nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-018-0226-x

Inhibitory connectivity defines the realm of excitatory plasticity

Gianluigi Mongillo[1,2]*, Simon Rumpel[3] and Yonatan Loewenstein[4]*

Recent experiments demonstrate substantial volatility of excitatory connectivity in the absence of any learning. This challenges the hypothesis that stable synaptic connections are necessary for long-term maintenance of acquired information. Here we measure ongoing synaptic volatility and use theoretical modeling to study its consequences on cortical dynamics. We show that in the balanced cortex, patterns of neural activity are primarily determined by inhibitory connectivity, despite the fact that most synapses and neurons are excitatory. Similarly, we show that the inhibitory network is more effective in storing memory patterns than the excitatory one. As a result, network activity is robust to ongoing volatility of excitatory synapses, as long as this volatility does not disrupt the balance between excitation and inhibition. We thus hypothesize that inhibitory connectivity, rather than excitatory, controls the maintenance and loss of information over long periods of time in the volatile cortex.

Experiments in recent years have provided a direct link between synaptic changes and memories: memory formation has been shown to be correlated with a transient increase in the density of spines (a proxy for synapse formation[1–4]). Moreover, the specific erasure of the spines formed during training results in a specific deletion of the corresponding memory[5]. If the patterns of excitatory synaptic connections are the physical correlate of long-term memories in the neocortex[6] then the lifetime of stored memories is expected to be directly tied to the lifetime of the underlying synaptic changes. Puzzling, therefore, is the observation that in the absence of explicit learning, excitatory connectivity is highly dynamic[7–11]. In fact, the rate of learning-driven spine formation and elimination is not much higher than the 'spontaneous' rate observed during basal conditions[6,12,13]. Cortical volatility manifests not only in the high rate of spine turnover but also in changes in the sizes of the spines[8,14], indicative of changes in the efficacies of the corresponding excitatory synapses[15].

These observations raise a fundamental question. How can memories be maintained in the presence of the substantial ongoing excitatory volatility? To address this question, we quantify the remodeling of cortical excitatory connectivity and use a cortical network model to study the effect of this remodeling on the patterns of firing rates. We show that these activity patterns are primarily determined by the inhibitory connectivity, even though the majority of neurons and synapses in the model are excitatory[16]. We also show that the inhibitory network can store many more memories than its excitatory counterpart, despite being endowed with a smaller number of synapses. These two results are a direct consequence of the differences in the firing-rate distributions of excitatory and inhibitory neurons[17]. Finally, we show that, in contrast to spontaneous remodeling, selective learning-like excitatory plasticity[2–5] has a substantial effect on the pattern of network activity by transiently disrupting the balance between excitation and inhibition. Taken together, our results imply that excitation and inhibition play fundamentally different roles in controlling the stability of activity patterns in the cortex. The inhibitory network, rather than providing 'blanket inhibition'[18], has the potential to control the stability of memory patterns for long periods of time.

Results

The volatility of spines and network activity. We imaged 3,688 dendritic spines of eight layer V pyramidal neurons in the mouse auditory cortex in six imaging sessions at an interval of 4 d[9,15] (see Methods). Individual dendritic spines exhibited substantial volatility. Less than half of the spines present in the first imaging session were still present in the last imaging session[9]. Moreover, even the 'stable' spines, those present in all imaging sessions, were highly volatile, such that over the course of 20 d, >70% of the spines changed their size by at least 50%[15]. Nevertheless, density and size distributions of the spines remained stationary across sessions. Previous studies have reported a wide range of turnover rates, and the rates we measured are at the higher end. Differences between rates have been attributed to cortical region, age of the animal, and differences in imaging techniques and analysis[3,12,19,20].

We constructed a biologically constrained model of a cortical area composed of 80% excitatory and 20% inhibitory randomly and sparsely connected spiking neurons (Fig. 1a and Methods). In addition to these connections, excitatory (E) and inhibitory (I) neurons also received feedforward input, constant in time and identical for all neurons in the population (E and I). Each recurrent excitatory to excitatory (E → E) connection in the network model is associated with a single spine measured in the first imaging session, and we use its size as a measure of the efficacy of the corresponding synapse (see Methods)[20]. The efficacies of the remaining synapses E → I, I → E, and I → I are drawn from log-normal distributions whose parameters are taken from cortical measurements[21] (Supplementary Table 1).

The model captures basic features of cortical dynamics. First, it operates in the asynchronous regime, in which spiking is temporally irregular (average coefficients of variation for the excitatory and inhibitory neurons are 1.07 and 1.22, respectively; Fig. 1b). Second, the distributions of firing rates of the excitatory and inhibitory

[1]Centre National de la Recherche Scientifique (CNRS), Paris, France. [2]Centre de Neurophysique, Physiologie et Pathologie (CNPP), Université Descartes, Paris, France. [3]Institute of Physiology, Focus Program Translational Neuroscience, University Medical Center, Johannes Gutenberg University, Mainz, Germany. [4]Department of Neurobiology, the Federmann Center for the Study of Rationality and the Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, Jerusalem, Israel. *e-mail: gianluigi.mongillo@univ-paris5.fr; yonatan@huji.ac.il

