Effects of homeostatic constraints on associative memory storage and synaptic connectivity of cortical circuits

Julio Chapeton, Rohan Gala and Armen Stepanyants*

Department of Physics and Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, MA, USA

The impact of learning and long-term memory storage on synaptic connectivity is not completely understood. In this study, we examine the effects of associative learning on synaptic connectivity in adult cortical circuits by hypothesizing that these circuits function in a steady-state, in which the memory capacity of a circuit is maximal and learning must be accompanied by forgetting. Steady-state circuits should be characterized by unique connectivity features. To uncover such features we developed a biologically constrained, exactly solvable model of associative memory storage. The model is applicable to networks of multiple excitatory and inhibitory neuron classes and can account for homeostatic constraints on the number and the overall weight of functional connections received by each neuron. The results show that in spite of a large number of neuron classes, functional connections between potentially connected cells are realized with less than 50% probability if the presynaptic cell is excitatory and generally a much greater probability if it is inhibitory. We also find that constraining the overall weight of presynaptic connections leads to Gaussian connection weight distributions that are truncated at zero. In contrast, constraining the total number of functional presynaptic connections leads to non-Gaussian distributions, in which weak connections are absent. These theoretical predictions are compared with a large dataset of published experimental studies reporting amplitudes of unitary postsynaptic potentials and probabilities of connections between various classes of excitatory and inhibitory neurons in the cerebellum, neocortex, and hippocampus.

Keywords: perceptron, associative memory, l_0 norm, l_1 norm, inhibitory, critical capacity, synaptic weight, connection probability

Introduction

It has long been known that learning and long-term memory formation in the brain are accompanied with changes in the patterns and weights of synaptic connections (see Bailey and Kandel, 1993; Chklovskii et al., 2004; Holtmaat and Svoboda, 2009 for review). Yet, a detailed understanding of the effects of learning on synaptic connectivity is still hindered by an insufficient account of network activity patterns and cell-type specific, experience-dependent learning rules. Thus, it is currently not feasible to directly relate the learning experience of an animal to

specific changes in its synaptic connectivity. As an alternative, one may look for basic statistical features of synaptic connectivity which are catalyzed by the learning process, develop over time, and are present in adult circuits. In this study, we examine a biologically motivated, exactly solvable model of associative learning in an attempt to identify such connectivity features in local cortical circuits. Inspired by the ideas introduced by Gardner and Derrida (1988) and further developed by Brunel et al. (2004), we hypothesized that a given local circuit of the adult cortex is functioning in a steady-state. In this state the associative memory storage capacity of the circuit is maximal (critical) (Cover, 1965; Hopfield, 1982; Gardner, 1988; Gardner and Derrida, 1988), and learning new associations is accompanied with forgetting some of the old ones (**Figure 1**).

The steady-state learning hypothesis is supported by computational studies conducted in the cerebellar (Brunel et al., 2004; Barbour et al., 2007) and cerebral (Chapeton et al., 2012) cortices. It is also consistent with recent experimental evidence from human subjects, showing that new learning and memory
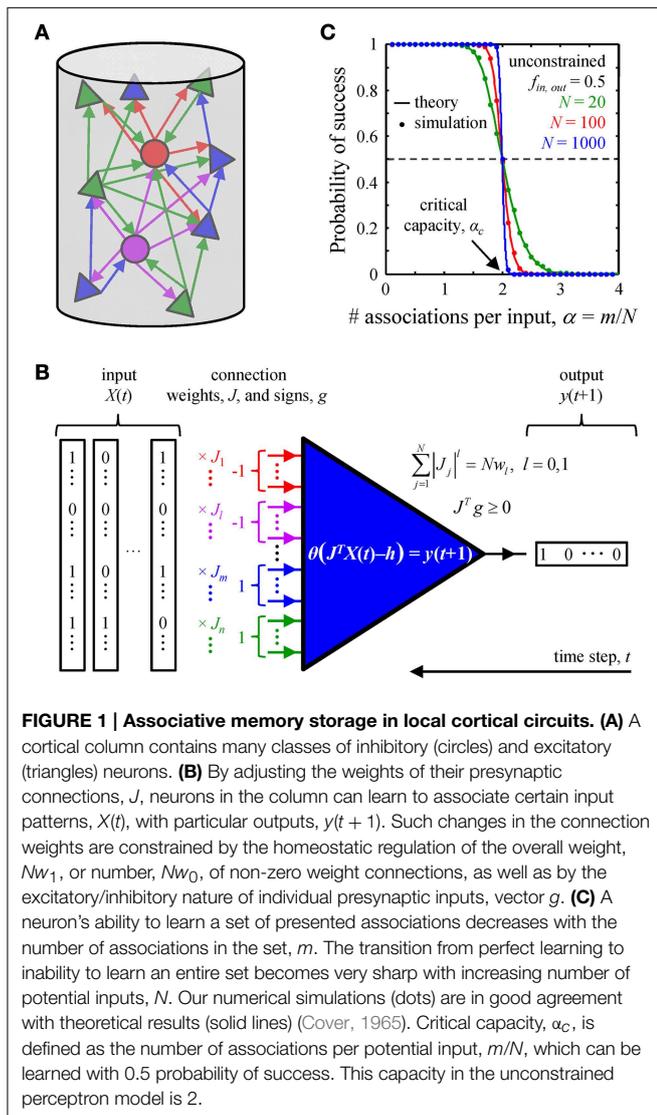
retrieval can be accompanied by forgetting (Kuhl et al., 2010; Wimber et al., 2015). Thus, ongoing activity in the brain is likely to present neurons with sets of associations that are much larger than the learning capacity of the neurons. Consequently, the neurons will learn (presumably in development) as much as they possibly can and from then on (throughout adulthood) will remain at their critical memory storage capacity. This seemingly trivial hypothesis has very powerful implications. Because the properties of the network in the steady-state are independent of the learning path taken by the network to reach that state, one can analyze the steady-state connectivity in the absence of detailed knowledge of the animal's experience or the learning rules involved. This will remain true as long as the animal's experience is rich enough to present the network with a number of associations that exceeds the network's capacity, and the learning rules are versatile enough to learn the critical number of associations within the developmental period. Due to the fundamental nature of this hypothesis its predictions are expected to hold for many species, brain areas, and learning conditions.

In what follows, we extend the steady-state learning model described in Chapeton et al. (2012) by considering multiple classes of excitatory and inhibitory neurons and by incorporating biologically motivated homeostatic constraints. There is emerging evidence suggesting that in spite of circuit changes which accompany learning, individual cells may regulate (*i*) the total number (l_0 norm) and/or (*ii*) the overall weight (l_1 norm) of their presynaptic inputs. For example, it has been shown that the numbers of excitatory and inhibitory synapses onto excitatory cells in the adult cortex remain constant over periods of many days to weeks (Holtmaat et al., 2005, 2006; Fuhrmann et al., 2007; Brown et al., 2009; Hofer et al., 2009; Kim and Nabekura, 2011; Chen et al., 2012). Similarly, it has been shown that synapse loss can be counterbalanced by enlargement of other synapses, such that the summed synaptic surface area per length of dendritic segment remains constant across time and conditions (Bourne and Harris, 2011). In addition, long-term imaging studies have reported that total spine volume, as measured by normalized brightness, remains constant over days (Holtmaat et al., 2006; Kim and Nabekura, 2011). Because spine volume is correlated with synaptic weight (Matsuzaki et al., 2004; Arellano et al., 2007; Harvey and Svoboda, 2007; Zito et al., 2009), these findings suggest that the overall weight of presynaptic inputs remains constant throughout learning.

Below, we provide a detailed formulation of the homeostatically constrained steady-state learning model. The model was solved analytically and the solution was validated numerically. The results were compared with a large number of published experimental studies reporting probabilities of connections and distributions of connection weights for various classes of excitatory and inhibitory neurons in the cerebellum, neocortex, and hippocampus.

Materials and Methods

In this section we formulate a theoretical model of steady-state learning, which incorporates various classes of neurons and

FIGURE 1 | Associative memory storage in local cortical circuits. (A) A cortical column contains many classes of inhibitory (circles) and excitatory (triangles) neurons. (B) By adjusting the weights of their presynaptic connections, J, neurons in the column can learn to associate certain input patterns, $X(t)$, with particular outputs, $y(t + 1)$. Such changes in the connection weights are constrained by the homeostatic regulation of the overall weight, Nw_1, or number, Nw_0, of non-zero weight connections, as well as by the excitatory/inhibitory nature of individual presynaptic inputs, vector g. (C) A neuron's ability to learn a set of presented associations decreases with the number of associations in the set, m. The transition from perfect learning to inability to learn an entire set becomes very sharp with increasing number of potential inputs, N. Our numerical simulations (dots) are in good agreement with theoretical results (solid lines) (Cover, 1965). Critical capacity, α_c, is defined as the number of associations per potential input, m/N, which can be learned with 0.5 probability of success. This capacity in the unconstrained perceptron model is 2.

a number of biologically inspired constraints. Related models, which only include some of the constraints considered here, were previously described in a number of studies (e.g., Cover, 1965; Edwards and Anderson, 1975; Sherrington and Kirkpatrick, 1975; Gardner, 1988; Gardner and Derrida, 1988; Amit et al., 1989; Viswanathan, 1993; Brunel et al., 2004; Chapeton et al., 2012). Detailed description of theoretical and numerical methods can be found in **Text S1**.

Biologically Constrained Model of Steady-state Learning

Networks in the cortex are thought to be organized in columnar units. Such units may include various functional (Hubel and Wiesel, 1963, 1977) and structural columns (Lübke and Feldmeyer, 2007; Stepanyants et al., 2008), which are typically a few hundred micrometers in radius. Analyses of neuron morphology (Kalisman et al., 2003; Binzegger et al., 2004; Stepanyants and Chklovskii, 2005; Stepanyants et al., 2008) have shown that the mesh created by the axonal and dendritic arbors of cells within such units contains numerous micron-size axo-dendritic appositions, which are called potential synapses. A pair of potentially connected cells can form a synaptic connection through local structural synaptic plasticity (Stepanyants et al., 2002; Trachtenberg et al., 2002; Escobar et al., 2008). Though nearby neurons (e.g., separated by less than $50\,\mu$m) within cortical units are typically interconnected in terms of potential synapses, functional synaptic connectivity is invariably sparse (Thomson and Lamy, 2007). For the purpose of this study we consider two cells to be functionally connected if an action potential fired by the presynaptic cell elicits a detectable response in the cell body of the postsynaptic neuron. Such a response, measured as a deviation of the membrane potential from its resting value, is referred to as a unitary postsynaptic potential (uPSP). The sign of a uPSP in a cortical neuron is dependent on the class of the presynaptic cell; it is positive if the presynaptic cell is excitatory (uEPSP) and negative if it is inhibitory (uIPSP).

