Noise in neurons and synapses enables reliable associative memory storage in local cortical circuits

[1]Chi Zhang, [1,2,3]Danke Zhang, and [1]Armen Stepanyants

[1]Department of Physics and Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, MA, USA; [2]The Brain Cognition and Brain Disease Institute, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences; [3]Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen, 518055, China.

ABSTRACT

Neural networks in the brain can function reliably despite various sources of errors and noise present at every step of signal transmission. These sources include errors in the presynaptic inputs to the neurons, noise in synaptic transmission, and fluctuations in the neurons' postsynaptic potentials. Collectively they lead to errors in the neurons' outputs which are, in turn, injected into the network. Does unreliable network activity hinder fundamental functions of the brain, such as learning and memory retrieval? To explore this question, this article examines the effects of errors and noise on properties of biologically constrained networks of inhibitory and excitatory neurons involved in associative sequence learning. The associative learning problem is solved analytically and numerically, and it is also shown how memory sequences can be loaded into the network with a more biologically plausible perceptron-type learning rule. Interestingly, the results reveal that errors and noise during learning increase the probability of memory recall. There is a tradeoff between the capacity and reliability of stored memories, and, noise during learning is required for optimal retrieval of stored information. What is more, networks loaded with associative memories to capacity display many structural and dynamical features observed in local cortical circuits. Due to the similarities between the associative and brain networks, this article predicts that the connections originating from unreliable inhibitory and excitatory neurons or neuron classes in the cortex must be depressed or eliminated during learning, while the connections onto noisy neurons or neuron classes must have lower probabilities and higher weights.

INTRODUCTION

Brain networks can reliably store and retrieve long-term memories despite the facts that various sources of errors and noise accompany every step of signal transmission through the network (1), synaptic connectivity changes over time (2-4), and extraneous sensory inputs are usually present during memory recall. The brain can reduce the effects of noise and extraneous sensory inputs by attending to the memory retrieval process (5, 6), but such hinderances cannot be eliminated entirely. Therefore, the reliability required for memory retrieval must be built into the network during learning. This proposal presents an interesting challenge. Traditional supervised learning models, such as the ones that rely on the perceptron rule (7, 8), modify connectivity only when a neuron's output deviates from its target output. Thus, in such models learning stops as soon as the neuron produces the desired response and, subsequently, there is no possibility for improving the response reliability. The network connection weights in such models may end up near the boundary of the solution region, and a small amount of noise or extraneous input during memory retrieval can lead to errors or completely disrupt the retrieval process. More reliable solutions are located farther away from the solution region boundary, but the perceptron rule is not guaranteed to find them. Thus, it is not clear how the neural networks in the brain manage not only to learn but also to do it reliably.

In the case of associative memory storage, reliability can be incorporated into the perceptron learning rule by means of a generic robustness parameter [see e.g. (9, 10)]. This simplified description, however, is neither biologically motivated nor does it reflect the recurrent cycle of errors and noise present during learning and memory retrieval (Figure 1A). A more comprehensive account must include errors in the input to the neurons, combine them with fluctuations in the neurons' presynaptic weights and intrinsic sources of noise, and produce spiking errors in the neurons' outputs. The latter injected back into the network, would give rise to input errors for the next time step, completing the error propagation cycle. The recurrence of errors presents a clear challenge for the retrieval of associative memory sequences considered in this study. If not corrected at every step of the retrieval process, errors in the network activity can amplify over time and lead to an irreversible deviation of the retrieved trajectory from the loaded sequence, i.e. a partially retrieved memory.

The premise of this article is that errors and noise are essential components of the reliable learning mechanism implemented in the brain. As different fluctuations accompany the presentation of the same learning example to a neuron on different trials, the neuron in effect never stops learning. Its connection weights move further away from the solution region boundary every time a progressively larger fluctuation is encountered. This process increases the reliability of the loaded memory which can later be retrieved in the presence of noise. Similar ideas have been successfully used in machine learning where an augmentation of training examples with noise (11) and dropping out neurons and connections (12) during training have been shown to significantly reduce both overfitting and training time. There are many other examples in which noise is put to constructive use to improve various functions of physical and neural systems [for reviews see (13-16)]. Therefore, the hypothesis that errors and noise are exploited by the brain for reliable memory storage is not entirely surprising. Still, this hypothesis requires careful quantitative evaluation and validation with experimental data which is the focus of this study.

To that end, we examined a network model of biologically constrained inhibitory and excitatory McCulloch and Pitts neurons (17) loaded to capacity with associative memory sequences (18) in the presence of errors and noise. First, we show that there is a tradeoff between the capacity and reliability of stored sequences which is controlled by the levels of errors and noise present during learning. For optimal tradeoff, as judged by the amount of information contained in the retrieved sequences, noise must be present during learning. Second, as synaptic connectivity of neurons changes during learning (3), it is not unreasonable to expect that the requirement of reliable memory retrieval is reflected in the properties of network connectivity and, consequently, the activity of neurons in the brain. Interestingly, local neural networks in the mammalian cortical areas have many common features of connectivity and network activity (18). We show that these network properties in the model emerge all at once during reliable memory storage. Third, as levels of errors and noise can differ across individual neurons or neuron classes, we examined the properties of model networks composed of heterogeneous neurons and make two salient predictions regarding the connectivity of noisy neurons.

We would like to mention other studies that incorporated sources of noise into the associative learning model and examined the effects of learning on neural network properties. In these studies, the basic associative learning model (7, 19-24) was extended to include biologically inspired

elements, such as sign-constrained postsynaptic connections (inhibitory and excitatory) [see e.g. (9, 25, 26)], homeostatically constrained presynaptic connections [see e.g. (27)], and robustness to noise [see e.g. (22, 23)]. In particular, Brunel et al. (9, 28) calculated neuron capacity in the presence of intrinsic noise and output spiking errors [see Supplementary Material in (9)] and showed that sparse excitatory connectivity and certain two- and three-neuron motifs develop in networks loaded to capacity. Rubin et al. (10) considered presynaptic and intrinsic noise and showed that the balance of inhibitory and excitatory currents emerges at capacity. Zhang et al. (18) showed that many structural and dynamical properties of local cortical networks emerge in associative networks robustly loaded to capacity. These studies significantly differ from the present work both in terms of the models and the results. First, the model introduced in this article includes the complete error propagation cycle with presynaptic spiking errors combining with synaptic and intrinsic noises to produce errors in the neurons' outputs. Second, the model allows for the possibility of having different levels of errors and noise during learning and memory retrieval. Third, the model makes it possible to analyze networks of neurons with heterogeneous properties. In terms of model results, we first show how errors and noise during learning facilitate reliable memory retrieval and next produce a comprehensive list of results related to network structure and dynamics that are then compared to the data from local cortical networks to validate the model and make predictions.

The manuscript is organized as follows. In Results, we first provide a description of an associative sequence learning model by a network of inhibitory and excitatory neurons in the presence of errors and noise in signal transmission and biologically inspired constraints on connectivity. Second, we solve the model numerically with nonlinear optimization for finite networks and analytically with the replica method in the thermodynamic limit and show how various sources of errors and noise affect the memory storage capacity of the neurons. Third, we describe the tradeoff between the capacity and reliability of stored memories and show that to retrieve the maximum amount of information memories must be loaded into the network in the presence of noise. Fourth, we examine the properties of connectivity and dynamics in model networks loaded to capacity with associative memory sequences in the presence of errors and noise and show that there is a region of parameters in which these properties agree with the reported experimental measurements. Fifth, we derive a perceptron-type learning rule that can be used to store associative memories into

the network in an online manner and show that this biologically more plausible method leads to similar network properties as those obtained with the nonlinear optimization and replica methods. Finally, we examine the properties of networks of heterogeneous neurons and make predictions regarding network connectivity. The details of the replica calculation and numerical solutions of the associative memory storage model are provided in SI.

RESULTS

A. Network model of associative memory storage in the presence of errors and noise

We modeled associative sequence learning by a local (~100 μm in size), all-to-all potentially (structurally) connected (29, 30) cortical network. The model network consisted of N_{inh} inhibitory and ($N - N_{inh}$) excitatory McCulloch and Pitts neurons (17) (Figure 1A) and was faced with a task of learning a sequence of consecutive network states, $X^1 \rightarrow X^2 \rightarrow ... X^{m+1}$, in which X^μ is a binary vector representing target activities of all neurons at a time step μ, and the ratio m/N is referred to as the memory load. During learning, individual neurons had to independently learn to associate inputs they received from the network with the corresponding target outputs derived from the associative memory sequence. The neurons learned these input-output associations by adjusting the weights of their input connections, J_{ij} (weight of connection from neuron j to neuron i). In contrast to previous studies, we accounted for the fact that learning in the brain is accompanied by several sources of errors and noise. Within the model, these sources are divided into three categories (orange lightning signs in Figure 1A): (1) spiking errors, or errors in X^μ, (2) synaptic noise, or noise in J_{ij}, and (3) intrinsic noise, which combines all other sources of noise affecting the neurons' postsynaptic potentials. The last category includes background synaptic activity and the stochasticity of ion channels. In the model, this category is equivalent to noise in the neurons' thresholds of firing, h_i (for neuron i). In the following, asterisks are used to denote quantities containing errors or noise (e.g. $X^{*\mu}$), whereas symbols without asterisks represent the mean (for h_i and J_{ij}) or target (for X^μ) values.

The target neuron activities (e.g. binary scalar X_i^{μ}) were independently drawn from neuron-dependent Bernoulli probability distributions: 0 with probability $1 - f_i$ and 1 with probability f_i. Spiking errors in neuron activity states were introduced with the Bernoulli trials by making independent and random 1 to 0 changes with probabilities $P\left(X_i^{*\mu} = 0 \mid X_i^{\mu} = 1\right) = p_i^-$ for spike failures and 0 to 1 changes with probabilities $P\left(X_i^{*\mu} = 1 \mid X_i^{\mu} = 0\right) = p_i^+$ for erroneous spikes. Without loss of generality, we assumed that these two types of spiking errors are balanced, $f_i p_i^- = \left(1 - f_i\right) p_i^+$, and do not affect the neuron's firing probability, f_i. This relation allowed us to describe both types of spiking errors in terms of the neuron's overall spiking error probability, $r_i = f_i p_i^- + \left(1 - f_i\right) p_i^+$:

$$p_i^+ = \frac{r_i}{2\left(1 - f_i\right)}; \quad p_i^- = \frac{r_i}{2 f_i} \tag{1}$$

To describe synaptic noise, we followed the basic model of quantal synaptic transmission (31) and assumed that the variance of a given connection weight, J_{ij}^*, is proportional to its mean, $\mathrm{var}\left(J_{ij}^*\right) = \frac{h_i \beta_{syn,i}}{N}\left|J_{ij}\right|$. The dimensionless coefficient $\beta_{syn,\,i}$ is referred to as the synaptic noise strength of neuron i, and the factor of h_i/N was introduced for convenience. We assumed that the intrinsic noise is Gaussian distributed across trials with the mean $\left\langle h_i^* \right\rangle = h_i$ and variance $\mathrm{var}\left(h_i^*\right) = \frac{h_i^2 \beta_{int,i}^2}{N}$. Here, $\beta_{int,i}$ is a dimensionless coefficient called the intrinsic noise strength of neuron i, and, as before, a factor of h_i^2 / N was introduced for convenience.

Similar to (27), two biologically inspired constraints were imposed on the learning process. First, the l_1-norm of input connection weights of each neuron was fixed during learning, $\frac{1}{N}\sum_{j=1}^{N}\left|J_{ij}\right| = w_i$. Here, parameter w_i is referred to as the average absolute connection weight of neuron i. Second, the signs of output connection weights of every neuron (inhibitory or excitatory) were fixed during

learning, $J_{ij} g_j \geq 0$. In these N^2 inequalities, parameter $g_j = 1$ if neuron j is excitatory and –1 if it is inhibitory. Biological motivations for these constraints were previously discussed (27).

Individual neurons (e.g. neuron i) learned independently to associate noisy inputs they received from the network, $X^{*\mu}$, with the corresponding target outputs derived from the associative memory sequence, $X_i^{\mu+1}$. Neuron i is said to have learned the presented set of associations successfully if, in the presence of input spiking errors, synaptic and intrinsic noise, the fractions of its erroneous and failed spikes do not exceed its assigned spiking error probabilities, p_i^+ and p_i^- (Figure 1B). The above-described model for neuron i can be summarized as follows:

$$
\begin{aligned}
&P\left(\theta\left(\sum_{j=1}^{N} J_{ij}^* X_j^{*\mu} - h_i^*\right) = 0 \middle| X_i^{\mu+1} = 1\right) \leq \frac{r_i}{2 f_i}; \quad \mu = 1,\ldots,m, \quad j = 1,\ldots,N \\
&P\left(\theta\left(\sum_{j=1}^{N} J_{ij}^* X_j^{*\mu} - h_i^*\right) = 1 \middle| X_i^{\mu+1} = 0\right) \leq \frac{r_i}{2(1-f_i)} \\
&P\left(X_i^{\mu} = 1\right) = f_i; \quad P\left(X_i^{*\mu} = 0 \middle| X_i^{\mu} = 1\right) = \frac{r_i}{2 f_i}; \quad P\left(X_i^{*\mu} = 1 \middle| X_i^{\mu} = 0\right) = \frac{r_i}{2(1-f_i)} \\
&\left\langle J_{ij}^* \right\rangle = J_{ij}; \quad \mathrm{var}\left(J_{ij}^*\right) = \frac{h_i \beta_{syn,i}}{N} \left| J_{ij} \right| \\
&\left\langle h^* \right\rangle = h_i; \quad \mathrm{var}\left(h_i^*\right) = \frac{h_i^2 \beta_{int,i}^2}{N} \\
&\frac{1}{N} \sum_{j=1}^{N} \left| J_{ij} \right| = w_i \\
&J_{ij} g_j \geq 0
\end{aligned}
\tag{2}
$$

We note that, depending on the loaded associative memory sequence, Eqs. (2) may have multiple solutions if the learning problem faced by the neuron is feasible or no solution if the problem is not feasible. The neuron's success probability in learning associative sequences of a given length is defined as the average of such binary outcomes over the sequences (Figure 1C). This probability is a decreasing function of the memory load and levels of errors and noise.

At the network level, the described associative memory storage model is governed by the network-related parameters N and $\{g_i\}$, the memory load m/N, and the neuron-related parameters $\{h_i\}$, $\{w_i\}$,

$\{f_i\}$, $\{r_i\}$, $\{\beta_{int,i}\}$, and $\{\beta_{syn,i}\}$, and the task is to find connection weights, $\{J_{ij}\}$, that satisfy the requirements of Eqs. (2) for all neurons. In the following, we examine the properties of associative networks composed of inhibitory and excitatory neurons governed by the same and distributed sets of neuron-related parameters. We refer to these networks as homogeneous and heterogeneous.

