths within each 10 ms interval. Dashed red curve is best-fit trend using scaled square root function (see Supp. Info). b. Green dots correspond to observed overlap widths, defined as the weighted sample standard deviation of $m_1(t)$ (see Supp. Info). Red dots display the analytically predicted overlap width of $\sqrt{\tau_l}$, where $\tau = 10$ ms and $l > 1$ (see Supp. Info). c. Cumulative percentage of peak times for single neurons (blue) and overlaps (grey). The dashed black line represents a uniform distribution. All parameters are as in Fig. 1.
null
Fig. 4. Sparse sequences with a nonlinear learning rule. 

- **a**: Probability density of Gaussian input pattern $\xi$ (black). Step function $f$ binarizing the input patterns prior to storage in the connectivity matrix (blue).
- **b**: Firing rate of several representative units as a function of time.
- **c**: Firing rates of 4000 neurons (out of 40,000) as a function of time, with ‘silent’ neurons shown on top, and active neurons on the bottom, sorted by time of peak firing rate.
- **d**: Correlation of network activity with each stored pattern.
- **e**: Average population rate as a function of ‘coding level’ (probability that input pattern is above $x_f$). The average of $f(x)$ is maintained by varying $q_f$ with $x_f$. In red, the average of $f(x)$ is constrained to $-0.15$, the value in panels (a-d). In black, the average of $f(x)$ is fixed to zero. All other parameters are as in panels (a-d).

For panels (a-d), parameters of the learning rule were $x_f = 1.645$, $x_g = 1.645$, $q_f = 0.8$, and $q_g = 0.95$. $S = 1$, $P = 30$.

Fig. 5. Diverse selectivity properties emerge after learning two random input sequences. Single-trial raster plots display units that are active only during the right trial (left column), left trial (middle column), and at similar times during both trials (right column). Activity is sorted by the time of maximal activity, and this sorting is fixed across rows. Numbers indicate amount of neurons passing selection criteria (total neuron count is 40,000). $S = 2$, and all other parameters are as in Fig. 4.

random uncorrelated component ($J^{\text{rand}}$):

$$J_{ij}^{\text{rand}}(n) = J_{ij}^{\text{seq}} + J_{ij}^{\text{rand}}(n)$$

$$J_{ij}^{\text{seq}} = A^C g(\xi_{ij}^{s,\mu+1})$$

$$J_{ij}^{\text{rand}}(1) = \sigma_z \xi_{ij}^1,$$

$$J_{ij}^{\text{rand}}(n) = \lambda J_{ij}^{\text{rand}}(n-1) + \sqrt{1-\lambda^2} \sigma_z \xi_{ij}^n, \quad n > 1$$

where $\lambda$ controls the decay of random perturbations, and $\sigma_z$ controls their amplitude.

We explored the possibility that these changes in single neuron selectivity might be due to ongoing changes in synaptic connectivity, consistent with spine dynamics observed in cortex and hippocampus (36, 37). To probe whether dynamic reorganization of sequential activity is consistent with such ongoing synaptic dynamics, we simulated the effects of such ongoing synaptic dynamics by introducing random weight perturbations. We simulated the same network over a fixed period of days, where the weights on day $n$ are now defined as a sum of a fixed, sequential component ($J^{\text{seq}}$) as in Eq. (3), and a random uncorrelated component ($J^{\text{rand}}$):

$$J_{ij}(n) = J_{ij}^{\text{seq}} + J_{ij}^{\text{rand}}(n)$$

$$J_{ij}^{\text{seq}} = A^C g(\xi_{ij}^{s,\mu+1})$$

$$J_{ij}^{\text{rand}}(1) = \sigma_z \xi_{ij}^1,$$

$$J_{ij}^{\text{rand}}(n) = \lambda J_{ij}^{\text{rand}}(n-1) + \sqrt{1-\lambda^2} \sigma_z \xi_{ij}^n, \quad n > 1$$

where $\lambda$ controls the decay of random perturbations, and $\sigma_z$ controls their amplitude.

In Fig. 6a-c, we examine the reorganization of activity during a 30-day period. In Fig. 6a, sorted sequential activity is plotted for retrieval on day 1, followed by that during the 30 day mark. We find that sequential activity shows progressive reorganization, with the sequence on day 30 being only weakly correlated (but significantly, with a Pearson correlation of $\sim 0.4$) with the sequence on day 1. Critically, the underlying stored sequence can be read out equally well on both days, as shown by the pattern correlation values plotted below. In Fig. 6b, representative single unit activity is shown on the first, second, and final simulated day for four neurons, showing the emergence and disappearance of activity. To measure how quickly sequential activity decorrelated across time, for each day we plotted the average correlation of single unit activity profiles with those on day 1 or 30 (Fig. 6c). For the parameters chosen here, sequences decorrelate in a few days, such that by day 15, the correlation has reached a stable baseline. This qualitatively matches the ensemble activity correlation times reported in mouse CA1 (38).

**Sequence retrieval in excitatory-inhibitory spiking networks.**

For simplicity, we had previously chosen to neglect two key
features of neuronal circuits: the separation of excitation and inhibition, and action potential generation. To investigate whether our results are valid in more realistic networks, we developed a procedure to map the dynamics of a single population rate network onto a network of current-based leaky-integrate-and-fire (LIF) neurons with separate populations of excitatory and inhibitory units. To implement sequence learning in two population networks, we made the following assumptions: (1) learning takes place only between excitatory-to-excitatory recurrent connections; (2) to implement the sign constraint in those connections, we apply to the Hebbian connectivity matrix (Eq. 3) a non-linear synaptic transfer function that imposes a non-negativity constraint - specifically, we use a simple rectification (see Supp. Info); (3) inhibition is faster than excitation. Under these assumptions, we reformulate the two population network into an equivalent single population network with two sets of connections: one representing the excitatory-to-excitatory inputs, and the other representing the inhibitory feedback. We compute the effective inhibitory drive that balance the average positive excitatory recurrent drive arising from rectification, and in turn excitatory-to-inhibitory and inhibitory-to-excitatory connections strengths in the original two population model that achieve this balance.

Figure 7 shows the results of applying this procedure in two steps (from a one population to two population rate model, and then from a two population rate model to a two population LIF network - see Supp. Info for details). In Fig. 7a, sequential activity is shown from a single population rate network of size $N = 20,000$ with a rectified linear rate transfer function. In Fig. 7b, we transform this population into excitatory units by rectifying their synaptic weights, and add an additional population of $N_I = 5,000$ inhibitory units, with random sparse connections to and from the excitatory units. Finally, in Fig. 7c, we construct a current-based spiking network using parameters derived from LIF transfer functions. These LIF transfer functions were fit to a bounded region of the rectified linear transfer functions in panel (b) (see Supp. Info). Connectivity matrices are identical for panels (b) and (c) up to a constant scaling factor. Raster plots of excitatory neurons reveal that activity is sparse, with individual neurons firing strongly for short periods of time. Inhibitory neurons fire at rates that are much less temporally modulated than excitatory neurons, and inherit their fluctuations in activity from random connections from the excitatory neurons.

Discussion

We have investigated a simple temporally asymmetric Hebbian learning rule in rate and spiking networks models, and developed a theory of the transient activity observed in these networks. Using this theory, we have computed analytically the storage capacity for sequences in the case of a bilinear learning rule, and derived parameters of the neuronal transfer function (threshold and gain) that maximize capacity.

We have found that a variety of temporal features of population activity in the network model match experimental observations, including those made from single trials of activity and across multi-day recording sessions. A bilinear learning rule produces sequences that qualitatively match recorded distributions of single neuron preferred times and tuning widths in cortical and hippocampal areas. Tuning widths increase over the course of retrieval, and activity peaks are more concentrated at earlier segments of retrieval (5, 27, 28). With a nonlinear learning rule, sequential activity is temporally sparse and more uniform in time, consistent with findings in Nucleus HVC of zebra finch (39, 40). To mimic the effects of synaptic turnover and ongoing learning across multi-day recording sessions, we continually perturbed synaptic connectivity across repeated trials of activity. We show that single unit activity can dramatically reorganize over this time period while maintaining a stable readout of sequential activity at the population level, consistent with recent experimental findings (29, 30). Finally, we have developed a procedure to map the sequential activity in simple rate networks to excitatory-inhibitory spiking networks.

Comparison with previous models. A large body of work has explored the generation and development of sequential activity in network models (11–20, 41–43). While many of these models focus on the reproduction of temporally sparse synfire chain activity, few address the growing number of observations that (outside of specialized neural systems such as Nucleus HVC) sequences are rarely consistent with such a simplified description. Our modeling approach departs from this previous work in that we seek to provide a unified framework to account for the diversity of observed dynamics. Recent modeling efforts to reproduce quantitatively experimentally observed dynamics have used supervised learning algorithms, in which initially randomly connected networks are trained to reproduce exactly activity patterns collected from experiment (18). However, these learning rules are non-local, and there are currently no known biophysical mechanisms that would allow them to be implemented in brain networks. Here, we have shown that a much simpler and biophysically realistic learning rule is sufficient to reproduce many of experimentally
observed features of sequences.

The analytical methods we have used are generalizations of mean-field methods used extensively both in networks of binary neurons (44, 45) and in networks of rate units (35, 46). Using these methods, we were able, to our knowledge for the first time, to compute analytically the sequence storage capacity in rate models using a TAH learning rule. This allowed us to characterize how the storage capacity depends on the parameters of the transfer function, and to obtain the threshold and gain of the transfer function that optimizes this storage capacity. We found, as in fixed-point attractor networks (35, 44, 47), and networks of binary neurons storing sequences (45), that capacity scales linearly with the number of stored patterns. While our result is derived in the large N limit, we have shown that it predicts well the measured capacity for finite-size networks.