We consider a local cortical network involved in an associative learning task (**Figure 1A**). The network may contain various excitatory and inhibitory neuron classes which are characterized by distinct firing probabilities. The state of the network at time t, $X(t)$, is described by the binary (0 or 1) activities of all neurons. The network must learn to associate this state with the subsequent network state $X(t+1)$, and that to the state at the following time step, $X(t+2)$, etc., thus learning a chain of associated network states, $X(t) \rightarrow X(t+1) \rightarrow \ldots X(t+m)$. Assuming that the successive network states are uncorrelated (see the next subsection) one can reduce the problem of network learning to the problem of learning by individual neurons (**Figure 1B**).

Thus, we consider a single model neuron, which receives N potential inputs from N potentially presynaptic partners and is faced with a task of learning a set of m input-output associations. The inputs, enumerated with index j, may come from various excitatory and inhibitory neuron classes which have characteristic firing probabilities, f_j. The model neuron is motivated by the McCulloch and Pitts model (McCulloch and Pitts, 1943):

$$\theta\left(\sum_{j=1}^{N} J_j X_j^{\mu} - h\right) = y^{\mu}, \quad \mu = 1, \ldots, m \tag{1}$$

Here, J_j is the weight of presynaptic input j, h is the firing threshold of the neuron, and θ denotes the Heaviside function. The inputs, X_j^{μ} ($\mu = 1, \ldots, m$), and outputs, y^{μ}, are binary and their values are randomly drawn from neuron-class dependent probability distributions: 0 with probability $1 - f_j$ and 1 with probability f_j. The term $\sum_{j=1}^{N} J_j X_j^{\mu}$ plays the role of the postsynaptic potential, and the neuron fires when this potential exceeds h. Equation (1) can be rewritten as a set of inequalities:

$$(2y^{\mu} - 1)\left(\sum_{j=1}^{N} J_j X_j^{\mu} - h\right) > 0, \quad \mu = 1, \ldots, m \tag{2}$$

In this study we impose the following biologically inspired constraints on the learning of associations $\left\{X_j^{\mu}, y^{\mu}\right\}$:

(1) The weights of presynaptic inputs, J_j, are sign-constrained in a way that is determined by the class of individual inputs,

$$J_j g_j \geq 0, \quad j = 1, \ldots, N \tag{3}$$

In these inequalities, $g_j = 1$ if input j is excitatory and $g_j = -1$ if it is inhibitory.

(2) The weights of input connections are also constrained to have a fixed norm. In the following we restrict the analysis to two cases: (i) l_0 norm constraint, which corresponds to learning with a fixed number of non-zero weight inputs and is defined in the limit, $\lim_{l \to 0} \sum_{j=1}^{N} |J_j|^l = Nw_0$, and (ii) l_1 norm constraint, which corresponds to learning with a fixed overall magnitude of the input weights, $\sum_{j=1}^{N} |J_j| = Nw_1$. In these expressions, w_0 is referred to as the overall connection probability, while w_1 is the average absolute connection weight. For conciseness, the l_0 and l_1 norm constraint can be combined into a single equation:

$$\sum_{j=1}^{N} |J_j|^l = Nw_l, \quad l = 0, 1 \tag{4}$$

(3) The firing threshold of the neuron, h, is fixed and does not change during learning.

(4) Associations, $\left\{X_j^{\mu}, y^{\mu}\right\}$, must be learned robustly, which means that the postsynaptic potential must be somewhat above (below) the firing threshold if $y^{\mu} = 1$ (0). This imposed minimal deviation from the threshold is referred to as the robustness parameter, $\kappa \geq 0$. To incorporate this parameter, we modify the r.h.s of Equation (2), making the inequalities more stringent:

$$(2y^{\mu} - 1)\left(\sum_{j=1}^{N} J_j X_j^{\mu} - h\right) > \kappa, \quad \mu = 1, \ldots, m \tag{5}$$

To summarize, the full model can be reduced to the following:

$$
\begin{aligned}
& (2y^\mu - 1)\left(\sum_{j=1}^{N} J_j X_j^\mu - h\right) > \kappa, \quad \mu = 1, \dots, m \\
& J_j g_j \geq 0, \quad j = 1, \dots, N \\
& \sum_{j=1}^{N} |J_j|^l = N w_l, \quad l = 0, 1 \\
& \text{Prob}\left(X_j^\mu\right) = \begin{cases} 1 - f_j, & X_j^\mu = 0 \\ f_j, & X_j^\mu = 1 \end{cases} \\
& \text{Prob}\left(y^\mu\right) = \begin{cases} 1 - f_{out}, & y^\mu = 0 \\ f_{out}, & y^\mu = 1 \end{cases}
\end{aligned}
\tag{6}
$$

Any set of connection weights, J_j, which satisfy Equation (6) is a valid solution of model.

Model Assumptions and Approximations

The steady-state learning model relies on several assumptions and approximations. Here we describe these assumptions and provide experimental evidence supporting the approximations made:

(1) We discretized time into finite-size bins and describe the activity of neurons in the network with binary values: 1 if a neuron is firing and 0 if a neuron is silent. This approximation is reasonable so long as one can choose an integration window which is larger than the duration of a typical uPSP (τ), yet small enough not to encompass successive action potentials fired by any given cell. Denoting the typical firing rate for a cell class with r, such binning of activity should be possible when $r \times \tau$ is smaller than one. In fact, many classes of neurons maintain *in vivo* firing rates that are low enough for this condition to be valid. Specifically, uEPSPs and uIPSPs in pyramidal cells typically have τ in the 40–60 ms range (Sayer et al., 1990; Markram et al., 1997; Gonzalez-Burgos et al., 2005; Sun et al., 2006; Lefort et al., 2009), while the spontaneous firing rates of these cells *in vivo* are $r \sim$ 1–3 Hz (Csicsvari et al., 1999; Puig et al., 2003; Hromadka et al., 2008; Yazaki-Sugiyama et al., 2009, also see Barth and Poulet, 2012 for review). These observations put cortical pyramidal cells well within the range of validity of the above approximation. For inhibitory cells the data is more variable, but generally also supports the approximation. For example, the reported spontaneous firing rates *in vivo* are 9.2 Hz for FS cells in mouse visual cortex (Yazaki-Sugiyama et al., 2009), 7.6 Hz for FS cells in cat striate cortex (Azouz et al., 1997), about 3 Hz for PV cells and <1 Hz for SOM cells in mouse visual cortex (Ma et al., 2010), and 13–14 Hz in CA1 interneurons of rat hippocampus (Csicsvari et al., 1999). We note that it is not clear if the activity of neurons during associative learning resembles low firing rate spontaneous activity, or whether it is similar to the bursting activity of subsets of neurons recorded in animals actively engaged in trained behaviors. Nonetheless, because the fraction of bursting neurons at any given time is small (Barth and Poulet, 2012), the average network firing rate is expected to be low. For example, *in vivo* imaging studies, in which the activities of large ensembles of cortical neurons are monitored over time, have reported population average firing rates of <1 Hz (Kerr et al., 2005; Greenberg et al., 2008; Golshani et al., 2009).

(2) We used linear summation to approximate integration of uPSPs in the cell body. This has been shown to be a good approximation in the neocortex (Tamas et al., 2002; Leger et al., 2005; Araya et al., 2006), cerebellum (Brunel et al., 2004), and hippocampus (Cash and Yuste, 1998, 1999).

(3) We assumed that the threshold of each neuron remains fixed throughout learning. This assumption was motivated by the fact that coefficients of variation in the values of firing thresholds of cortical excitatory and inhibitory neurons are several-fold smaller than the corresponding numbers for the uPSP amplitudes. For example, coefficients of variation for the numerous cortical projections summarized in Supporting Tables 1 and 2 of Chapeton et al. (2012) have the following average values: 0.17 ± 0.02, (mean $\pm$ SE, $n = 9$ systems) for firing thresholds and 0.94 ± 0.03 ($n = 52$ systems) for connection weights.

(4) We followed Dale's principle (Dale, 1935) and assumed that the weights of excitatory/inhibitory inputs remain positive/negative throughout learning.

(5) The activities of all neurons in the network ($j = 1, \dots, N$) at every time step, μ, were randomly drawn from neuron-class specific probability distributions, $\text{Prob}\left(X_j^\mu\right)$, leading to successive network states that are (*i*) independent and (*ii*) random. With this approximation, the problem of learning by the network was decoupled and reduced to the problem of learning by N independent neurons. This approximation is supported by the following experimental observations. (*i*) For cortical neurons *in vivo*, serial correlation coefficients of inter-spike intervals are known to be small. For example, correlations of all lags greater than one are not significantly different from zero (Nawrot et al., 2007; Engel et al., 2008). Although small, but significant, lag one correlations (~ -0.2) are observed at high firing rates (Nawrot et al., 2007), these correlations vanish at <2 Hz (Engel et al., 2008). (*ii*) Correlations between the activities of pairs of cells *in vivo* are known to be small. For example, low pairwise correlations have been reported for pyramidal cells in rat olfactory (~ 0.05) (Miura et al., 2012) and visual (~ 0.033) (Greenberg et al., 2008) cortices. Weak pairwise correlations have also been found in the sensorimotor cortex of behaving monkeys and humans (Truccolo et al., 2010). In addition, extracellular recordings from L2/3 of somatosensory cortex have shown that correlation coefficients between regular spiking cells are small during periods of spontaneous and evoked activity (0.04 and 0.02) (Middleton et al., 2012). Similar results have been obtained for the correlations between regular spiking and fast spiking cells (0.11 and 0.01) (Middleton et al., 2012).

