nature neuroscience

Is cortical connectivity optimized for storing information?

Nicolas Brunel[1,2]

Cortical networks are thought to be shaped by experience-dependent synaptic plasticity. Theoretical studies have shown that synaptic plasticity allows a network to store a memory of patterns of activity such that they become attractors of the dynamics of the network. Here we study the properties of the excitatory synaptic connectivity in a network that maximizes the number of stored patterns of activity in a robust fashion. We show that the resulting synaptic connectivity matrix has the following properties: it is sparse, with a large fraction of zero synaptic weights ('potential' synapses); bidirectionally coupled pairs of neurons are over-represented in comparison to a random network; and bidirectionally connected pairs have stronger synapses on average than unidirectionally connected pairs. All these features reproduce quantitatively available data on connectivity in cortex. This suggests synaptic connectivity in cortex is optimized to store a large number of attractor states in a robust fashion.

Local networks of pyramidal cells in neocortex are extensively interconnected through recurrent collaterals[1]. Anatomical studies have shown that these networks could be potentially fully connected, in the sense that axons of presynaptic pyramidal cells pass within a micrometer of dendrites of all neighboring pyramidal cells[2]. Because dynamically growing or retracting dendritic spines can bridge this distance[3], it is thought that this geometry gives the local network a property of being potentially fully connected[4], even though electrophysiological studies detect connections only in approximately 10% of pairs of neurons[5–10]. Investigations of the statistical properties of the resulting recurrent connectivity have revealed a number of nonrandom features, such as an over-representation of bidirectionally connected pairs or of specific higher order network motifs[10–14].

This recurrent connectivity has long been hypothesized to allow cortical networks to store an extensive amount of information. However, there is a debate over the format of stored information. One popular theory is that information is stored in the form of attractor states[15,16]: specific attractors of the network dynamics represent learned internal representations of external stimuli that have been repeatedly presented to the network. In the simplest type of attractors, a subset of neurons is active at elevated firing rates, while the rest stay at background activity levels[17]. This is consistent with a wide array of experimental data on persistent activity during delayed response tasks in monkeys in various cortical areas[18–21]. Another popular theory posits that information is stored in the form of sequences of activity[22,23]. Such a theory is consistent with sequential activity observed, for example, in the posterior parietal cortex[24] and hippocampus[25].

Theoretical models that describe persistent activity, or sequences of activity, in large networks of neurons typically assume very specific connectivity rules. It is unclear whether cortical connectivity obeys the rules postulated by these models. Here we adopt a completely different approach that was pioneered by Gardner[26] (**Fig. 1**). The idea is to consider the space of all possible connectivity matrices, without making any assumption on the specific form of the learning rule, except that it should be able to learn specific patterns of activity. Specific patterns of activity that need to be learned by the network impose specific constraints on the synaptic connectivity. The more patterns are learned by the network, the smaller the set of synaptic weights that stabilize all the patterns as attractors of the dynamics. As shown by Gardner, the maximal capacity of the network (maximal number of patterns that can be stored) can be computed as the number of patterns for which the 'typical' volume of the set of weights that satisfies all the constraints imposed by learning goes to zero. Here we characterize the statistics of connectivity matrices satisfying all constraints imposed by learning when the network is close to maximal capacity.

RESULTS

Model and approach

Cortical networks are composed of excitatory and inhibitory neurons. Some evidence suggests that the excitatory subnetwork is highly structured, while connections involving inhibitory interneurons are more stereotyped, with close to full local connectivity[27,28]. We therefore focused on the excitatory connectivity and first incorporated inhibition in the thresholds of excitatory neurons (see below for a discussion of models in which inhibitory neurons are explicitly modeled). We considered a large, potentially fully connected network of N excitatory neurons, with a connectivity matrix $w_{ij} \geq 0$ ($i,j = 1, \ldots, N$) (**Fig. 1a**). For simplicity, neurons were taken to be binary, with two possible states, 0 (inactive) or 1 (active). At each time step, neurons were active (inactive), $S_i(t + 1) = 1$ (0), if their summed synaptic inputs

$$\sum_{j=1,\ldots,N} w_{ij} S_j(t)$$

were above (below) a threshold T, incorporating in an effective way inhibitory neurons. The connectivity matrix was assumed to

[1]Department of Statistics, The University of Chicago, Chicago, Illinois, USA. [2]Department of Neurobiology, The University of Chicago, Chicago, Illinois, USA. Correspondence should be addressed to N.B. (nbrunel@uchicago.edu).

Received 11 January; accepted 14 March; published online 11 April 2016; doi:10.1038/nn.4286

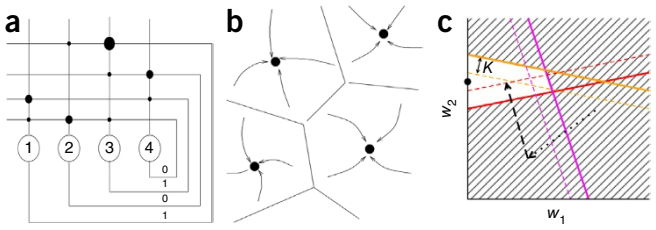

Figure 1 Model and space of synaptic weights. (**a**) Sketch of the network model. N neurons (here, $N = 4$) are connected through recurrent synaptic connections (black circles with a diameter proportional to synaptic strength, at the intersections of vertical lines representing dendrites and horizontal lines representing axons). A particular pattern that needs to be stored is "1010," in which neurons 1 and 3 need to be active while neurons 2 and 4 need to be silent. For this particular pattern to be stored, connections between neurons 1 and 3 need to be strong. (**b**) Sketch of 'attractor landscape' in the space of network states. Black circles are attractors of the network dynamics (here, $p = 4$). Lines indicate boundaries of basins of attraction around each attractor. The larger these basins, the more robust the attractors are against noise. The size of the basins of attraction is measured by the parameter K. (**c**) Sketch of the space of couplings: three patterns to be stored constrain the synaptic weights to be in the white region. Constraints corresponding to the three patterns are shown as the three colored lines, while the $w_1 = 0$ and $w_2 = 0$ axes are shown as black lines. Arrows indicate possible learning dynamics through the space of weights that reach the white region. The black circle indicates the synaptic connectivity that maximizes robustness, as it lies at a maximal distance from the colored lines. Note that in this example this point lies on the $w_1 = 0$ axis.

have been shaped by the learning of p random uncorrelated binary patterns of activity η_i^μ ($i = 1, \ldots, N$, $\mu = 1, \ldots, p$), drawn randomly with a coding level f (for all i and μ, $\eta_i^\mu = 1$ with probability f and 0 with probability $1 - f$). These patterns of activity were defined as being learned only when they were attractors of the dynamics of the network. Furthermore, we imposed a minimal size K of the basins of attraction of each attractor, by requiring that the summed synaptic inputs were above $T + K$ (below $T - K$) for neurons that should be active (inactive) in a given attractor, to ensure robustness of these attractors in the face of noise. Attractors and their basins of attraction are sketched schematically in **Figure 1b**.