Storage and retrieval. In the present work, we have used a simplified version of a TAH rule in which only nearest neighbor pairs in the sequence of uncorrelated patterns modify the weights. This learning rule assumes that network activity is ‘clamped’ during learning by external inputs, and that recurrent inputs do not interfere with such learning. We found for the bilinear rule that, following learning, the speed of retrieval (i.e., the inverse time between recall of consecutive patterns) matches the time constant of the rate dynamics. Moreover, we showed empirically that this speed does not depend on the number of stored patterns, nor on the particulars of the neural transfer function. This suggests that if patterns are presented during learning at this same timescale, one that is consistent with STDP, then they can be retrieved at the same speed as they were presented, without any additional mechanisms (48). If patterns are presented on slower timescales, then retrieval of input patterns could proceed on a faster speed than in which they were originally presented. Such a mechanism could underlie the phenomenon of hippocampal replay during sharp-wave ripple events, in which previously-experienced sequences of place cell activity can reactivate on significantly temporally-compressed timescales (49). Whether additional cell-intrinsic or network mechanisms allow for dynamic control of retrieval speed remains an open question.

Retrieval in the present network model bears similarities to that of the functionally feedforward rate networks analyzed by Goldman (50), in which recurrent connectivity was constructed by using an orthogonal random matrix to rotate a feedforward chain of unity synaptic weights. Retrieval was initiated by briefly stimulating an input corresponding to the first orthogonal pattern in the chain. The retrieval dynamics in those networks correspond exactly to those of the overlap dynamics for the bilinear rule in the presently studied network if a constant gain of $G = 1$ is assumed (see Supp. Info). The key difference in our model is that the gain $G$ is dynamic (Eq. 5), as it depends on the strength of network activity and quality of retrieval (Eq. 4), and acts in concert with network dynamics as a homeostatic-like mechanism to maintain a value slightly above one (see Supp. Info for detailed discussion). This mechanism solves the fine-tuning problem noted in reference (50), in which small perturbations away from $G = 1$ in the form of a uniform weight perturbation result in either

![Fig. 7. Sequence retrieval in excitatory-inhibitory networks.](image-url)
exponential growth or decay of activity. Weight values do not need to be constrained in our network to achieve successful recall, as an appropriate $G$ is arrived at in a dynamic fashion.

We also found that retrieval is highly robust to initial Gaussian perturbations. Contrary to expectation, storage of additional non-retrieved patterns/sequences confers significantly increased robustness against a perturbation of the same magnitude (Supplementary Figure 2). This can be understood by examining the mean-field equations, and noticing that a larger $\alpha$ provides a greater range of convergence in initial overlap values to a gain that reaches an equilibrium value around one (see Supp. Info for detailed discussion).

While we have made several simplifying assumptions in the interest of analytical tractability, many of these assumptions do not hold in real neural networks. Biological neurons receive spatially and temporally correlated, continuous input, and STDP rules at the synapse are unlikely to differentiate between externally-derived and recurrent input. Plasticity rules also operate over a range of temporal offsets, and not at a single discrete time interval as was assumed here (51). Understanding how each of these properties impact the findings presented will be the subject of future work.

Temporal characteristics of activity. We have found two emergent features of retrieval with a bilinear learning rule that are consistent with experiment: a broadening of single unit activity profiles with time, and an over-representation of peaks at earlier times of the sequence. These features have been reported in lateral and medial PFC, and in hippocampal CA3 (5, 27, 28). Our theory provides a simple explanation for both of these features. Mean-field theory analysis shows that the temporal profile of the overlaps control to a large extent the profiles of single units. Specifically, the mean input current of a single unit is a linear combination of overlaps, with coefficients equal to the pattern values shown during learning. For Gaussian patterns, the overlaps rise transiently through an effective delay-line system, where the activation of each overlap is fed as input to drive the next. As the overlaps decay on the same timescale that they rise, the effective decay time grows for inputs later in the sequence, resulting in more broadly tuned profiles. This is also reflected in the longer decay time in the two-point rate correlation function at later times (Supplementary Figure 5b). The over-representation of peaks at earlier times can be accounted for by a decrease in overlap magnitude and population firing rate throughout recall, as fluctuations in activity earlier in recall are more likely to rise above threshold and establish a peak.

We have also investigated the effects of storing sparse patterns using a nonlinear learning rule. The form of this rule bears similarities to that inferred from electrophysiological data in inferior temporal cortex (34), and has been shown to be optimal for storing fixed point attractors in the space of separable, sigmoidal learning rules (35). Sparse patterns lead to sequence statistics that are consistent instead with song-specific sequential activity in Nucleus HVC of zebra finch (39, 40).

Dynamics of sequential activity on long time scales. We showed that simulating the effects of synaptic turnover by adding random weight perturbations can account for sequences that are unstable at the single unit level, but which still maintain information as a population, consistent with experiment (29, 30, 38). We assumed that synaptic connectivity has both a stable component (storing the memory of sequences) and an unstable component that decays over time scales of days. While this is roughly consistent with observations of spine dynamics (e.g. (52)), it is unclear what mechanisms would give rise to such a separation of stable and unstable components. We also have not addressed the learning process that would lead to stable sequence retrieval. However, in many systems sequences appear to be learned from initially unstructured activity through activity-dependent mechanisms. In motor cortex, task-specific sequential activity becomes more reproducible with repeated exposure to the task (53). In Nucleus HVC of juvenile zebra finch, activity progresses through distinct stages before arriving at a stable sequence (54). Future work should explore online TAH learning, and study how the temporal dynamics of activity change over the course of learning.

From rate to spiking network model. While our initial results were derived in rate models that do not respect Dale’s law, we have shown that it is possible to build a network of spiking excitatory and inhibitory neurons that stores and retrieves sequences in a qualitatively similar way as in the rate model. In the spiking network, the Hebbian rule operates in excitatory-to-inhibitory connections only, while all connections involving inhibitory neurons are fixed. With this architecture, most E neurons show strong temporal modulations in their firing rate, while inhibitory neurons show much smaller temporal modulations around an average firing rate. These firing patterns are consistent with observations in multiple neural systems of smaller temporal modulations in firing of inhibitory neurons, compared to excitatory cells (9, 55, 56).

Overall, our paper demonstrates that a simple unsupervised Hebbian learning rule leads to storage and retrieval of sequences in large networks, with a phenomenology that mimics multiple experimental observations in multiple neural systems. The next step will be to investigate how the connectivity matrix used in this study can be obtained through biophysically realistic online learning dynamics.

Materials and Methods
See Supplementary Information for detailed materials and methods.

Simulation details. Rate and spiking network simulations were performed using custom C++ and Python routines. Code to generate the figures is available upon request.

Author contributions. MG, UP and NB designed the study. MG, UP and NB performed analytical calculations. MG performed numerical simulations of the model and analyzed the results. MG and NB wrote the paper with contributions from UP.

Correspondence should be addressed to nicolas.brunel@duke.edu

ACKNOWLEDGMENTS. This work was supported by NIH R01 EB022891 and ONR N00014-16-1-2327. We thank Henry Greenside and Harel Shouval for their comments on a previous version of the manuscript, and all members of the Brunel group for discussions.

1. C. D. Harvey, P. Coen, and D. W. Tank. Choice-specific sequences in parietal cortex during a virtual navigation decision task. Nature, 484:62–68, 2012.
2. J. Wang, D. Narain, E. A. Hosseini, and M. Jazayeri. Flexible timing by temporal scaling of cortical responses. Nat. Neurosci., 21(1):102–110, Jan 2018.
3. A. K. Lee and M. A. Wilson. Memory of sequential experience in the hippocampus during slow wave sleep. Neuron, 36(6):1183–1194, Dec 2002.
4. E. Pastalkova, V. Itskov, A. Amarasingham, and G. Buzsaki. Internally generated cell assembly sequences in the rat hippocampus. Science, 321:1322–1327, 2008.
5. D. M. Saltz, Z. Tigany, S. Khansabish, A. Kohley, D. Sheehan, M. W. Howard, and H. Eichenbaum. Time cells in Hippocampal Area CA3. J. Neurosci., 36(28):7476–7484, 07 2016.

6. J. G. Heys and D. A. Dombeck. Evidence for a subcircuit in medial entorhinal cortex representing elapsed time during immobility. Nat. Neurosci., 21(11):1574–1582, 11 2018.

7. G. Busak. Hippocampal sharp waves: their origin and significance. Brain Res., 398:242–252, 1986.

8. H. Ahkilaghour, J. Wiskerke, J. Y. Choi, J. P. Tallarida, J. Au, and I. B. Witten. Dissociated sequential activity and stimulus encoding in the dorsomedial striatum during spatial working memory. Elife, 5, 09 2016.

9. R. H. Hahnloser, A. A. Kozhevnikov, and M. S. Fee. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature, 419(6902):65–70, 2002.

10. M. M. Churchland and K. V. Shenoy. Temporal complexity and heterogeneity of single-neuron activity in prefrontal and motor cortex. J. Neurophysiol., 97(5):4235–4257, Jun 2007.

11. S.-I. Amari. Learning patterns and pattern sequences by self-organizing nets of threshold elements. IEEE Trans. Comput., 21:1197–1206, 1972.

12. A. Herz, B. Sulzer, R. Kühn, and J. L. van Hemmen. The Hebb rule: Storing static and dynamic objects in an associative neural network. Europhys. Lett., 4:167–173, 1987.

13. H. Gutfreund and M. Mezard. Processing of temporal sequences in neural networks. Phys. Rev. Lett., 61(2):235–238, Jul 1988.

14. B. Tirozzi and M. Tsodyks. Chaos in highly diluted neural networks. EPL (Europhysics Letters), 14:727, April 1991.

15. D. J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys. Rev. Lett., 55:1530–1531, 1985.

16. W. Gerstner, R. N. van der Toorn, and J. L. van Hemmen. Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. Biol. Cybern., 65(6):503–515, 1993.

17. Z. Nadaday, H. Hirase, A. Czurko, J. Coicovari, and G. Bizusaki. Replay and time compression of recurring spike sequences in the hippocampus. J. Neurosci., 19(21):9497–9507, Nov 1999.

18. M. S. Goldman. Memory without feedback in a neural network. Neural, 61:621–634, 2009.

19. L. F. Abbott and S. B. Nelson. Synaptic plasticity: taming the beast. Nat. Neurosci., 3 Suppl:1178–1183, Nov 2000.

20. J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Walker, and K. Svoboda. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. Nature, 428(6971):788–794, 2004.