(6) The l_0 and l_1 norm constraints were motivated by the following experimental evidence. (*i*) The density of spines on excitatory neuron dendrites remains constant over days to weeks in many areas of the adult cortex (Holtmaat et al., 2005, 2006; Fuhrmann et al., 2007; Brown et al., 2009; Hofer et al., 2009; Kim and Nabekura, 2011). Likewise, the number of inhibitory synapses onto excitatory dendrites (Chen et al., 2012) and the number of spines on some inhibitory cell dendrites (Keck et al., 2011) remain nearly constant over days. Together, these studies suggest that homeostatic mechanisms may regulate the number of synapses received by excitatory and inhibitory neurons (l_0 norm constraint). (*ii*) It has been reported that the total size

of spines remains constant over several days as measured by the normalized spine brightness (Holtmaat et al., 2006; Kim and Nabekura, 2011). Because the normalized spine brightness is correlated with spine volume (Holtmaat et al., 2005) and the latter is correlated with synaptic weight (Matsuzaki et al., 2004; Arellano et al., 2007; Harvey and Svoboda, 2007; Zito et al., 2009), the overall weight of the presynaptic inputs of a pyramidal cell may be conserved. Another study (Bourne and Harris, 2011) has reported that by 2 h after LTP induction dendrites of CA1 pyramidal neurons in the hippocampus lose some of their small dendritic spines. However, this loss is balanced by an enlargement of the surface area of other excitatory synapses in such a way that the summed surface area of excitatory synapses remained constant across time and conditions. A similar trend was observed for the inhibitory synapses (Bourne and Harris, 2011). These observations imply that dendrites may use local protein synthesis to maintain the overall weight of excitatory and inhibitory inputs (l_1 norm constraint).

(7) We assumed that associative memories can be recalled robustly in the presence of small noise in synaptic transmission, e.g., failures in generation or propagation of presynaptic action potentials, spontaneous neural activity, synaptic failure, and fluctuations in synaptic weight. In order to incorporate this feature into the model we assumed that an association was robustly learned by a neuron if it could be correctly recalled even in the presence of fluctuations in postsynaptic potential of size κ.

Theoretical Solution of the Model

The theoretical solution of the model, Equation (6), is governed by four variables (u_+, u_-, z, and x), which are implicitly defined by the following system of equations:

$$f_{out}I_1\left(-u_-, 0\right) = (1 - f_{out})I_1\left(-u_+, 0\right)$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{f_i(1-f_i)}}\right)^l I_l\left(g_i z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x - hz)}{Nw_1\sqrt{f_i(1-f_i)}}\delta_{l,1},\right.$$
$$\left.\sqrt{\frac{2(\sqrt{N}\kappa x - hz)Q}{hw_0}}\delta_{l,0}\right) = \left(\frac{N}{h}\right)^l w_l Q^l$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{f_i g_i}{\sqrt{f_i(1-f_i)}}I_1\left(g_i z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x - hz)}{Nw_1\sqrt{f_i(1-f_i)}}\delta_{l,1},\right.$$
$$\left.\sqrt{\frac{2(\sqrt{N}\kappa x - hz)Q}{hw_0}}\delta_{l,0}\right) = Q \qquad (7)$$

$$\frac{1}{N}\sum_{i=1}^{N}I_2\left(g_i z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x - hz)}{Nw_1\sqrt{f_i(1-f_i)}}\delta_{l,1},\right.$$
$$\left.\sqrt{\frac{2(\sqrt{N}\kappa x - hz)Q}{hw_0}}\delta_{l,0}\right) = \frac{2Q^2 N\kappa^2}{(u_+ + u_-)^2 h^2}$$

$$Q = 2h\frac{(u_+ + u_-)}{\sqrt{N}\kappa}\frac{f_{out}I_0(-u_-,0) + (1-f_{out})I_0(-u_+,0)}{f_{out}I_1(-u_-,0) + (1-f_{out})I_1(-u_+,0)}x$$

$$u_+ + u_- \geq 0; \quad x \geq 0; \quad \left(\sqrt{N}\kappa x - hz\right)\delta_{l,0} \geq 0$$

Detailed derivation of these equations, together with the definitions of the special functions $I_{0,1,2}$, can be found in **Text S1**.

The critical capacity of a neuron, the probabilities of its non-zero weight connections for different input classes (denoted with i), P_i^{con}, and the probability density functions for its non-zero

input weights, $p_i(J)$, can be expressed in terms of these four variables:

$$\alpha_c = 2x^2\frac{\left(f_{out}I_2\left(-u_-, 0\right) + (1 - f_{out})I_2\left(-u_+, 0\right)\right)}{\left(f_{out}I_1\left(-u_-, 0\right) + (1 - f_{out})I_1\left(-u_+, 0\right)\right)^2}$$

$$P_i^{con} = I_0\left(g_i z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x - hz)}{Nw_1\sqrt{f_i(1-f_i)}}\delta_{l,1} + \sqrt{\frac{2(\sqrt{N}\kappa x - hz)Q}{hw_0}}\delta_{l,0}, 0\right)$$

$$p_i(J) = \frac{\sqrt{f_i(1-f_i)}}{2\sqrt{\pi}I_0\left(g_i z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x-hz)}{Nw_1\sqrt{f_i(1-f_i)}}\delta_{l,1} + \sqrt{\frac{2(\sqrt{N}\kappa x-hz)Q}{hw_0}}\delta_{l,0}, 0\right)}$$

$$\times \left(Q - \frac{2(\sqrt{N}\kappa x - hz)}{f_i(1-f_i)w_0}\frac{h\delta_{l,0}}{N^2J^2}\right)_+$$

$$\times e^{-\left(\frac{\sqrt{f_i(1-f_i)}Q}{2}\frac{N}{h}J + z\sqrt{\frac{f_i}{1-f_i}} + \frac{(\sqrt{N}\kappa x-hz)}{h\sqrt{f_i(1-f_i)}}\left(\frac{hg_i\delta_{l,1}}{Nw_1} + \frac{h\delta_{l,0}}{w_0 NJ}\right)\right)^2} \qquad (8)$$

Plus-sign in the subscript of the last equation denotes the positive part function. Corresponding results for the unconstrained case are included in **Text S1**. Equations (7) and (8) were solved with custom MatLab code (**Theoretical_Results.m** of Supplementary Materials) to produce the results shown in **Figures 2**, **3**, and **Figure S1**.

Model Parameters

Results of the model, Equation (8), depend on the following dimensionless parameters: fraction of potential inputs of each class, N_i/N, firing probabilities of these input classes, f_i, robustness of the postsynaptic neuron, $\sqrt{N}\kappa/h$, and the values of norm constraints, w_0 and Nw_1/h. In Results, we only consider two classes of inputs, inhibitory and excitatory, and thus, the number of independent parameters in the unconstrained model reduces to four (N_{inh}/N, f_{inh}, f_{exc}, $\sqrt{N}\kappa/h$). An additional parameter, w_0 (Nw_1/h), is present in the l_0 (l_1) norm constrained case.

The fraction of potential inhibitory inputs received by a neuron in the network, N_{inh}/N, can be approximated by the average fraction of inhibitory neurons in the cortical column. The latter is known to be in the 0.11–0.20 range (Braitenberg and Schüz, 1998; Lefort et al., 2009; Meyer et al., 2011; Sahara et al., 2012). Thus, we used $N_{inh}/N = 0.15$ in **Figures 3**, **4**. Firing probabilities can be estimated based on the expression $f = r \times \tau$. Numerical values of firing rates, r, and integration windows, τ, for excitatory and inhibitory neurons are given in point 1 of Model Assumptions and Approximations subsection. Based on these numbers we estimated that $f_{exc} \approx 0.1$, while f_{inh} is expected to be larger due to generally higher firing rates of inhibitory neuron classes. However, because the exact values of firing probabilities are not known, in Results we decided to adopt $f_{inh} = f_{exc} = 0.15$ (**Figures 2–4**), while in Supplementary Materials we show the results for the unbiased case, $f_{inh} = f_{exc} = 0.5$ (**Figure S1**). We did not find a clear way to determine the value of robustness parameter, $\sqrt{N}\kappa/h$, from experimental data. This is why, in **Figures 2**, **3** we first show that the results of the model
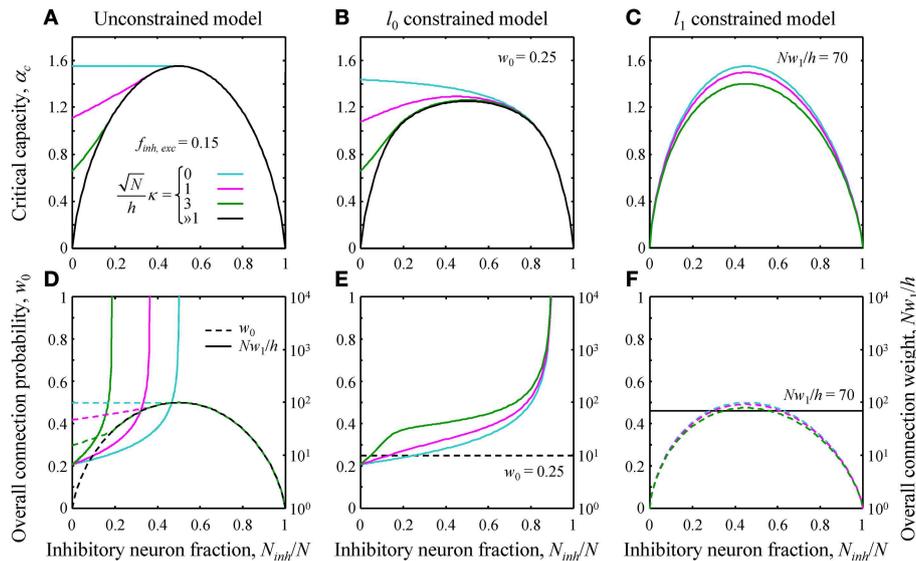
FIGURE 2 | (A–C) Critical capacity as a function of the fraction of potential inhibitory inputs for $f_{inh} = f_{exc} = 0.15$ and various values of the robustness parameter κ. **(A)** The unconstrained model. At certain values of N_{inh}/N, the curves merge with the asymptotic solution (black curve) corresponding to the limit of large robustness, $\sqrt{N}\kappa/h \gg 1$. For smaller fractions of inhibitory inputs, the critical capacity is a decreasing function of κ and an increasing function of N_{inh}/N. **(B)** Qualitatively similar results were obtained in the l_0 constrained model. **(C)** In the l_1 constrained

model the critical capacity curves are slightly skewed to the left and have a maximum at $N_{inh}/N < 0.5$ for all values of κ. Values of the constraints are $w_0 = 0.25$ in B and $Nw_1/h = 70$ in **(C)**. **(D–F)** Overall connection probability (dashed lines) and overall connection weight (solid lines) as a function of N_{inh}/N. Note the different y-axis scales for w_0 (linear, left) and Nw_1/h (logarithmic, right). In the constrained models, the overall connection probability **(E)** or the overall connection weight **(F)** is fixed for all values of N_{inh}/N (horizontal lines).

depend on the value of this parameter in a predictable way, and then set $\sqrt{N}\kappa/h = 3$ in **Figure 4**.