We used two approaches to characterize the statistics of network connectivity. The first approach uses the 'cavity method' from statistical physics[29]. The idea of the cavity method is to compute self-consistently the distributions of 'local fields' (that is, the distributions of total inputs to neurons when the network is in an attractor state corresponding to one of the learned patterns) and synaptic weights, in the subspace of weights satisfying all constraints imposed by learning (**Fig. 1c**). These constraints take the form of hyperplanes in the spaces of input connectivity of each input neuron (Online Methods, equations (1) and (2)), that separate regions of synaptic connectivity that satisfy the constraints from regions that do not. Increasing the robustness parameter K means that the constraints are more stringent and, consequently, that the subspace of weights satisfying all constraints becomes smaller. The calculation involves performing two types of averages: an average over all weights in this subspace and an average over random patterns (see **Supplementary Note**). We used this approach to compute the distribution of synaptic weights in optimal networks, as well as the joint distributions of pairs of weights. The second approach was to use numerical simulations, using the perceptron learning algorithm[30], adapted to networks with non-negative weights[31], independently for each neuron to learn the randomly generated patterns. The advantage of using this algorithm is that it is guaranteed to find a solution to the learning problem, provided such a solution exists.

Distribution of synaptic weights

We first computed the distribution of synaptic weights, in a network that maximizes the number of stored patterns for a fixed robustness level or, equivalently, maximizes the robustness level for a fixed number of stored patterns, using the cavity method. We found this distribution to be identical to the distribution in perceptrons with excitatory weights at maximal storage capacity[32]: it is composed of a delta function at zero weight ('potential' synapses) and a truncated Gaussian at positive weights (**Fig. 2**). The fraction of zero-weight synapses is large (50% or more) for any parameter of the model. The connectivity matrix of a network close to its maximal storage capacity is therefore a sparse matrix. The connection probability is determined primarily by the robustness of the attractors, as measured by a rescaled robustness parameter

$$\rho = K \Big/ \left(w \sqrt{f(1-f)N} \right),$$

where w is the average synaptic weight (**Supplementary Note**): it is 50% when $\rho = 0$, but goes to zero in the limit of a large ρ (**Fig. 2b**). Hence, the more robust the information storage, the sparser the network. This is because the boundary of the subspace of connectivity matrices satisfying all the constraints includes two types of hyperplanes: those associated with the learned patterns and those that enforce the sign constraints on the weights; that is, $w_{ij} = 0$ for some i,j (**Fig. 1c**). Increasing the robustness implies increasing the distance from the pattern-associated hyperplanes, but not from the sign-constraint-associated ones. Therefore, networks maximizing robustness typically lie on a large fraction of $w_{ij} = 0$ hyperplanes (**Fig. 1c**).

This distribution of synaptic weights is in good agreement with experimental data in neocortex. Anatomical studies have found that

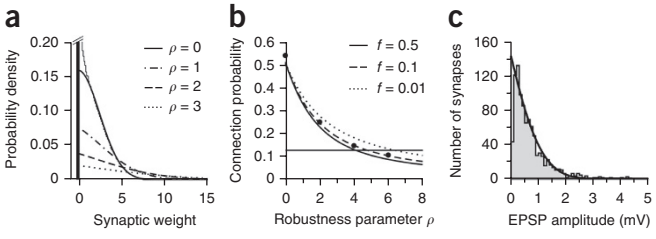

Figure 2 Distribution of synaptic weights of an optimal network. (**a**) Optimal distribution of weights at maximal capacity, for different values of the robustness parameter ρ. The thin black line shows the result of a numerical simulation using the perceptron learning algorithm in a network of $N = 800$ neurons, for $\rho = 0$, $f = 0.5$. Note that the distribution of weights depends only on two parameters, the mean weight and ρ, which determines its shape (and in particular the fraction of zero-weight synapses). (**b**) Connection probability ($P(w_{ij} > 0) = 1 - P(w_{ij} = 0)$) at maximal capacity as a function of robustness parameter, for different values of the coding level f. Circles show the result of simulations ($N = 800$, $f = 0.5$). The horizontal line shows the observed connection probability in the data of ref. 11. (**c**) Distribution of synaptic weights of optimal network (black curve) versus experimentally recorded distribution of weights (gray histogram)[11].

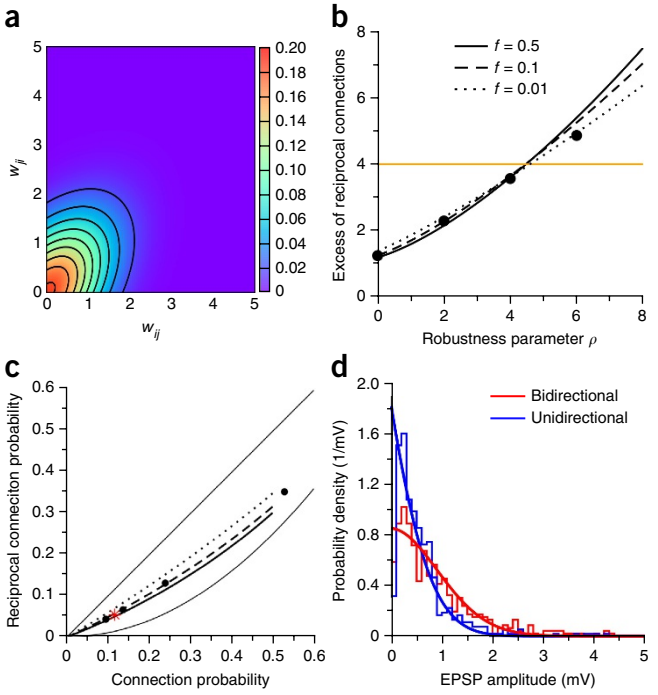

Figure 3 Joint distribution of synaptic weights connecting a pair of neurons. (**a**) The joint distribution is a correlated two-dimensional Gaussian, truncated on the upper right quadrant. (**b**) Excess of reciprocally connected pairs (compared to a random Erdos-Renyi network), versus the robustness parameter ρ. Black curves show the result of the calculation for various coding levels f; black circles, numerical simulations using the perceptron learning algorithm ($N = 800$); orange line, experimentally observed excess. (**c**) Reciprocal connection probability–connection probability plane. The thin black lines show reciprocal connection probability in fully symmetric ($y = x$) and random asymmetric ($y = x^2$) networks. Black curves show the results of the calculation for various robustness levels, parameterized by ρ ($\rho = 0$ at the rightmost point of the curves; $\rho \to \infty$ at the origin), for different values of f (see **b**). Black circles, numerical simulations; red star, experimental data[11]. (**d**) Distribution of weights for bidirectionally connected pairs (red) and unidirectionally connected pairs (blue). Smooth curves, theory; histograms, experimental data.