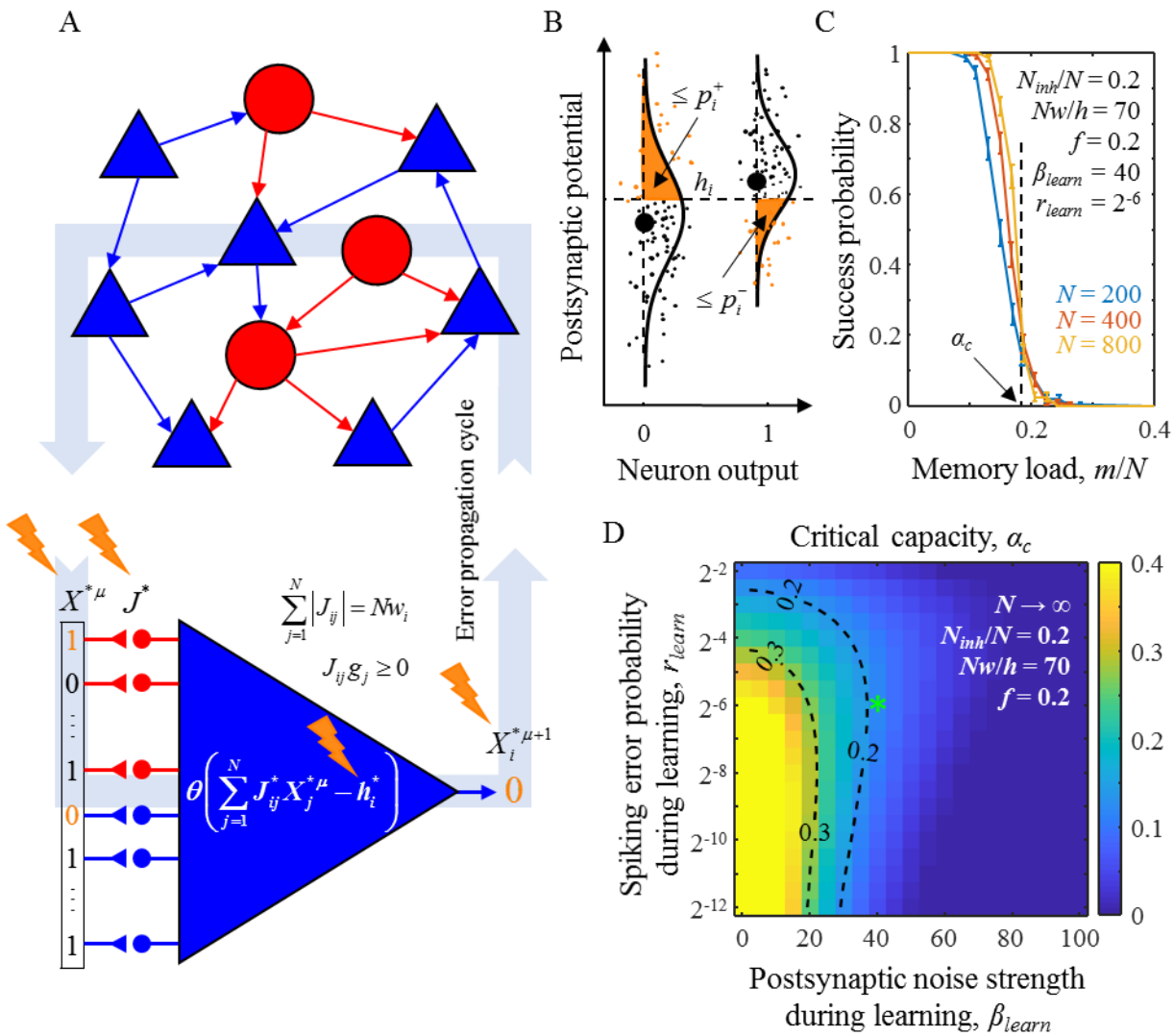

Figure 1: Associative memory storage in a recurrent network of inhibitory and excitatory neurons in the presence of errors and noise. **A.** A cycle of error propagation through the network. Inhibitory neurons (red circles) and excitatory neurons (blue triangles) form an all-to-all potentially (structurally) connected network. Red and blue arrows represent actual (functional) connections. Spiking errors, X^*, synaptic noise, J_{ij}^*, and intrinsic noise, h_i^*, accompany signal transmission

(orange lightning signs). Errors in the neurons' outputs at a given time step become the spiking errors for the next time step. **B.** Fluctuations in postsynaptic potentials for two associations with target neuron outputs 0 (left) and 1 (right). Large black dots denote postsynaptic potentials in the absence of errors and noise. Small dots represent postsynaptic potentials on different trials in the presence of errors and noise. Orange areas to the left of the postsynaptic potential probability densities (solid lines) represent the probabilities of erroneous spikes (left) and spike failures (right). **C.** The probability of successful learning by a neuron is a sharply decreasing function of associative sequence length, m (or memory load m/N). Solid curves represent the probabilities of successful learning obtained with nonlinear optimization (see section I of SI) for neurons receiving N = 200, 400, and 800 homogeneous inputs (same f and r). The values of β_{learn} and r_{learn} are provided in the figure. The numerical values of all other parameters of the model were adapted from (27). At 0.5 success probability, the neuron is said to be loaded to capacity, α. The dashed black line represents the theoretical (critical) capacity, α_c, obtained with the replica method in the $N \rightarrow \infty$ limit. **D.** Map of α_c for a neuron receiving homogeneous input as a function of β_{learn} and r_{learn}. The map was obtained with the replica method (see sections A-H of SI). The green asterisk indicates the values of β_{learn} and r_{learn} used in (C). Dashed isocontours are drawn as a guide to the eye.

B. Solutions of the model

Eqs. (2) were solved with the replica method, nonlinear optimization, and perceptron-type learning rule (see SI). Each of these methods has its advantages and drawbacks, and, consequently, all three methods were used in this study. The replica method (32, 33) provides an analytical solution in the $N \rightarrow \infty$ limit. Though neuron networks in the brain are finite, they are thought to be large enough to have many properties that are well described by this limit (18). More importantly, the analytical solution of the replica method reveals the dependence of the results on combinations of network parameters that can be explored with other methods. The downside of the replica solution is that it does not provide the full connectivity matrix, J_{ij}, but instead gives the connectivity statistics that cannot be readily used to calculate all relevant network properties. Nonlinear optimization can be used to solve Eqs. (2). This method is fast and accurate for small networks, yielding the full

connectivity matrix, but is impractical for large networks ($N \sim 1{,}000$). As the replica and nonlinear optimization solutions cannot be readily implemented by neural networks in the brain, we also developed a biologically more realistic perceptron-type learning rule that can be used to approximate the solution of Eqs. (2). Because simulations based on the perceptron-type learning rule are time-consuming, results for varying levels of errors and noise were obtained with the replica and nonlinear optimization methods, while the perceptron-type learning rule was used only for a biologically realistic set of parameters in order to confirm that all three methods lead to similar results.

In the $N \to \infty$ limit, the associative memory storage problem for a neuron loaded to capacity was solved with the replica method (see sections E-G of SI). This solution for a neuron in a homogeneous network ($f_i = f$, $r_i = r$, $\beta_{syn,i} = \beta_{syn}$, and $\beta_{int,i} = \beta_{int}$) depends on the following combination of the intrinsic and synaptic noise strengths (see section H of SI):

$$\beta = \sqrt{\beta_{int}^2 + \beta_{syn} f N \frac{w}{h}} \tag{3}$$

The latter is referred to as the postsynaptic noise strength. In the following, we assume that the postsynaptic noise strength, β, and the spiking error probability, r, can differ between the times of learning and memory retrieval and add subscripts "*learn*" and "*retr*" to these parameters to distinguish among the two phases.

Figure 1C shows that when the memory load is relatively low, the probability of successful learning by a neuron is close to 1. With increasing load, the learning problem becomes more difficult, and the success probability undergoes a smooth transition from 1 to 0. Memory load corresponding to the success probability of 0.5 is referred to as the neuron's associative memory storage capacity, α. This capacity depends on the levels of errors and noise accompanying learning and other parameters of the model. With increasing network size, N, the transition from successful learning to inability to accurately learn the complete memory sequence becomes sharper, and the neuron's capacity monotonically approaches its $N \to \infty$ limit, which is referred to as the critical capacity, α_c. Figure 1D illustrates the dependence of a single-neuron critical capacity on the postsynaptic noise strength and spiking error probability. As expected, because noise makes the

learning problem more challenging, α_c is a decreasing function of β_{learn}. Dependence of α_c on r_{learn} is more complex as this parameter controls both the presynaptic and output spiking errors of the neuron. On the one hand, an increase in r_{learn} makes the learning problem more challenging as the presynaptic errors increase and on the other, the problem becomes easier due to a higher tolerance of output errors.

C. The tradeoff between capacity and reliability of loaded memories

Can memories loaded into individual neurons be successfully recalled at the network level? To answer this question, we loaded neurons in the network to capacity with associations derived from a single associative sequence by solving Eqs. (2). The postsynaptic noise and spiking errors during learning were set at the levels β_{learn} and r_{learn} described by the green asterisk in Figure 1D. During memory retrieval, the network was initialized at the beginning of the loaded sequence, X^1, and no additional spiking errors, beyond those produced by the network at subsequent steps, were added during memory playout. At each step of memory playout, synaptic and intrinsic noise were added independently to every connection and every neuron in the network at strengths governed by β_{retr}, which can differ from $\beta_{learn.}$

The sequence is said to be retrieved completely if the network states during the retrieval do not deviate substantially from the target states. Otherwise, the sequence is said to be retrieved partially, and the retrieved sequence length is defined by the number of steps taken to the point where the network states begin to deviate substantially from the target states (Figure 2A). In practice, there is no need to precisely define the threshold amount of deviation. This is because for large networks the fraction of errors in a retrieved network state either fluctuates around $r_{learn} \pm \sqrt{r_{learn}\left(1-r_{learn}\right)/N}$ (mean ± SD) or diverges to $2f\left(1-f\right) \pm \sqrt{2f\left(1-f\right)\left(1-2f\left(1-f\right)\right)/N}$ (expected fraction of differences in two random network states of firing probability f), which is significantly greater for the chosen values of parameters r_{learn} and f. Figure 2B shows the probability of retrieving the complete loaded sequence and the fraction of retrieved sequence length for different values of β_{learn}. It illustrates that memory sequences can be reliably retrieved if they were loaded with the

postsynaptic noise strength that is slightly higher than that during memory retrieval. Likewise, the averaged retrieved sequence length fraction increases with β_{learn} and approaches one as the latter exceeds the noise strength present during retrieval. A similar conclusion can be drawn from Figure 2C which shows the map of the retrieval probability as a function of β_{learn} and r_{learn}. Errors and noise during learning make memory retrieval more reliable. However, the reliability of loaded memories comes at the expense of the memory storage capacity, α. Figure 2D shows the tradeoff between the retrieval probability and capacity of loaded associative memories in which higher levels of errors and noise during learning enable reliable memory retrieval but reduce α.

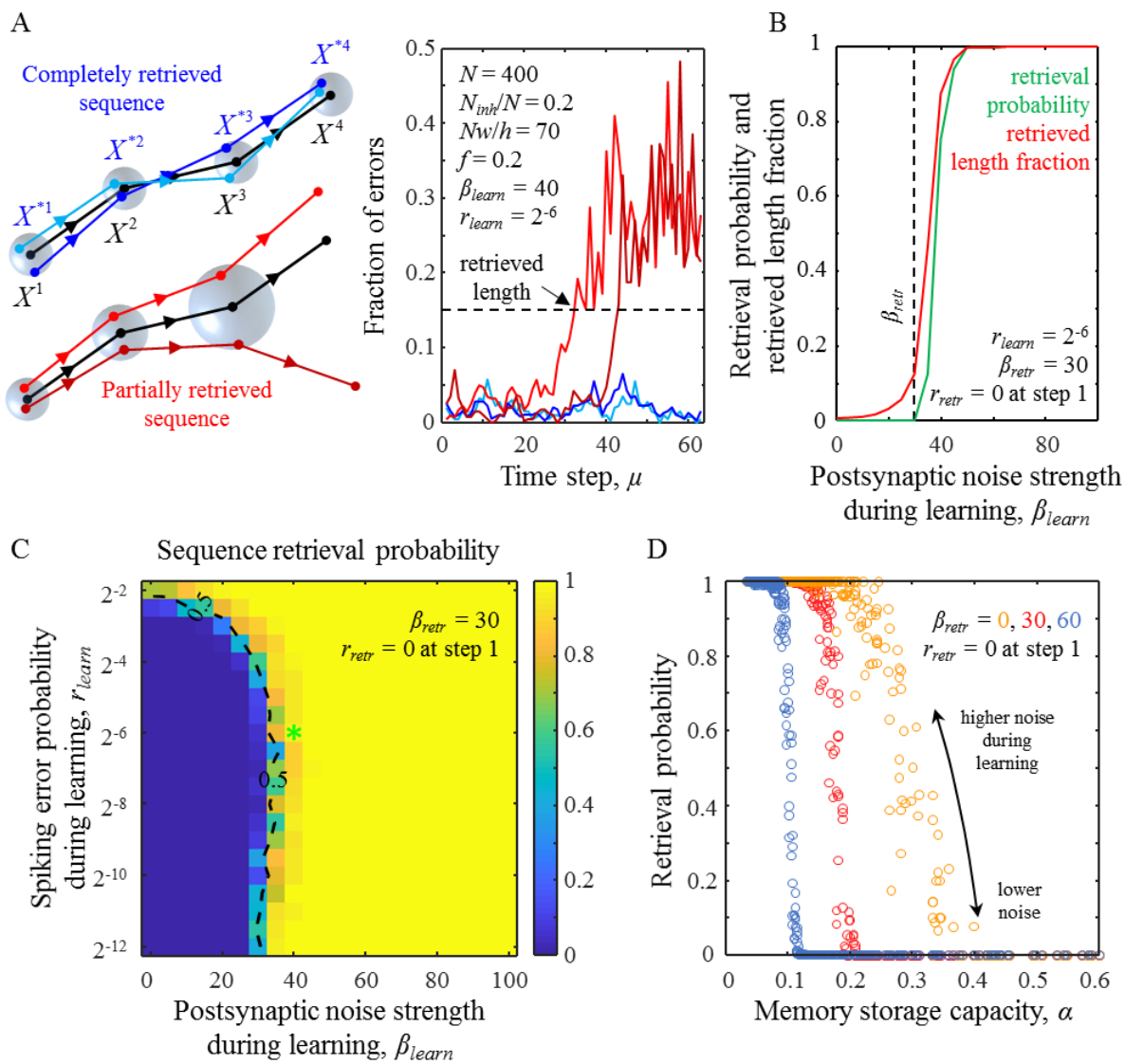

Figure 2: Retrieval of loaded associative memory sequences and the tradeoff between capacity and reliability of loaded memories. **A.** Illustration of memory playout during complete and partial memory retrieval (left). The target memory sequence is shown in black, while the sequences retrieved on different trials are in blue and red. Memory retrieval is incomplete when the retrieved sequence deviates significantly from the target sequence (see text for details). Radii of blue spheres illustrate the root-mean-square Euclidean distances between the retrieved and target states. The fraction of errors as a function of time step during sequence retrieval (right). Successfully retrieved sequences do not deviate from the loaded sequences by more than a threshold amount (dashed line). The parameters of the associative network are provided in the figure. The values of β_{learn} and r_{learn} correspond to the green asterisk from Figure 1D. **B.** The probability of successful memory retrieval (green) and the retrieved fraction of loaded sequence length (red) as a function of β_{learn}. The postsynaptic noise strength $\beta_{retr} = 30$ (dashed line) at every step of memory retrieval and r_{retr} was set to 0 at the first step. **C**. Map of retrieval probability as a function of β_{learn} and r_{learn}. Dashed isocontour is drawn as a guide to the eye. The location of the green asterisk is the same as in Figure 1D. **D.** The tradeoff between memory retrieval probability and α. Individual points correspond to all values of β_{learn} and r_{learn} considered in (C). Higher errors and noise during learning result in lower α and higher retrieval probability regardless of the noise strength during memory retrieval (different colors). The results shown in (A-D) were obtained with the nonlinear optimization method described in section I of SI. For every parameter setting, the results shown in (B-D) were averaged over 100 networks and 1,000 retrievals of the loaded sequence in each network.

D. Noise during learning is required for optimal retrieval of stored information

Figures 3A, B show the maps of expected retrieved information per sequence playout calculated in two different ways. In the first calculation, the contribution of partially retrieved sequences to the expected retrieved information was set to zero, while in the second, partially retrieved sequences contributed in proportion of the retrieved sequence length (see section K of SI for details). Both maps illustrate that optimal retrieval of stored information is achieved when memories are stored in the presence of noise, $\beta_{learn} > 0$. This conclusion is independent of the postsynaptic noise strength during memory retrieval, which was set to $\beta_{retr} = 30$ in Figures 3A, B.

To illustrate this finding, we averaged the maps over the r_{learn} dimension and determined β_{learn} that correspond to the maxima of the retrieved information. Figure 3C illustrates the results of this procedure for different values of β_{retr}, showing that the optimal β_{learn} is greater than zero even when there is no noise during memory retrieval. The optimal β_{learn} increases with β_{retr}, and when noise is high the two noise strengths approximately equal.