The biologically plausible ranges for the dimensionless constraints, w_0 and Nw_1/h, were approximated based on their definitions (see **Text S1**). For two classes of presynaptic inputs these definitions yield:

$$w_0 = \frac{N_{inh}}{N} P_{inh}^{con} + \left(1 - \frac{N_{inh}}{N}\right) P_{exc}^{con}$$
$$\frac{Nw_1}{h} = \frac{N}{h}\left(\frac{N_{inh}}{N} P_{inh}^{con} \langle J_{inh}\rangle + \left(1 - \frac{N_{inh}}{N}\right) P_{exc}^{con} \langle J_{exc}\rangle\right) \quad (9)$$

Here $P_{inh,exc}^{con}$ and $\langle J_{inh,exc}\rangle$ are the connection probabilities and the average uPSP amplitudes of inhibitory and excitatory inputs. To estimate the values of the constraints we combined the dataset compiled in Chapeton et al. (2012) with a recent study of inhibitory connectivity (Packer and Yuste, 2011) and then restricted the analysis to neocortical systems. The 95% confidence intervals were then obtained using bootstrap sampling with replacement ($n = 10,000$ samples). Parameters h, $P_{inh,exc}^{con}$, and $\langle J_{inh,exc}\rangle$ were sampled with weights proportional to the numbers of experimental counts, whereas N and N_{inh}/N were sampled uniformly from 5,000 to 10,000 (Lefort et al., 2009; Meyer et al., 2010) and 0.11–0.20 (Braitenberg and Schüz, 1998; Lefort et al., 2009; Meyer et al., 2011; Sahara et al., 2012) intervals. This procedure resulted in 95% confidence intervals of [0.1–0.4] for w_0 and [20–190] for Nw_1/h. In **Figure 3** we show how results of the model depend on the values of these constraints, while in

Figures 2, **5** we opted to use the average values obtained from the bootstrap sampling, $w_0 = 0.25$ and $Nw_1/h = 70$.
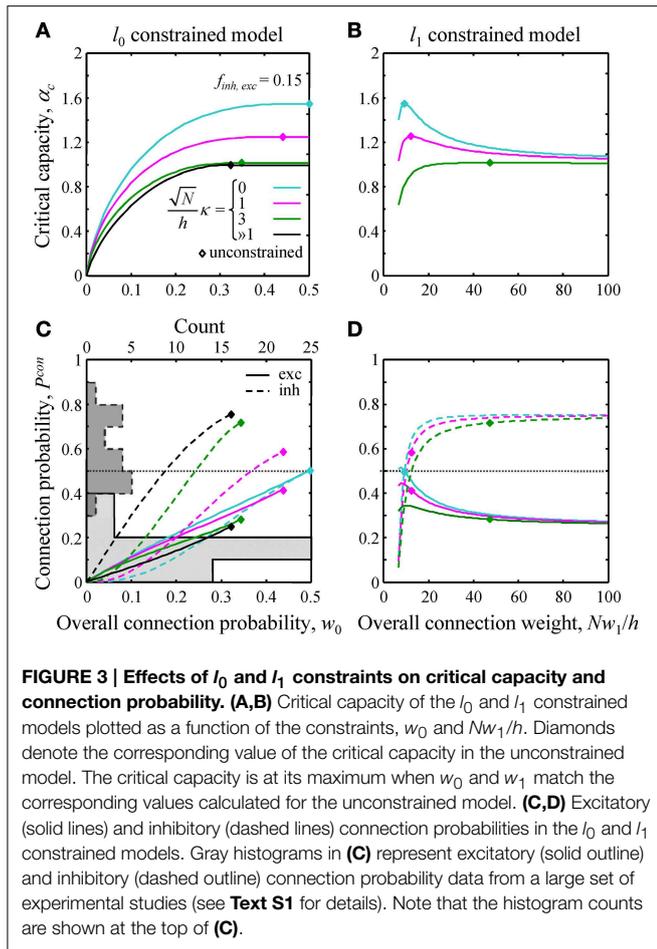
Numerical Solutions of the Model

Since the analytical calculations used to produce the results of this study are very involved we used numerical simulations as an additional validation step. Details of these numerical algorithms can be found in **Text S1**. Standard convex optimization methods were used to produce numerical solutions for the unconstrained and l_1 norm constrained problems. Since the l_0 norm constrained problem is non-convex, numerical solutions in this case were performed with a modified perceptron learning rule. The critical capacity (**Figures S1A–C**) and the distributions of connection weights (**Figures S1D–F**) resulting from these simulations are in good agreement with the theoretical calculations.

Numerical simulations were also used to illustrate plausibility of the steady-state learning hypothesis, which relies on the assumption that a network is able to reach the state of maximum memory storage capacity. To this end, we used perceptron-type learning rules and biologically plausible model parameters to reproduce theoretical results of all three cases (**Figure 4**).

Fitting Distributions of Connection Weights

Theoretical probability density functions of Equation (8) were used to fit simulated and experimental distributions of connection weights. To this end, these equations were rewritten

FIGURE 3 | Effects of l_0 and l_1 constraints on critical capacity and connection probability. (A,B) Critical capacity of the l_0 and l_1 constrained models plotted as a function of the constraints, w_0 and Nw_1/h. Diamonds denote the corresponding value of the critical capacity in the unconstrained model. The critical capacity is at its maximum when w_0 and w_1 match the corresponding values calculated for the unconstrained model. **(C,D)** Excitatory (solid lines) and inhibitory (dashed lines) connection probabilities in the l_0 and l_1 constrained models. Gray histograms in **(C)** represent excitatory (solid outline) and inhibitory (dashed outline) connection probability data from a large set of experimental studies (see **Text S1** for details). Note that the histogram counts are shown at the top of **(C)**.

by using the experimentally determined values of P_i^{con} and introducing two parameters: σ_i, which describes the width of the distribution and G_i, which is the magnitude of the minimum non-zero connection weight present in the l_0 model. The resulting probability density functions are governed by two parameters (σ_i and G_i) in the l_0 case and one parameter (σ_i) in the l_1 case:

$$\sigma_i = \frac{\sqrt{2}h}{\sqrt{f_i(1-f_i)QN}}; \quad G_i = \sqrt{\frac{2(\sqrt{N}\kappa x - hz)}{f_i(1-f_i)w_0 Q}} \frac{h}{N^2}$$

$$p_i(J) = \frac{1}{\sqrt{2\pi}P_i^{con}} \frac{h}{N\sigma_i} \left(1 - \frac{G_i^2}{J^2}\delta_{l,0}\right)_+ \times$$

$$e^{-\left(\frac{g_i J}{\sqrt{2}\sigma_i} + \text{erfinv}(1-2P_i^{con}) + \frac{\sqrt{2}G_i}{\sigma_i}\left(\frac{G_i}{2g_i J}-1\right)\delta_{l,0}\right)^2} \quad (10)$$

Fitting the simulated distributions of inhibitory/excitatory connection weights shown in **Figure 4** was done with one parameter ($\sigma_{inh}/\sigma_{exc}$) in the unconstrained and l_1 constrained models, and two parameters ($\sigma_{inh}/\sigma_{exc}$ and G_{inh}/G_{exc}) in the l_0 case. Fitting was done in MatLab by using non-linear least squares fit. We note that the functional form of the distribution of connection weights in the unconstrained model (Chapeton et al., 2012), written in the notation of Equation (10), is identical to that

of the l_1 case. For this reason, the unconstrained and l_1 models produce identical fits.

A similar fit of experimental distributions of uEPSP amplitudes is shown in **Figure 5**. Fitting with the l_0 model produced the following best fit parameters: $\sigma = 1.06$ [0.88–1.23] mV (mean [95% confidence interval]), $G = 0.055$ [0.045–0.064] mV in **Figure 5A** and $\sigma = 0.79$ [0.71–0.88] mV, $G = 0.051$ [0.041–0.060] mV in **Figure 5B**. For the l_1 model we discarded weak, unreliable connections (gray regions in **Figure 5**) and thus introduced a normalization factor A as an additional fitting parameter. Hence, fitting in this case was also performed with two parameters. The resulting best fit values of σ were: 1.02 [0.92–1.13] mV in **Figure 5A** and 0.81 [0.70–0.92] mV in **Figure 5B**.

Results

Effects of Homeostatic Constraints on Network Capacity and Connectivity

The general solution of the model is described in **Text S1**. Since this solution is very involved, theoretical results were validated with numerical simulations (see **Figure S1**). Here we illustrate the theoretical results by considering a single cell receiving inputs from two classes of presynaptic neurons, one inhibitory and one excitatory. The critical associative memory storage capacity of this cell, α_c, was calculated by solving the system of Equations (7) and (8) (**Theoretical_Results.m** of Supplementary Materials). **Figures 2A–C** show the dependence of α_c in the unconstrained and $l_{0,1}$ norm constrained models on the fraction of potential inhibitory inputs, N_{inh}/N, and the robustness parameter, κ. Though the results for the three models are distinctly different, there are notable common trends. First, in all three models, the critical capacity is a decreasing function of κ, indicating the trade-off between the maximum number of associations a neuron can learn and the robustness of the learned associations. Second, in the case of robust memory storage ($\kappa > 0$), adding a small fraction of inhibitory neurons increases α_c. This, however, comes at the expense of the total number of non-zero weight connections, Nw_0, and/or the overall connection weight, Nw_1 (**Figures 2D–F**).