21. A. J. Peters, S. X. Chen, and T. Korniayeva. Emergence of reproducible spatiotemporal activity during motor learning. Nature, 510(7504):263–267, Jun 2014.

22. T. S. Okubo, E. L. Mackenzie, H. L. Payne, G. F. Lynch, and M. S. Fee. Growth and splitting of neural sequences in songbird vocal development. Nature, 528(7582):352–357, Dec 2015.

23. G. Kosche, D. Vallentin, and M. A. Long. Interplay of inhibition and excitation shapes a premotor neural sequence. J. Neurosci., 35(3):1217–1227, Jan 2015.

24. S. Lim, J. L. McKee, L. Woloszyn, Y. Amit, D. J. Freedman, D. L. Sheinberg, and N. Brunel. Novel activity patterns in Parietal Cortex. J Cogn Neurosci, 27(12):5663–5671, 12 2017.

25. M. A. Picardo, J. Merel, K. A. Katlowitz, D. Vallentin, D. E. Okobi, S. E. Benezra, R. C. Clary, E. A. Pnevmatikakis, L. Paninski, and M. A. Long. Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch. Neuron, 90(4):877–892, 05 2016.
Supplementary Information

Contents

1 Rate network .......................... 2
  1.1 The network model .................. 2
  1.2 Learning rule ........................ 2

2 Mean-field theory of sequential activity 3
  2.1 Overlap with network activity 3
  2.2 Average squared firing rate 5
  2.3 Retrieval properties 6
  2.4 Sequence capacity 7
    2.4.1 Conditions on transfer function for successful retrieval 7
    2.4.2 Maximum load 7
  2.5 Sequence robustness 8

3 Excitatory-inhibitory rate network 9
  3.1 Simulation procedure ............... 10

4 Spiking network ....................... 11
  4.1 Simulation procedure ............... 11

5 Supplemental Procedures .......... 12

6 Supplemental Figures .......... 15

7 Parameter tables ..................... 20
1 Rate network

1.1 The network model

We consider a network of $N$ randomly connected rate units, each of which make an average of $K$ connections with other units. Unless otherwise noted, for all rate network simulations, $N = 40,000$ and $K = 200$. In analytical calculations, we take $N \to \infty$ and $K \to \infty$ while $K/N \to 0$. In all but Figure 7, units are governed by the following dynamics

$$\tau \frac{dr_i}{dt} = -r_i + \phi (h_i), \quad h_i = \sum_{j \neq i} J_{ij} r_j$$

where $\tau$ is the time constant of the rate dynamics, and $h_i$ is the total synaptic input to neuron $i$. Synaptic input is transformed through a sigmoidal neural transfer function

$$\phi(x) = \frac{r_{\text{max}}}{2} \left( 1 + \text{erf} \left( \frac{x - \theta}{\sigma \sqrt{2}} \right) \right)$$

where $r_{\text{max}}$ is the maximal firing rate, $\theta$ is the input at which the firing rate is $r_{\text{max}}/2$, and $\sigma$ controls the gain of the transfer function. In particular, for $\sigma \to 0$ the transfer function becomes the Heaviside function, $\phi(x) = 1$ for $x > \theta$, and zero otherwise.

1.2 Learning rule

The temporally asymmetric Hebbian learning rule produces the following connectivity matrix $J_{ij}$:

$$J_{ij} = A \frac{c_{ij}}{K} \sum_{s=1}^{S} \sum_{\mu=1}^{P-1} f(\xi_{s,\mu}^{i+1}) g(\xi_{s,\mu}^j)$$

where $A$ is a learning rate, $c_{ij}$ are i.i.d. Bernoulli random variables ($c_{ij} = 1$ with probability $c$ and 0 otherwise) and $K = Nc$ represents the average in-degree of a neuron. This connectivity matrix stores $S$ sequences of $P$ patterns $\{\xi_{s,\mu}\}$, where $\xi_{s,\mu}$ can be thought as the input to neuron $i (i = 1, \ldots, N)$ in pattern $\mu (\mu = 1, \ldots, P)$ of sequence $s (s = 1, \ldots, S)$, using a temporally asymmetric Hebbian rule. With such a rule, synapses are modified for each pair of successive patterns in the sequence by an amount $f(\xi_{s,\mu}^{i+1}) g(\xi_{s,\mu}^j)$, where the function $f$ describes the dependence of the learning rule on input to the post-synaptic neuron, and $g$ describes the dependence on input to the pre-synaptic neuron. The patterns are identically and independently distributed (i.i.d.) as $\xi_{s,\mu} \sim \mathcal{N}(0,1)$. In the first part of the paper (Figures 1-3 and Supplementary Figures 1-3,5-7) we use a bilinear rule, $f(x) = g(x) = x$. In Figure 4-7 and Supplementary Figure 4 we use a non-linear learning rule, with

$$f(x) = \begin{cases} q_f & \text{if } x > x_f \\ q_f - 1 & \text{if } x \leq x_f \end{cases}$$

$$g(x) = \begin{cases} q_g & \text{if } x > x_g \\ q_g - 1 & \text{if } x \leq x_g \end{cases}$$

The variable $x_f$ ($x_g$) defines the threshold separating potentiation and depression when post (pre) synaptic firing rates are varied, while $q_f$ ($q_g$) controls the strength of plasticity at high post (pre) firing rate. In order to keep the average sum over incoming connection strengths zero, we set $q_g = F_z(x_f)$, where $F_z$ is the cumulative distribution function of a standard Gaussian [1].
# 2 Mean-field theory of sequential activity

## 2.1 Overlap with network activity

In this section we derive a mean field theory for a network where stored patterns are Gaussian and the learning rule is bilinear, i.e. \( f(x) = x \) and \( g(y) = y \). With this choice, the connectivity matrix is given by

\[
J_{ij} = \frac{c_{ij}}{Nc} \sum_{s=1}^{S} \sum_{\mu=1}^{P-1} \xi_{s,\mu+1}^{s} s_{j}. \tag{6}
\]

where the learning rate \( A \) in Eq. (3) has been absorbed in the parameters of the transfer function (Eq. 2), by rewriting \( \sigma* = A^\sigma \) and \( \theta* = A^\theta \). The input current to neuron \( i \) at time \( t \) is given by a weighted sum of the firing rates of all its pre-synaptic neurons:

\[
h_i(t) = \sum_{j \neq i} J_{ij} r_j(t). \tag{7}
\]

We start by assuming that the number of sequences \( S \) is large, \( S \gg 1 \), and the number of patterns per sequence is much smaller than the in-degree, \( P \ll cN \). Assuming that the dynamics start from an initial condition that is correlated with the first pattern of sequence \( s \), i.e. \( \vec{\xi}_s,1 \), the input current can be re-written as

\[
h_i(t) = \sum_{\mu=1}^{P-1} \xi_{i}^{\mu+1} \frac{1}{Nc} \sum_{j \neq i} c_{ij} \xi_{j}^{\mu} r_j(t) + Y_i(t) \tag{8}
\]

where \( Y_i \) describes the ‘noise’ term,

\[
Y_i(t) = \frac{1}{Nc} \sum_{l \neq s} \sum_{\mu=1}^{P-1} \xi_{l}^{\mu+1} \sum_{j \neq i} c_{ij} \xi_{j}^{\mu} r_j(t). \tag{9}
\]

In the large \( cN \) limit, due to the law of large numbers, and using the fact that \( P \ll cN \), the first term in Eq. (8) converges in probability to

\[
\sum_{\mu=1}^{P-1} \xi_{i}^{\mu+1} q_{i}^{\mu}(t), \tag{10}
\]

where the \( q_{i}^{\mu} \) are given by

\[
q_{i}^{\mu}(t) = \frac{1}{N} \sum_{j=1}^{N} \xi_{j}^{\mu} r_j(t). \tag{11}
\]

The overlaps \( \{q_{i}^{\mu}(t)\}_{\mu=1}^{P} \) are our first \( P \) order parameters. They describe how correlated the network state is with the stored patterns \( \xi_{i}^{1}, \xi_{i}^{2}, \ldots, \xi_{i}^{P} \). We assume that the network state is uncorrelated with the rest of stored patterns since \( q_{l}^{\mu}(t) \sim O(1/\sqrt{N}) \) for \( l \neq s \). The ‘noise term’ \( Y_i \) then has mean zero, and variance

\[
\text{Var}(Y_i(t)) = \alpha M(t) \tag{12}
\]

where the sequential load is defined by

\[
\alpha \equiv \frac{S(P-1)}{Nc} \tag{13}
\]

while \( M \), the mean of the squared firing rate, is an additional order parameter defined by

\[
M(t) = \frac{1}{N} \sum_{j=1}^{N} r_j^2(t). \tag{14}
\]
In this theory we assume that the total number of stored patterns is much larger than the number of patterns in a sequence, i.e. $S \gg 1$. We can now plug Eqs. (8-14) in Eq. (1) to yield

$$\tau \frac{dr_i}{dt} = -r_i + \phi \left( \sum_{\mu=1}^{P-1} \xi_{i+1}^{\mu} q_{\mu}(t) + Y_i(t) \right). \quad (15)$$

Since all sequences are statistically equivalent, we have dropped the index $s$ corresponding to the particular sequence of concatenated patterns. The variable $Y_i(t)$ corresponds to the interference produced by stored patterns that do not belong to the sequence being retrieved (in this case, sequence $s$). By the central limit theorem, the variables $Y_i$ are approximately i.i.d. Gaussian random variables across neurons, i.e. $Y_i(t) \sim \mathcal{N}(0, \alpha M(t))$. Using equation (11) we get the following dynamical equations for the overlaps:

$$\tau \frac{dq_l}{dt} = -q_l + \int D\xi D y \xi^l \phi \left( \sum_{\mu=1}^{P-1} \xi_{l+1}^{\mu} q_{\mu}(t) + \sqrt{\alpha M(t)} y \right) \quad (16)$$

where $D y = e^{-y^2/2}/\sqrt{2\pi}$ and $D\xi = \prod_{i=2}^{P} D\xi^i$. The dynamical equation for the first overlap (i.e. $q_1$) simplifies to

$$\tau \frac{dq_1}{dt} = -q_1. \quad (17)$$

where we have used the fact that $\xi^1$ does not appear in the argument of $\phi$ in the r.h.s. of Eq. (16).