local networks of pyramidal cells in neocortex could potentially be fully connected[2]. However, electrophysiological studies have consistently found connection probabilities of order 10% (refs. 5–10). This discrepancy between anatomy and electrophysiology is entirely consistent with the attractor model close to maximal capacity. Furthermore, published distributions of synaptic weights are also consistent with the theory. **Figure 2c** shows a comparison between the distribution of synaptic weights recorded by Sjöström and colleagues (paired recordings of layer 5 pyramidal neurons[7,11]) and the analytically calculated distribution, whose parameters are fixed in order to reproduce the observed connection probability ($c = 0.116$) and average synaptic weight. The theoretical distribution is in good agreement with the experimental one, except that it overestimates the fraction of very weak synapses (those with excitatory postsynaptic potential amplitudes smaller than 0.1 mV) and underestimates the fraction of very strong synapses (those with amplitudes larger than 3 mV). The overestimation of very weak synapses could be due to under-reporting of very weak synapses owing to noise[32]. Alternatively, very weak synapses might lack stability, leading to retraction of the corresponding spine. Likewise, various factors could explain the underestimation of strong weights: experimental artifacts, finite-size effects[33], supra-linear EPSP summation[32], heterogeneities in single neuron thresholds. Note also that distributions of synaptic weights far from capacity lack the large fraction of zero-weight synapses and therefore cannot reproduce well the experimentally observed distribution (see **Supplementary Fig. 1**).

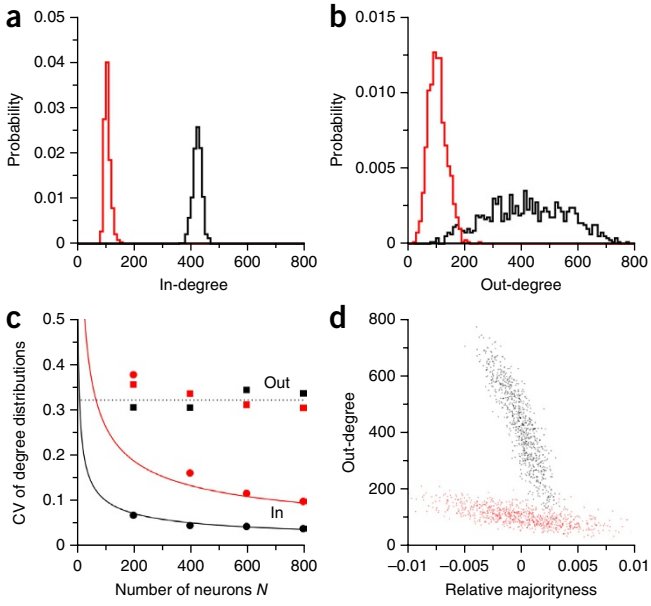

Figure 4 Distributions of degrees in optimal network ($N = 800$, $f = 0.5$, for two values of ρ, $\rho = 0$ (black) and $\rho = 4$ (red). (**a**) Distribution of in-degrees. (**b**) Distribution of out-degrees. (**c**) Coefficient of variation (CV; s.d. divided by the mean) of the distribution of degrees as a function of network size. The CV of in-degrees (circles) scales as $1/\sqrt{N}$ for large N, consistent with a random Erdos-Renyi graph (solid lines), while the CV of out-degrees (open squares) is independent of N (dashed line). (**d**) Anticorrelation between out-degree and majorityness, average number of active neurons in patterns in which a neuron is active. For $\rho = 0$, the Pearson correlation coefficient is $r = -0.87$ ($P < 2 \times 10^{-16}$), while for $\rho = 4$, $r = -0.69$ ($P < 2 \times 10^{-16}$).

Statistics of two-neuron bidirectional connectivity

We next turned to joint distributions of pairs of synaptic weights, again in networks optimizing storage capacity. We first considered the joint distribution of synaptic weights connecting two neurons. The calculation, using the cavity method (**Supplementary Note**), yields a correlated two-dimensional Gaussian distribution, truncated in the upper right quadrant (**Fig. 3a**). The intuition underlying this correlation is the following: in a given attractor, a subset of neurons needs to remain active. To allow such a state to be an attractor of the dynamics, synapses connecting neurons in this subset need to be strengthened. Thus, pairs of neurons belonging to this subset tend to develop reciprocal connections. This leads to an over-representation of reciprocally connected neurons, in agreement with cortical slice data[11,12,34] (but see ref. 10). This over-representation can be quantified by r, the ratio of the probability of finding reciprocally connected neurons divided by the probability of reciprocal connections in a random Erdos-Renyi graph[11]. At maximal capacity, this parameter is again strongly dependent on the robustness parameter (**Fig. 3b**). It increases both with the number of patterns stored in the connectivity matrix and with robustness of information storage. For the experimentally observed connection probability ($c = 0.116$), r is close to the