Next, we evaluated the effects of the $l_{0,1}$ norm constraints on the critical capacity and connection probabilities for various input classes. **Figure 3** shows the results for two classes of inputs, excitatory and inhibitory, with $N_{inh}/N = 0.15$. This numerical value, as well as the values of other parameters of the theory, is based on published experimental data (see Materials and Methods for details). As expected, the critical capacity of the constrained models is maximal when w_0 (**Figure 3A**) and w_1 (**Figure 3B**) match the corresponding values of these parameters in the unconstrained case (diamonds in the figure). This is because at these exact values of w_0 and w_1 the norm constraints are effectively removed, and the solutions of the constrained models reduce to that of the unconstrained case, which naturally has the maximum critical capacity.

Increasing w_0 beyond this point in the l_0 model has no effect on critical capacity (**Figure 3A**) because it is always possible to start with the connectivity of the unconstrained network, and
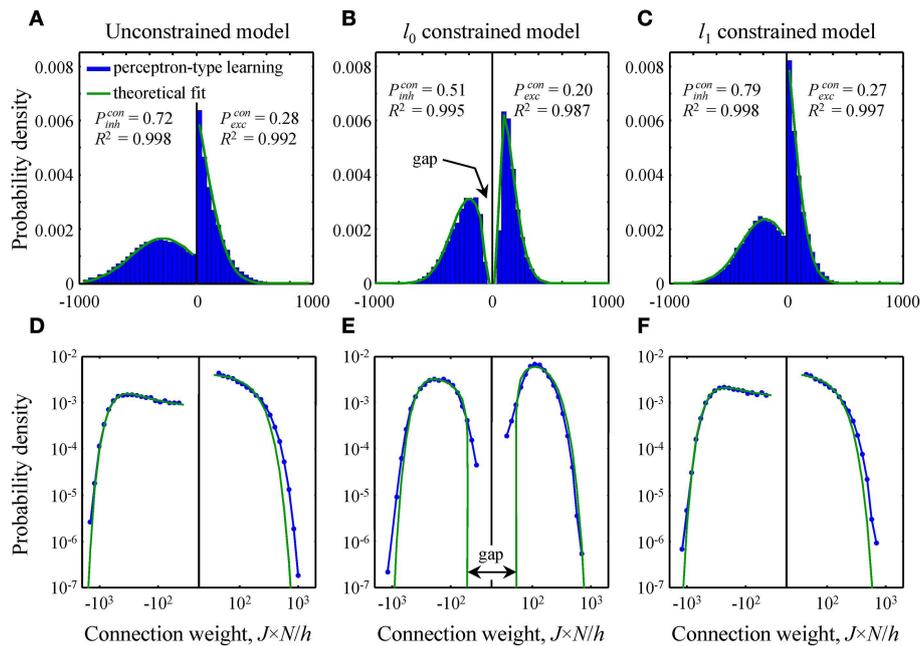
FIGURE 4 | Comparison of theoretical distributions of connection weights to numerical simulations. (A–C) Probability density functions obtained with perceptron-type learning rules for $N = 500$, $N_{inh}/N = 0.15$, $\sqrt{N}\kappa/h = 3$, and $f_{inh,exc} = 0.15$ are shown with blue bars. Values of the constraints are $w_0 = 0.25$ in B and $Nw_1/h = 70$ in C. Green lines show theoretical fits of these probability density functions with Equation (10). Goodness of the fits is captured by the high adjusted R^2 coefficients. The theoretical distributions of excitatory and inhibitory connection weights in the unconstrained **(A)** and l_1 constrained **(C)** models consist of Gaussians truncated at $J = 0$ and finite fractions of zero-weight connections. The distribution in the l_0 constrained model **(B)** also contains a finite fraction of zero-weight connections, but is non-Gaussian. This distribution has gaps between zero and non-zero connection weights for excitatory and inhibitory inputs. Parameters $P^{con}_{inh,exc}$ give theoretical fractions of inhibitory and excitatory non-zero weight connections. **(D–F)** Same probability density functions plotted on a log-log scale show deviations between theory and numerical simulations in the head and tail regions of the distributions.
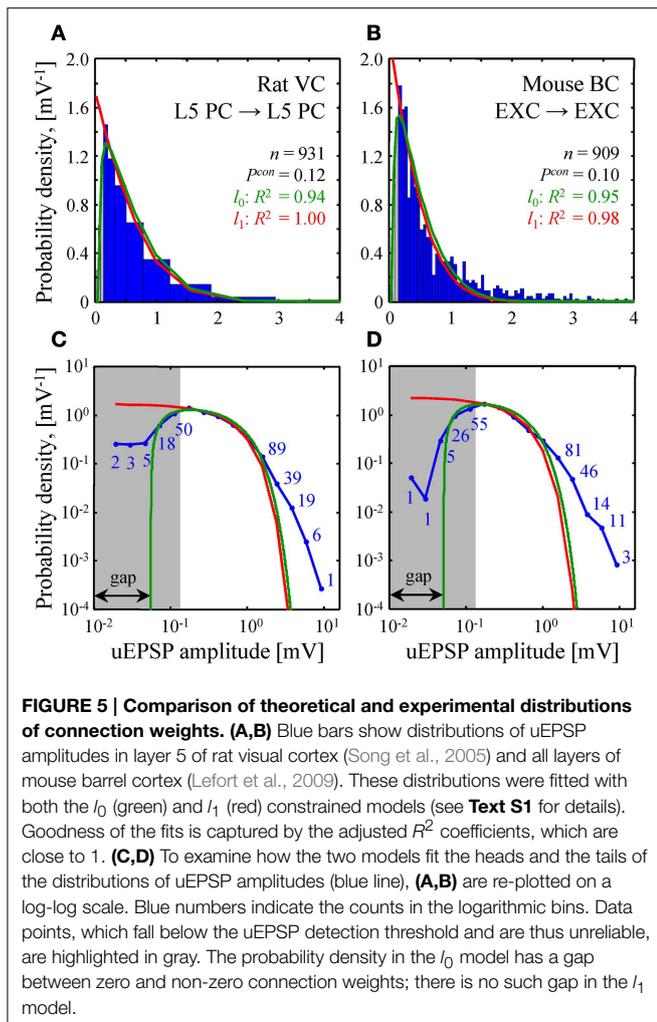
then add a small number of infinitesimally weak connections which will have no effect on the learned associations, but will increase the overall connection probability to the desired value. As the capacity of the l_0 model cannot be greater than the capacity of the unconstrained model, solutions constructed in this way have the maximum possible capacity, and thus are valid for large values of w_0 (to the right of the diamonds in **Figure 3A**). However, because multiple solutions of this type exist, excitatory and inhibitory connection probabilities cannot be defined uniquely.

Numerous experimental studies have shown that the probabilities of local excitatory and inhibitory connections onto the principal cortical neurons (pyramidal and spiny stellate cells in the cerebral cortex and Purkinje cells in the cerebellum) are distinctly different. In particular, excitatory connections are sparse, with connection probabilities well below 0.5, while the inhibitory connection probabilities are generally much higher. This difference can be seen in the histograms of **Figure 3C** which summarize connection probabilities compiled in Chapeton et al. (2012) together with the data from a large study of parvalbumin positive to pyramidal cell connectivity (Packer and Yuste, 2011). Consistent with these observations, the probabilities of excitatory and inhibitory connections in the unconstrained model have been shown to be distinctly different (Chapeton et al., 2012), P^{con}_{exc}

< 0.5 and P^{con}_{inh} > 0.5 (diamonds in **Figures 3C,D**). Therefore, we decided to examine if the constrained models considered in this study produce a similar trend. **Figure 3C** shows P^{con}_{exc} and P^{con}_{inh} in the l_0 norm constrained model plotted as functions of the overall connection probability, w_0. Both P^{con}_{exc} and P^{con}_{inh} increase with w_0, however, P^{con}_{exc} always remains below 0.5, while P^{con}_{inh} exceeds 0.5 beyond certain values of w_0. The range of w_0 values estimated for excitatory cells in the neocortex is 0.1–0.4 (see Materials and Methods). In this range P^{con}_{exc} < 0.5, while P^{con}_{inh} is higher than P^{con}_{exc} in the case of robust memory storage ($\kappa >$ 1). Connection probabilities in the l_1 norm constrained model depend on the value of Nw_1/h (**Figure 3D**). This parameter, estimated from the experimental data, is in the 20–190 range (see Materials and Methods). In this range P^{con}_{exc} < 0.5 and P^{con}_{inh} > 0.5 for all values of robustness. Thus, for biologically realistic values of w_0 and Nw_1/h, the connection probabilities produced by the homeostatically constrained models are consistent with the experimentally observed difference in probabilities of excitatory and inhibitory connections onto principal cells.

Distribution of Connection Weights

In this subsection we compare and contrast the probability densities of input connection weights at critical capacity for the unconstrained and $l_{0,1}$ norm constrained models [see **Text S1**

FIGURE 5 | Comparison of theoretical and experimental distributions of connection weights. (A,B) Blue bars show distributions of uEPSP amplitudes in layer 5 of rat visual cortex (Song et al., 2005) and all layers of mouse barrel cortex (Lefort et al., 2009). These distributions were fitted with both the l_0 (green) and l_1 (red) constrained models (see **Text S1** for details). Goodness of the fits is captured by the adjusted R^2 coefficients, which are close to 1. **(C,D)** To examine how the two models fit the heads and the tails of the distributions of uEPSP amplitudes (blue line), **(A,B)** are re-plotted on a log-log scale. Blue numbers indicate the counts in the logarithmic bins. Data points, which fall below the uEPSP detection threshold and are thus unreliable, are highlighted in gray. The probability density in the l_0 model has a gap between zero and non-zero connection weights; there is no such gap in the l_1 model.

and Equations (8) and (10)]. In the unconstrained and l_1 norm constrained cases these probability densities consist of Gaussian functions, truncated at zero, and finite fractions of zero-weight connections (**Figures 4A,C**). The distribution of connection weights in the l_0 norm constrained model also contains a finite fraction of zero-weight connections. However, the probability density function for non-zero connection weights is non-Gaussian (**Figure 4B**). Interestingly, this function has a gap for weak input connections, i.e., it does not contain non-zero connection weights below a certain threshold. We would like to point out that this feature of connection weights constrained by the l_0 norm was previously reported by Bouten et al. (1990), who considered associative learning by a neuron receiving a single class of sign-unconstrained inputs, i.e., the inputs were not constrained to be excitatory or inhibitory.