To simplify the equations for the other overlaps, we define

$$R^2_l(t) = \sum_{k \neq l}^{P-1} q_k^2(t) + \alpha M(t). \quad (18)$$

Since the stored patterns are Gaussian, we write Eq. (16) as

$$\tau \frac{dq_l}{dt} = -q_l + \int D\xi D x \xi^l \phi \left( \xi^l q_{l-1}(t) + R_{l-1}(t) x \right) \quad l = 2, \ldots, P \quad (19)$$

where $\xi^l$ and $x$ are independent standard normal random variables. Using the change of coordinates

$$v = \frac{\xi^l q_{l-1} + x R_{l-1}}{\sqrt{q_{l-1}^2 + R_{l-1}^2}}$$
$$u = \frac{\xi^l R_{l-1} - x q_{l-1}}{\sqrt{q_{l-1}^2 + R_{l-1}^2}}$$

where $u$ and $v$ are also uncorrelated standard normal random variables, equation (19) becomes

$$\tau \frac{dq_l}{dt} = -q_l + q_{l-1} G \left( \sum_{\mu}^{P-1} q_{\mu}^2(t) + \alpha M(t) \right) \quad l = 2, \ldots, P \quad (20)$$

where

$$G(x) \equiv \int_{\mathbb{R}} Dv D\phi(v\sqrt{x}) \sqrt{x}. \quad (21)$$

Using the neural transfer function of Eq. (2), we can use integration by parts to simplify:

$$G(x) \equiv \frac{1}{\sqrt{2\pi(\sigma^2+x)}} \exp \left( -\frac{\theta^2}{2(\sigma^2+x)} \right). \quad (22)$$
By defining the ‘delay line’ matrix as $L_{ij} = \delta_{i,j+1}$ we can also write equation (21) in vectorial form

$$\tau \frac{d\vec{q}}{dt} = -\vec{q} + G \left( \sum_{\mu} q_{\mu}^{2}(t) + \alpha M(t) \right) L\vec{q}. \quad (23)$$

When $S$ is of order 1, and $P$ is of order $Nc$, the variance of the two terms in the r.h.s. of Eq. (8) are of the same order, and in particular the variance of the first term no longer vanishes. In this scenario, we assume that at any given time during retrieval of a sequence, the network state has a finite overlap with only a small fraction of the patterns in the retrieved sequence. The ‘signal’ term in Eq. (8) then needs to include only those patterns, while the noise term $Y_i(t)$ includes all the patterns that are far from the patterns that have currently a finite overlap with the network state. All the resulting equations are therefore the same as the ones derived above in the case $S \gg 1$.

### 2.2 Average squared firing rate

The system of equations describing the dynamics of the overlaps, Eq. (23), depends on $M(t)$. The next step is therefore to find a self-consistent mean-field equation for $M$:

$$M(t) = \frac{1}{N} \sum_{i} r_i^2(t). \quad (24)$$

Taking the derivative with respect to time, we find

$$\tau \frac{dM}{dt} = \tau \frac{2}{N} \sum_{i} r_i(t) \frac{dr_i(t)}{dt} = -2M + \frac{2}{N} \sum_{i} r_i(t) \phi(h_i(t)) \quad (25)$$

$$= -2M + e^{-t/\tau} \frac{2}{N} \left[ \sum_{i} r_i(0) \phi(h_i(t)) + \sum_{i} \int_{0}^{t} du \frac{d}{\tau} e^{u/\tau} \phi[h_i(u)] \phi(h_i(t)) \right] \quad (26)$$

where we have used the general solution of the ODE for $r_i$:

$$r_i(t) = e^{-t/\tau} \left( r_i(0) + \int_{0}^{t} du \frac{d}{\tau} e^{u/\tau} \phi[h_i(u)] \right). \quad (27)$$

Recalling Eqs. (8-11), the field $h_i(t)$ in the first sum in the r.h.s. of Eq. (27) can be described as a random Gaussian variable, whose variance $\sigma_h^2(t)$ is the sum of the norm of the vector of overlaps, plus the noise term due to interference with the other stored sequences,

$$\sigma_h^2(t) = \sum_{\mu} q_{\mu}^{2}(t) + \alpha M(t). \quad (29)$$

In the large $N$ limit, this sum can be replaced by an integral over the Gaussian distribution of $h_i(t)$. We turn now to the other sum in the r.h.s. of Eq. (27). This sum can also be replaced in the large $N$ limit by an integral over the joint distribution of $h(u)$ and $h(t)$, which is a correlated bivariate Gaussian distribution. Here we write $h(u)$ and $h(t)$ in terms of 3 uncorrelated Gaussian variables, where each are the sum of an independent and shared variable. Specifically, we write $h_i(u) = \sigma_a(u)(a(u,t)x + b(u,t)z)$, and $h_i(t) = \sigma_b(t)(a(t,u)y + b(t,u)z)$. It remains now to find the time-dependent parameters $a$ and $b$ in terms of the original order parameters. By computing the variance and covariance of the fields using both the original mean-field description and our Gaussian formulation, we can solve for $a$ and $b$.
\begin{align*}
\langle h_i(t)^2 \rangle &= \sum_{k}^{P-1} q_k^2(t) + \alpha M(t) = \sigma_h(t)^2(a_i^2(u, t) + b_i^2(u, t)) \tag{30} \\
\langle h_i(t) h_i(u) \rangle &= \sum_{k}^{P-1} q_k(t) q_k(u) + \alpha C(t, u) = \sigma_h(u) \sigma_h(t) \tag{31}
\end{align*}

where in Eqs. (30,31) we have used the fact that patterns \( \{\xi^u\} \) are independent, and that \( x, y, \) and \( z \) are independent and uncorrelated. We have also defined \( C(t, u) = \frac{1}{N} \sum_i r_i(u) r_i(t) \). Solving Eqs. (30,31), we find

\begin{align*}
a(t, u) &= \sqrt{1 - \rho(u, t)} \\
b(t, u) &= \sqrt{\rho(u, t)} \tag{32}
\end{align*}

where we have defined

\[ \rho(u, t) = \frac{\sum_k^{P-1} q_k(t) q_k(u) + \alpha C(t, u)}{\sigma_h(u) \sigma_h(t)}. \]

Averaging over the statistics of \( x, y, \) and \( z \) we obtain:

\[ \tau \frac{dM}{dt} = -2M + 2e^{-t/\tau} \int D\xi Dz \phi(\xi) \phi(\sigma_h(t) z) \]

\[ + 2 \int_0^t du \frac{e^{-u/\tau}}{\tau} \int DxDyDz \phi(\sigma_h(u) \sqrt{1 - \rho(u, t)} x + \sigma_h(u) \sqrt{\rho(u, t)} z) \]

\[ \phi(\sigma_h(t) \sqrt{1 - \rho(t, u)} y + \sigma_h(t) \sqrt{\rho(t, u)} z). \tag{34} \]

The time evolution of \( C(t, u) \) can be derived similarly as for Eq. (27),

\[ \tau \frac{dC(t, u)}{dt} = -C(t, u) + e^{-u/\tau} \int D\xi Dz \phi(\xi) \phi(\sigma_h(t) z) \]

\[ + \int_0^u dv \frac{e^{(v-u)/\tau}}{\tau} \int DxDyDz \phi(\sigma_h(v) \sqrt{1 - \rho(v, t)} x + \sigma_h(v) \sqrt{\rho(v, t)} z) \]

\[ \phi(\sigma_h(t) \sqrt{1 - \rho(t, u)} y + \sigma_h(t) \sqrt{\rho(t, u)} z). \tag{35} \]

The dynamics of the average firing rate \( \bar{r}(t) = \frac{1}{N} \sum_i r(t) \) are given by:

\[ \tau \frac{d\bar{r}}{dt} = \int Dz \phi(\sigma_h(t) z) + e^{(-t/\tau)} \left[ -\bar{r}(0) - \int_0^t du \frac{e^{(u/\tau)}}{\tau} \int Dz \phi(\sigma_h(u) z) \right]. \tag{36} \]

In Supplementary Figure 5, solutions to equations (34-35) are plotted along with numerically computed values from a full network simulation of size \( N = 50,000 \).

\subsection{2.3 Retrieval properties}

To study the time-dependent properties of retrieval, we can analyze Eq. (20) for the case in which the gain is constant: \( G = 1 + \epsilon \). With this approximation, it is straightforward to analytically derive several properties of the recalled sequence, including retrieval speed and the scaling of overlap widths.

For \( l = 2, \ldots, P \), we have:

\[ \tau \frac{dq_l}{dt} = -q_l + (1 + \epsilon) q_{l-1} \]

which leads by recursion to

\[ q_l(t) = q_l(t=0)(1 + \epsilon)^{l-1} \frac{\exp(-t/\tau)}{\tau^{l-1}(l-1)!} \tag{37} \]
From this equation, we can easily see that for \( \epsilon > 0 \), sequences grow unbounded, while for \( \epsilon < 0 \) sequences decay. Furthermore, by computing the derivative of \( q_l(t) \) with respect to time, we find that the overlaps \( q_l \) peak at \( t = \tau(l-1) \), which shows that the sequence progresses at a speed proportional to \( \tau \).

To determine the widths of the overlaps vs time curves, we compute the standard deviation of the distribution given by \( q_l(t)/\int q_l(u) du \). We find that the mean is equal to \( \tau l \), while the standard deviation is \( \tau \sqrt{l} \). Thus, the width of the overlap vs time curves is proportional to the square root of the peak time. This prediction agrees well with the empirically measured values for full network simulations (see Figure 2 of main article).

2.4 Sequence capacity

In the following sections, we will calculate the maximum number of sequences that a network can store and successfully retrieve as a function of network parameters.