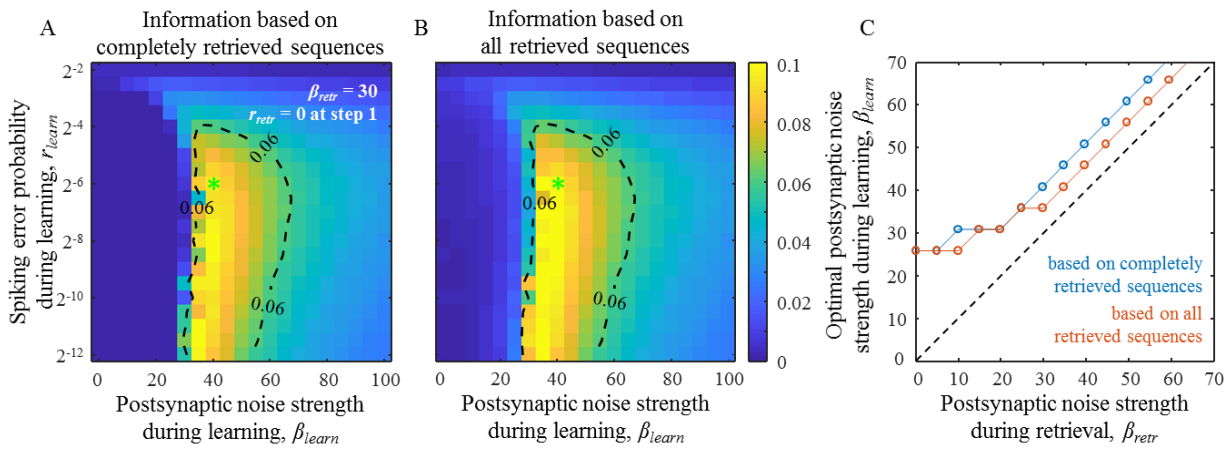

Figure 3: Postsynaptic noise during learning is required for optimal retrieval of stored of information. **A-B.** Maps of expected retrieved information per memory playout calculated based on completely retrieved sequences (A) and completely and partially retrieved sequence (B) in bits×N^2 as functions of β_{learn} and r_{learn}. β_{retr} = 30 at every step of memory retrieval and r_{retr} was set to 0 at the first step. Dashed isocontours are drawn as guides to the eye. The locations of the green asterisks are the same as in Figure 1D. **C.** The maximum of retrieved information is achieved when β_{learn} is greater than zero regardless of the value of β_{retr}. The optimal postsynaptic noise strengths were calculated based on the averages of the results from (A) (blue line) and (B) (orange line) over the range of r_{learn} values from (A, B). All results were obtained with the nonlinear optimization method described in section I of SI and averaged over 100 networks and 1,000 retrievals of the loaded sequence in each network for every parameter setting.

E. Neuron-to-neuron connectivity in associative networks of homogeneous inhibitory and excitatory neurons

One of the most salient features of sign-constrained associative learning models, such as the one described in this study, is that finite fractions of inhibitory and excitatory connections assume zero weights at capacity (25), mirroring the trend observed in many local cortical networks. We compared the connection probabilities (P_{con}) and the coefficients of variation (CV) of non-zero connection weights in associative networks at capacity to the connection probabilities and CVs of unitary postsynaptic potentials (uPSP) obtained experimentally. To that end, we used the dataset compiled in (18) based on 87 electrophysiological studies describing neuron-to-neuron connectivity for 420 local cortical projections (lateral distance between neurons < 100 μm). Figure 4A shows that the average inhibitory P_{con} (38 studies, 9,522 connections tested) is significantly larger ($p < 10^{-10}$, two-sample t-test) than the average excitatory P_{con} (67 studies, 63,020 connections tested). Associative networks exhibit a similar trend in the entire region of considered β_{learn} and r_{learn} values (Figures 4B, C). What is more, in the (β_{learn}, r_{learn}) parameter region demarcated with the dashed isocontours and arrows in Figures 4B, C, the model results are consistent with the middle 50% of the experimentally measured P_{con} values for inhibitory and excitatory connections.

Figure 4D shows that the average CV of inhibitory uPSP (10 studies, 503 connections recorded) is slightly lower than that for excitatory (36 studies, 3,956 connections recorded), and this trend is also reproduced by the associative networks in the entire region of considered β_{learn} and r_{learn} values (Figures 4E, F). As before, there are (β_{learn}, r_{learn}) parameter regions in these maps in which the results of the model are consistent with the middle 50% of the CV of uPSP measurements for inhibitory and excitatory connections.

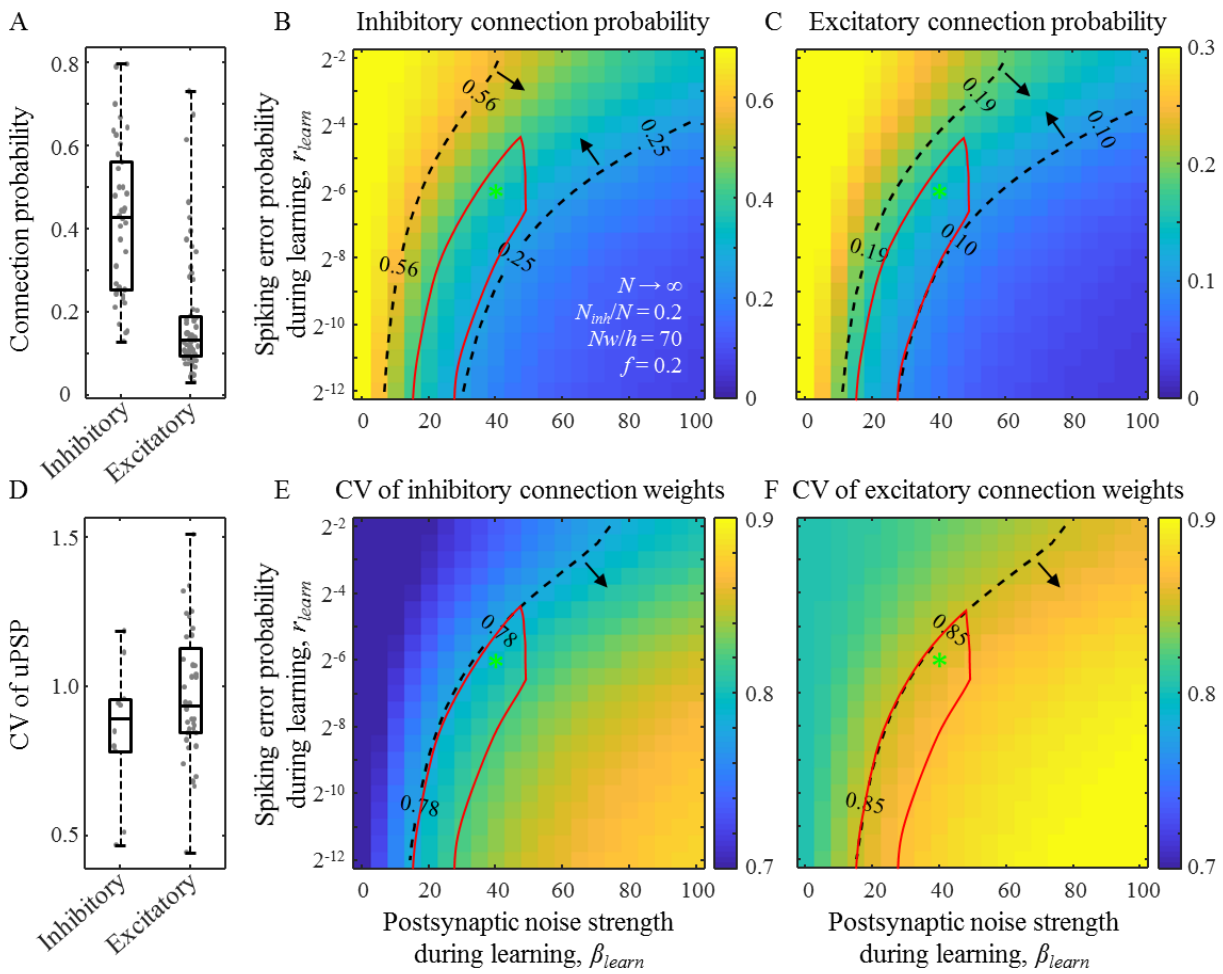

Figure 4: Comparison of structural properties of the model and cortical networks. **A.** Inhibitory and excitatory connection probabilities reported in 87 studies describing 420 local cortical projections. Each dot represents the result of a single study/projection. **B, C.** Maps of inhibitory and excitatory connection probabilities as functions of β_{learn} and r_{learn}. The results are based on the replica method (see section H of SI). Dashed isocontours and arrows illustrate the interquartile ranges of the experimentally observed connection probabilities from (A). The red contour outlines a region of parameters that is consistent with all structural and dynamical measurements in cortical networks considered in this study. The locations of the green asterisks are the same as in Figure 1D. **D-F.** Same for the CV of inhibitory and excitatory connection weights. (A) and (D) were adapted from (18).

F. Spontaneous dynamics in associative networks of homogeneous inhibitory and excitatory neurons

The model associative networks can exhibit irregular and asynchronous spiking activity like that observed in cortical networks. To analyze such spontaneous (not learned) network dynamics, we used associative networks loaded to capacity, initialized them at random states of firing probability $f = 0.2$, and followed their activity for 1,000 time-steps. Because the number of available network states, which is exponential in N, is much larger than the number of loaded states, αN, the spontaneous network activity in the numerical simulations never passed through any of the loaded states.

To quantify the degree of similarity in the dynamics of the model and brain networks we compared the CV of inter-spike-intervals (ISI) and the cross-correlation coefficient of spiking neuron activity in the model to those measurements obtained experimentally. Dashed isocontour in Figure 5A outlines (β_{learn}, r_{learn}) parameter region in which the model CV of ISI is consistent with the 0.7-1.1 range measured in different cortical systems (34-38). Similarly, Figure 5B shows that there is a (β_{learn}, r_{learn}) parameter region in which the calculated spike cross-correlation coefficients are in agreement with the interquartile range of the corresponding cortical measurements, 0.04-0.15 (39). The degree of asynchrony in spontaneous spiking activity in associative networks increases with the postsynaptic noise strength, which can be explained by the decrease in connection probability (Figures 4B, C) and, consequently, a reduction in the amount of common input to the neurons.

It was shown that irregular and asynchronous activity can result from the balance of inhibitory and excitatory postsynaptic inputs to individual cells (40, 41). In a balanced state, the magnitudes of these inputs are much greater than the threshold of firing, but, due to a high degree of anti-correlation, these inputs nearly cancel, and firing is driven by fluctuations. Figure 5C shows a region of parameters in which neurons in the associative model function in a balanced regime. Because it is difficult to simultaneously measure inhibitory and excitatory postsynaptic inputs to a neuron, the anti-correlation of inhibitory and excitatory inputs has only been measured in nearby cells, averaging to about 0.4 (42, 43). As within-cell anti-correlations are expected to be stronger than between-cell anti-correlations, 0.4 was used as a lower bound for the former (dashed isocontour and arrow in Figure 5C).

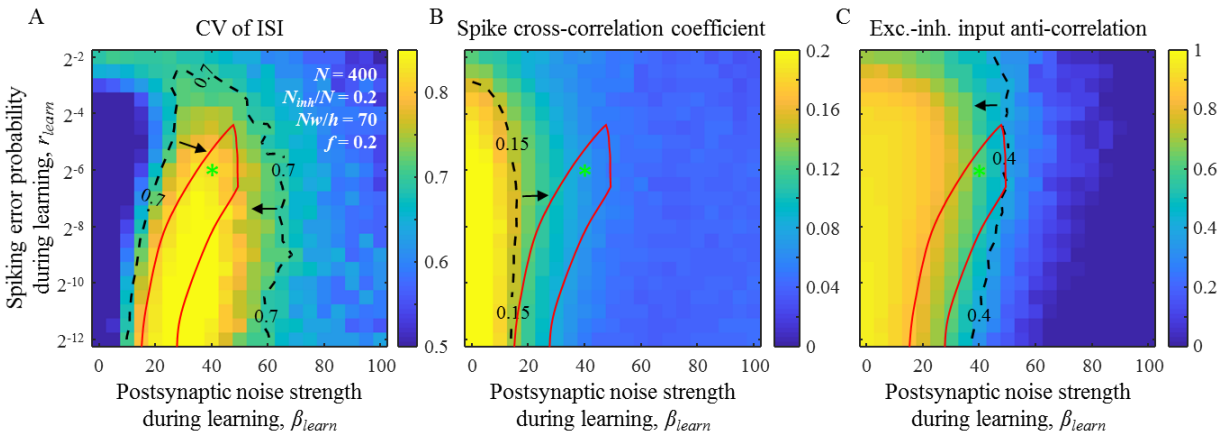

Figure 5: Comparison of dynamical properties of the model and cortical networks. **A.** The CV of ISI for spontaneous (not learned) activity as a function of β_{learn} and r_{learn}. Results were obtained with nonlinear optimization (see section I of SI). Dashed isocontour and arrows demarcate a region of CV values that is in general agreement with experimental measurements. **B.** Same for the cross-correlation coefficient of neuron spike trains. **C.** Same for the anti-correlation coefficient of inhibitory and excitatory postsynaptic inputs to a neuron. The red contour outlines a region of parameters that is consistent with the considered structural and dynamical measurements. The locations of the green asterisk are the same as in Figure 1D. All results were obtained with the nonlinear optimization method described in section I of SI and averaged over 100 networks and 100 runs for each network for every parameter setting.

The seven error-noise regions obtained based on the properties of neuron-to-neuron connectivity (Figure 4) and network dynamics (Figure 5) have a non-empty intersection (red contour in Figures 4 and 5). In this biologically plausible region of parameters, the considered properties of the associative networks are consistent with the corresponding experimental measurements. This observation suggests that β_{learn} must lie in the 20-50 range and r_{learn} must be less than 0.06. While we are not aware of direct experimental measurements of these parameters, the low value of r_{learn} is in qualitative agreement with the reliability of firing patterns evoked by time-varying stimuli *in vivo* (38) and *in vitro* (44).

G. Solution of the model with a perceptron-type learning rule

As the replica and nonlinear optimization solutions of Eqs. (2) cannot be easily implemented by neural networks in the brain, we set out to develop a biologically more plausible online solution to the associative learning problem. The following perceptron-type learning rule was devised to approximate the solution of Eqs. (2) (see section J of SI). At each learning step, e.g. μ, a neuron receives an input containing spiking errors, $X^{*\mu}$, combines it with postsynaptic noise, and produces an output, $y^{*\mu}$. If this output differs from the neuron's target output, y^{μ}, the neuron's input connection weights are updated in four consecutive steps:

$$\begin{aligned}
&J_j \mapsto J_j + \frac{\gamma h}{N}\left(2y^{\mu}-1\right)X_j^{*\mu}, \quad j=1,\ldots,N \\
&J_j \mapsto J_j\theta\left(J_j g_j\right) \\
&J_j \mapsto J_j + \left(w - \frac{1}{N}\sum_{j=1}^{N}\left|J_j\right|\right)g_j \\
&J_j \mapsto J_j\theta\left(J_j g_j\right)
\end{aligned} \tag{4}$$

The first line in Eqs. (4) is a stochastic perceptron learning step (45), in which parameter γ is referred to as the learning rate. The second line enforces the sign constraints, while the last two lines combined implement the homeostatic l_1-norm constraint and are equivalent to the soft thresholding used in LASSO regression (46). In contrast to the standard perceptron learning rule, Eqs. (4) utilize noisy inputs and enforce sign and homeostatic constraints at every learning step.

By including presynaptic spiking errors, synaptic and intrinsic noise in the condition that triggers the learning steps outlined in Eqs. (4), the learning rule implicitly depends on the model parameters $\{r_j\}$, $\{\beta_{syn,\,j}\}$ describing the fluctuations in the neuron's inputs (indexed with j), and the parameter β_{int} which describes the neuron's intrinsic noise. Because Eqs. (4) are designed to approximately minimize the neuron's output spiking error probability for a given memory load (see section J of SI), which at capacity matches the desired output error probability of the neuron r, the learning rule also depends implicitly on fluctuations in the neuron's output.

Figure 6 compares the theoretical solution obtained with the replica method in the $N \rightarrow \infty$ limit with numerical solutions for networks of N = 200, 400, and 800 neurons obtained with nonlinear optimization and the perceptron-type learning rule. Figure 6A shows that the perceptron-type learning rule sometimes fails to find a solution to a feasible learning problem, i.e. a problem that can be solved with nonlinear optimization. Yet, even in such cases, the perceptron connection weights in a steady state (after 10^6 learning epochs) are well-correlated with the nonlinear optimization weights (Figure 6B). Therefore, though the perceptron-type learning rule is not as efficient as nonlinear optimization, it can find an approximate solution to the learning problem. Consistent with this conclusion, the associative memory storage capacity of a neuron loaded with the perceptron-type learning rule is 15% - 18% lower than that loaded with nonlinear optimization, and the two methods lead to similar structural and dynamical network properties (red and blue bars in Figure 6C).