Since the steady-state learning hypothesis relies on the assumption that the network can achieve the state of maximum associative memory storage capacity, we set out to show that this can be done with a biologically plausible learning rule. To this end, we attempted to reproduce the theoretical critical capacities and the shapes of theoretical connection weight distributions in

the three models by using modified perceptron learning rules (see Materials and Methods). The simulations were performed for biologically plausible values of model parameters (see Materials and Methods) and about 95% of theoretical, maximum memory storage capacity was reached in all three models: $\alpha_c = 0.95/0.99$ (numerical/theoretical) in the unconstrained model, 0.94/0.99 in the l_0 model, and 0.97/1.01 in the l_1 model.

The overall shape of numerical distributions generated at theoretical critical capacity (**Figures 4A–C**) was in good agreement with the theory, Equation (10). However, small deviations between theory and numerical simulations were observed in the head and tail regions of these distributions. To examine these deviations in more detail we re-plotted the distributions of connection weights on a log-log scale (**Figures 4D–F**). In the unconstrained and l_1 norm constrained models the tails of the numerical distribution appear to be slightly heavier than theoretically predicted Gaussian tails (**Figures 4D,F**), while in the l_0 case (**Figure 4E**) there is a slight deviation in the regions of weak inhibitory and excitatory connections. It is likely that this deviation results from the fact that the numerical simulations were performed for a large, yet finite number of potential inputs, $N = 500$, while theoretical distributions were obtained in the limit of infinite N.

Comparison of Experimental and Theoretical Connection Weight Distributions

The two homeostatically constrained models produce distinctly different distributions of connection weights. Below, we investigate how these distributions compare with the distributions of uPSP amplitudes reported in experimental studies. For this purpose, we selected two studies with very high counts of recorded uEPSPs, one performed in rat visual cortex ($n = 931$, layer 5) (Song et al., 2005) and the other in mouse barrel cortex ($n = 909$, all layers) (Lefort et al., 2009). The two models (green and red lines in **Figures 5A,B**) were used to fit both experimental distributions (blue histograms in **Figures 5A,B**). In spite of the fact that the goodness of these fits was high as measured by the adjusted R^2 coefficients, significant deviations were observed between the distributions ($P < 10^{-12}$ for both models in **Figures 5A,B**; Kolmogorov-Smirnov test), most noticeably in the head and tail regions. To focus on these differences, **Figures 5C,D** show the distributions on a log-log scale.

Due to the fluctuations in electrophysiological recordings, very weak connections between neurons cannot be detected reliably. Such connections are often missed or ignored, leading to a systematic underestimate of weak connection counts. The gray regions in **Figure 5** highlight the values of uEPSP amplitudes which fall below the reliable uEPSP detection threshold [0.1–0.25 mV in rodent neocortex (Mason et al., 1991; Markram et al., 1997; Feldmeyer et al., 1999; Berger et al., 2009)]. Unfortunately, the difference between the distributions produced by the two models only becomes apparent inside the gray regions, and thus, it cannot be directly tested. In these regions of small uEPSP amplitudes the l_0 model provides seemingly better fits to the experimental distributions, but the statistical significance of these results could not be verified based on the available data.

Both models of steady-state learning underestimate the counts of strong synaptic connections (uEPSP amplitudes > 1 mV). Though not very numerous (blue numbers in **Figures 5C,D**), these strong connections appear to be a characteristic feature of cortical connectivity. Sub-criticality of neural networks in the brain, non-linearity in the summation of presynaptic inputs (Brunel et al., 2004), as well as the effects of the finite size of local cortical networks (Chapeton et al., 2012) (e.g., **Figure 4D**, right) have been proposed as possible explanations for the discrepancy between the Gaussian tails of the theoretical probability density functions and the much heavier distribution tails observed experimentally. At present, we are unable to differentiate between these explanations.

Discussion

The study of associative learning by artificial neural networks has a long history dating back to the work of McCulloch and Pitts who introduced one of the first binary neuron models (McCulloch and Pitts, 1943). Rosenblatt later showed that such a binary neuron, now termed the perceptron, can solve classification problems by learning to associate certain input patterns with specific outputs (Rosenblatt, 1957, 1958). The memory storage capacity of a simple perceptron (no constraints, $h = 0$ or learnable threshold, $\kappa = 0$) was calculated by Cover (1965), who used geometrical arguments to show that a simple perceptron can learn to associate $2N$ unbiased input-output patterns ($f_{in} = f_{out} = 0.5$). It was Hopfield who recognized that stable states in recurrent networks of binary neurons can be used as a mechanism for memory storage and recall (Hopfield, 1982). Subsequently, a general framework for the analysis of memory storage capacity was established by Gardner (1988), Gardner and Derrida (1988) who used the replica theory, originally developed for spin glass applications (Edwards and Anderson, 1975; Sherrington and Kirkpatrick, 1975), to solve the problem of robust learning ($\kappa > 0$) of arbitrarily biased associations. To model granule to Purkinje cell connectivity in the cerebellum, Brunel and colleagues constrained Gardner's solution by fixing the firing threshold and forcing the inputs to be all excitatory ($J_j \geq 0$) (Brunel et al., 2004). These results were then extended by Chapeton et al. (2012) on the case of excitatory and inhibitory inputs and applied to cortical circuits.

In this study, we generalize the model of Chapeton et al. (2012) by incorporating multiple classes of excitatory and inhibitory neurons and including two types of experimentally motivated homeostatic constraints. The constraints were designed to ensure that individual neurons maintain a fixed total number or a fixed overall weight of their non-zero inputs throughout learning. Both constrained models were solved analytically by using the replica theory. The results were validated with numerical simulations and compared to the available data on cortical connectivity.

Our theory produced two specific results regarding the connectivity among potentially connected cells in steady-state networks. First, we showed that functional excitatory connections onto principal cells should be realized with less than 50% probability, while the probabilities of inhibitory connections should be higher (**Figures 3C,D**). Because in cortical systems inhibitory cells account for only 11–20% of all neurons, functional connectivity in a steady-state network is expected to be sparse, i.e., it must contain a large fraction of zero-weight connections. This theoretical finding is in qualitative agreement with a dataset compiled from 38 published studies (62 projections in total) in which connection probabilities have been measured in various cortical systems (histograms in **Figure 3C**). It is important to note that a zero-weight connection between neurons does not necessarily imply that the structural connection is absent, as the neurons may still be connected with synapses that are silent (synapses devoid of AMPA receptors Malinow and Malenka, 2002). This detail, however, did not factor into the comparison because neither the theory nor the electrophysiological recordings discriminate between the two alternatives. Furthermore, it has been shown that the fraction of silent synapses in adult cortex is low (e.g., Busetto et al., 2008).

Second, we derived the shapes of the connection weight distributions in a steady-state [see Equation (10) and **Figures 4, 5**]. Similar to the unconstrained model, distribution in the l_1 case consists of Gaussians truncated at zero and a finite fraction of zero-weight connections. The distribution in the l_0 model also contains a finite fraction of zero-weight connections, but the shape of the distribution for non-zero connection weights is no longer Gaussian. It is characterized by a gap between zero and non-zero connection weights. Hence, we predict that a network operating in a steady-state, subject to a constraint on the number of functional connections cannot contain arbitrarily small connection weights. Rather, there should be no functional connections weaker than a certain threshold (∼0.05 mV for uEPSP amplitudes in the neocortex, **Figure 5**). It is not impossible to envision a biological mechanism by which weak connections can be silenced and/or completely eliminated (Oh et al., 2013). In practice, however, it may be difficult to test this hypothesis because the connection weight threshold is too small to be measured reliably with current experimental techniques (Mason et al., 1991; Markram et al., 1997; Feldmeyer et al., 1999; Berger et al., 2009). Interestingly, the value of the connection weight threshold obtained in this study is in agreement with the smallest quantal sizes and miniature EPSP amplitudes recorded in cortical pyramidal neurons [∼0.1 mV (Hardingham et al., 2010)], responses believed to be produced by the release of neurotransmitter from a single presynaptic vesicle.

On the whole, the shapes of model distributions for non-zero connection weights are consistent with the experimental distributions of uEPSP amplitudes in rodent neocortex (**Figure 5**). However, significant discrepancies were observed in the head and tail regions of the distributions ($P < 10^{-12}$, Kolmogorov-Smirnov test). Due to the uncertainties in electrophysiological recordings of very weak connections (below 0.1–0.25 mV in the neocortex), there is still controversy regarding the shape of the uPSP amplitude distribution in this region. Does the distribution smoothly approach zero with decreasing uPSP amplitude as suggested by the log-normal fit performed in Song et al. (2005) and predicted based on the model of multiplicative spine size dynamics (Loewenstein et al., 2011), does it increase as

implied by the exponential distributions of spine head volumes (Mishchenko et al., 2010; Stepanyants and Escobar, 2011) and predicted by the unconstrained and l_1 norm constrained models, or does it have a gap near zero as predicted by the l_0 model? New, more sensitive experimental measurements are required to provide a definitive answer to these questions.

Both models of steady-state learning described in this study predict Gaussian decay of connection weight distributions in the region of strong connections. Contrary to this, distributions of uEPSP amplitudes in **Figure 5** exhibit much heavier tails in the region beyond 1 mV. Several explanations have been previously proposed in order to account for this feature theoretically. Neurons may be operating below their critical capacity, or individual inputs to a neuron may be non-linearly transformed in the dendrites before they are summed in the cell body (Brunel et al., 2004). Heavy distribution tails have also been attributed to the fact that the number of potential presynaptic connections received by cortical neurons, though large [$N \sim 5,000$–$10,000$ (Lefort et al., 2009; Meyer et al., 2010)], is finite, while the theoretical results of this study were obtained in the $N \to \infty$ limit. In fact, numerical simulations performed for $N = 500$ potential inputs can lead to heavier tails of connection weight distributions, supporting this interpretation (e.g., Figure 6D in Chapeton et al., 2012).

It was previously shown (see Figure 5C in Chapeton et al., 2012) that the unconstrained model provides a good fit to the IPSP distribution reported in Holmgren et al. (2003). Unfortunately, due to generally low counts in published PSP data for inhibitory neurons we had to refrain from examining such connections in this study. With future advances in optical methods such as glutamate uncaging (Packer and Yuste, 2011) and optogenetic tagging of genetically defined interneurons (Kvitsiani et al., 2013), it should be possible to study connections between multiple inhibitory and excitatory classes. Here, significant deviations in connection probabilities and shapes of PSP distributions from what is predicted by the theory could reveal connection types that are not directly involved in associative learning.