2.4.1 Conditions on transfer function for successful retrieval

In the previous section we found that recalled sequences with a constant gain above one grow unbounded, while those with gain below one decay. To make the notion of retrieval more precise, we can define a sequence of asymptotically long size as being successfully retrieved if the gain converges to a value larger or equal to one during the sequence.

The shape of the gain function \( G \) in Eq. (22) dictates whether it can take values that are greater than one, and its form depends on the transfer function parameters \( \theta \) and \( \sigma \). By examining the dependence of \( G \) on these two parameters, we can bound the region for which successful retrieval is possible. Note that the shape of \( G \) only determines whether it is in principle possible to successfully retrieve a sequence. Whether or not the temporal trajectory of the gain rises above one during recall depends on the initial condition of the overlaps and mean squared firing rate, as well as the number of stored sequences. We find that \( G \) is either a monotonically decreasing function (if \( G'(0) < 0 \)), or that it is a monotonically increasing function for small values of its argument, reaching a maximum \( G_{\text{max}} \), and then monotonically decaying towards zero (see Fig. 6c for examples). Thus, successful retrieval is possible if at least one of the following conditions is satisfied:

1. \( G(0) \) is larger than one
2. \( G'(0) \) is positive, and the maximum of \( G \), \( G_{\text{max}} \), is larger than one

These criteria lead to the following conditions for \( \theta \) and \( \sigma \), using \( G \) defined in Eq. (22):

1. If \( |\theta| < \sigma \sqrt{-\log(\sqrt{2\pi}\sigma^2)} \), then \( G(0) > 1 \)
2. If \( |\theta| > \sigma \), then \( G'(0) > 0 \)
3. If \( |\theta| < 1/\sqrt{2\pi} \), then \( G_{\text{max}} > 1 \).

These conditions define 6 possible regions, that are plotted in Supplementary Figure 6a. In region D and E (see figure), condition 1 is satisfied (\( G(0) > 1 \)), so retrieval is possible for vanishingly small initial overlaps, provided \( \alpha \) is sufficiently small. In region F, conditions 2 and 3 are satisfied but not 1 (\( G(0) < 1 \), but \( G \) initially increases with its argument and reaches a peak \( G_{\text{max}} > 1 \)), so retrieval is possible for initial overlaps of small but finite size, again provided \( \alpha \) is small enough. In regions A, B, and C retrieval is not possible, as both \( G(0) < 0 \) and \( G_{\text{max}} < 1 \).

2.4.2 Maximum load

As already mentioned, a necessary condition for sequence retrieval is that the function \( G(x) \) has a maximum that is larger than 1. If this condition is satisfied, then let \( x_c \) be the largest value of \( x \) such that \( G(x) = 1 \). For \( \alpha \rightarrow 0 \), the condition \( G_{\text{max}} > 1 \) guarantees that a sequence can be retrieved, provided the initial overlap
with the first pattern is large enough. If $G\left(\sum_{\mu=1}^{P-1} q_{\mu}^2(0)\right) > 1$, the overlaps initially grow until the norm of the overlap vector stabilizes at a value where $G\left(\sum_{\mu=1}^{P-1} q_{\mu}^2(t)\right) \sim 1$.

When $\alpha > 0$, the gain now depends on the additional ‘noise term’ $\alpha M(t)$, since the argument of $G$ is now $\sum_{\mu=1}^{P-1} q_{\mu}^2(t) + \alpha M(t)$. If this noise term is larger than $x_c$, then the gain will be smaller than one for any $\bar{q}(t)$, and therefore sequences will decay starting from any initial condition. The maximal value of $\alpha$ for which sequences can be retrieved is therefore given by

$$G(\alpha_c M) = 1,$$

where $M$ is given by its steady state value when $\bar{q} = 0$:

$$M = \int Dv \phi^2 \left( v \sqrt{\alpha_c M} \right).$$

When $\alpha < \alpha_c$, there exists an overlap vector for which $\sum_{\mu=1}^{P-1} q_{\mu}^2(0) + \alpha M = x_c$. With suitable initial conditions, the dynamics of the network will converge to a vector with such a norm, and sequences will be retrieved. When $\alpha > \alpha_c$, $\sum_{\mu=1}^{P-1} q_{\mu}^2(0) + \alpha M > x_c$ and hence $G < 1$ for any $\bar{q}$, and therefore stored sequences cannot be retrieved.

Equations (38-39) were solved numerically. Solutions are plotted in Fig. 3 of the main text as a function of $\sigma$ for a few values of $\theta$, and in Supplementary Fig. 7 in the whole $\sigma$-$\theta$ plane.

2.5 Sequence robustness

We assessed sequence robustness by perturbing the initial condition of the rates by a Gaussian pattern: $r(t = 0) = \phi(x^{1.1} + \sigma z_0)$, where $\sigma$ controls the standard deviation of the standard Gaussian perturbation $z_0 \sim \mathcal{N}(0, 1)$. We focused specifically on transfer function parameters in region F (see section 2.4.1), as sequences with parameters falling in regions D and E will still be retrieved for arbitrarily small initial overlaps (not shown). Surprisingly, we find in region F that increasing $\alpha$ results in higher robustness to perturbations (Supplementary Figure 2). Mean-field analysis of the perturbation provides an explanation for this effect. The perturbed initial conditions for $q$ and $M$ are given by:

$$q_1(t = 0) = \int Dv Dz \phi(v + \sigma z)$$
$$M(t = 0) = \int Dv \phi(v \sqrt{1 + \sigma^2})^2$$

with all other overlaps equal to zero, $q(t = 0) = 0, t \neq 1$. Increasing perturbation strength increases slightly $M(0)$ while more dramatically decreasing $q_1(0)$. Looking at the mean-field dynamics of $M$ in Eqs. (34) and (35), we can also see that the effect of the perturbation decreases exponentially with $\tau$.

As explained in the previous section, successful recall depends on the argument of $G$: $\sum_{\mu=1}^{P-1} q_{\mu}^2(t) + \alpha M$, specifically on whether or not this argument stabilizes to a value around $x_c$. For small $\alpha$, this argument is dominated by the overlap norm, and as the perturbation strength increases, the initial value of this norm decreases. When the initial norm becomes too small, the dynamics of $q$ and $M$ decay to a region where $G(\sum_{\mu=1}^{P-1} q_{\mu}^2(0) + \alpha M) < 1$, and the sequence is not retrieved. For example, for $S = 1$ and $P = 16$ corresponding to an $\alpha = 0.08$, when $\sigma = 2.5$ the initial argument of $G$ is too small to converge to $x_c$ (Supplementary Figure 2, left).

For large $\alpha$, the initial argument is dominated by $\alpha M(0)$. Increasing perturbation strength decreases the initial norm of the overlaps, as in the case for small $\alpha$, but $\alpha M(0)$ can remain large enough such that the argument converges in time to $x_c$. This phenomenon ensures that the critical initial overlap below which a sequence is not retrieved (due to the dynamics of the argument not converging to $x_c$) is smaller, and thus confers a higher tolerance for perturbation strength.
3 Excitatory-inhibitory rate network

So far, we have analyzed a simplified rate network that ignores the separation between excitation and inhibition. To assess whether our results hold also in networks that obey Dale’s law, we developed a procedure to build an E/I network that can store and retrieve sequences. Our goal is to transform a rate network of the form

\[ \tau \frac{dh_i}{dt} = -h_i + \sum_{j=1}^{N} J_{ij} \phi(h_j) \]  \hspace{1cm} (40)

where \( J_{ij} \) are unconstrained, into the following two-population network, composed of \( N_E \) E neurons and \( N_I \) I neurons:

\[ \tau_E \frac{dh_i^E}{dt} = -h_i^E + \sum_{j=1}^{N_E} J^{EE}_{ij} \phi^E(h_j^E) - \sum_{j=1}^{N_I} J^{EI}_{ij} \phi^I(h_j^I) \]  \hspace{1cm} (41)

\[ \tau_I \frac{dh_i^I}{dt} = -h_i^I + \sum_{j=1}^{N_I} J^{IE}_{ij} \phi^E(h_j^E) \]  \hspace{1cm} (42)

such that the excitatory population in Eq. (41) shares the same pattern overlap dynamics as those in Eq. (40), and excitatory (inhibitory) units send only positive (negative) projections. Note that for simplicity we ignore inhibitory to inhibitory connections.

The four connectivity matrices are given by

\[ J^{EE}_{ij} = \frac{c^{EE}_{ij}}{J^{EE}_{ij}} \omega \left( \frac{A_{EE}}{\sqrt{K_{EE}}} \sum_{s=1}^{P-1} \sum_{\mu=1}^{S} f(\xi_{s,\mu}^{E,s}) g(\xi_{j,\mu}^{E}) \right) \]  \hspace{1cm} (43)

\[ J^{EI}_{ij} = \frac{c^{EI}_{ij}}{J^{EI}_{ij}} \]  \hspace{1cm} (44)

\[ J^{IE}_{ij} = \frac{c^{IE}_{ij}}{J^{IE}_{ij}} \]  \hspace{1cm} (45)

where the synaptic transfer function \( \omega \) has zero support at negative values, and all types of connections have sparse connectivity (i.e. \( \mathbb{E}(c_{ij}^{EE} N_E) = K_{EE} \ll N_E, \mathbb{E}(c_{ij}^{IE} N_I) = K_{IE} \ll N_I, \mathbb{E}(c_{ij}^{EI} N_I) = K_{EI} \ll N_I \)). Note that in this model, both excitatory and inhibitory synaptic efficacies onto excitatory neurons scale as \( 1/\sqrt{K} \), as in the balanced network model \([2, 3]\) and other associative memory models with separate E and I populations \([4, 5]\), while excitatory synapses onto inhibitory neurons scale as \( 1/K \).