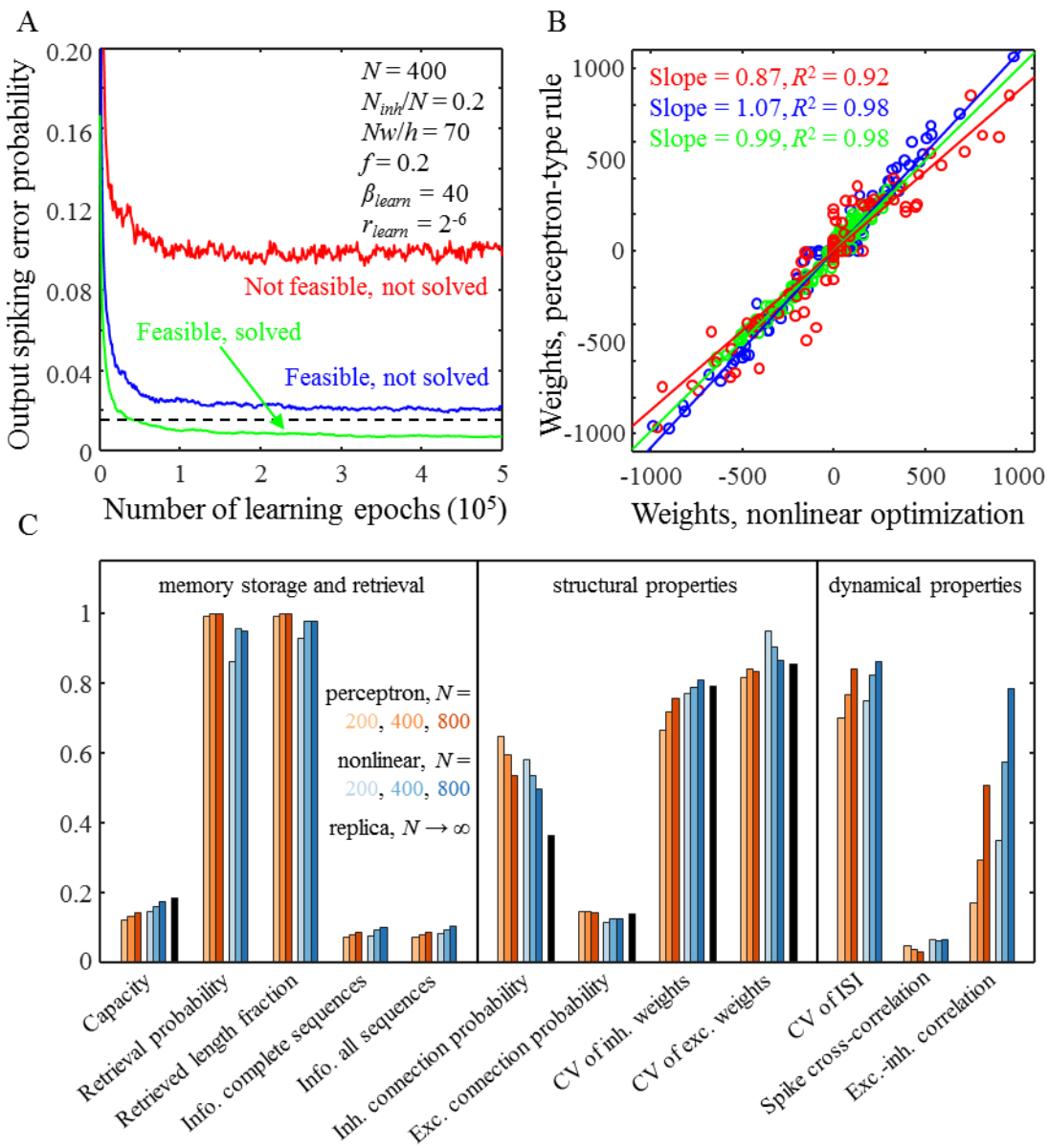

Figure 6: Comparison of solutions obtained with the perceptron-type learning rule, nonlinear optimization, and replica method. **A.** Output error probability as a function of the number of learning epochs for the perceptron-type learning rule. The black dashed line indicates the target output error probability. Results for three different cases are shown: a not feasible problem (red line), a feasible problem which was not solved with the perceptron-type learning rule (blue line), and a feasible problem which was solved with the perceptron-type learning rule (green line). The parameters of the associative network are provided in the figure. The values of β_{learn} and r_{learn} correspond to the green asterisk from Figure 1D. **B.** Comparisons of connection weights obtained

with the perceptron-type learning rule and nonlinear optimization for the three cases shown in (A). Straight lines are the best linear fits. **C.** Comparisons of memory storage capacity, retrieval, structural, and dynamical properties in networks of N = 200, 400, and 800 neurons obtained with the perceptron-type learning rule (red colors) and nonlinear optimization (blue colors). The memory storage capacity and structural properties calculated with the replica method in the $N \to \infty$ limit are shown in black.

H. Properties of heterogeneous associative networks

The associative learning model, Eqs. (2), makes it possible to investigate the properties of networks composed of heterogeneous populations of inhibitory and excitatory neurons. Specifically, we examined the effects of distributed spiking error probabilities and distributed synaptic and intrinsic noise strengths on properties of connectivity at critical capacity. Figures 7A-C show that in networks of neurons with heterogeneous spiking error probabilities (homogeneous in all other parameters), the probabilities and weights of inhibitory and excitatory connections monotonically decrease with increasing r_{learn}. Therefore, as may have been expected, connections originating from unreliable presynaptic neurons (high r_{learn}) are depressed and/or eliminated during learning. Properties of networks of neurons with distributed synaptic and intrinsic noise strengths (homogeneous otherwise) depend on the combination of these parameters in the form of the postsynaptic noise strengths, β_{learn}. Figures 7D-F show how connection probabilities and average connection weights depend on β_{learn}. Similar to the previous case, connections onto noisy neurons (high β_{learn}) are less probable. Here, however, the average inhibitory and excitatory connection weights increase with β_{learn} due to the homeostatic l_1-norm constraint, Eqs. (2).

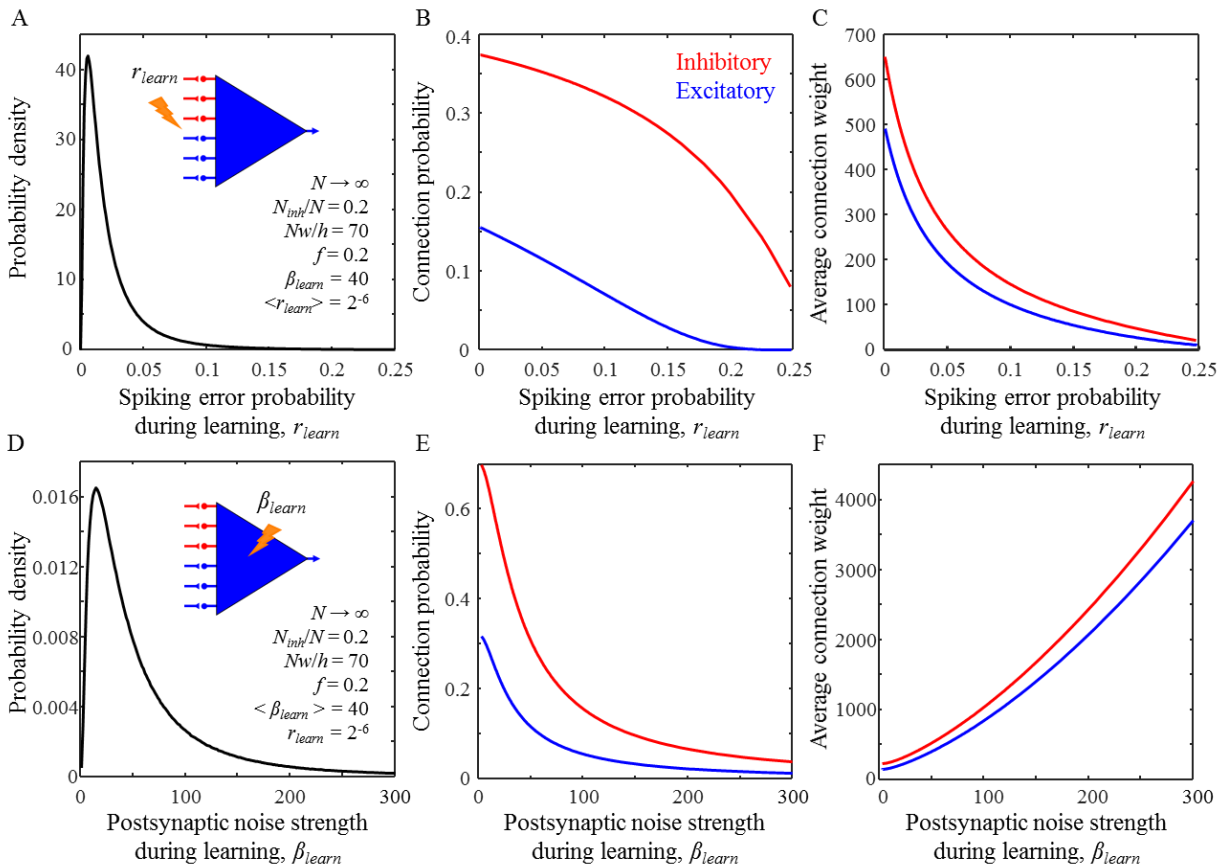

Figure 7: Properties of connections in associative networks of heterogeneous neurons. **A-C.** Connection probability (B) and average non-zero connection weight (C) for inhibitory (red) and excitatory (blue) connections in a network of neurons with distributed spiking error probabilities and homogeneous in all other parameters. The spiking error probabilities of inhibitory and excitatory inputs during learning were randomly drawn from the log-normal distribution shown in (A). Unreliable inputs have lower probabilities and weights. The parameters of the associative network are shown in (A). The values of β_{learn} and $<r_{learn}>$ correspond to the green asterisk from Figure 1D. **D-F.** Same for a network of neurons with heterogeneous postsynaptic noise strengths. The postsynaptic noise strengths of neurons during learning were randomly drawn from the log-normal distribution shown in (D). Noisy neurons receive stronger but fewer inhibitory and excitatory inputs.

Motivated by the agreement between the results of the associative learning model and cortical measurements, we put forward two predictions that can be tested in future experiments. First, we predict that in cortical networks, inhibitory and excitatory connections originating from unreliable neurons or neuron classes must have lower connection probabilities and average uPSPs (Figures 7B, C). Second, we predict that connections onto noisy neurons or neuron classes must have lower connection probabilities but higher average uPSPs (Figures 7E, F).

DISCUSSION

This study incorporates a comprehensive description of the error propagation cycle into the model of associate sequence learning by recurrent networks of neurons with biologically inspired constraints. It shows that errors and noise during learning can be beneficial, as they can increase the reliability of loaded memories to fluctuations during memory retrieval. Because errors and noise are both free and unavoidable harnessing their power, rather than trying to suppress it, may be an efficient way of improving the reliability of memories in the brain. This mechanism is illustrated in Figure 8. When the associative memories are loaded at a below capacity level, the solution region of Eqs. (2) is relatively large. A solution, e.g. a vector of connection weights of a neuron obtained with a perceptron-type learning rule, may be located near the solution region boundary. Such a solution is deemed unreliable because a small amount of noise during memory retrieval can move it outside the solution region, resulting in spiking errors that can disrupt the associative sequence retrieval process (Figures 8A). By adding noise during learning, the solution can be forced to move away from the boundary, thus making it more reliable (Figures 8B). However, increasing the noise strength reduces the neuron's capacity, and at a certain strength, the capacity and memory load are guaranteed to match (Figure 8C). A further increase in noise strength can improve the reliability even more, but at the expense of the memory load as the latter must remain at or below capacity (Figure 8D). An alternative way of improving reliability is by suppressing noise during memory retrieval (Figure 8E). Interestingly, it has been shown that visual attention that improves behavioral performance reduces the variability in spike counts of individual neurons in Macaque V4 (5, 6). Though significant, the amount of reduction is relatively small, suggesting that this mechanism has physical limitations. Therefore, using noise during

learning can enhance the reliability of stored memories beyond what can be accomplished by attending to the memory retrieval process.

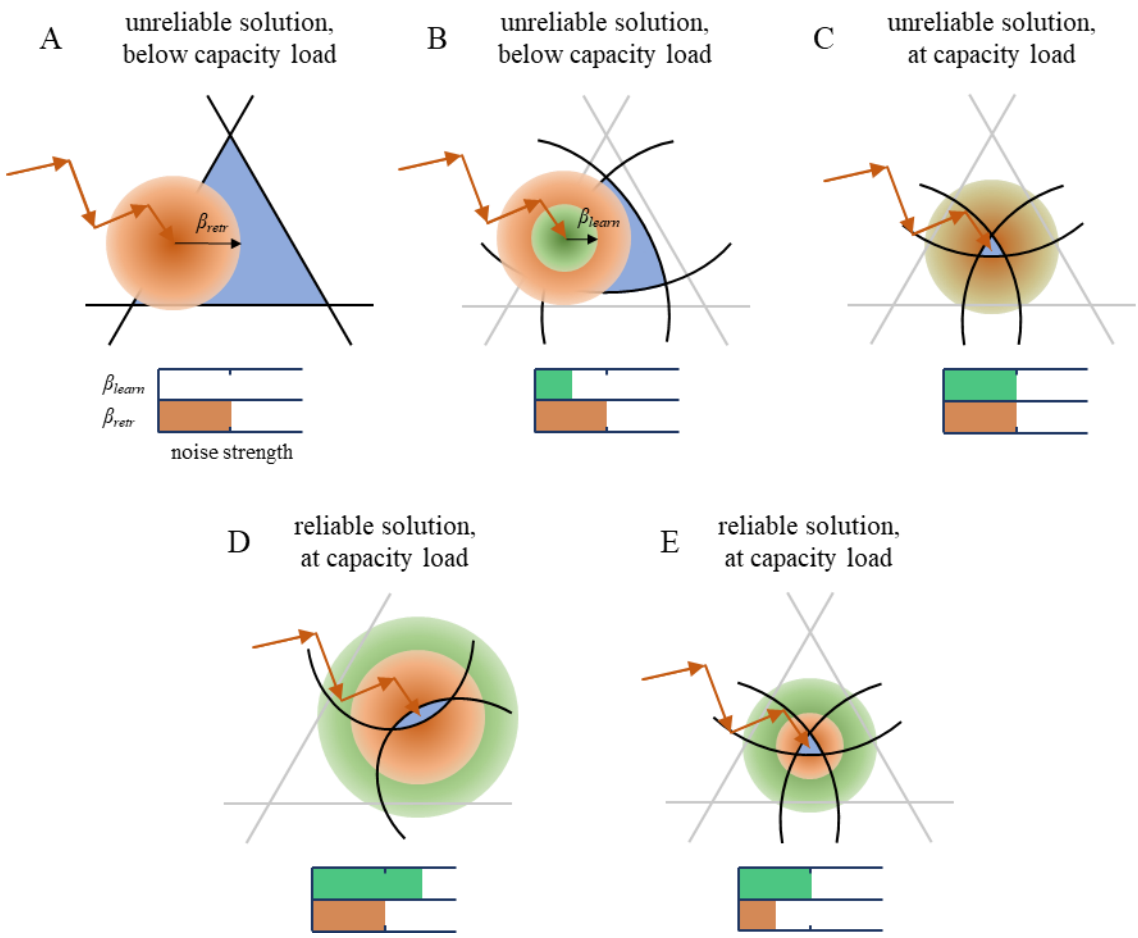

Figure 8: Increasing the noise strength during learning or decreasing it during memory recall leads to more reliable solutions. **A.** The associative learning problem for a below capacity load in the absence of noise during learning, $\beta_{learn} = 0$. The solution region (blue) is bounded by hyperplanes corresponding to the individual associations (black lines). The learning phase (red arrows) ends as the connection weight vector enters the solution region. The solution shown in (A) is unreliable because noise during memory retrieval (red cloud of radius β_{retr}) can move it outside the solution region with high probability. **B.** Adding noise during learning (green cloud of radius β_{learn}) transforms the association hyperplanes (gray lines) into hypersurfaces (black lines), Eqs. (2), reducing the solution region and forcing the connection weight vector further away from the hyperplanes. This increases solution reliability. **C.** The continued increase of the noise strength

improves reliability as the solution region shrinks to zero. At this noise strength, the memory load is at capacity. A further increase in reliability can be achieved by increasing the noise strength during learning (D) or decreasing it during retrieval (E). In the former case, the memory load must be reduced to match the reduction in capacity.