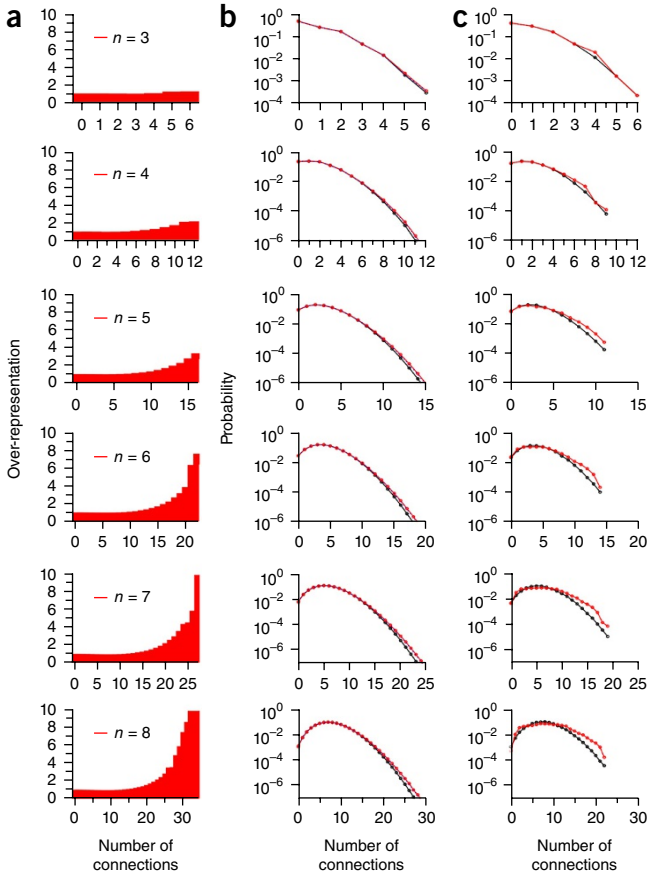

Figure 5 Higher order motifs: probabilities of observing k connections in n-neuron subnetworks, as a function of k, for different values of n ($3 \le n \le 8$; $N = 800$, $f = 0.5$, $\rho = 4$). (**a**) Probability of observing k connections, divided by the probability to observe k connections in a random network with identical pair statistics. (**b**) Observed probability (red) and expected probability in a random network with observed pair statistics (black) are plotted together in logarithmic scale. The figure also shows the probability of observing k connections in a random network with the same pair statistics and the same distribution of in- and out-degrees as observed in the network at maximal capacity (dotted blue). The almost perfect overlap between curves shows that the pattern of over-representation is largely due to the wide distribution of out-degrees. (**c**) Observed probability (red) and expected probability in a random network with observed pair statistics (black) in cortical slices (data from ref. 13).

experimentally observed value ($r \approx 4$) in a large range of coding levels f (**Fig. 3c**). For $c = 0.116$, r increases monotonically from $r = 3.46$ at $f = 0.5$ to $r = 4.84$ at $f = 0.01$. The experimentally observed value of r is obtained for a coding level of $f \approx 0.1$. **Figure 3c** shows the reciprocal connection probability r as a function of the connection probability c. For a wide range of parameters, optimal networks lie approximately halfway between a perfectly symmetric network ($r = c$) and a random asymmetric network ($r = c^2$). This leads to a 'symmetry degree' (correlation between the two weights connecting a pair of neurons) that is close to 0.5, similar to that which was found in a previous study in a network with unconstrained weights[35].

Song *et al.*[11] also reported that bidirectionally connected pairs tend to contain stronger weights than unidirectionally connected pairs. Networks at maximal capacity also reproduce quantitatively this feature (**Fig. 3d**). In our theory, distributions of synaptic weights for bidirectionally connected pairs, and for unidirectionally connected pairs, are entirely determined by three parameters: c, r and the average synaptic weight. When we fix these three parameters at their experimentally observed value, the two distributions reproduce well the data (**Fig. 3d**).

Degree distributions

We also analyzed other statistical properties of the connectivity matrix of optimal networks, using simulations with the perceptron learning rule (Online Methods). In particular, we found a striking difference between distributions of in- and out-degrees (numbers of input and output synapses per neuron; **Fig. 4**). While the distribution of in-degrees was not significantly different from that of a random Erdos-Renyi network (**Fig. 4a**), the distribution of out-degrees was much wider, with a s.d. that increases linearly with network size N, instead of the expected $\sqrt{N}$ dependence. Out-degrees were strongly anticorrelated with 'majorityness'. Majorityness of neuron i is defined as the average number of active neurons in patterns in which neuron i is active; that is,

$$\left(\sum_{\mu,j}\eta_i^\mu\eta_j^\mu\right)\Big/\left(fN\sum_\mu\eta_i^\mu\right).$$

This anticorrelation can be understood intuitively as follows. For neurons that are active in patterns that involve fewer neurons than average, synaptic weights need to be strengthened more than average to compensate for the smaller size of the sets of active neurons. This does not have a pronounced effect on incoming connections because of the constraint that the average incoming weights should be equal to the threshold divided by the mean number of active neurons in a pattern $T/(fN)$ (up to fluctuations of order $1/\sqrt{N}$; **Supplementary Note**), but it does strongly affect outgoing connections, which do not collectively obey such a constraint. This leads to the predictions that cortical neurons should have very broad out-degree but not in-degree distributions and, counterintuitively, that 'hub' neurons that project to the largest number of neurons should be 'minority' neurons, in the sense that they tend to be active in patterns involving smaller-than-average sets of active neurons.

Statistics of higher order motifs

We next considered higher order motifs. Perin *et al.*[13] considered subsets of n of neurons ($3 \le n \le 8$) and showed that the distribution of the total number of connections in such n-neuron networks departs from the distribution of a random network with the observed pair statistics. In particular, strongly connected subsets of n neurons are significantly more represented than in such random networks. In networks whose connectivity matrix is built using the perceptron rule, we find qualitatively the same trend in a network close to maximal capacity (**Fig. 5**), though the over-representation of highly connected subnetworks is smaller than in the data (**Fig. 5b**,**c**). This pattern of over-representation of highly connected subnetworks is largely due to the wide distribution of out-degrees (**Fig. 5b**).

Impact of inhibition on statistics of excitatory connectivity

We have so far focused on a situation in which inhibition is incorporated into the threshold of excitatory neurons. We now turn to a more realistic implementation where inhibition is modeled explicitly, using simulations in which the connectivity is obtained using the perceptron learning rule. We first start with a model in which inhibition provides a feedback onto the excitatory population that is proportional to the global activity of the excitatory population. For simplicity, we take this feedback to be linearly proportional to the

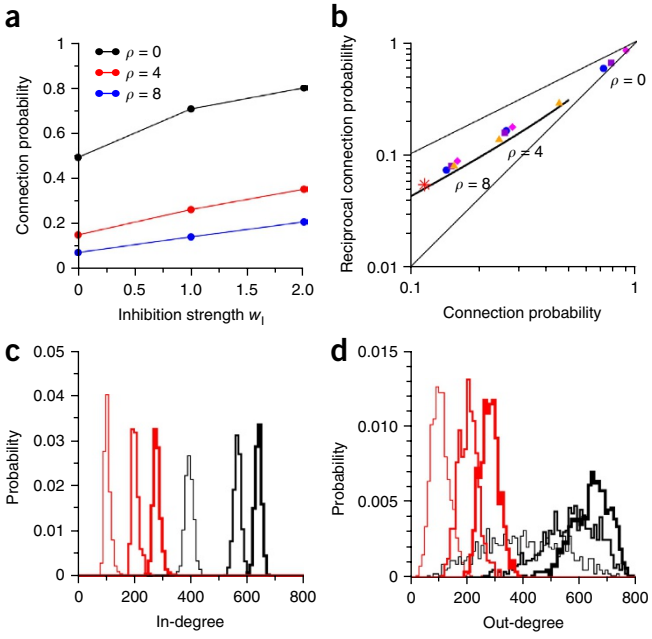

Figure 6 Effect of inhibition on statistics of connectivity. (**a**) Connection probability versus inhibition strength w_I, for three values of ρ, for $N = 800$ and $f = 0.5$, in the model with global linear inhibitory feedback. (**b**) Bidirectional connection probability versus connection probability, for the same three values of ρ, with $w_I = 1$, in different models with inhibition. Blue circles, model with global inhibitory feedback. All other symbols show models in which individual inhibitory neurons are modeled explicitly: for violet squares, all connections involving inhibition are fixed; for magenta diamonds, I→E connections are plastic and other connections are fixed; for orange triangles, all connections are plastic. Other curves and symbols as in **Figure 3**. (**c**) Effect of global feedback inhibition on in-degree distributions. Red curves correspond to $\rho = 4$, black to $\rho = 0$. For each color, three distributions are shown: $w_I = 0$ (thin), 1 (intermediate) and 2 (thick). (**d**) Effect of global feedback inhibition on out-degree distributions. Same parameters as in **c**.