Note also that we have assumed that TAH learning takes place only in excitatory recurrent weights \( J_{ij}^{EE} \), and that all other connection strengths are fixed. For the sake of simplicity, we make the following additional assumptions:

1. Inhibition is fast \( (\tau_I \ll \tau_E) \)
2. Inhibitory firing rates depend linearly on their input (i.e. \( \phi^I(h) \approx g(h)^I \))

Using these assumptions, we can reduce Eqs. (41) and (42) to the following equation, that now only describes inputs to E neurons:

\[ \tau_E \frac{dh_i^E}{dt} = -h_i^E + \sum_{j=1}^{N_E} J_{ij}^{IE} \phi^E(h_j^E) - \sum_{j=1}^{N_E} J_{ij}^{EI} \phi^I(h_j^I) \]  \hspace{1cm} (46)

where \( J_{ij}^{I} \) represents an effective inhibitory connectivity matrix, given by:

\[ J_{ij}^{I} = g_i \sum_{k=1}^{N_I} J_{ik}^{EI} J_{kj}^{IE} = \frac{g_i J_{IE}^{IE} J_{IE}^{EI}}{K_{IE} \sqrt{K_{EI}}} \sum_{k=1}^{N_I} c_{ik}^{IE} c_{kj}^{IE} \]  \hspace{1cm} (47)
Taylor expanding the excitatory connectivity in Eq. (43) around all but the first stored sequence, we find that only the first and second order terms will contribute to the total synaptic inputs in the $K \to \infty$ limit:

$$J_{ij}^{EE} \approx \frac{c_{ij}}{\sqrt{K_{EE}}} \left( \omega \left( \frac{A_{EE}}{\sqrt{K_{EE}}} \sum_{s>1} \sum_{\mu=1}^{P-1} f(\xi^{\mu+1}_s/g(\xi^{\mu}_s)) \right) + \frac{A_{EE}}{\sqrt{K_{EE}}} \omega' \left( \frac{A_{EE}}{\sqrt{K_{EE}}} \sum_{s>1} \sum_{\mu=1}^{P-1} f(\xi^{\mu+1}_s/g(\xi^{\mu}_s)) \right) \right).$$

Plugging the above into Eq. (46) and averaging the field over sequences $s \neq 1$, we get:

$$E \left( \sum_{j=1}^{N_E} J_{ij}^{EE} \phi(h^E_j) - \sum_{j=1}^{N_E} J_{ij}^I \phi(h^I_j) \right) = \left( \sqrt{K_{EE}} \omega_{EE} - \sqrt{K_{EI}^g} g_J^I J_{IE} \right) R_E + A_{EE} \omega_{EE} \sum_{\mu=1}^{P-1} f(1^{\mu+1} m^{\mu})$$

where $E$ denotes an average over patterns in all sequences except the retrieved one, and the random structural connectivity matrix $c_{ij}$, and in addition we have defined

$$\omega_{EE} = \int_{-\infty}^{\infty} Dx \omega(A_{EE} \sqrt{\alpha} x)$$
$$\omega'_{EE} = \int_{-\infty}^{\infty} Dx \omega'(A_{EE} \sqrt{\alpha} x)$$
$$\gamma = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Dx Dy f(x) g(y)^2$$

and introduced the following order parameters:

$$R_E = E \left( \phi(h^E) \right)$$
$$q^\mu = E \left( g(\xi^\mu) \phi(h^E) \right).$$

$R_E$ represents the mean firing rate of the excitatory population, while $q^\mu$s are the overlaps with the patterns of the retrieved sequence.

The mean field, Eq. (49), is composed of two terms: The first term in the r.h.s. is proportional to the mean firing rate of the excitatory population. Such a term does not appear in the one population network. It scales as $\sqrt{K}$, and therefore diverges in the large $K$ limit, unless there is a balance between excitation and inhibition, $\sqrt{K_{EE} \omega_{EE}} = \sqrt{K_{EI}^g g_J^I J_{IE}}$. The second term contains the sum over overlaps with the patterns in the sequence. This term is the same as in the one population network, except for the additional factor $\omega_{EE}$. Thus, for the E-I network to be described by the same equations as the one population network, a balance between excitation and inhibition is required [2, 3, 4, 5].

Thus, in the E-I network, we impose the condition

$$\sqrt{K_{EE} \omega_{EE}} = \sqrt{K_{EI}^g g_J^I J_{IE}}.$$  

In numerical simulations we used a rectified linear transformation $\omega(x) = [g_\omega \cdot x + o_\omega]_+$. Other parameters are specified in Table 7b.

### 3.1 Simulation procedure

To construct the excitatory-inhibitory rate network, we take the following steps:

1. We begin by simulating the single population rate network of Eq. (40) with connectivity specified by Eq. (3). We use a rectified linear rate transfer function, $\phi(h) = [g_\phi \cdot h]_+$, and the threshold plasticity rule of Eqs. (6-7) in the main text. We fix $x_J$ and $q_0$ to a desired coding level, and find values for $A$ and $g_\phi$ that lead to sequence recall for a given sequence length.
2. We next construct a two-population rate network (Eqs. 41-42). We start by defining $N_E$ excitatory neurons, each with the same rectified linear rate transfer function as in the previous step, $\phi^E(h) = [g_\phi \cdot h]_+$. We then construct the recurrent excitatory connectivity as specified in Eq. (43), with all plasticity-related free parameters fixed as before. To impose non-negative weights, we choose a rectified linear transformation for the synaptic transfer function $\omega(x) = [g_\omega \cdot x + o_\omega]_+$. 

3. We next add $N_I$ inhibitory neurons, each with a rectified linear rate transfer function, $\phi^I(h) = [g_\phi \cdot h]_+$. For sparse E-I and I-E connectivity, the weights $J^E_{ij}$ and $J^I_{ij}$ are constrained by our choice of $\omega$, and Eq. 50. The initial condition for the inhibitory neurons is fixed to the initial excitatory population average firing rate (i.e. $E(\phi^E((\xi_1^{(1)}, 1)))$).

4. **Spiking network**

   To transform the rate network of Eqs. (41,42) into a spiking network, we mapped dynamics to a current-based leaky integrate-and-fire network. Single unit dynamics were governed by the following current-based equations. For $\alpha, \beta \in \{E, I\}$:

   \[
   \tau_m^\alpha \frac{dV_i^\alpha}{dt} = \Theta(V_i^\alpha - V_{i\text{floor}}^\alpha) \left(-V_i^\alpha + V_{\text{rest}}^\alpha + \sum_{\beta} K_{\alpha\beta} S_{ij}^\alpha \delta(t - t_j^\beta - D) + I_\alpha + \sigma_\alpha \sqrt{\tau_m^\alpha} W_\alpha(t)\right)
   \]

   (51)

   \[
   \tau_s^\alpha \frac{dS_{ij}^\alpha}{dt} = -S_{ij}^\alpha + J^E_{i\beta} \tau_s^\alpha \sum_{t_j^\beta} \delta(t - t_j^\beta - D)
   \]

   (52)

   where $D$ controls the synaptic delay, $I_\alpha$ controls the external input drive, and $\tau_{rp}$ controls the refractory period. $\sigma_\alpha$ controls the strength of the stochastic fluctuations induced by a white noise input $W_\alpha(t)$ with unit variance density. The Heaviside function $\Theta$ sets a lower bound on the attainable voltage, so that the membrane potential cannot be more hyperpolarized than a ‘floor’ $V_{i\text{floor}}^\alpha$. This lower bound captures in a simplified fashion the inhibitory reversal potential that prevents the neuronal membrane potential from going to arbitrarily hyperpolarized values. Without this lower bound, many neurons have voltages with unrealistically large hyperpolarizing deflections, as large as 100 mV below the resting potential. Note that retrieval still occurs without the implementation of this lower bound.

   We simulated a spiking network with $N_E = 20,000$ excitatory units, and $N_I = 5,000$ inhibitory units. $V_{i\text{floor}}^\alpha$ was set to $-80$ mV for excitatory units and $-\infty$ mV for inhibitory units, with $V_{\text{rest}}^\alpha = -70$ mV. We set $D = 1$ ms, $\tau_{rp} = 1$ ms, the reset potential $V_{\text{reset}}^\alpha = V_{\text{rest}}^\alpha$, and the input drive $I_\alpha = 0$ mV. A full list of parameters can be found in Table 7c.

4.1 **Simulation procedure**

   To construct the full spiking network, we take the following steps:

   1. We start by simulating an excitatory-inhibitory rate network, following the steps as outlined in section 4.2.

   2. We then match both rectified linear rate transfer functions to those derived from leaky-integrate and fire (LIF) units operating in the presence of noise. We use the following equation for the transfer function of unit activity in population $\alpha$ [6]:

   \[
   \nu_\alpha(x) = \tau_{rp} + \tau_m^\alpha \sqrt{\pi} \int_{V_{\text{reset}}^\alpha - x}^{V_{\text{threshold}}^\alpha - x} \frac{d\nu}{\sigma_\alpha} \left(1 + \text{erf}(u)\right).
   \]

   To fit the LIF transfer function, we fix the reset potential to zero, and the refractory period and membrane time constant to desired values. The desired membrane time constant should be less than
the synaptic time constant (see final step). We leave as free parameters the threshold and strength of external white noise ($\sigma_\alpha$). We minimize the Euclidean distance between the two transfer functions over the interval $x \in \{x_{\text{lower}}, x_{\text{upper}}\}$ using the L-BFGS-B optimization algorithm, bounding both the threshold and noise strength from below at zero. Starting from many random initial conditions we find that a global minimum is reached by the procedure.

3. To produce threshold, reset, and resting membrane potentials within a physiological range, we apply a linear transformation $\phi_V(x) = \lambda_\alpha^V x + \Delta_\alpha^V$ that shifts and scales these parameters. All connectivity weights are also scaled by the same factor. To impose a nonzero resting membrane potential, we fix $V_\alpha^{\text{rest}} = V_\alpha^{\text{reset}} = \Delta_\alpha^V$. Note that all weights $J_{ij}^{\alpha\beta}$ taken from the rate network are also divided by the synaptic time constant $\tau_\alpha$. The new parameters and connectivities are therefore:

\[
V_\alpha^{\text{thresh}}^* = \lambda_\alpha^V V_\alpha^{\text{thresh}} + \Delta_\alpha^V \\
V_\alpha^{\text{reset}}^* = \Delta_\alpha^V \\
J_{ij}^{\alpha\beta} = \lambda_\alpha^V J_{ij}^{\alpha\beta} / \tau_\alpha \\
\sigma_\alpha = \lambda_\alpha^V \sigma_\alpha.
\]

The initial conditions are $S_{EE}(0) = \phi^E(\lambda_\alpha^E f(\xi_{1,1}^E))$, $S_{IE}(0) = \mathbb{E}[S_{EE}(0)]$, $S_{EI}(0) = 0$, $V_E(0) = \Delta_\alpha^V$, and $V_I(0) = \Delta_\alpha^V$, where $z$ is a standard Gaussian variable.