Our results show how neural networks may benefit from the presence of a small fraction of inhibitory neurons and connections. **Figure 2** illustrates that a small fraction of inhibitory connections increases the capacity of the neurons for robust associative memory storage. This increase, however, comes at the expense of the overall connection weight in the l_0 model or total number of functional connections in the l_1 model, quantities that are likely to be directly related to the metabolic cost of the brain. It would be interesting to find out how different cortical areas balance network performance, measured in terms of information flow or memory storage capacity, with the metabolic cost associated with neuron firing, and number and weight of functional synaptic connections.

In this study, it was assumed that memory recall is a dynamic event in which activity of the network steps through a chain of predefined states. It is well known that in binary neural networks, such as the ones considered here, a chain of network states will inevitably terminate at an attractor in the form of a fixed point or a limit cycle. Hence, memory recall is bound to lead to a frozen network state, or states of cycling activity. With no external input, a network that is robust to small fluctuations could remain in the attractor state over a prolonged period of time, which is unrealistic. Thus, we think that external input, delivered via inter-areal, inter-hemispheric, and subcortical projections, must be responsible for reinitializing the network activity and initiating the recall of new associative memories.

The theoretical model presented in this study describes the effects of associative learning on synaptic connectivity in a steady-state of maximum memory storage capacity. We would like to emphasize that network connectivity in such a state need not be static. As the network is continually learning new associative memories (while forgetting some of the old ones) individual synaptic connections may potentiate or depress, new synaptic connections may be created, and the existing connections may be eliminated. Yet, the average features of steady-state connectivity must remain constant. What is more, these features are independent of the path taken by the network to reach the steady-state, making the problem theoretically tractable. Clearly, not all changes in synaptic connectivity can be described within the steady-state framework presented here. Changes which occur during development (Van Ooyen, 2011), follow injury or lesion (Butz and Van Ooyen, 2013), or accompany learning of new skills may perturb the network from the steady-state for prolonged periods of time (Ruediger et al., 2011). It is more difficult to model these non-equilibrium processes, as they require detailed knowledge of animal's experience and the learning rules involved.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fncom.2015.00074/abstract

Text S1 | Detailed analytical and numerical solutions of the homeostatically constrained model of associative memory storage.

Figure S1 | Validation of theoretical results with numerical simulations.

Theoretical_Results.m. | Custom MatLab code for solving the unconstrained and $l_{0,1}$ constrained models.

References

Amit, J. A., Campbell, C., and Wong, K. Y. M. (1989). The interaction space of neural networks with sign_constrained synapses. *J. Phys. A Math. Gen.* 22, 4687–4693. doi: 10.1088/0305-4470/22/21/030

Araya, R., Eisenthal, K. B., and Yuste, R. (2006). Dendritic spines linearize the summation of excitatory potentials. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18799–18804. doi: 10.1073/pnas.0609225103

Arellano, J. I., Benavides-Piccione, R., Defelipe, J., and Yuste, R. (2007). Ultrastructure of dendritic spines: correlation between synaptic and spine morphologies. *Front. Neurosci.* 1, 131–143. doi: 10.3389/neuro.01.1.1.010.2007

Azouz, R., Gray, C. M., Nowak, L. G., and McCormick, D. A. (1997). Physiological properties of inhibitory interneurons in cat striate cortex. *Cereb. Cortex* 7, 534–545. doi: 10.1093/cercor/7.6.534

Bailey, C. H., and Kandel, E. R. (1993). Structural changes accompanying memory storage. *Annu. Rev. Physiol.* 55, 397–426. doi: 10.1146/annurev.ph.55.030193.002145

Barbour, B., Brunel, N., Hakim, V., and Nadal, J. P. (2007). What can we learn from synaptic weight distributions? *Trends Neurosci.* 30, 622–629. doi: 10.1016/j.tins.2007.09.005

Barth, A. L., and Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.* 35, 345–355. doi: 10.1016/j.tins.2012.03.008

Berger, T. K., Perin, R., Silberberg, G., and Markram, H. (2009). Frequency-dependent disynaptic inhibition in the pyramidal network: a ubiquitous pathway in the developing rat neocortex. *J. Physiol.* 587, 5411–5425. doi: 10.1113/jphysiol.2009.176552

Binzegger, T., Douglas, R. J., and Martin, K. A. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* 24, 8441–8453. doi: 10.1523/JNEUROSCI.1400-04.2004

Bourne, J. N., and Harris, K. M. (2011). Coordination of size and number of excitatory and inhibitory synapses results in a balanced structural plasticity along mature hippocampal CA1 dendrites during LTP. *Hippocampus* 21, 354–373. doi: 10.1002/hipo.20768

Bouten, M., Engel, A., Komoda, A., and Serneels, R. (1990). Quenched versus annealed dilution in neural networks. *J. Phys. A Math. Gen.* 23, 4643–4657. doi: 10.1088/0305-4470/23/20/025

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity.* Berlin; New York, NY: Springer.

Brown, C. E., Aminoltejari, K., Erb, H., Winship, I. R., and Murphy, T. H. (2009). *In vivo* voltage-sensitive dye imaging in adult mice reveals that somatosensory maps lost to stroke are replaced over weeks by new structural and functional circuits with prolonged modes of activation within both the peri-infarct zone and distant sites. *J. Neurosci.* 29, 1719–1734. doi: 10.1523/JNEUROSCI.4249-08.2009

Brunel, N., Hakim, V., Isope, P., Nadal, J. P., and Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron* 43, 745–757. doi: 10.1016/j.neuron.2004.08.023

Busetto, G., Higley, M. J., and Sabatini, B. L. (2008). Developmental presence and disappearance of postsynaptically silent synapses on dendritic spines of rat layer 2/3 pyramidal neurons. *J. Physiol.* 586, 1519–1527. doi: 10.1113/jphysiol.2007.149336

Butz, M., and Van Ooyen, A. (2013). A simple rule for dendritic spine and axonal bouton formation can account for cortical reorganization after focal retinal lesions. *PLoS Comput. Biol.* 9:e1003259. doi: 10.1371/journal.pcbi.1003259

Cash, S., and Yuste, R. (1998). Input summation by cultured pyramidal neurons is linear and position-independent. *J. Neurosci.* 18, 10–15.

Cash, S., and Yuste, R. (1999). Linear summation of excitatory inputs by CA1 pyramidal neurons. *Neuron* 22, 383–394. doi: 10.1016/S0896-6273(00)81098-3

Chapeton, J., Fares, T., Lasota, D., and Stepanyants, A. (2012). Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3614–E3622. doi: 10.1073/pnas.1211467109

Chen, J. L., Villa, K. L., Cha, J. W., So, P. T., Kubota, Y., and Nedivi, E. (2012). Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron* 74, 361–373. doi: 10.1016/j.neuron.2012.02.030

Chklovskii, D. B., Mel, B. W., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* 431, 782–788. doi: 10.1038/nature03012

Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. EC* 14, 326–334. doi: 10.1109/pgec.1965.264147

Csicsvari, J., Hirase, H., Czurko, A., Mamiya, A., and Buzsaki, G. (1999). Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat. *J. Neurosci.* 19, 274–287.

Dale, H. (1935). Pharmacology and nerve-endings. *Proc. R. Soc. Med.* 28, 319–332. doi: 10.1097/00005053-193510000-00018

Edwards, S. F., and Anderson, P. W. (1975). Theory of spin glasses. *J. Phys. F Metal Phys.* 5, 965–974. doi: 10.1088/0305-4608/5/5/017

Engel, T. A., Schimansky-Geier, L., Herz, A. V., Schreiber, S., and Erchova, I. (2008). Subthreshold membrane-potential resonances shape spike-train patterns in the entorhinal cortex. *J. Neurophysiol.* 100, 1576–1589. doi: 10.1152/jn.01282.2007

Escobar, G., Fares, T., and Stepanyants, A. (2008). Structural plasticity of circuits in cortical neuropil. *J. Neurosci.* 28, 8477–8488. doi: 10.1523/JNEUROSCI.2046-08.2008

Feldmeyer, D., Egger, V., Lubke, J., and Sakmann, B. (1999). Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J. Physiol.* 521(Pt 1), 169–190. doi: 10.1111/j.1469-7793.1999.00169.x

Fuhrmann, M., Mitteregger, G., Kretzschmar, H., and Herms, J. (2007). Dendritic pathology in prion disease starts at the synaptic spine. *J. Neurosci.* 27, 6224–6233. doi: 10.1523/JNEUROSCI.5062-06.2007

Gardner, E. (1988). The space of interactions in neural network models. *J. Phys. A Math. Gen.* 21, 257–270. doi: 10.1088/0305-4470/21/1/030

Gardner, E., and Derrida, B. (1988). Optimal storage properties of neural network models. *J. Phys. A Math. Gen.* 21, 271–284. doi: 10.1088/0305-4470/21/1/031

Golshani, P., Goncalves, J. T., Khoshkhoo, S., Mostany, R., Smirnakis, S., and Portera-Cailliau, C. (2009). Internally mediated developmental desynchronization of neocortical network activity. *J. Neurosci.* 29, 10890–10899. doi: 10.1523/JNEUROSCI.2012-09.2009

Gonzalez-Burgos, G., Krimer, L. S., Povysheva, N. V., Barrionuevo, G., and Lewis, D. A. (2005). Functional properties of fast spiking interneurons and their synaptic connections with pyramidal cells in primate dorsolateral prefrontal cortex. *J. Neurophysiol.* 93, 942–953. doi: 10.1152/jn.00787.2004

Greenberg, D. S., Houweling, A. R., and Kerr, J. N. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11, 749–751. doi: 10.1038/nn.2140

Hardingham, N. R., Read, J. C., Trevelyan, A. J., Nelson, J. C., Jack, J. J., and Bannister, N. J. (2010). Quantal analysis reveals a functional correlation between presynaptic and postsynaptic efficacy in excitatory connections from rat neocortex. *J. Neurosci.* 30, 1441–1451. doi: 10.1523/JNEUROSCI.3244-09.2010