global activity of the excitatory population and to be instantaneous. This leads to an effective connectivity between excitatory neurons that is given by $w_{ij} - w_I T/(fN)$, where w_I represents the strength of inhibitory feedback (Online Methods and ref. 36). This effective connectivity between neurons is not bounded by zero, but rather by a negative number proportional to $-w_I$. As a result, the connection probability is no longer bounded by above by 0.5, but can now be larger than 0.5 (**Fig. 6a**). For a fixed value of ρ, both connection probability and reciprocal connection probability are an increasing function of w_I. However, arbitrarily low values of the connection probability can still be obtained at arbitrary values of w_I, through an appropriate increase of ρ. In particular, connection and reciprocal connection probabilities close to those in experimental data can be achieved at any value of w_I by an appropriate increase of ρ (for example, for $w_I = 1$, ρ has to be increased from $\rho \approx 4$ to $\rho \approx 8$ to reproduce the data). The distributions of degrees are qualitatively unaffected by the presence of inhibition (**Fig. 6c**,**d**).

We also considered a model in which individual inhibitory neurons are modeled explicitly as binary units. We considered three variants of such a model: (i) a model in which all synapses involving inhibition (E→I, I→I, I→E) are random but fixed, (ii) a model in which I→E synapses are plastic but synapses onto inhibitory neurons are fixed and random, and (iii) a model in which all synapses are plastic. Synapses that were fixed were drawn randomly according to a uniform distribution between zero and twice the average value. This led to non-sparse connectivity for those synapses, consistent with *in vitro* studies that show close to full connectivity between some types of interneurons and pyramidal cells[27]. In models (i) and (ii), the pattern of activity in the inhibitory subnetwork was determined from the pattern of activity in the excitatory subnetwork and the input connectivity of interneurons; in model (iii), the pattern of activity in the inhibitory subnetwork was random, with the same coding level as the excitatory subnetwork. In all cases, we used the perceptron learning algorithm to learn a set of connectivity matrices that satisfied all constraints imposed by the patterns. Models (i) and (ii) produced high connection probabilities at low values of the robustness parameter ρ (**Fig. 6b**), as in the simpler model with linear inhibitory feedback. Model (iii), by contrast, was similar to the purely excitatory network: it produced excitatory connection probabilities smaller than 0.5 for all values of parameters, while the inhibitory connection probabilities were higher than 0.5, consistent with ref. 33. In all models, E→E connection probabilities decreased as ρ increased, and the over-representation of bidirectionally connected pairs was similar in all models (**Fig. 6b**).

Statistics of connectivity in networks storing sequences

One might wonder whether the observed statistical properties of optimal synaptic matrices are specific to the format of the stored information. We therefore considered an alternative model in which information is stored in the form of sequences of activity: if the network is at some point in time in a given pattern η^μ, then it has to recall the next pattern in the stored sequence $\eta^{\mu+1}$ at the next time step. Successive patterns in the sequence were again taken to be independent for simplicity. We found that the distribution of synaptic weights for such a network is identical to the distribution in networks storing fixed-point attractors (**Supplementary Note**). By contrast, the joint distribution of weights of pairs of cells in such a network, computed using the cavity method, is drastically different from the one in a network storing fixed-point attractors: it simply factors into a product of distributions of single weights. In other words, there are no

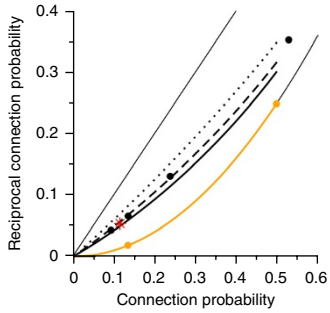

Figure 7 Reciprocal connection probability as a function of connection probability: Fixed-point attractors versus sequences of activity. The black curves show the analytical calculations for networks storing information as fixed-point attractors, as in **Figure 3** for different coding levels (same conventions as in **Fig. 3**: full line, $f = 0.5$; dashed line, $f = 0.1$; dotted line, $f = 0.01$); The orange curve shows the calculation for networks storing sequences of activity, for which the reciprocal connection probability is equal to the square of the connection probability, for any value of the coding level f. Filled circles indicate the results of numerical simulations using the perceptron algorithm (black: fixed point attractors, $\rho = 0$, 2, 4, 6; orange: sequences, $\rho = 0$, 4).

correlations between weights in a pair of neurons, and no over-representations of reciprocally connected pairs of neurons (**Fig. 7**).

The intuitive reason why the two synaptic weights of a pair of neurons are uncorrelated in a network storing sequences of activity is that the two synaptic weights are driven by independent sources: for the transition from pattern μ to pattern $\mu + 1$, the connection from neuron i to neuron j has to associate η_i^μ with $\eta_j^{\mu+1}$, while the connection from j to i has to associate η_j^μ to $\eta_i^{\mu+1}$, which are uncorrelated with the values of η seen by the connection in the opposite direction.

DISCUSSION

We have analyzed the statistics of connectivity in a large network of neurons connected by a plastic excitatory connectivity matrix that maximizes information storage. We considered two scenarios: one in which the network is optimized for information storage in the form of fixed point attractors and another in which it is optimized for storage of sequences of activity. In both types of networks, we found that optimal information storage leads to a large fraction of zero-weight synapses, which can be interpreted either as silent or potential synapses. The resulting connection probability (probability that a synapse has a strictly positive weight) is necessarily smaller or equal than 0.5 in networks with fixed inhibition, but can be larger in models in which inhibitory neurons are modeled explicitly. In all cases, the connection probability decreases monotonically as the robustness level (size of the basin of attraction of stored patterns) increases. We also find in both types of networks a broad distribution of out-degrees, whose width scales with network size N, rather than the expected $\sqrt{N}$. A major difference between fixed point attractor and sequence scenarios is the joint distribution of the two synaptic weights connecting a pair of neurons: in the fixed point attractor scenario, we find a strong over-representation of reciprocally connected pairs, while no such over-representation is found in the sequence scenario.