4. We aim to keep the range of neural firing rates below saturation, and adjust several parameters to this effect:

(a) The dynamic range of excitatory firing can be adjusted by scaling the gain of the excitatory linear transfer function, $g_{\phi}^E$. The dynamic range of inhibitory firing can be adjusted by rescaling $J_{IE}$ and $J_{EI}$ while keeping their product constant.

(b) The gain of the excitatory LIF transfer function is fixed by the gain of the original linear transfer function, $g_{\phi}$. The gain of the inhibitory LIF transfer function can be controlled by adjusting the gain of the corresponding linear transfer function, while scaling $J_{IE}$ inversely.

5. Finally, we build the LIF spiking network of Eqs. (51,52). We now have two timescales, one for the synapses ($\tau_s$) and one for the single units ($\tau_m$). The time constants of the rate network are mapped to the synaptic time constants. All other parameters, including weight matrices, are taken from the previous two steps. Note that if the fitted $\sigma_\alpha^V$ is unrealistically large and disrupts retrieval, we reduce it to a more realistic value. This is the case for the parameters of Figure 7 in the main article, and so we lower it to a value equal to half the spiking threshold (see Table 7c). An alternative approach would be to fix $\sigma_\alpha$ to a desired value, and fit the LIF transfer function using only the threshold as a free parameter.

5 Supplemental Procedures

Measuring robustness

To measure robustness in the top row of Supplementary Figure 2, we computed the difference in the time-averaged norm of the overlaps between perturbed and unperturbed trials:

\[
k(t_0, t_1) = \mathbb{E}_{t_0 < t < t_1} (||\vec{m}^{\text{unperturbed}}(t)||_2) - \mathbb{E}_{t_0 < t < t_1} (||\vec{m}^{\text{perturbed}}(t)||_2)
\]

where the expectation is over the time interval starting at $t_0$ and ending at $t_1$. Unperturbed overlaps $\vec{m}^{\text{unperturbed}}(t)$ in Supplementary Figure 2 correspond to those in Figure 1 of the main article. To generate the perturbed overlaps $\vec{m}^{\text{perturbed}}(t)$, we fixed the initial condition of rate units to $r(t = 0) = \phi(\xi_{1,1} + \sigma_z z_0)$, where $\sigma_z$ controls the standard deviation of the standard Gaussian perturbation $z_0 \sim \mathcal{N}(0, 1)$.

We measured $k$ in two time intervals: 1) in the time interval leading up to the observed retrieval time, where $t_0 = 0$ and $t_1 = \arg\max_t m_P(t)$, and 2) in the latter half of this interval, where $t_0 = \arg\max_t m_P(t)/2$ and $t_1 = \arg\max_t m_P(t)$. 12
Retrieval time ratio
To compute the retrieval time ratio (RTR) in Supplementary Figure 3, we divided the observed retrieval time by the predicted retrieval time (see Section 2.3.1): \( RTR = \argmax_t m_P(t) / (\tau(P - 1)) \).

Peak distributions

Neurons
To find firing rate peaks for a given unit during a single trial of recall, we computed the average firing rate \( \bar{r}_i \) for unit \( i \) and selected all continuous intervals of firing occurring one standard deviation above this threshold. The \( n \)th peak midpoint time for unit \( i \) was computed using a weighted average,

\[
t_{i,n}^{\text{midpoint}} = \frac{\sum_{t_n^0 < t < t_n^1} r(t) \cdot t}{\sum_{t_n^0 < t < t_n^1} r(t)}.
\]

where \( t_n^0 \) and \( t_n^1 \) mark the beginning and end of the \( n \)th continuous interval. The width \( \sigma_{i,n}^{\text{width}} \) of a peak was defined as the length of the continuous interval: \( t_n^1 - t_n^0 \). Single unit peaks that had maximal firing rate at \( t=0 \) were excluded from the analysis, but do not qualitatively alter the distribution if included.

Overlaps
To measure the width of each overlap, we used the following equation:

\[
\sigma_{\mu}^{\text{width}} = \sqrt{\frac{\sum_t m_{\mu}(t) \cdot (t - t_{\mu}^{\text{mean}})^2}{\sum_t m_{\mu}(t)}}
\]

where \( t_{\mu}^{\text{mean}} \) is defined as:

\[
t_{\mu}^{\text{mean}} = \frac{\sum_t t \cdot m(t)}{\sum_t m(t)}.
\]

To compute the cumulative density of peak times for single neurons and overlaps, we measured the time interval starting at \( t = 0 \) and lasting up to the retrieval time (defined in the previous sections).

Measuring capacity in a finite network
To compute the mean-field correlation in Fig 3b, we use both \( M \) and \( \bar{r} \) to normalize the overlaps by the standard deviation of the firing rates (see section 2.2).

To compute the mean-field capacity curve in Fig. 3c (dashed curves), we use a bisection method to converge on the smallest \( \alpha \) for which the maximal overlap values in the sequence \( \{ \argmax_t q_1(t), \argmax_t q_2(t), ..., \argmax_t q_S(t) \} \) are a monotonically decaying function.

Retrieval with nonlinear rule
We initialize the network to \( r(t = 0) = \phi(f(\xi^1)) \). To measure the pattern correlations in Figure 4d, we compute the Pearson correlation coefficient between \( r(t) \) and \( g(\xi^l) \) for each pattern \( \xi^l \).

Selectivity criterion
To generate the sorted raster plots of Fig. 5, we first divided units into active (silent) pools according to their maximum firing rate: \( \argmax_t r(t) > \theta \) (\( \argmax_t r(t) < \theta \)), where \( \theta \) defines a minimal activity threshold. We set \( \theta = 0.05 \cdot r_{\text{max}} \). Selective units were defined by the union or intersection of these pools in different stimulus contexts.
**Turn selective units**

Right (left) selective units were those that had membership in the active pool during recall of the right (left) sequence, and were in the silent pool during recall of the left (right) sequence.

**Non-specific units**

To determine non-specific units, we first sought neurons that were active in both left and right stimulus contexts. We then computed the mean absolute difference (MD) in activity between the left and right trial for each of these units: \( \text{MD} = \frac{1}{NT} \sum_i \sum_t \left[ r_{i}^{\text{left}}(t) - r_{i}^{\text{right}}(t) \right] \). Units with \( \text{MD} < 0.075 \) were identified as non-specific.

**Decaying weight perturbations**

To generate the sorted raster plots of Fig. 6a, we included units that had a maximal activity of at least \( 0.15 \cdot r_{\text{max}} \) on day 1 (for day 1 sorted) or day 30 (for day 30 sorted). These accounted for roughly half of the total population of neurons.

**Activity profile correlation**

To measure the similarity of activity across days in Fig. 6c, we computed the Pearson correlation coefficient between rate trajectories \( \vec{r}_n \) and \( \vec{r}_m \), where \( n \) and \( m \) are simulation days, and \( i \) is the index of the neuron. We averaged this quantity across all neurons.

**Spiking network**

**Quantifying correlations**

To compute pattern correlations in Figure 7c, we transformed spiking activity into rates by convolving spikes with a Gaussian kernel (20 ms standard deviation). We then computed the Pearson correlation between rates and transformed patterns as in the rate network case with a nonlinear rule.
Figure 1: Single unit retrieval examples for the network in Figure 1 in the main text. Units are randomly selected from the whole population.
Figure 2: Robustness to initial perturbations of a network storing sequence length $P = 16$ for $S = 1$ (left column), and $S = 4$ (right column) sequence number. Top row: The difference in the time-averaged norm of the overlaps between perturbed and unperturbed trajectories, given as a function of the standard deviation of the initial Gaussian perturbation $\sigma_z$. This perturbation is added to the initial condition ($\xi_1$). In green, the difference is computed using the time interval leading up to retrieval time (see Supp. Procedures). In blue, the difference is computed using only the latter half of this interval. The green and blue dashed lines display the average overlap norms of the unperturbed trajectory for the full and latter half of retrieval, respectively. Bottom row: Overlaps during retrieval following an initial perturbation of strength $\sigma_z$. All other parameters are as in Figure 1 of the main text.

Figure 3: The ratio of the observed and predicted retrieval times (see Supplemental Procedures), for varying $P$, as a function of rate transfer function parameters $\theta$ and $\sigma$. All other parameters are as in Figure 1 of the main text.
Figure 4: Temporal characteristics of a retrieved sequence stored with a nonlinear learning rule. 

\( a \). Distribution of single neuron peak widths, defined as continuous firing intervals occurring one standard deviation above the time-averaged firing rate. Black error bars denote mean and standard deviation of widths within each 10 ms interval.  

\( b \). Green dots indicate observed overlap widths (see Supp. Procedures). Analytic predictions for Gaussian patterns stored with bilinear learning rule (see Figure 2 of main text) are shown in red.  

\( c \). Cumulative percentage of peak times for single neurons (blue) and overlaps (grey). The dashed black line represents a uniform distribution. All parameters are as in Figure 4 of the main text.

Figure 5: Comparison of mean-field with full network simulations

\( a \). Solid lines show overlaps computed using a simulation of the full network. Dashed lines are solutions to the mean-field equations. Top: Overlaps \( \{ q_{\mu} \} \) of network activity with stored patterns. Bottom: Average squared firing rates, \( M \).  