The model described in this study assumes that individual neurons learn independently from one another and are loaded with memories to capacity. There is no direct support for these assumptions, but they have been shown to lead to structural and dynamical network properties that are consistent with experimental data (9, 18, 26, 28). This study corroborates these assumptions by matching a variety of experimental results with a single set of model parameters. The derived perceptron-type rule can mediate learning by modifying connection weights based on local activities of pre- and postsynaptic neurons in the presence of errors and noise. This is biologically feasible; however, a supervision signal must be fed to each neuron during learning. This is a major drawback of the presented rule, and the supervised learning rules in general because the origins of this signal in the brain remain unknown.

ACKNOWLEDGMENTS

This work was supported by the AFOSR grant FA9550-15-1-0398 and the NSF grant IIS-1526642.

SUPPORTING INFORMATION

This Supporting Information describes a model of associative memory storage by a McCulloch and Pitts neuron (17) receiving inhibitory and excitatory inputs (Figure 1A, bottom). In contrast to previous studies [see e.g. (9, 10, 18, 22, 23, 26-28)], the model explicitly accounts for several distinct sources of errors and noise present during learning and incorporates two biologically inspired constraints on connectivity. The model was solved theoretically with the replica method in the limit of infinite network size (32, 33) and numerically, with nonlinear optimization and perceptron-type learning rule, for large but finite networks.

A. Perceptron model with biologically inspired constraints in the presence of errors and noise

We consider a single perceptron-like neuron that receives N_{inh} inhibitory and ($N - N_{inh}$) excitatory input connections (Figure 1A). The neuron is faced with a task of learning a set of m input-output associations $\left\{X^{\mu}, y^{\mu}\right\}$, where X^{μ} and y^{μ} are a binary vector and a scalar describing the neuron's input and target output at time step μ. The neuron must learn m input-output associations by adjusting the weights of its connections in the presence of two constraints. First, the l_1-norm of connection weights must remain fixed during learning. Second, the signs of inhibitory and excitatory connection weights must not change during learning. In the following, J_j denotes the weight of connection j.

Learning in the model is accompanied by four types of errors and noise. These include presynaptic and output spiking errors, or errors in X^{μ} and y^{μ}, synaptic noise, or noise in J, and intrinsic noise, or noise in the neuron's threshold of firing, h. In the following, we use asterisks to denote quantities containing errors or noise (e.g. $X^{*\mu}$), whereas variables without asterisks represent the mean (for h and J_j) or target (for X^{μ} and y^{μ}) values.

Individual elements of X^{μ} (X_j^{μ}), are independently drawn from connection-dependent Bernoulli probability distributions: 0 with probability $1 - f_j$ and 1 with probability f_j. Errors in X^{μ} are introduced with Bernoulli trials by making independent and random changes in X_j^{μ} with

probabilities $P\left(X_j^{*\mu}=0\,|\,X_j^{\mu}=1\right)=p_j^-$ and $P\left(X_j^{*\mu}=1\,|\,X_j^{\mu}=0\right)=p_j^+$. Similarly, the target outputs, y^{μ}, are independently drawn from a Bernoulli probability distribution: 0 with probability $1-f_{out}$ and 1 with probability f_{out}. The probabilities of error in the neuron's output are bounded, $P\left(y^{*\mu}=0\,|\,y^{\mu}=1\right)\le p_{out}^-$ and $P\left(y^{*\mu}=1\,|\,y^{\mu}=0\right)\le p_{out}^+$ (Figure 1B). In the following, we only consider the output spiking error probabilities in the ranges $p_{out}^+ < f_{out}$ and $p_{out}^- < 1-f_{out}$, which is required for the stability of the replica solution. Noises in the neuron's connection weights and postsynaptic potential are described in the main text. The resulting model is equivalent to Eqs. (2) with minor changes in the notation: $X_i^{\mu+1}$ is replaced with y^{μ}, f_i with f_{out}, $p_i^{\pm}$ with $p_{out}^{\pm}$, and index i is dropped. The model can be summarized as follows:

$$
\begin{aligned}
&\theta\left(\sum_{j=1}^{N} J_j^* X_j^{*\mu} - h^*\right) = y^{*\mu}; \quad \mu=1,\ldots,m, \quad j=1,\ldots,N \\
&\frac{1}{N}\sum_{j=1}^{N} J_j g_j = w \\
&J_j g_j \ge 0 \\
&P\left(X_j^{\mu}=1\right)=f_j; \quad P\left(X_j^{*\mu}=0\,|\,X_j^{\mu}=1\right)=p_j^-; \quad P\left(X_j^{*\mu}=1\,|\,X_j^{\mu}=0\right)=p_j^+ \\
&P\left(y^{\mu}=1\right)=f_{out}; \quad P\left(y^{*\mu}=0\,|\,y^{\mu}=1\right)\le p_{out}^-; \quad P\left(y^{*\mu}=1\,|\,y^{\mu}=0\right)\le p_{out}^+ \\
&\left\langle J_j^*\right\rangle = J_j; \quad \mathrm{var}\left(J_j^*\right)=\frac{h\beta_{syn}}{N}\left|J_j\right| \\
&\left\langle h^*\right\rangle = h; \quad \mathrm{var}\left(h^*\right)=\frac{h^2\beta_{int}^2}{N}
\end{aligned}
\tag{S1}
$$

In these expressions, θ is the Heaviside step-function. To enforce sign-constraints on connection weights we introduced scalars g_j, equaling 1 if the input j is excitatory and –1 if it is inhibitory. Parameter w is referred to as the average absolute connection weight. The neuron is faced with the task of finding the connection weights, $\{J_j\}$, that satisfy Eqs. (S1) for a given set of model parameters: $N, m/N, h, w, \{g_j\}, \{f_j\}, f_{out}, \{p_j^-\}, \{p_j^+\}, p_{out}^-, p_{out}^+, \beta_{syn}, \beta_{int}$.

B. Reformulation of the model in the large *N* limit

In the limit of large N, the Central Limit Theorem ensures that the neuron's postsynaptic potential (PSP), $\sum_{j=1}^{N} J_j^* X_j^{*\mu}$, is Gaussian distributed at every time step. Therefore, the deviation of PSP from the threshold of firing, $I^{*\mu} = \sum_{j=1}^{N} J_j^* X_j^{*\mu} - h^*$, is also Gaussian distributed with the mean and standard deviation given by the following expressions:

$$I^{\mu} = \sum_{j=1}^{N} J_j \left[\left(1 - p_j^-\right) X_j^{\mu} + p_j^+ \left(1 - X_j^{\mu}\right) \right] - h$$

$$\sigma^{\mu} = \left(\sum_{j=1}^{N} \left[J_j^2 \left(p_j^- \left(1 - p_j^-\right) X_j^{\mu} + p_j^+ \left(1 - p_j^+\right)\left(1 - X_j^{\mu}\right) \right) + \frac{h\beta_{syn}}{N} J_j g_j \left(\left(1 - p_j^-\right) X_j^{\mu} + p_j^+ \left(1 - X_j^{\mu}\right) \right) + \frac{h^2 \beta_{int}^2}{N^2} \right] \right)^{\frac{1}{2}}$$

(S2)

As a result, the inequality constraints on the probabilities of output spiking errors [line five in Eqs. (S1)] can be expressed in terms of I^{μ} and σ^{μ}:

$$\begin{aligned} I^{\mu} &\geq \sqrt{2}\,\mathrm{erf}^{-1}\left(1 - 2p_{out}^-\right)\sigma^{\mu}, \quad y^{\mu} = 1 \\ I^{\mu} &\leq -\sqrt{2}\,\mathrm{erf}^{-1}\left(1 - 2p_{out}^+\right)\sigma^{\mu}, \quad y^{\mu} = 0 \end{aligned} \tag{S3}$$

The above two inequalities can be combined into a single expression that must hold for a successfully learned association μ:

$$\left(2y^{\mu} - 1\right) I^{\mu} \geq \sqrt{2}\left(\mathrm{erf}^{-1}\left(1 - 2p_{out}^-\right) y^{\mu} + \mathrm{erf}^{-1}\left(1 - 2p_{out}^+\right)\left(1 - y^{\mu}\right)\right)\sigma^{\mu} \tag{S4}$$

C. Additional assumptions required for the replica calculation

Following the procedure outlined in (9, 26), we assumed that the model parameters m/N, $\{f_j\}$, f_{out}, $\{p_j^-\}$, $\{p_j^+\}$, p_{out}^-, p_{out}^+, β_{syn}, and β_{int} are intensive, or of order 1 with respect to N. In addition, we assumed that the connection weights are inversely proportional to the system size, $\left\{ J_j = \frac{h}{N} \tilde{J}_j \right\}$,

and refer to $\left\{\tilde{J}_j\right\}$ as scaled connection weights. This particular scaling is traditionally used in associative memory models (9), and it has been shown that in the biologically plausible high-weight regime, $Nwf \gg h$, many model results become independent of this assumption (18). It follows from the second line of Eqs. (S1) that $w = \frac{h}{N}\tilde{w}$, and we refer to $\tilde{w}$ as scaled average absolute connection weight.

The model, rewritten in terms of the scaled variables, contains one equality and $m + N$ inequality constraints:

$$\begin{aligned}
&\left(2y^{\mu}-1\right)\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]-1\right)\geq\sqrt{\frac{2}{N}}\left(\mathrm{erf}^{-1}\left(1-2p_{out}^-\right)y^{\mu}+\mathrm{erf}^{-1}\left(1-2p_{out}^+\right)\left(1-y^{\mu}\right)\right)\times\\
&\left(\frac{1}{N}\sum_{j=1}^{N}\left[\tilde{J}_j^2\left(p_j^-\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-p_j^+\right)\left(1-X_j^{\mu}\right)\right)+\beta_{syn}\tilde{J}_j g_j\left(\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right)+\beta_{int}^2\right]\right)^{\frac{1}{2}};\quad \mu=1,\ldots,m\\
&\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j g_j=\tilde{w}\\
&\tilde{J}_j g_j\geq 0;\quad j=1,\ldots,N\\
&P\left(X_j^{\mu}=1\right)=f_j;\quad P\left(y^{\mu}=1\right)=f_{out}
\end{aligned}$$

(S5)

D. Replica theory solution of the model

We begin by calculating the volume of the connection weight space, $\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)$, in which Eqs. (S5) hold for a given set of associations, $\left\{X^{\mu},y^{\mu}\right\}$:

$$\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)=\int\prod_{j=1}^{N}d\tilde{J}_{j}\prod_{j=1}^{N}\theta\left(\tilde{J}_{j}g_{j}\right)\delta\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_{j}g_{j}-\tilde{w}\right)\prod_{\mu=1}^{m}\theta\left(\left(2y^{\mu}-1\right)\times\right.$$

$$\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_{j}\left[\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-X_{j}^{\mu}\right)\right]-1\right)-\sqrt{\frac{2}{N}}\left(\mathrm{erf}^{-1}\left(1-2p_{out}^{-}\right)y^{\mu}+\mathrm{erf}^{-1}\left(1-2p_{out}^{+}\right)\left(1-y^{\mu}\right)\right)\times$$

$$\left.\left(\frac{1}{N}\sum_{j=1}^{N}\left[\tilde{J}_{j}^{2}\left(p_{j}^{-}\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-p_{j}^{+}\right)\left(1-X_{j}^{\mu}\right)\right)+\beta_{syn}\tilde{J}_{j}g_{j}\left(\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-X_{j}^{\mu}\right)\right)+\beta_{int}^{2}\right]\right)^{\frac{1}{2}}\right)$$

(S6)

The typical volume of this solution space, $\Omega_{typical}$, is defined through the averaging of $\ln\left(\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)\right)$ over the set of associations $\left\{X^{\mu},y^{\mu}\right\}$, and is calculated by introducing n replica systems:

$$\ln\left(\Omega_{typical}\right)=\left\langle\ln\left(\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)\right)\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}}=\lim_{n\to 0}\frac{\left\langle\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^{n}\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}}-1}{n} \tag{S7}$$

The quantity $\left\langle\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^{n}\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}}$ can be rewritten as a single multidimensional integral:

$$\left\langle\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^{n}\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}}=\left\langle\int\prod_{j,a=1}^{N,n}d\tilde{J}_{j}^{a}\prod_{j,a=1}^{N,n}\theta\left(\tilde{J}_{j}^{a}g_{j}\right)\prod_{a=1}^{n}\delta\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_{j}^{a}g_{j}-\tilde{w}\right)\prod_{\mu,a=1}^{m,n}\theta\left(\left(2y^{\mu}-1\right)\times\right.\right.$$

$$\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_{j}^{a}\left[\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-X_{j}^{\mu}\right)\right]-1\right)-\sqrt{\frac{2}{N}}\left(\mathrm{erf}^{-1}\left(1-2p_{out}^{-}\right)y^{\mu}+\mathrm{erf}^{-1}\left(1-2p_{out}^{+}\right)\left(1-y^{\mu}\right)\right)\times$$

$$\left.\left.\left(\frac{1}{N}\sum_{j=1}^{N}\left[\left(\tilde{J}_{j}^{a}\right)^{2}\left(p_{j}^{-}\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-p_{j}^{+}\right)\left(1-X_{j}^{\mu}\right)\right)+\beta_{syn}\tilde{J}_{j}^{a}g_{j}\left(\left(1-p_{j}^{-}\right)X_{j}^{\mu}+p_{j}^{+}\left(1-X_{j}^{\mu}\right)\right)+\beta_{int}^{2}\right]\right)^{\frac{1}{2}}\right)\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}}$$

(S8)

This integral was calculated by following a previously established procedure (9, 26), and below, we only outline the main steps of this calculation.