Some of our results are in notable agreement with published data on the statistics of connectivity of various cortical areas. Virtually all electrophysiological studies *in vitro*[5–7,10–12,14] find connection probabilities of order 0.1–0.2 for pairs of nearby pyramidal cells. This is consistent with networks with a highly robust storage of information. Furthermore, the distribution of synaptic weight at maximal capacity is in good agreement with data, once the empirical connection probability and mean synaptic weights are fixed (**Fig. 2c**). Published data show a diversity of degrees of over-representation of reciprocal connections, from $r \approx 4$ in prefrontal cortex[12] and visual cortex[11], $r \approx 3$ in somatosensory cortex[6], $r \approx 2$ in visual cortex[12] and $r \approx 1$ in barrel cortex[10]. $r \approx 4$ is close to the value that is predicted for a network optimizing storage of fixed point attractors, given the empirical connection probability, for a wide range of coding levels f. Conversely, $r \approx 1$ is the value that is predicted for a network optimizing storage of sequences. Our model also reproduces quantitatively the distribution of synaptic weights for both bidirectionally and unidirectionally connected pairs[11]. Finally, we found that the statistics of higher order motifs was in qualitative agreement with the data of refs. 11,13. Note that the results from our model are also consistent with data from Mrsic-Flogel and colleagues, who have documented that pairs of neurons that share the same selectivity properties have an greater connection probability than pairs that do not[14], and with other data showing that excitatory cortical neurons form fine-scale functional networks[37]. Overall, a surprisingly large amount of experimental data on short-range synaptic excitatory connectivity in cortex can be reproduced by a network designed to store highly robust fixed-point attractors close to its maximal storage capacity. This scenario fits well with increasing amount of experimental evidence for attractor dynamics in various cortical areas: auditory cortex[38], areas of the temporal lobe[19,39,40], and parietal and prefrontal cortices[18,20,21]. Interestingly, the data of ref. 12 indicate a larger proportion of reciprocally connected pairs of neurons in prefrontal cortex, the area most associated with persistent activity and attractor dynamics, than in visual cortex. The data of ref. 10, recorded in barrel cortex, do not show any over-representation of reciprocally connected pairs, consistent with the fact that barrel cortex experiences highly dynamic patterns of activity *in vivo*.

There are a number of caveats that need to be considered when comparing our model to experimental data. First, space is ignored in our model, whereas connection probability decays with distance in cortical networks[13]. However, the data of ref. 11, which are in quantitative agreement with our theory, take into consideration only pairs separated by short distances (typically less than 100 μm), for which connection probability has been shown explicitly not to depend appreciably on distance[11]. For data sets in which connection probability depends appreciably on distance[13], we expect spatial effects to lead to an increase in the over-representation of highly connected subsets of neurons. This could potentially explain the quantitative differences we find between our simulations and the data of ref. 13. Second, our model ignores functional subdivisions between pyramidal cells, which might affect the statistics of their connectivity. For example, it has been suggested that 'sister' pyramidal cells arising from the same mother cell in the developing neocortex have a larger than average connection probability[41]. However, the possibility cannot be excluded that this preferential connectivity is developed through learning in a network in which sister cells share similar selectivity properties, which would lead to a scenario consistent with the model presented here.

The analysis presented here suggest that a distribution of weights in which the majority are zero is a universal feature of neural systems close to their storage capacity, since such a distribution has been shown to be optimal in all neural architectures and models considered to date[31–33,42]. By contrast, higher order statistics turn out to be specific to the format of stored information. Storage of fixed-point attractors leads to connectivity matrices with a strong degree of symmetry, while storage of sequences leads to asymmetric connectivity matrices. Therefore, our theory predicts a strong link between the joint statistics of connections in a pair of neurons and the spatio-temporal statistics of the information stored by the network. A similar conclusion was reached by Clopath *et al.*[43], who investigated how a specific synaptic plasticity rule based on presynaptic spike trains and filtered post-synaptic membrane potentials sculpts the connectivity of a recurrent network.

There are a number of testable predictions that follow from our results. (i) The analysis predicts a wide distribution of out-degrees. A definitive test of this prediction will become possible when cortical circuits at the scale of the local cortical connectivity are fully reconstructed using serial scanning electron microscopy techniques[44]. (ii) It predicts that out-degrees are anticorrelated with the majorityness of neurons. This would be the hardest prediction to test, since it would involve measuring a large number of patterns of activity *in vivo* and correlating majorityness with out-degrees, presumably measured *in vitro* or using dense reconstruction. (iii) It predicts that in diseased states in which information storage is impaired, statistics of connectivity should depart from the ones derived here. In particular, the over-representation of reciprocal connections in pairs of neurons should decrease in such conditions. Finally, specific predictions about statistics of connectivity depend on specific assumptions on the time scale of the learning process that allows cortical networks

to reach their optimal point. The most straightforward assumption would be that during development both the number of stored patterns of activity and their robustness increase progressively. This scenario would predict that connection probability decreases with development, consistent with synaptic pruning observed in the late stage of development[45]. It would also predict that the over-representation of bidirectional connections grows progressively with development, a prediction that could be tested by comparing data at different stages of development.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

I warmly thank J. Sjöström for providing his data, and Y. Amit, C. Baldassi, B. Barbour, V. Hakim, D. Marti, G. Mongillo, S. Ostojic, A. Roxin, J. Sjöström and R. Zecchina for discussions and/or comments on previous versions of the manuscript.

COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

1. Braitenberg, V. & Schütz, A. *Anatomy of the Cortex* (Springer, 1991).
2. Kalisman, N., Silberberg, G. & Markram, H. The neocortical microcircuit as a tabula rasa. *Proc. Natl. Acad. Sci. USA* **102**, 880–885 (2005).
3. Trachtenberg, J.T. *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
4. Stepanyants, A., Hof, P.R. & Chklovskii, D.B. Geometry and structural plasticity of synaptic connectivity. *Neuron* **34**, 275–288 (2002).
5. Mason, A., Nicoll, A. & Stratford, K. Synaptic transmission between individual pyramidal neurons of the rat visual cortex *in vitro*. *J. Neurosci.* **11**, 72–84 (1991).
6. Markram, H., Lübke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* **500**, 409–440 (1997).
7. Sjöström, P.J., Turrigiano, G.G. & Nelson, S.B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
8. Holmgren, C., Harkany, T., Svennenfors, B. & Zilberter, Y. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol. (Lond.)* **551**, 139–153 (2003).
9. Thomson, A.M. & Lamy, C. Functional maps of neocortical local circuitry. *Front. Neurosci.* **1**, 19–42 (2007).
10. Lefort, S., Tomm, C., Floyd Sarria, J.C. & Petersen, C.C. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009).
11. Song, S., Sjöström, P.J., Reigl, M., Nelson, S. & Chklovskii, D.B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
12. Wang, Y. *et al.* Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.* **9**, 534–542 (2006).
13. Perin, R., Berger, T.K. & Markram, H. A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. USA* **108**, 5419–5424 (2011).
14. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).
15. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
16. Amit, D.J. The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav. Brain Sci.* **18**, 617 (1995).
17. Amit, D.J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**, 237–252 (1997).
18. Fuster, J.M. & Alexander, G.E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
19. Miyashita, Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* **335**, 817–820 (1988).
20. Funahashi, S., Bruce, C.J. & Goldman-Rakic, P.S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
21. Romo, R., Brody, C.D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
22. Abeles, M. *Corticonics* (Cambridge Univ. Press, 1991).
23. Goldman, M.S. Memory without feedback in a neural network. *Neuron* **61**, 621–634 (2009).
24. Harvey, C.D., Coen, P. & Tank, D.W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
25. Pfeiffer, B.E. & Foster, D.J. Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science* **349**, 180–183 (2015).
26. Gardner, E.J. The phase space of interactions in neural network models. *J. Phys. Math. Gen.* **21**, 257–270 (1988).
27. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).
28. Hofer, S.B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**, 1045–1052 (2011).
29. Mézard, M., Parisi, G. & Virasoro, M.A. *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
30. Rosenblatt, F. *Principles of Neurodynamics* (Spartan, New York, 1962).
31. Clopath, C., Nadal, J.P. & Brunel, N. Storage of correlated patterns in standard and bistable Purkinje cell models. *PLoS Comput. Biol.* **8**, e1002448 (2012).
32. Brunel, N., Hakim, V., Isope, P., Nadal, J.P. & Barbour, B. Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron* **43**, 745–757 (2004).
33. Chapeton, J., Fares, T., LaSota, D. & Stepanyants, A. Efficient associative memory storage in cortical circuits of inhibitory and excitatory neurons. *Proc. Natl. Acad. Sci. USA* **109**, E3614–E3622 (2012).
34. Markram, H. A network of tufted layer 5 pyramidal neurons. *Cereb. Cortex* **7**, 523–533 (1997).
35. Gardner, E.J., Gutfreund, H. & Yekutieli, I. The phase space of interactions in neural network models with definite symmetry. *J. Phys. Math. Gen.* **22**, 1995–2008 (1989).
36. Alemi, A., Baldassi, C., Brunel, N. & Zecchina, R. A three-threshold learning rule approaches the maximal capacity of recurrent neural networks. *PLoS Comput. Biol.* **11**, e1004439 (2015).
37. Yoshimura, Y., Dantzker, J.L.M. & Callaway, E.M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
38. Bathellier, B., Ushakova, L. & Rumpel, S. Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* **76**, 435–449 (2012).
39. Fuster, J.M. & Jervey, J.P. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* **212**, 952–955 (1981).
40. Nakamura, K. & Kubota, K. Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J. Neurophysiol.* **74**, 162–178 (1995).
41. Yu, Y.C., Bultje, R.S., Wang, X. & Shi, S.H. Specific synapses develop preferentially among sister excitatory neurons in the neocortex. *Nature* **458**, 501–504 (2009).
42. Clopath, C. & Brunel, N. Optimal properties of analog perceptrons with excitatory weights. *PLoS Comput. Biol.* **9**, e1002919 (2013).
43. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* **13**, 344–352 (2010).
44. Kasthuri, N. *et al.* Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
45. Bourgeois, J.P. & Rakic, P. Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *J. Neurosci.* **13**, 2801–2820 (1993).

ONLINE METHODS

Network storing fixed-point attractors. We consider a fully connected network of N neurons, connected through an excitatory synaptic weight matrix $w_{ij} \geq 0$ for all i,j. The network has to learn to stabilize p random patterns, $\{\eta_i^\mu\}$, for $\mu = 1, \ldots, p$, with robustness κ, where η_i^μ are random independent binary variables, such that $\eta_i^\mu = 1$ with probability f and $\eta_i^\mu = 0$ with probability $1 - f$, for all i,μ, where f is the coding level ($0 < f \leq 0.5$). This leads to the following constraints:

- when neuron i should be active in pattern μ (that is, $\eta_i^\mu = 1$):

$$\sum_j w_{ij}\eta_j^\mu > T + K \tag{1}$$

- when neuron i should be inactive in pattern μ (that is, $\eta_i^\mu = 0$):

$$\sum_j w_{ij}\eta_j^\mu < T - K \tag{2}$$

where T is the threshold and K measures the size of the basins of attraction of the patterns.

Networks storing sequences. We consider sequences of p random patterns, $\{\eta_i^\mu\}$, for $\mu = 1, \ldots, p$, with robustness κ, where η_i^μ are again random independent binary variables, such that $\eta_i^\mu = 1$ with probability f and $\eta_i^\mu = 0$ with probability $1 - f$ for all i,μ, where f is the coding level ($0 < f \leq 0.5$). The network should learn to retrieve robustly the whole sequence, starting from the first pattern. This leads to the following constraints:

- when neuron i should be active in pattern $\mu + 1$ (that is, $\eta_i^{\mu+1} = 1$):

$$\sum_j w_{ij}\eta_j^\mu > T + K \tag{3}$$

- when neuron i should be inactive in pattern $\mu + 1$ (that is, $\eta_i^{\mu+1} = 0$):

$$\sum_j w_{ij}\eta_j^\mu < T - K \tag{4}$$

Models with inhibition. *Model with global linear feedback inhibition.* In the model with instantaneous feedback inhibition, we take the inhibitory feedback to be instantaneous and linearly proportional to the average activity in the excitatory network. The total inputs to neuron i become

$$\sum_j w_{ij}S_j(t) - \frac{w_I T}{fN}\sum_j S_j(t) \tag{5}$$

where w_I measures the strength of inhibition, such that when the network is in a fixed point attractor the average inhibitory input is $w_I T$. For example, $w_I = 1$ means that the average inhibitory input is equal to the threshold. This means that the excitatory inputs have to be equal (on average) to twice the threshold for the neuron to be active.

Equation (5) is equivalent to a model in which the connectivity matrix is $w_{ij} - w_I T/(fN)$. In such a model, synaptic weights are no longer bounded by 0, but can become negative and are bounded by below by $-w_I T/(fN)$ (ref. 36).