\( b \). Mean-field two-point rate correlation function \( C \), as defined in section 2.2. All parameters are as in Figure 1 of the main text, except \( N = 50,000 \). Discrepancies in the mean-field are finite-size effects, and decrease with larger \( N \).
Figure 6: Gain function behavior. a. Conditions for successful retrieval. Sequences can be retrieved for appropriate α and q(0) in the green regions, and cannot be retrieved in the blue regions. Within D and E, retrieval is possible for vanishingly small initial overlaps. In region F, retrieval is possible for initial overlaps of small but finite size. In regions A, B, and C retrieval is not possible, as both G(0) < 1 and G_{max} < 1. The blue line corresponds to condition 1, the orange line to condition 2, and the green line to condition 3 (see section 2.4.1). The maximal capacity within the green region is shown in Supplementary Figure 7. The red dot corresponds to the parameters in panel b. b. Overlaps as a function of time (top), average squared rate (middle), and gain function (bottom) for full network (left), and overlap dynamics with a constant M and G approximation (right). In the constant gain case, G has been fixed to the average value of G during retrieval in the dynamic gain case: 1.0718, and M is shown purely for illustration. All parameters in the "dynamic gain" case are as in Figure 1 of the main text. c. Solid lines are profiles of gain function G(x) as a function of x, for three sets of parameters corresponding to the three possible regions of successful retrieval. Dashed lines indicate threshold at one. The parameters chosen for region F correspond to the the red dot in panel a.
Figure 7: Maximal capacity, computed analytically using Eqs. (38,39) as a function of $\theta$ and $\sigma$, for the bilinear plasticity rule and the error-function transfer function. The white boundary corresponds to the successful retrieval region in Figure 6. Note that the full curves in Fig. 3c of the main text correspond to horizontal cuts in this plane.
### Figure 1, 2: Retrieval of a stored sequence

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $N$       | 40,000| Neuron count             |
| $c$       | 0.005 | Connection probability   |
| $\theta$  | 0.22  | Rate transfer function offset |
| $\sigma$  | 0.1   | Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant            |
| $S$       | 1     | Sequence number          |
| $P$       | 16    | Sequence length          |

### Figure 3a: Sequence capacity

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $N$       | 40,000| Neuron count             |
| $c$       | 0.005 | Connection probability   |
| $\theta$  | 0.22  | Rate transfer function offset |
| $\sigma$  | 0.1   | Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant            |
| $S$       | 2     | Sequence number          |
| $P$       | 16    | Sequence length          |

### Figure 3b: Sequence capacity

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $\theta$  | 0.22  | Rate transfer function offset |
| $\sigma$  | 0.1   | Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant            |

### Figure 3c: Sequence capacity

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $N$       | 40,000| Neuron count             |
| $c$       | 0.005 | Connection probability   |
| $\theta$  | 0.2   | Rate transfer function offset |
| $\sigma$  | 0.1   | Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant            |
### Figure 4: Retrieval with nonlinear learning rules

| Parameter | Value | Comment                   |
|-----------|-------|---------------------------|
| $N$       | 40,000| Neuron count              |
| $c$       | 0.005 | Connection probability    |
| $\theta$  | 0.005 | Rate transfer function offset |
| $\sigma$  | 0.00357| Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant             |
| $S$       | 1     | Sequence number           |
| $P$       | 30    | Sequence length           |
| $x_f$     | 1.645 | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.645 | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8   | Plasticity rule parameter |
| $q_g$     | 0.95  | Plasticity rule parameter |

### Figure 5: Selectivity emerges from random input patterns

| Parameter | Value | Comment                   |
|-----------|-------|---------------------------|
| $N$       | 40,000| Neuron count              |
| $c$       | 0.005 | Connection probability    |
| $\theta$  | 0.005 | Rate transfer function offset |
| $\sigma$  | 0.00357| Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant             |
| $S$       | 2     | Sequence number           |
| $P$       | 30    | Sequence length           |
| $x_f$     | 1.645 | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.645 | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8   | Plasticity rule parameter |
| $q_g$     | 0.95  | Plasticity rule parameter |
Figure 6: Changes in synaptic connectivity preserve collective sequence retrieval

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $N$       | 40,000| Neuron count             |
| $c$       | 0.005 | Connection probability   |
| $\theta$  | 0.005 | Rate transfer function offset |
| $\sigma$  | 0.00357 | Rate transfer function inverse gain |
| $\tau$    | 10 ms | Time constant            |
| $S$       | 1     | Sequence number          |
| $P$       | 30    | Sequence length          |
| $x_f$     | 1.645 | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.645 | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8   | Plasticity rule parameter |
| $q_g$     | 0.95  | Plasticity rule parameter |
| $\lambda$ | 0.85  | Decay rate               |
| $\sigma_z$ | 0.03  | Perturbation strength    |

Figure 7a: One population rate network

| Parameter | Value | Comment                  |
|-----------|-------|--------------------------|
| $N$       | 20,000| Neuron count             |
| $c$       | 0.04  | Connection probability   |
| $g_\phi$  | 12    | Rate transfer function gain |
| $\tau$    | 20 ms | Time constant            |
| $S$       | 1     | Sequence number          |
| $P$       | 32    | Sequence length          |
| $A$       | 6.3   | Learning strength        |
| $x_f$     | 1.5   | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.5   | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8   | Plasticity rule parameter |
| $q_g$     | 0.933 | Plasticity rule parameter |
### Figure 7b: Two population rate network

| Parameter | Value | Comment |
|-----------|-------|---------|
| $N_E$     | 20,000| Excitatory neuron count |
| $N_I$     | 5,000 | Inhibitory neuron count |
| $c_{EE}$  | 0.04  | Connection probability |
| $c_{IE}$  | 0.04  | Connection probability |
| $c_{EI}$  | 0.04  | Connection probability |
| $g_E^E$   | 12    | Rate transfer function gain, excitatory |
| $\tau_E$ | 20 ms | Time constant, excitatory |
| $g_I^I$   | 20    | Rate transfer function gain, inhibitory |
| $\tau_I$ | 5 ms  | Time constant, inhibitory |
| $S$       | 1     | Sequence number |
| $P$       | 32    | Sequence length |
| $A_{EE}$  | 6.3   | Learning strength |
| $x_f$     | 1.5   | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.5   | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8   | Plasticity rule parameter |
| $q_g$     | 0.933 | Plasticity rule parameter |
| $g_\omega$ | 1 | Synaptic transfer function gain |
| $o_\omega$ | -0.0038 | Synaptic transfer function offset |
| $J^{IE}$  | 0.0673 | Synaptic weight |
| $J^{EI}$  | -0.0250 | Synaptic weight |
| Parameter | Value     | Comment                                |
|-----------|-----------|----------------------------------------|
| $N_E$     | 20,000    | Excitatory neuron count                |
| $N_I$     | 5,000     | Inhibitory neuron count                |
| $c_{EE}$  | 0.04      | Connection probability                 |
| $c_{IE}$  | 0.04      | Connection probability                 |
| $c_{EI}$  | 0.04      | Connection probability                 |
| $\tau^E_m$ | 10 ms     | Membrane time constant, excitatory     |
| $\tau^E_s$ | 20 ms     | Synaptic time constant, excitatory     |
| $V_{thresh}^E$ | -50 mV | Spiking threshold, excitatory         |
| $V_{reset}^E$ | -70 mV  | Voltage reset, excitatory             |
| $V_{rest}^E$ | -70 mV  | Voltage resting potential, excitatory  |
| $V_{reversal}^E$ | -80 mV | Voltage floor potential, excitatory   |
| $\tau^E_{rp}$ | 1 ms     | Refractory period, excitatory         |
| $\sigma^E$ | 10 mV    | White noise strength (std dev), excitatory |
| $\tau^I_m$ | 2 ms     | Membrane time constant, inhibitory     |
| $\tau^I_s$ | 5 ms     | Synaptic time constant, inhibitory     |
| $V_{thresh}^I$ | -50 mV  | Spiking threshold, inhibitory         |
| $V_{reset}^I$ | -70 mV  | Voltage reset, inhibitory             |
| $V_{rest}^I$ | -70 mV  | Voltage floor potential, inhibitory    |
| $V_{reversal}^I$ | $-\infty$ mV | Voltage reversal potential, inhibitory |
| $\tau^I_{rp}$ | 1 ms     | Refractory period, inhibitory         |
| $\sigma^I$ | 10 mV    | White noise strength (std dev), inhibitory |
| $S$       | 1         | Sequence number                        |
| $P$       | 32        | Sequence length                        |
| $A$       | 6.3       | Learning strength                      |
| $x_f$     | 1.5       | Post-synaptic threshold of plasticity rule |
| $x_g$     | 1.5       | Pre-synaptic threshold of plasticity rule |
| $q_f$     | 0.8       | Plasticity rule parameter              |
| $q_g$     | 0.933     | Plasticity rule parameter              |
| $g_\omega$ | 1        | Synaptic transfer function gain        |
| $o_\omega$ | -0.0038   | Synaptic transfer function offset      |
| $J^{IE}/K_{IE}$ | 0.208 mV | Synaptic weight                        |
| $J^{EI}/\sqrt{K_{EI}}$ | -0.228 mV | Synaptic weight                        |
| $\lambda^E$ | 5.17     | Excitatory rescaling factor            |
| $\lambda^I$ | 3.09     | Inhibitory rescaling factor            |
References

[1] U. Pereira and N. Brunel. Attractor Dynamics in Networks with Learning Rules Inferred from In Vivo Data. *Neuron*, 99(1):227–238, Jul 2018.

[2] C. van Vreeswijk. Partial synchronization in populations of pulse-coupled oscillators. *Phys. Rev. E*, 54:5522–5537, 1996.

[3] C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10:1321–1371, 1998.

[4] R. Rubin, L. F. Abbott, and H. Sompolinsky. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proc. Natl. Acad. Sci. U.S.A.*, 114(44):E9366–E9375, 10 2017.

[5] G. Mongillo, S. Rumpel, and Y. Loewenstein. Inhibitory connectivity defines the realm of excitatory plasticity. *Nat. Neurosci.*, 21(10):1463–1470, 10 2018.

[6] L. M. Ricciardi. *Diffusion processes and Related topics on biology*. Springer-Verlag, Berlin, 1977.