To calculate the average over the associations, we first decoupled X^{μ} and y^{μ} by introducing two new sets of variables:

$$\frac{\lambda^{a,\mu}}{\sqrt{N}} = \frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j^a\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]-1$$
$$\nu^{a,\mu} = \frac{1}{N}\sum_{j=1}^{N}\left[\left(\tilde{J}_j^a\right)^2\left(p_j^-\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-p_j^+\right)\left(1-X_j^{\mu}\right)\right)+\beta_{syn}\tilde{J}_j^a g_j\left(\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right)+\beta_{int}^2\right]$$

(S9)

These variables were next incorporated into Eq. (S8) with the help of Dirac δ-functions:

$$\left\langle\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^n\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}} = \int\prod_{j,a=1}^{N,n}d\tilde{J}_j^a\prod_{j,a=1}^{N,n}\theta\left(\tilde{J}_j^a g_j\right)\prod_{a=1}^{n}\delta\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j^a g_j-\tilde{w}\right)\prod_{\mu,a=1}^{m,n}\frac{d\lambda^{a,\mu}}{\sqrt{N}}\prod_{\mu,a=1}^{m,n}d'\nu^{a,\mu}\times$$
$$\left\langle\prod_{\mu,a=1}^{m,n}\theta\left(\left(2y^{\mu}-1\right)\lambda^{a,\mu}-\sqrt{2}\left(\mathrm{erf}^{-1}\left(1-2p_{out}^-\right)y^{\mu}+\mathrm{erf}^{-1}\left(1-2p_{out}^+\right)\left(1-y^{\mu}\right)\right)\left(\nu^{a,\mu}\right)^{\frac{1}{2}}\right)\right\rangle_{\left\{y^{\mu}\right\}}\times$$
$$\left\langle\prod_{\mu,a=1}^{m,n}\delta\left(1+\frac{\lambda^{a,\mu}}{\sqrt{N}}-\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j^a\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]\right)\times\right.$$
$$\left.\prod_{\mu,a=1}^{m,n}\delta\left(\nu^{a,\mu}-\frac{1}{N}\sum_{j=1}^{N}\left[\left(\tilde{J}_j^a\right)^2\left(p_j^-\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-p_j^+\right)\left(1-X_j^{\mu}\right)\right)+\beta_{syn}\tilde{J}_j^a g_j\left(\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right)+\beta_{int}^2\right]\right)\right\rangle_{\left\{X^{\mu}\right\}}$$

(S10)

Symbol d' in this expression and thereafter is designated for 0 to ∞ integration, whereas d is used for integration from $-\infty$ to ∞. Next, the Heaviside step-functions and the δ-functions were replaced with their Fourier representations, making it possible to perform the averaging over the associations:

$$\left\langle\Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^n\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}} = \int\prod_{j,a=1}^{N,n}d\tilde{J}_j^a\prod_{\mu,a=1}^{m,n}\frac{d\lambda^{a,\mu}d\hat{\lambda}^{a,\mu}}{2\pi}\prod_{\mu,a=1}^{m,n}\frac{d'u^{a,\mu}d\hat{u}^{a,\mu}}{2\pi}\prod_{\mu,a=1}^{m,n}\frac{d'\nu^{a,\mu}d\hat{\nu}^{a,\mu}}{2\pi}\prod_{a=1}^{n}\frac{d\hat{k}^a}{2\pi/N}\times$$
$$\prod_{j,a=1}^{N,n}\theta\left(\tilde{J}_j^a g_j\right)\prod_{a=1}^{n}e^{i\hat{k}^a\left(\sum_{j=1}^{N}\tilde{J}_j^a g_j-N\tilde{w}\right)}\prod_{\mu,a=1}^{m,n}e^{-i\hat{u}^{a,\mu}u^{a,\mu}}\prod_{\mu,a=1}^{m,n}\left(f_{out}e^{i\hat{u}^{a,\mu}\left(\lambda^{a,\mu}-C_{out}^-\left(\nu^{a,\mu}\right)^{\frac{1}{2}}\right)}+(1-f_{out})e^{-i\hat{u}^{a,\mu}\left(\lambda^{a,\mu}+C_{out}^+\left(\nu^{a,\mu}\right)^{\frac{1}{2}}\right)}\right)\times$$
$$\prod_{\mu,a=1}^{m,n}e^{i\hat{\lambda}^{a,\mu}\left(\sqrt{N}+\lambda^{a,\mu}\right)+i\hat{\nu}^{a,\mu}\nu^{a,\mu}-i\hat{\nu}^{a,\mu}\beta_{int}^2}\prod_{\mu,j=1}^{m,N}e^{-\frac{iA_j}{\sqrt{N}}\sum_{a=1}^{n}\hat{\lambda}^{a,\mu}\tilde{J}_j^a-\frac{i\beta_{syn}g_jA_j}{N}\sum_{a=1}^{n}\hat{\nu}^{a,\mu}\tilde{J}_j^a-\frac{iB_j}{N}\sum_{a=1}^{n}\hat{\nu}^{a,\mu}\left(\tilde{J}_j^a\right)^2-\frac{C_j}{2N}\left(\sum_{a=1}^{n}\hat{\lambda}^{a,\mu}\tilde{J}_j^a\right)^2}$$

(S11)

The following notation was used in the above expression:

$$
\begin{aligned}
A_j &= \left(1-p_j^-\right) f_j + p_j^+\left(1-f_j\right) \\
B_j &= f_j p_j^-\left(1-p_j^-\right)+\left(1-f_j\right) p_j^+\left(1-p_j^+\right) \\
C_j &= f_j\left(1-f_j\right)\left(1-p_j^+-p_j^-\right)^2 \\
D_{out}^{\pm} &= \sqrt{2}\,\mathrm{erf}^{-1}\left(1-2p_{out}^{\pm}\right)
\end{aligned}
\tag{S12}
$$

We note that parameters A_j, B_j, and C_j are nonnegative.

Next, we decoupled the products containing indices j and μ by introducing three sets of order parameters and embedding them into Eq. (S11) with the help of Dirac δ-functions as was done in Eqs. (S10, S11):

$$
\frac{1}{N}\sum_{j=1}^{N} A_j \tilde{J}_j^a = 1+\frac{s^a}{\sqrt{N}}, \quad \frac{1}{N}\sum_{j=1}^{N}\left(\beta_{syn} A_j g_j \tilde{J}_j^a + B_j\left(\tilde{J}_j^a\right)^2+\beta_{int}^2\right)=\kappa^a \geq 0, \quad \frac{1}{N}\sum_{j=1}^{N} C_j \tilde{J}_j^a \tilde{J}_j^b = q^{a,b}
$$

$$
\begin{aligned}
&\left\langle \Omega\left(\left\{X^\mu, y^\mu\right\}\right)^n\right\rangle_{\left\{X^\mu, y^\mu\right\}} = \int \prod_{a=1}^{n}\frac{d\hat{k}^a}{2\pi/N}\prod_{a=1}^{n}\frac{ds^a d\hat{s}^a}{2\pi/\sqrt{N}}\prod_{a=1}^{n}\frac{d\kappa^a d\hat{\kappa}^a}{2\pi/N}\prod_{a,b=1}^{n}\frac{dq^{a,b}d\hat{q}^{a,b}}{2\pi/N}\times \\
&\prod_{a=1}^{n} e^{i\hat{s}^a\left(N+\sqrt{N}s^a\right)}\prod_{a=1}^{n} e^{iN\kappa^a\hat{\kappa}^a}\prod_{a,b=1}^{n,n} e^{iN\hat{q}^{a,b}q^{a,b}}\prod_{a=1}^{n} e^{-iN\hat{k}^a\tilde{w}}\times \\
&\left[\prod_{a=1}^{n}\frac{d\lambda^a d\hat{\lambda}^a}{2\pi}\prod_{a=1}^{n}\frac{d'u^a d\hat{u}^a}{2\pi}\prod_{a=1}^{n}\frac{d'\nu^a d\hat{\nu}^a}{2\pi}\left(f_{out}e^{i\sum_{a=1}^{n}\hat{u}^a\left(\lambda^a-D_{out}^-\sqrt{\nu^a}\right)}+(1-f_{out})e^{-i\sum_{a=1}^{n}\hat{u}^a\left(\lambda^a+D_{out}^+\sqrt{\nu^a}\right)}\right)\times\right. \\
&\left. e^{i\sum_{a=1}^{n}\left(\hat{\lambda}^a\left(\lambda^a-s^a\right)+\hat{\nu}^a\left(\nu^a-\kappa^a\right)-\hat{u}^a u^a\right)-\frac{1}{2}\sum_{a,b=1}^{n,n}\hat{\lambda}^a\hat{\lambda}^b q^{a,b}}\right]^m\times \\
&\prod_{j,a=1}^{N,n} d\tilde{J}_j^a \prod_{j,a=1}^{N,n}\theta\left(\tilde{J}_j^a g_j\right)\prod_{a=1}^{n} e^{i\hat{k}^a\sum_{j=1}^{N}\tilde{J}_j^a g_j - i\hat{s}^a\sum_{j=1}^{N}A_j\tilde{J}_j^a - i\hat{\kappa}^a\sum_{j=1}^{N}\left(\beta_{syn}A_j g_j\tilde{J}_j^a+B_j\left(\tilde{J}_j^a\right)^2+\beta_{int}^2\right)}\prod_{a,b=1}^{n,n} e^{-i\hat{q}^{a,b}\sum_{j=1}^{N}C_j\tilde{J}_j^a\tilde{J}_j^b}
\end{aligned}
$$

(S13)

After integrating over ν^a, $\hat{\nu}^a$, λ^a, and $\hat{\lambda}^a$ we obtained:

$$\left\langle \Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^{n}\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}} = \int\prod_{a=1}^{n}\frac{d\hat{k}^{a}}{2\pi/N}\prod_{a=1}^{n}\frac{ds^{a}d\hat{s}^{a}}{2\pi/\sqrt{N}}\prod_{a=1}^{n}\frac{d'\kappa^{a}d\hat{\kappa}^{a}}{2\pi/N}\prod_{a,b=1}^{n}\frac{dq^{a,b}d\hat{q}^{a,b}}{2\pi/N}\times$$

$$e^{N\left(i\sum_{a=1}^{n}\left(\hat{s}^{a}-\hat{k}^{a}\tilde{w}\right)+i\sum_{a=1}^{n}\kappa^{a}\hat{\kappa}^{a}+i\sum_{a,b=1}^{n,n}q^{a,b}\hat{q}^{a,b}+n\alpha\tilde{G}_{E}\left(\left\{s^{a}\right\},\left\{q^{a,b}\right\}\right)+n\tilde{G}_{S}\left(\left\{\hat{k}^{a}\right\},\left\{\hat{s}^{a}\right\},\left\{\hat{q}^{a,b}\right\}\right)\right)}$$

$$\tilde{G}_{E}\left(\left\{s^{a}\right\},\left\{\kappa^{a}\right\},\left\{q^{a,b}\right\}\right)=\frac{1}{n}\ln\left(\int\prod_{a=1}^{n}\frac{d'u^{a}d\hat{u}^{a}}{2\pi}e^{-\frac{1}{2}\sum_{a,b=1}^{n}\hat{u}^{a}\hat{u}^{b}q^{a,b}}\left(f_{out}e^{i\sum_{a=1}^{n}\hat{u}^{a}\left(s^{a}-u^{a}-D_{out}^{-}\sqrt{\kappa^{a}}\right)}+(1-f_{out})e^{-i\sum_{a=1}^{n}\hat{u}^{a}\left(s^{a}+u^{a}+D_{out}^{+}\sqrt{\kappa^{a}}\right)}\right)\right)$$

$$\tilde{G}_{S}\left(\left\{\hat{k}^{a}\right\},\left\{\hat{s}^{a}\right\},\left\{\hat{\kappa}^{a}\right\},\left\{\hat{q}^{a,b}\right\}\right)=\frac{1}{n}\frac{1}{N}\sum_{j=1}^{N}\ln\left(\int\prod_{a=1}^{n}d'\tilde{J}^{a}e^{i\sum_{a=1}^{n}\left(\hat{k}^{a}g_{j}-\hat{s}^{a}A_{j}\right)g_{j}\tilde{J}^{a}-i\sum_{a=1}^{n}\hat{\kappa}^{a}\left(\beta_{syn}A_{j}\tilde{J}^{a}+B_{j}\left(\tilde{J}^{a}\right)^{2}+\beta_{int}^{2}\right)-iC_{j}\sum_{a,b=1}^{n}\tilde{J}^{a}\tilde{J}^{b}\hat{q}^{a,b}}\right)$$

(S14)

The above integral was calculated by using the steepest descent method (47) under the assumption of a replica symmetric saddle point, $s^{a}=s$, $\kappa^{a}=\kappa\geq 0$, $q^{a,a}=q_{0}$, $q^{a\neq b}=q$, $\hat{k}^{a}=\hat{k}$, $\hat{s}^{a}=\hat{s}$, $\hat{\kappa}^{a}=\hat{\kappa}$, $\hat{q}^{a,a}=\hat{q}_{0}$, and $\hat{q}^{a\neq b}=\hat{q}$:

$$\left\langle \Omega\left(\left\{X^{\mu},y^{\mu}\right\}\right)^{n}\right\rangle_{\left\{X^{\mu},y^{\mu}\right\}} \sim e^{Nn\left(i\left(\hat{s}-\hat{k}\tilde{w}\right)+i\kappa\hat{\kappa}+iq_{0}\hat{q}_{0}-iq\hat{q}+\alpha G_{E}\left(s,\kappa,q_{0},q\right)+G_{S}\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_{0},\hat{q}\right)\right)}$$

$$G_{E}\left(s,\kappa,q_{0},q\right)=\frac{1}{n}\ln\left(\int\prod_{a=1}^{n}\frac{d'u^{a}d\hat{u}^{a}}{2\pi}e^{-\frac{1}{2}q_{0}\sum_{a=1}^{n}\left(\hat{u}^{a}\right)^{2}-\frac{1}{2}q\sum_{a\neq b=1}^{n}\hat{u}^{a}\hat{u}^{b}}\left(f_{out}e^{i\sum_{a=1}^{n}\hat{u}^{a}\left(s-u^{a}-D_{out}^{-}\sqrt{\kappa}\right)}+(1-f_{out})e^{-i\sum_{a=1}^{n}\hat{u}^{a}\left(s+u^{a}+D_{out}^{+}\sqrt{\kappa}\right)}\right)\right)$$

$$G_{S}\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_{0},\hat{q}\right)=\frac{1}{n}\frac{1}{N}\sum_{j=1}^{N}\ln\left(\int\prod_{a=1}^{n}d'\tilde{J}^{a}e^{i\left(\hat{k}g_{j}-\hat{s}A_{j}\right)g_{j}\sum_{a=1}^{n}\tilde{J}^{a}-i\hat{\kappa}\sum_{a=1}^{n}\left(\beta_{syn}A_{j}\tilde{J}^{a}+B_{j}\left(\tilde{J}^{a}\right)^{2}+\beta_{int}^{2}\right)-iC_{j}\hat{q}_{0}\sum_{a=1}^{n}\left(\tilde{J}^{a}\right)^{2}-iC_{j}\hat{q}\sum_{a\neq b=1}^{n}\tilde{J}^{a}\tilde{J}^{b}}\right)$$

(S15)

The non-redundant, replica symmetric saddle point coordinates $\left(s,\kappa,q_{0},q,\hat{k},\hat{s},\hat{\kappa},\hat{q}_{0},\hat{q}\right)$ satisfy the following system of nine equations and one inequality:

$$\frac{\partial G_E(s,\kappa,q_0,q)}{\partial s}=0;\quad \frac{\partial G_E(s,\kappa,q_0,q)}{\partial \kappa}=-\frac{i\hat{\kappa}}{\alpha};\quad \frac{\partial G_E(s,\kappa,q_0,q)}{\partial q_0}=-\frac{i\hat{q}_0}{\alpha};\quad \frac{\partial G_E(s,\kappa,q_0,q)}{\partial q}=\frac{i\hat{q}}{\alpha};$$

$$\frac{\partial G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)}{\partial \hat{k}}=i\tilde{w};\quad \frac{\partial G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)}{\partial \hat{s}}=-i;\quad \frac{\partial G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)}{\partial \hat{\kappa}}=-i\kappa;$$

$$\frac{\partial G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)}{\partial \hat{q}_0}=-iq_0;\quad \frac{\partial G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)}{\partial \hat{q}}=iq;\ \ \kappa\geq 0$$

(S16)

To simplify the expressions for G_E and G_S we employed the Hubbard-Stratonovich transformation [see e.g. (27)] and took the $n\to 0$ limit:

$$G_E(s,\kappa,q_0,q)=\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\left[f_{out}\ln\left(\int\frac{d'ud\hat{u}}{2\pi}e^{-\frac{1}{2}q_0\hat{u}^2+2x\sqrt{-\frac{1}{2}}q\hat{u}+\frac{1}{2}q\hat{u}^2+i\hat{u}\left(s-u-D_{out}^{-}\sqrt{\kappa}\right)}\right)+\right.$$

$$\left.(1-f_{out})\ln\left(\int\frac{d'ud\hat{u}}{2\pi}e^{-\frac{1}{2}q_0\hat{u}^2+2x\sqrt{-\frac{1}{2}}q\hat{u}+\frac{1}{2}q\hat{u}^2-i\hat{u}\left(s+u+D_{out}^{+}\sqrt{\kappa}\right)}\right)\right] \tag{S17}$$

$$G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)=\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\ln\left(e^{-i\hat{\kappa}\beta_{int}^2}\int d'\tilde{J}e^{\left(i\left(\hat{k}g_j-\hat{s}A_j\right)g_j+2x\sqrt{-iC_j\hat{q}}-i\hat{\kappa}\beta_{syn}A_j\right)\tilde{J}-i\left(C_j\left(\hat{q}_0-\hat{q}\right)+B_j\hat{\kappa}\right)\tilde{J}^2}\right)$$