Models with individual inhibitory interneurons. We also studied a model in which $N_I = 0.25N$ individual binary inhibitory interneurons are coupled to the N excitatory neurons. The states of the inhibitory neurons are denoted $S_i^I(t)$. Connectivity matrices of E→I neurons are denoted w_{ij}^{IE}. Inhibitory neurons are connected together through I→I connections, which are denoted w_{ij}^{II}. They project back onto excitatory neurons with synaptic connections w_{ij}^{EI}. The dynamics of the network obey the equations

$$S_i(t+1) = \Theta\left(\sum_j w_{ij}S_j(t) - \sum_j w_{ij}^{EI}S_j^I(t) - T\right) \tag{6}$$

$$S_i^I(t+1) = \Theta\left(\sum_j w_{ij}^{IE}S_j(t) - \sum_j w_{ij}^{II}S_i^I(t) - T\right) \tag{7}$$

where Θ is the Heaviside function. We considered three variants of such a model, depending on which connections involving interneurons are plastic.

(i) Model with fixed inhibitory connectivity. In this model, all connections involving inhibition were drawn randomly and independently from a uniform distribution from zero to twice the mean of the distribution. The means of the distributions were taken to be $(1 + w_I)T/(fN)$ for E→I connections, $w_I T/(f_I N_I)$ for I→I connections, and $w_I T/(f_I N_I)$ for I→E connections, where we used $f_I = 0.5$. This choice ensured that the mean total inputs to inhibitory neurons were equal to the threshold when the excitatory subnetwork was in a state corresponding to one of the patterns to be stored, with coding level f. For each of the patterns to be stored, we then ran the dynamics of the inhibitory subnetwork with fixed excitatory inputs until it reached a fixed point. This fixed point had on average half the inhibitory neurons active. It determined the inhibitory inputs to the excitatory neurons, which was maintained fixed during the learning process that involved only excitatory to excitatory weights.

(ii) Model with plastic I→E connections but fixed connections to inhibitory neurons. In this model, all connections involving inhibition were initialized randomly as in model (i). The state of the inhibitory subnetwork was also determined as in model (i) and remained fixed during the learning process. However, during the learning process both E→E and I→E weights were allowed to change. We also imposed a constraint that the mean excitatory and inhibitory weights remain constant during the learning process, to avoid unrealistic growth of both types of weights that typically occurs in models in which both excitatory and inhibitory plastic weights obey a perceptron learning rule[33].

(iii) Model in which all connections are plastic. In this model, the state of the inhibitory network in each pattern was drawn randomly. Each inhibitory neuron was active with probability $f_I = 0.5$ in all patterns and inactive with probability $1 - f_I$. During the learning process, all connections were allowed to change in order to satisfy the constraints imposed by the patterns, in both the excitatory and the inhibitory subnetworks. We again imposed a constraint that the mean weights do not change during the learning process.

Analytical calculations. Details of the calculations can be found in the **Supplementary Note**.

Experimental data. We use data recorded by J. Sjöström in cortical slices using quadruple recordings, which has already been analyzed in detail[11]. It contains 997 nonzero synaptic connections, out of 8,596 possible connections (connection probability 0.116). The histogram of synaptic weights is shown in **Figure 2c** (gray histogram). The histograms of synaptic weights of bidirectionally connected pairs, as well as unidirectionally connected pairs, are shown in **Figure 3c**.

Theory versus experiment. To compare experimental and theoretical distributions, we considered only the part of the distribution above 0.1 mV (owing to possible under-reporting of weak connections due to noise). The theoretical distribution is then uniquely determined by two numbers extracted from data: the connection probability gives the parameter B in equation (28) of the **Supplementary Note** and the mean synaptic weight gives w_s through equation (29) of the **Supplementary Note**. Comparison of the theoretical curve with the data is shown in **Figure 2c**.

To obtain the distributions of bidirectionally/unidirectionally connected pairs, we also need a third parameter, λ. This parameter depends on the coding level f. We chose this parameter to give the observed over-representation of bidirectionally connected pairs r. The value of f that gives $r = 4$ is $f \approx 0.1$. The comparison with the data is shown in **Figure 3c**.

Numerical simulations. We applied the perceptron learning algorithm, modified to take into account the fact that synapses are sign-constrained[31]. In the simulation, both synapses and thresholds were integer-valued. The initial value of the threshold was N, while synapses were initially uniformly distributed between 0 and twice the average value predicted by the theory. Patterns were generated randomly one after the other. After each pattern generation, we used the perceptron learning algorithm to modify the synaptic weights until the patterns were learned correctly by all neurons in the network. Modifications were only made if equations (1) and (2) were not satisfied. In that event, active synapses were updated by +1 if the neuron was supposed to be active in that pattern or by −1 if the neuron was supposed to be silent in that pattern. If all patterns

doi:10.1038/nn.4286

could be learned correctly after a number of iterations smaller than 4,000 per patterns, a new pattern was generated and added to the set. If not, for all neurons that could not learn all patterns, we multiplied all synapses and the threshold by a factor of 2 and continued the learning process. We continued this iteration until the threshold was equal to 4,096*N*. For each neuron, we defined the maximal capacity as the size of the largest set that could be learned using this procedure. **Supplementary Figure 2** shows the fraction of neurons that stored correctly all patterns as a function of the number of patterns, for different values of N and two values of ρ, with $f = 0.5$. The comparison with the analytical prediction for the storage capacity (1 for $\rho = 0$, 0.14 for $\rho = 4$) shows that the algorithm comes very close to the theoretical storage capacity.

To analyze the connectivity matrix at maximal storage capacity, we used for each neuron the synaptic weights obtained at its maximal capacity. We computed the histogram of synaptic weights and compared them with the analytical prediction (**Fig. 2a**). The connectivity matrix was composed of a majority of zero or extremely low weights, and a minority of large weights. We binarized the connectivity matrix $w_{ij}^{\mathrm{b}} = \Theta\left(w_{ij} - w_{\mathrm{c}}\right)$, where w_{c} was taken to be 0.1 times the average weight. We then analyzed the statistics of the binarized matrix: connection probability (**Fig. 2b**), probabilities of bidirectional connections (**Fig. 3b**,**c**), distributions of degrees (**Fig. 4**) and probabilities of higher order motifs (**Fig. 5**). We also analyzed finite-size effects, plotting observed probabilities as a function of $1/N$ (**Supplementary Fig. 3**). Correlation coefficients and associated P-values reported in **Figure 4** were computed using R.

Code availability. The code for solving the equations presented in the **Supplementary Note** and the code for simulations with the perceptron algorithm were written in C; both are available on request.

A **Supplementary Methods Checklist** is available.