Harvey, C. D., and Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* 450, 1195–1200. doi: 10.1038/nature06416

Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T., and Hubener, M. (2009). Experience leaves a lasting structural trace in cortical circuits. *Nature* 457, 313–317. doi: 10.1038/nature07487

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol.* 551, 139–153. doi: 10.1113/jphysiol.2003.044784

Holtmaat, A., and Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* 10, 647–658. doi: 10.1038/nrn2699

Holtmaat, A., Wilbrecht, L., Knott, G. W., Welker, E., and Svoboda, K. (2006). Experience-dependent and cell-type-specific spine growth in the neocortex. *Nature* 441, 979–983. doi: 10.1038/nature04783

Holtmaat, A. J., Trachtenberg, J. T., Wilbrecht, L., Shepherd, G. M., Zhang, X., Knott, G. W., et al. (2005). Transient and persistent dendritic spines in the neocortex *in vivo. Neuron* 45, 279–291. doi: 10.1016/j.neuron.2005.01.003

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Hromadka, T., Deweese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6:e16. doi: 10.1371/journal.pbio.0060016

Hubel, D. H., and Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *J. Physiol.* 165, 559–568. doi: 10.1113/jphysiol.1963.sp007079

Hubel, D. H., and Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B. Biol. Sci.* 198, 1–59. doi: 10.1098/rspb.1977.0085

Middleton, J. W., Omar, C., Doiron, B., and Simons, D. J. (2012). Neural correlation is stimulus modulated by feedforward inhibitory circuitry. *J. Neurosci.* 32, 506–518. doi: 10.1523/JNEUROSCI.3474-11.2012

Kalisman, N., Silberberg, G., and Markram, H. (2003). Deriving physical connectivity from neuronal morphology. *Biol. Cybern.* 88, 210–218. doi: 10.1007/s00422-002-0377-3

Keck, T., Scheuss, V., Jacobsen, R. I., Wierenga, C. J., Eysel, U. T., Bonhoeffer, T., et al. (2011). Loss of sensory input causes rapid structural changes of inhibitory neurons in adult mouse visual cortex. *Neuron* 71, 869–882. doi: 10.1016/j.neuron.2011.06.034

Kerr, J. N., Greenberg, D., and Helmchen, F. (2005). Imaging input and output of neocortical networks *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14063–14068. doi: 10.1073/pnas.0506029102

Kim, S. K., and Nabekura, J. (2011). Rapid synaptic remodeling in the adult somatosensory cortex following peripheral nerve injury and its association with neuropathic pain. *J. Neurosci.* 31, 5477–5482. doi: 10.1523/JNEUROSCI.0328-11.2011

Kuhl, B. A., Shah, A. T., Dubrow, S., and Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* 13, 501–506. doi: 10.1038/nn.2498

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498, 363–366. doi: 10.1038/nature12176

Lefort, S., Tomm, C., Floyd Sarria, J. C., and Petersen, C. C. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* 61, 301–316. doi: 10.1016/j.neuron.2008.12.020

Leger, J. F., Stern, E. A., Aertsen, A., and Heck, D. (2005). Synaptic integration in rat frontal cortex shaped by network activity. *J. Neurophysiol.* 93, 281–293. doi: 10.1152/jn.00067.2003

Loewenstein, Y., Kuras, A., and Rumpel, S. (2011). Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex *in vivo*. *J. Neurosci.* 31, 9481–9488. doi: 10.1523/JNEUROSCI.6130-10.2011

Lübke, J., and Feldmeyer, D. (2007). Excitatory signal flow and connectivity in a cortical column: focus on barrel cortex. *Brain Struct. Funct.* 212, 3–17. doi: 10.1007/s00429-007-0144-2

Ma, W. P., Liu, B. H., Li, Y. T., Huang, Z. J., Zhang, L. I., and Tao, H. W. (2010). Visual representations by cortical somatostatin inhibitory neurons–selective but with weak and delayed responses. *J. Neurosci.* 30, 14371–14379. doi: 10.1523/JNEUROSCI.3248-10.2010

Malinow, R., and Malenka, R. C. (2002). AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* 25, 103–126. doi: 10.1146/annurev.neuro.25.112701.142758

Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.* 500(Pt 2), 409–440. doi: 10.1113/jphysiol.1997.sp022031

Mason, A., Nicoll, A., and Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex *in vitro*. *J. Neurosci.* 11, 72–84.

Matsuzaki, M., Honkura, N., Ellis-Davies, G. C., and Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature* 429, 761–766. doi: 10.1038/nature02617

McCulloch, W., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biol.* 5, 115–133.

Meyer, H. S., Schwarz, D., Wimmer, V. C., Schmitt, A. C., Kerr, J. N., Sakmann, B., et al. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16807–16812. doi: 10.1073/pnas.1113648108

Meyer, H. S., Wimmer, V. C., Oberlaender, M., De Kock, C. P., Sakmann, B., and Helmstaedter, M. (2010). Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cereb. Cortex* 20, 2277–2286. doi: 10.1093/cercor/bhq067

Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K. M., and Chklovskii, D. B. (2010). Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron* 67, 1009–1020. doi: 10.1016/j.neuron.2010.08.014

Miura, K., Mainen, Z. F., and Uchida, N. (2012). Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* 74, 1087–1098. doi: 10.1016/j.neuron.2012.04.021

Nawrot, M. P., Boucsein, C., Rodriguez-Molina, V., Aertsen, A., Grun, S., and Rotter, S. (2007). Serial interval statistics of spontaneous activity in cortical neurons *in vivo* and *in vitro*. *Neurocomputing* 70, 1717–1722. doi: 10.1016/j.neucom.2006.10.101

Oh, W. C., Hill, T. C., and Zito, K. (2013). Synapse-specific and size-dependent mechanisms of spine structural plasticity accompanying synaptic weakening. *Proc. Natl. Acad. Sci. U.S.A.* 110, E305–E312. doi: 10.1073/pnas.1214705110

Packer, A. M., and Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *J. Neurosci.* 31, 13260–13271. doi: 10.1523/JNEUROSCI.3131-11.2011

Puig, M. V., Celada, P., Diaz-Mataix, L., and Artigas, F. (2003). *In vivo* modulation of the activity of pyramidal neurons in the rat medial prefrontal cortex by 5-HT2A receptors: relationship to thalamocortical afferents. *Cereb. Cortex* 13, 870–882. doi: 10.1093/cercor/13.8.870

Rosenblatt, F. (1957). *The Perceptron–a Perceiving and Recognizing Automaton*. Cornell Aeronautical Laboratory Report 85-460-1.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408. doi: 10.1037/h0042519

Ruediger, S., Vittori, C., Bednarek, E., Genoud, C., Strata, P., Sacchetti, B., et al. (2011). Learning-related feedforward inhibitory connectivity growth required for memory precision. *Nature* 473, 514–518. doi: 10.1038/nature09946

Sahara, S., Yanagawa, Y., O'Leary, D. D., and Stevens, C. F. (2012). The fraction of cortical GABAergic neurons is constant from near the start of cortical neurogenesis to adulthood. *J. Neurosci.* 32, 4755–4761. doi: 10.1523/JNEUROSCI.6412-11.2012

Sayer, R. J., Friedlander, M. J., and Redman, S. J. (1990). The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *J. Neurosci.* 10, 826–836.

Sherrington, D., and Kirkpatrick, S. (1975). Solvable model of a spin glass. *Phys. Rev. Lett.* 35, 1792–1796. doi: 10.1103/PhysRevLett.35.1792

Song, S., Sjostrom, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 3:e68. doi: 10.1371/journal.pbio.0030068

Stepanyants, A., and Chklovskii, D. B. (2005). Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* 28, 387–394. doi: 10.1016/j.tins.2005.05.006

Stepanyants, A., and Escobar, G. (2011). Statistical traces of long-term memories stored in strengths and patterns of synaptic connections. *J. Neurosci.* 31, 7657–7669. doi: 10.1523/JNEUROSCI.0255-11.2011

Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvarday, Z. F., Ferecsko, A. S., and Chklovskii, D. B. (2008). Local potential connectivity in cat primary visual cortex. *Cereb. Cortex* 18, 13–28. doi: 10.1093/cercor/bhm027

Stepanyants, A., Hof, P. R., and Chklovskii, D. B. (2002). Geometry and structural plasticity of synaptic connectivity. *Neuron* 34, 275–288. doi: 10.1016/S0896-6273(02)00652-9

Sun, Q. Q., Huguenard, J. R., and Prince, D. A. (2006). Barrel cortex microcircuits: thalamocortical feedforward inhibition in spiny stellate cells is mediated by a small number of fast-spiking interneurons. *J. Neurosci.* 26, 1219–1230. doi: 10.1523/JNEUROSCI.4727-04.2006

Tamas, G., Szabadics, J., and Somogyi, P. (2002). Cell type- and subcellular position-dependent summation of unitary postsynaptic potentials in neocortical neurons. *J. Neurosci.* 22, 740–747.

Thomson, A. M., and Lamy, C. (2007). Functional maps of neocortical local circuitry. *Front. Neurosci.* 1, 19–42. doi: 10.3389/neuro.01.1.1.002.2007

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., et al. (2002). Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* 420, 788–794. doi: 10.1038/nature01273

Truccolo, W., Hochberg, L. R., and Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nat. Neurosci.* 13, 105–111. doi: 10.1038/nn.2455

Van Ooyen, A. (2011). Using theoretical models to analyse neural development. *Nat. Rev. Neurosci.* 12, 311–326. doi: 10.1038/nrn3031

Viswanathan, R. R. (1993). Sign-constrained synapses and biased patterns in neural networks. *J. Phys. A Math. Gen.* 26, 6195–6203. doi: 10.1088/0305-4470/26/22/020

Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., and Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat. Neurosci.* 18, 582–589. doi: 10.1038/nn.3973

Yazaki-Sugiyama, Y., Kang, S., Cateau, H., Fukai, T., and Hensch, T. K. (2009). Bidirectional plasticity in fast-spiking GABA circuits by visual experience. *Nature* 462, 218–221. doi: 10.1038/nature08485

Zito, K., Scheuss, V., Knott, G., Hill, T., and Svoboda, K. (2009). Rapid functional maturation of nascent dendritic spines. *Neuron* 61, 247–258. doi: 10.1016/j.neuron.2008.10.054

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.