Integrals in the arguments of the natural logarithm functions were expressed in terms of the complementary error functions:

$$G_E(s,\kappa,q_0,q)=\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\left(f_{out}\ln\left(\operatorname{erfc}\left(\frac{-s+D_{out}^{-}\sqrt{\kappa}-x\sqrt{2q}}{\sqrt{2(q_0-q)}}\right)\right)+(1-f_{out})\ln\left(\operatorname{erfc}\left(\frac{s+D_{out}^{+}\sqrt{\kappa}-x\sqrt{2q}}{\sqrt{2(q_0-q)}}\right)\right)\right)-\ln 2$$

$$G_S\left(\hat{k},\hat{s},\hat{\kappa},\hat{q}_0,\hat{q}\right)=\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\ln\left(e^{-i\hat{\kappa}\beta_{int}^2}\frac{\sqrt{\pi}e^{\frac{\left(i\hat{k}-i\hat{s}g_jA_j-i\hat{\kappa}\beta_{syn}A_j+2x\sqrt{-iC_j\hat{q}}\right)^2}{4i\left(C_j\left(\hat{q}_0-\hat{q}\right)+B_j\hat{\kappa}\right)}}}{2\sqrt{i\left(C_j\left(\hat{q}_0-\hat{q}\right)+B_j\hat{\kappa}\right)}}\operatorname{erfc}\left(-\frac{i\hat{k}-i\hat{s}g_jA_j-i\hat{\kappa}\beta_{syn}A_j+2x\sqrt{-iC_j\hat{q}}}{2\sqrt{i\left(C_j\left(\hat{q}_0-\hat{q}\right)+B_j\hat{\kappa}\right)}}\right)\right)$$

(S18)

The following substitutions transform the replica symmetric saddle point coordinates into the real domain, $u_{\pm}=\frac{D_{out}^{\pm}\sqrt{\kappa}\pm s}{\sqrt{2q}}$, $\varepsilon=\frac{q_0-q}{q}\geq 0$, $t=-i\hat{q}$, $z=\frac{i\hat{s}}{2\sqrt{-i\hat{q}}}$, $\delta=-\frac{\hat{q}_0-\hat{q}}{\hat{q}}$, $\tau=\frac{\hat{\kappa}}{-\hat{q}}$, $\eta=-\frac{i\hat{k}}{\sqrt{-i\hat{q}}}$, where these variables can be determined by solving the following problem:

$$G_E\left(u_+,u_-,\varepsilon\right)=\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\left(f_{out}\ln\left(\operatorname{erfc}\left(\frac{u_- - x}{\sqrt{\varepsilon}}\right)\right)+(1-f_{out})\ln\left(\operatorname{erfc}\left(\frac{u_+ - x}{\sqrt{\varepsilon}}\right)\right)\right)-\ln 2$$

$$G_S\left(\eta,t,\tau,z,\delta\right)=\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}\frac{e^{-x^2}}{\sqrt{\pi}}dx\ln\left(e^{-\tau t\beta_{int}^2}\frac{\sqrt{\pi}e^{\left(\frac{\eta+2zg_jA_j+\tau\sqrt{t}\beta_{syn}A_j-2x\sqrt{C_j}}{2\sqrt{C_j\delta+B_j\tau}}\right)^2}}{2\sqrt{C_j\delta t+B_j\tau t}}\operatorname{erfc}\left(\frac{\eta+2zg_jA_j+\tau\sqrt{t}\beta_{syn}A_j-2x\sqrt{C_j}}{2\sqrt{C_j\delta+B_j\tau}}\right)\right)$$

$$\frac{\partial G_E\left(u_+,u_-,\varepsilon\right)}{\partial u_+}-\frac{\partial G_E\left(u_+,u_-,\varepsilon\right)}{\partial u_-}=0;\quad \frac{\partial G_E\left(u_+,u_-,\varepsilon\right)}{\partial\varepsilon}=\frac{t}{\alpha}\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^2\kappa}{2\left(u_-+u_+\right)^2};\quad \frac{\partial G_E\left(u_+,u_-,\varepsilon\right)}{\partial u_+}=-\frac{2t\tau\kappa}{\alpha}\frac{1}{\left(u_-+u_+\right)};$$

$$\frac{\partial G_E\left(u_+,u_-,\varepsilon\right)}{\partial u_+}=\frac{t}{\alpha}\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^2\kappa}{\left(u_-+u_+\right)^3}\left(\delta-\varepsilon\right);\quad \frac{\partial G_S\left(\eta,t,\tau,z,\delta\right)}{\partial\eta}=-\tilde{w}\sqrt{t};\quad \frac{\partial G_S\left(\eta,t,\tau,z,\delta\right)}{\partial z}=-2\sqrt{t};$$

$$\frac{\partial G_S\left(\eta,t,\tau,z,\delta\right)}{\partial\tau}=-t\kappa;\quad \frac{\partial G_S\left(\eta,t,\tau,z,\delta\right)}{\partial\delta}=-t\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^2\kappa}{2\left(u_-+u_+\right)^2}$$

$$\frac{\partial G_S\left(\eta,t,\tau,z,\delta\right)}{\partial t}=\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^2\kappa\left(\varepsilon-\delta\right)}{2\left(u_-+u_+\right)^2}-\frac{\tilde{w}\eta}{2\sqrt{t}}-\frac{z}{\sqrt{t}}-\kappa\tau;\quad \kappa\geq 0;\quad \varepsilon\geq 0;\quad u_++u_-\geq 0$$

(S19)

E. Replica theory solution at critical capacity

With an increasing number of associations *m*, $\Omega_{typical}$ shrinks and approaches zero at the maximum (critical) capacity of the neuron, $\alpha_c=\frac{m_c}{N}$. In this limit, $\left(q_0-q\right)$ goes to zero and Eqs. (S19) can be expanded asymptotically in terms of $1/\varepsilon$ and $1/\delta$. In the leading order these expansions yield:

$$
\begin{aligned}
&(1-f_{out})F\left(u_{+}\right)-f_{out}F\left(u_{-}\right)=0\\
&f_{out}D\left(u_{-}\right)+(1-f_{out})D\left(u_{+}\right)=\frac{\varepsilon^{2}t\kappa}{\alpha_{c}}\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}}{\left(u_{-}+u_{+}\right)^{2}}\\
&(1-f_{out})F\left(u_{+}\right)=\frac{2t\tau\varepsilon\kappa}{\alpha_{c}}\frac{1}{\left(u_{-}+u_{+}\right)}\\
&(1-f_{out})F\left(u_{+}\right)=-\frac{\varepsilon t\kappa\left(\delta-\varepsilon\right)}{\alpha_{c}}\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}}{\left(u_{-}+u_{+}\right)^{3}}\\
&\frac{1}{N}\sum_{j=1}^{N}\frac{\sqrt{C_{j}}}{C_{j}\delta+B_{j}\tau}F\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)=2\tilde{w}\sqrt{t}\\
&\frac{1}{N}\sum_{j=1}^{N}\frac{A_{j}g_{j}\sqrt{C_{j}}}{C_{j}\delta+B_{j}\tau}F\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)=2\sqrt{t}\\
&\frac{1}{2N}\sum_{j=1}^{N}\frac{B_{j}C_{j}}{\left(C_{j}\delta+B_{j}\tau\right)^{2}}D\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)+t\left(\beta_{int}^{2}-\kappa\right)=-\frac{1}{2N}\sum_{j=1}^{N}\frac{\sqrt{C_{j}}\sqrt{t}\beta_{syn}A_{j}}{C_{j}\delta+B_{j}\tau}F\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)\\
&\frac{1}{N}\sum_{j=1}^{N}\frac{C_{j}^{2}}{\left(C_{j}\delta+B_{j}\tau\right)^{2}}D\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)=\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}t\kappa}{\left(u_{-}+u_{+}\right)^{2}}\\
&\frac{1}{4N}\sum_{j=1}^{N}\frac{\tau\beta_{syn}A_{j}\sqrt{C_{j}}}{\sqrt{t}\left(C_{j}\delta+B_{j}\tau\right)}F\left(-\frac{\eta+2zg_{j}A_{j}+\tau\sqrt{t}\beta_{syn}A_{j}}{2\sqrt{C_{j}}}\right)+\tau\left(\beta_{int}^{2}-\kappa\right)=-\frac{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}\kappa\left(\varepsilon-\delta\right)}{2\left(u_{-}+u_{+}\right)^{2}}+\frac{\tilde{w}\eta+2z}{2\sqrt{t}}\\
&\kappa\geq 0;\quad \varepsilon\geq 0;\quad u_{+}+u_{-}\geq 0
\end{aligned}
$$

(S20)

Functions *E*, *F*, and *D*, in Eqs. (S20) are defined as follows:

$$
\begin{aligned}
E\left(x\right)&=\frac{1}{2}\left(1+\operatorname{erf}\left(x\right)\right)\\
F\left(x\right)&=\frac{1}{\sqrt{\pi}}e^{-x^{2}}+x\left(1+\operatorname{erf}\left(x\right)\right)\\
D\left(x\right)&=xF\left(x\right)+E\left(x\right)
\end{aligned}
\tag{S21}
$$

After replacing $\tau\sqrt{t}$ with y, δ/τ with *x*, and eliminating variables, ε, t, κ, τ, and δ, we arrived at the final system of six equations and one inequality. This system contains six latent variables $u_{\pm}$, x, η, y, and z which determine the critical capacity of the neuron, α_{c}:

$$
\begin{cases}
(1-f_{out})F\left(u_{+}\right)-f_{out}F\left(u_{-}\right)=0 \\
x=\dfrac{4\left(u_{-}+u_{+}\right)}{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}}\dfrac{f_{out}E\left(u_{-}\right)+(1-f_{out})E\left(u_{+}\right)}{f_{out}F\left(u_{-}\right)+(1-f_{out})F\left(u_{+}\right)} \\
\dfrac{1}{N}\sum_{j=1}^{N}\dfrac{\sqrt{C_j}}{C_j x+B_j}F\left(-\dfrac{\eta+2zA_j g_j+y\beta_{syn}A_j}{2\sqrt{C_j}}\right)=2\tilde{w}y \\
\dfrac{1}{N}\sum_{j=1}^{N}\dfrac{A_j g_j\sqrt{C_j}}{C_j x+B_j}F\left(-\dfrac{\eta+2zA_j g_j+y\beta_{syn}A_j}{2\sqrt{C_j}}\right)=2y \\
\dfrac{1}{N}\sum_{j=1}^{N}\dfrac{C_j}{\left(C_j x+B_j\right)^{2}}\left(\dfrac{B_j}{2}-\dfrac{\left(u_{-}+u_{+}\right)^{2}}{\left(D_{out}^{+}+D_{out}^{-}\right)^{2}}C_j\right)D\left(-\dfrac{\eta+2zA_j g_j+y\beta_{syn}A_j}{2\sqrt{C_j}}\right)=\beta_{int}^{2}y^{2}-y\tilde{w}\eta-2yz \\
\dfrac{1}{N}\sum_{j=1}^{N}\dfrac{\beta_{syn}A_j\sqrt{C_j}}{C_j x+B_j}F\left(-\dfrac{\eta+2zA_j g_j+y\beta_{syn}A_j}{2\sqrt{C_j}}\right)=2\tilde{w}\eta+4z-4\beta_{int}^{2}y \\
u_{+}+u_{-}\geq 0
\end{cases}
$$

$$
\alpha_c=x^{2}\frac{f_{out}D\left(u_{-}\right)+(1-f_{out})D\left(u_{+}\right)}{\left(f_{out}E\left(u_{-}\right)+(1-f_{out})E\left(u_{+}\right)\right)^{2}}\frac{1}{N}\sum_{j=1}^{N}\frac{C_j^{2}}{\left(C_j x+B_j\right)^{2}}D\left(-\frac{\eta+2zA_j g_j+y\beta_{syn}A_j}{2\sqrt{C_j}}\right)
$$

(S22)

We note that Eqs. (S22) are in agreement with the solution described in (9), where a simplified version of the model presented here was solved by minimizing the probability of output spiking errors for a given intrinsic noise strength. Eqs. (S22) account for additional features such as the homeostatic constraint, learning by inhibitory inputs, heterogeneity of inputs, synaptic noise, and presynaptic spiking errors.

F. Distribution of input weights at critical capacity

Connection probabilities, P^{con}, and probability densities for non-zero input weights, p^{PSP}, at critical capacity were calculated as previously described (9, 26). The result depends on the set of latent variables of Eqs. (S22):

$$
\begin{aligned}
P_i^{con} &= E\left(-\frac{\eta + 2zA_j g_j + y\beta_{syn} A_j}{2\sqrt{C_j}}\right) \\
p_i^{PSP}\left(\tilde{J}\right) &= \frac{\theta\left(g_i \tilde{J}\right)}{\sqrt{2\pi}\sigma_i \tilde{w} E\left(-\frac{\eta + 2zA_j g_j + y\beta_{syn} A_j}{2\sqrt{C_j}}\right)} e^{-\left(\frac{\tilde{J}}{\sqrt{2}\sigma_i \tilde{w}} + \frac{\eta + 2zA_i g_i + y\beta_{syn} A_j}{2\sqrt{C_i}} g_i\right)^2} \\
\sigma_i &= \frac{\sqrt{C_i}}{\sqrt{2}\tilde{w} y\left(C_i x + B_i\right)}
\end{aligned}
\tag{S23}
$$

A given input, *i*, has a non-infinitesimal probability of having a connection weight of zero, while its probability density for having non-zero connection weights is a truncated Gaussian with a standard deviation $\sigma_i \tilde{w}$.

G. The solution in the case of balanced spiking errors

Without loss of generality, we assumed that the expected rates of erroneous spikes and spike failures are equal, and, as a result, these spiking errors do not affect the input and output firing probabilities $\left\{f_j\right\}$ and f_{out}:

$$
\begin{aligned}
p_j^+\left(1 - f_j\right) &= p_j^- f_j, \quad j = 1, \ldots, N \\
p_{out}^+\left(1 - f_{out}\right) &= p_{out}^- f_{out}
\end{aligned}
\tag{S24}
$$

In this case, it is more convenient to express the results of the model in terms of spiking error probabilities:

$$
\begin{aligned}
r_j &= p_j^+\left(1 - f_j\right) + p_j^- f_j \\
r_{out} &= p_{out}^+\left(1 - f_{out}\right) + p_{out}^- f_{out}
\end{aligned}
\tag{S25}
$$

This change does not alter Eqs. (S22, S23), but the coefficients defined in Eqs. (S12) transform into:

$$
\begin{aligned}
A_j &= f_j \\
B_j &= r_j\left(1-\frac{r_j}{4f_j\left(1-f_j\right)}\right) \\
C_j &= f_j\left(1-f_j\right)\left(1-\frac{r_j}{2f_j\left(1-f_j\right)}\right)^2 \\
D_{out}^{+} &= \sqrt{2}\,\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{\left(1-f_{out}\right)}\right) \\
D_{out}^{-} &= \sqrt{2}\,\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{f_{out}}\right)
\end{aligned}
\tag{S26}
$$

Eqs. (S22, S23) were solved in MATLAB to produce the results for heterogeneous networks consisting of inhibitory and excitatory neurons with distributed spiking error probabilities (Figures 7A-C of the main text) and distributed intrinsic and synaptic noise strengths (Figures 7D-F of the main text). In both cases, the remaining model parameters were the same for all neurons. We note that in both cases the solutions of Eqs. (S22, S23) depend on β_{int} and β_{syn} only in a combination $\beta = \sqrt{\beta_{int}^2 + \tilde{w}\beta_{syn} f}$. This parameter is referred to as the postsynaptic noise strength.

The code is available at https://github.com/neurogeometry/Associative_Learning_with_Noise.

H. The solution in the case of two homogeneous classes of inputs and balanced spiking errors

In this case, all neuron inputs have the same firing probability, *f*, and the same spiking error probability, *r*. Eqs. (S22, S23) simplify significantly after the introduction of two new variables, $v_{\pm} = \dfrac{-\eta \pm 2zf - y\beta_{syn} f}{2\sqrt{C}}$:

$$
\begin{cases}
(1-f_{out})F(u_+)-f_{out}F(u_-)=0 \\
\frac{N_{inh}}{N}F(v_+)+\frac{N_{exc}}{N}F(v_-)=\frac{\sqrt{2}}{\sigma} \\
-\frac{N_{inh}}{N}F(v_+)+\frac{N_{exc}}{N}F(v_-)=\frac{\sqrt{2}}{\sigma\tilde{w}f} \\
\left((u_-+u_+)^2-2\xi\right)\left(\frac{N_{inh}}{N}D(v_+)+\frac{N_{exc}}{N}D(v_-)\right)=\frac{2\beta^2\zeta^2}{\sigma^2} \\
\sigma=\dfrac{\sqrt{2}\beta^2\zeta^2}{\left((u_-+u_+)\frac{f_{out}E(u_-)+(1-f_{out})E(u_+)}{f_{out}F(u_-)+(1-f_{out})F(u_+)}+\xi\right)\left(\frac{1}{\tilde{w}f}(v_+-v_-)-(v_++v_-)\right)} \\
u_++u_->0
\end{cases}
$$

$$
\alpha_c=\frac{\sigma^2(u_-+u_+)^2}{2\beta^4\zeta^4}\frac{f_{out}D(u_-)+(1-f_{out})D(u_+)}{\left(f_{out}F(u_-)+(1-f_{out})F(u_+)\right)^2}\left(\frac{1}{\tilde{w}f}(v_+-v_-)-(v_++v_-)\right)^2\left(\frac{N_{inh}}{N}D(v_+)+\frac{N_{exc}}{N}D(v_-)\right)
$$

$$
P_{inh/exc}^{con}=E(v_\pm)
$$

$$
p_{inh/exc}^{PSP}(\tilde{J})=\frac{\theta(\mp\tilde{J})}{\sqrt{2\pi}\sigma\tilde{w}E(v_\pm)}e^{-\left(\frac{\tilde{J}}{\sqrt{2}\sigma\tilde{w}}\pm v_\pm\right)^2}
$$

(S27)

The intrinsic and synaptic noises in Eqs. (S27) are entirely contained within the parameter β, while the spiking error probabilities r and r_{out} appear only in the parameters ξ and ζ:

$$
\begin{aligned}
\beta&=\sqrt{\beta_{int}^2+\tilde{w}\beta_{syn}f} \\
\xi&=\frac{r(4f(1-f)-r)}{2(2f(1-f)-r)^2}\left(\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{1-f_{out}}\right)+\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{f_{out}}\right)\right)^2 \\
\zeta&=\frac{\sqrt{2f(1-f)}}{\tilde{w}(2f(1-f)-r)}\left(\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{1-f_{out}}\right)+\mathrm{erf}^{-1}\left(1-\frac{r_{out}}{f_{out}}\right)\right)
\end{aligned}
\tag{S28}
$$

We note that in the absence of spiking errors in the input (r = 0), Eqs. (S27) are similar in structure to the solution of a simplified model considered by us previously (18). That model did not explicitly consider different sources of errors and noise, but instead, used a generic parameter ρ, referred to as the rescaled robustness, to ensure that memories are recalled reliably in the presence of intrinsic noise only. Solutions to both models become identical when $\rho=\beta\zeta$.

For associative networks considered in the main text, we set $r_{out} = r$ and $f_{out} = f$ in Eqs. (S27, S28) as these parameters in the homogeneous case must be the same for all neurons in the network. Figures 1, 4, and 6 of the main text show the results of the homogeneous model as functions of β and r.

I. Numerical solution of the model with nonlinear optimization

For a finite number of inputs, the solution to the problem outlined in Eqs. (S5) was obtained numerically. To that end, we made the problem feasible by introducing a slack variable $s^{\mu} \geq 0$ for every association and chose the solution that minimizes the sum of these variables:

$$
\begin{aligned}
&\underset{\{\tilde{J}_j\}}{\arg\min}\left(\sum_{\mu=1}^{m} s^{\mu}\right) \\
&\left(2y^{\mu}-1\right)\left(\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]-1\right) \geq -s^{\mu}+\frac{D_{out}^- y^{\mu}+D_{out}^+\left(1-y^{\mu}\right)}{\sqrt{N}}\times \\
&\left(\frac{1}{N}\sum_{j=1}^{N}\left[\tilde{J}_j^2\left(p_j^-\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-p_j^+\right)\left(1-X_j^{\mu}\right)\right)+\beta_{syn}\tilde{J}_j g_j\left(\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right)+\beta_{int}^2\right]\right)^{\frac{1}{2}}; \quad \mu=1,\ldots,m \\
&\frac{1}{N}\sum_{j=1}^{N}\tilde{J}_j g_j = \tilde{w} \\
&\tilde{J}_j g_j \geq 0; \quad j=1,\ldots,N \\
&s^{\mu} \geq 0; \quad \mu=1,\ldots,m
\end{aligned}
\tag{S29}
$$

Eqs. (S29) were solved by using the *fmincon* function of MATLAB and the results are shown in Figures 2, 3, 5, and 6 of the main text. The *fmincon* function utilizes the interior-point technique for finding solutions of constrained nonlinear optimization problems (48, 49). The code is available at https://github.com/neurogeometry/Associative_Learning_with_Noise.

J. Numerical solution of the model with a perceptron-type learning rule

In addition to the replica and nonlinear optimization solutions, a biologically more plausible online solution of Eqs. (S29) was devised by approximately stepping in the direction of the negative gradient of the sum of the slack variables. The latter is,

$$-\frac{\partial}{\partial \tilde{J}_j}\sum_{\mu=1}^{m} s^{\mu} = \sum_{\mu=1}^{m}\frac{1}{N}\left(2y^{\mu}-1\right)\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]-$$
$$\sum_{\mu=1}^{m}\frac{\left(D_{out}^- y^{\mu}+D_{out}^+\left(1-y^{\mu}\right)\right)\left(\tilde{J}_j\left(p_j^-\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-p_j^+\right)\left(1-X_j^{\mu}\right)\right)+\frac{1}{2}\beta_{syn}g_j\left(\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right)\right)}{N^{3/2}\left(\frac{1}{N}\sum_{k=1}^{N}\left[\tilde{J}_k^2\left(p_k^-\left(1-p_k^-\right)X_k^{\mu}+p_k^+\left(1-p_k^+\right)\left(1-X_k^{\mu}\right)\right)+\beta_{syn}\tilde{J}_k g_k\left(\left(1-p_k^-\right)X_k^{\mu}+p_k^+\left(1-X_k^{\mu}\right)\right)+\beta_{int}^2\right]\right)^{\frac{1}{2}}}$$

(S30)

The first approximation to this gradient was made by ignoring the second term in the r.h.s. of Eq. (S30). This was done because there is no clear way of calculating this term in an online, biologically plausible manner. The second approximation was made by noting that $\left[\left(1-p_j^-\right)X_j^{\mu}+p_j^+\left(1-X_j^{\mu}\right)\right]$ in the r.h.s. of Eq. (S30) is the average of $X_j^{*\mu}$ over the spiking errors, and therefore, a stochastic estimate of this gradient direction can be made in an online manner with a perceptron-type learning step $\left(2y^{\mu}-1\right)X_j^{*\mu}$ (45).

With these approximations, we proposed the following learning rule. At each learning step, e.g. μ, a neuron receives an input containing spiking errors, $X^{*\mu}$, combines it with its synaptic and intrinsic noise, and produces an output, $y^{*\mu}$. If this output differs from the neuron's target output, y^{μ}, its input connection weights are updated in four consecutive steps:

$$\begin{aligned}
&\tilde{J}_j \mapsto \tilde{J}_j + \gamma\left(2y^{\mu}-1\right)X_j^{*\mu}; \quad j=1,\ldots,N\\
&\tilde{J}_j \mapsto \tilde{J}_j\theta\left(\tilde{J}_j g_j\right)\\
&\tilde{J}_j \mapsto \tilde{J}_j + \left(\tilde{w}-\frac{1}{N}\sum_{j=1}^{N}\left|\tilde{J}_j\right|\right)g_j\\
&\tilde{J}_j \mapsto \tilde{J}_j\theta\left(\tilde{J}_j g_j\right)
\end{aligned} \tag{S31}$$

By including presynaptic spiking errors, synaptic and intrinsic noise in the condition that triggers the learning steps outlined in Eqs. (S31), the learning rule implicitly depends on the model

parameters $\{r_j\}$, $\{\beta_{syn,\,j}\}$, and β_{int} describing the neuron's intrinsic noise. By approximately minimizing the sum of the slack variables, the learning rule minimizes the neuron's output spiking error probability for a given memory load, and when the feasible solution of Eqs. (S29) is found (i.e. $s^{\mu}=0;\quad \mu=1,\ldots,m$), the neuron's output spiking error probability falls below the desired probability, r_{out} (Figure 6A). At capacity, the two probabilities must equal and, therefore, the learning rule also depends implicitly on r_{out}.

Unlike the standard perceptron learning rule, Eqs. (S31) are based on noisy input and enforce sign and homeostatic constraints during learning. The first update step in Eqs. (S31) is a stochastic perceptron learning step, in which parameter γ is referred to as the learning rate. The second step is introduced to enforce the sign constraints, while the last two steps combined implement the homeostatic l_1-norm constraint and are equivalent to the soft thresholding done in LASSO regression (46). A related rule, in the absence of errors, noise, and l_1-norm constraint, was previously described in (9, 50). In numerical simulations, we trained neurons on associations presented in the order of their appearance in the associative sequence, one at a time. This constitutes one learning epoch. We set γ = 0.1 and ran the algorithm until a solution was found or the maximum number of 10^6 epochs was reached. The results of this procedure are shown in Figure 6 of the main text.

K. Mutual information contained in retrieved associative sequences

The mutual information contained in one successfully retrieved association $\left(X^{\mu}\rightarrow X^{\mu+1}\right)$ can be calculated as a difference of marginal and conditional entropies,

$$I\left(X^{\mu};X^{\mu+1}\right)=H\left(X^{\mu+1}\right)-H\left(X^{\mu+1}\mid X^{\mu}\right) \tag{S32}$$

For homogeneous networks loaded with associations consisting of random and independent network states with balanced spiking errors the two entropies reduce to:

$$H\left(X^{\mu+1}\right) = -N\left[f\log_2 f + (1-f)\log_2(1-f)\right]$$

$$H\left(X^{\mu+1} \mid X^{\mu}\right) = -N\left[f\left(\frac{r}{2f}\log_2\frac{r}{2f} + \left(1-\frac{r}{2f}\right)\log_2\left(1-\frac{r}{2f}\right)\right) + \right. \tag{S33}$$

$$\left.(1-f)\left(\frac{r}{2(1-f)}\log_2\frac{r}{2(1-f)} + \left(1-\frac{r}{2(1-f)}\right)\log_2\left(1-\frac{r}{2(1-f)}\right)\right)\right]$$

As the length of a retrieved sequence may be shorter than the length of the loaded sequence, *m*, in the main text we considered two types of retrieved information. One type is defined as the expected retrieved information per memory playout in which contributions of partially retrieved sequences are set to zero. This information is based on completely retrieved sequences only and is equal to the product of the retrieval probability (Figure 2C) and *mI*. The other type of retrieved information is calculated based on completely and partially retrieved sequences and is equal to the product of the average retrieved sequence length and *I*. According to these definitions, the former is always less or equal to the latter.

REFERENCES

1. Faisal AA, Selen LP, & Wolpert DM (2008) Noise in the nervous system. *Nature reviews. Neuroscience* 9(4):292-303.
2. Trachtenberg JT, *et al.* (2002) Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* 420(6917):788-794.
3. Holtmaat A & Svoboda K (2009) Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature reviews. Neuroscience* 10(9):647-658.
4. Gala R, *et al.* (2017) Computer assisted detection of axonal bouton structural plasticity in in vivo time-lapse images. *eLife* 6.
5. Cohen MR & Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience* 12(12):1594-1600.
6. Mitchell JF, Sundberg KA, & Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63(6):879-888.
7. Minsky ML & Papert S (1988) *Perceptrons : an introduction to computational geometry* (MIT Press, Cambridge, Mass.) Expanded Ed pp xv, 292 p.
8. Hertz J, Palmer RG, & Krogh A (1991) *Introduction to the theory of neural computation* (Addison-Wesley Pub. Co., Redwood City, Calif.) pp xxii, 327 p.
9. Brunel N, Hakim V, Isope P, Nadal JP, & Barbour B (2004) Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron* 43(5):745-757.
10. Rubin R, Abbott LF, & Sompolinsky H (2017) Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proc Natl Acad Sci U S A* 114(44):E9366-E9375.
11. Bishop CM (1995) *Neural networks for pattern recognition* (Clarendon Press; Oxford University Press, Oxford, New York) pp xvii, 482 p.
12. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, & Salakhutdinov R (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res* 15:1929-1958.
13. Gammaitoni L, Hanggi P, Jung P, & Marchesoni F (1998) Stochastic resonance. *Rev Mod Phys* 70(1):223-287.

14. McDonnell MD & Abbott D (2009) What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *Plos Computational Biology* 5(5).
15. McDonnell MD & Ward LM (2011) The benefits of noise in neural systems: bridging theory and experiment. *Nature reviews. Neuroscience* 12(7):415-426.
16. Stein RB, Gossen ER, & Jones KE (2005) Neuronal variability: noise or part of the signal? *Nature reviews. Neuroscience* 6(5):389-397.
17. McCulloch W & Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol* 5(4):115 - 133.
18. Zhang D, Zhang C, & Stepanyants A (2019) Robust Associative Learning Is Sufficient to Explain the Structural and Dynamical Properties of Local Cortical Circuits. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 39(35):6888-6904.
19. Hebb DO (1949) *The organization of behavior; a neuropsychological theory* (Wiley, New York,) pp xix, 335 p.
20. Rosenblatt F (1957) The Perceptron--a perceiving and recognizing automaton. *Cornell Aeronautical Laboratory* Report 85-460-1.
21. Cover TM (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. EC* 14:326-334.
22. Gardner E (1988) The space of interactions in neural network models. *J. Phys. A: Math. Gen.* 21:257-270.
23. Gardner E & Derrida B (1988) Optimal storage properties of neural network models. *J. Phys. A: Math. Gen.* 21:271-284.
24. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 79(8):2554-2558.
25. Kohler HM & Widmaier D (1991) Sign-constrained linear learning and diluting in neural networks. *J. Phys. A: Math. Gen.* 24:L495-L502.
26. Chapeton J, Fares T, LaSota D, & Stepanyants A (2012) Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proc Natl Acad Sci U S A* 109(51):E3614-3622.
27. Chapeton J, Gala R, & Stepanyants A (2015) Effects of homeostatic constraints on associative memory storage and synaptic connectivity of cortical circuits. *Front Comput Neurosci* 9:74.

28. Brunel N (2016) Is cortical connectivity optimized for storing information? *Nature neuroscience* 19(5):749-755.
29. Stepanyants A & Chklovskii DB (2005) Neurogeometry and potential synaptic connectivity. *Trends Neurosci* 28(7):387-394.
30. Stepanyants A, *et al.* (2008) Local potential connectivity in cat primary visual cortex. *Cereb Cortex* 18(1):13-28.
31. Katz B (1969) *The release of neural transmitter substances* (Thomas, Springfield, Ill.,) pp ix, 60 p.
32. Edwards SF & Anderson PW (1975) Theory of spin glasses. *J. Phys. F: Metal Phys.* 5:965-974.
33. Sherrington D & Kirkpatrick S (1975) Solvable model of a spin glass. *Physical Review Letters* 35(26):1792-1796.
34. Shadlen MN & Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 18(10):3870-3896.
35. Stevens CF & Zador AM (1998) Input synchrony and the irregular firing of cortical neurons. *Nature neuroscience* 1(3):210-217.
36. Holt GR, Softky WR, Koch C, & Douglas RJ (1996) Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of neurophysiology* 75(5):1806-1814.
37. Softky WR & Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 13(1):334-350.
38. Buracas GT, Zador AM, DeWeese MR, & Albright TD (1998) Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* 20(5):959-969.
39. Cohen MR & Kohn A (2011) Measuring and interpreting neuronal correlations. *Nature neuroscience* 14(7):811-819.
40. van Vreeswijk C & Sompolinsky H (1998) Chaotic balanced state in a model of cortical circuits. *Neural computation* 10(6):1321-1371.

41. van Vreeswijk C & Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274(5293):1724-1726.

42. Graupner M & Reyes AD (2013) Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(38):15075-15085.

43. Okun M & Lampl I (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience* 11(5):535-537.

44. Mainen ZF & Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. *Science* 268(5216):1503-1506.

45. Rosenblatt F (1962) *Principles of neurodynamics; perceptrons and the theory of brain mechanisms* (Spartan Books, Washington,) p 616 p.

46. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1):267-288.

47. Arfken GB & Weber H-J (2005) *Mathematical methods for physicists* (Elsevier, Boston) 6th Ed pp xii, 1182 p.

48. Byrd RH, Gilbert JC, & Nocedal J (2000) A trust region method based on interior point techniques for nonlinear programming. *Mathematical programming* 89(1):149-185.

49. Byrd RH, Hribar ME, & Nocedal J (1999) An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization* 9(4):877-900.

50. Clopath C, Nadal J-P, & Brunel N (2012) Storage of correlated patterns in standard and bistable Purkinje cell models. *PLoS computational biology* 8(4):e1002448.