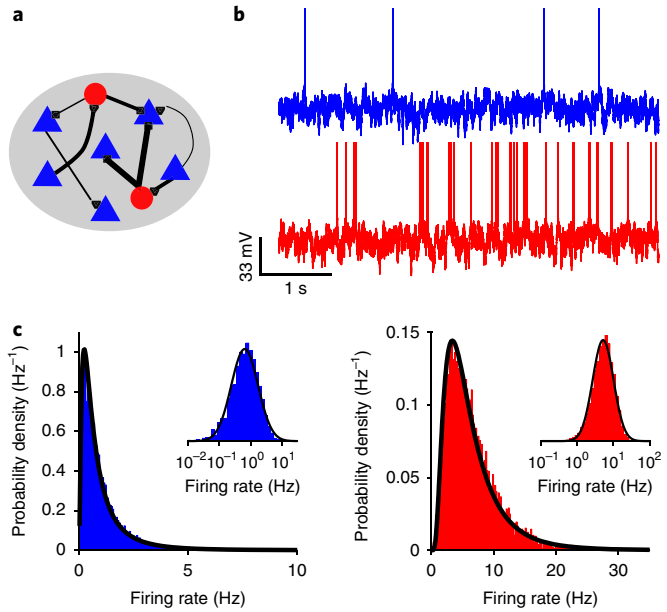

Fig. 1 | The spiking network model. a, A schematic drawing of the network of sparsely connected excitatory (blue triangles) and inhibitory (red circles) integrate-and-fire neurons. **b**, Membrane potential traces of randomly selected excitatory (blue) and inhibitory (red) neurons. **c**, The distributions of firing rates of the excitatory (blue) and inhibitory (red) neurons (estimated from a simulation of 3 min) in linear and logarithmic (inset) scales. Black lines are log-normal functions fitted to the (logarithmic) histograms (parameters of the fit: excitatory: mean, 1.01 Hz; variance, 1.03 Hz^2; inhibitory: mean, 6.41 Hz; variance, 16.27 Hz^2).

neurons are approximately log-normal (Fig. 1c), with means and variances that are comparable with those measured in the cortex[17,22]. These two features of the activity are a direct result of the fact that in such neuronal circuits, in which neurons are recurrently connected via strong synapses, the firing rates of the excitatory and inhibitory populations adjust dynamically and the time-averaged input to the neurons is subthreshold[23,24]. In this balanced excitation–inhibition regime, spiking is driven by temporal fluctuations of synaptic inputs, which results in Poisson-like timing of action potentials. The heterogeneity in the firing rates of the otherwise identical neurons results from the heterogeneity in their random synaptic inputs. Because the number of presynaptic inputs is large and these inputs are only weakly correlated, the distributions of time-averaged inputs onto the excitatory and inhibitory neurons are approximately normal. As in a noise-driven escape process, the firing rate of a neuron exponentially depends on its time-averaged input. This, combined with the normal distribution of inputs, results in a log-normal distribution of firing rates[25].

To study the consequences of the experimentally observed volatility of E → E connections, we generate six networks corresponding to the six consecutive imaging sessions, such that each E → E synapse in a network is matched to a single spine in the corresponding imaging session (Fig. 2a). Each formation, elimination, or change in the size of a spine manifest in the model as the formation, elimination, or change in the efficacy of the corresponding synapse (Methods).

Because the density and the size distribution of the spines were similar across sessions, the distributions of firing rates in the six networks are almost identical (Fig. 2b). We expected that the considerable changes in E → E connectivity (illustrated for ten randomly chosen pairs of excitatory neurons in Fig. 2a) would result in substantial changes in the firing rates of the individual neurons. This is because the specific firing rates of neurons reflect the particular realization of their synaptic inputs, and E → E connections constitute half of the synapses in our model. To our surprise, the similarities between the firing patterns are remarkably high, as demonstrated by the similarity of the firing rates of specific neurons in the six networks (Fig. 2b). The stability of network activity is quantified in Fig. 2c, where we plot the autocorrelogram of the matrices of E → E connectivity, together with the autocorrelograms of the vectors of firing rates of the excitatory and inhibitory neurons.

Sensitivity of network activity to various types of synaptic remodeling. The experimentally observed volatility did not correspond to a complete rewiring of the E → E connectivity. We wondered whether the remaining correlations in the E → E connectivity were sufficient to support the stability of the firing pattern. To test this, we remove all these residual correlations by generating a new E → E connectivity matrix with the same statistical properties. Unexpectedly, we find that even complete rewiring of the E → E connectivity has only a modest effect on the pattern of network activity (Fig. 3).

A complete rewiring of the network will necessarily result in vanishing correlation coefficients of the firing rates. Therefore, E → I, I → E, and/or I → I rewiring play a disproportionally large role in determining the firing pattern of the network. Indeed, we find that while the effect of rewiring of the E → I synapses is also modest (Fig. 3), the pattern of activity is much more sensitive to changes in the inhibitory synapses (Fig. 3). Particularly remarkable is its sensitivity to the rewiring of I → I connections, despite the fact that they constitute only 6% of the total number of connections in the network model.

To gain insight into the striking differential sensitivity of the pattern of firing rates to excitatory and inhibitory rewiring, we consider the network from the perspective of a single postsynaptic excitatory neuron, innervated by presynaptic excitatory and inhibitory neurons (Fig. 4a). Consider two synapses in this feedforward network. Trivially, swapping their efficacies will not change the excitatory input to the postsynaptic neuron if the two synapses have identical efficacies. Similarly, the swapping of their efficacies, even if they are different, will not affect the average input as long as the firing rates of the two corresponding presynaptic neurons are equal. In other words, the more different the efficacies of the two synapses are and the more different the firing rates of the corresponding presynaptic neurons are, the larger the expected effect of synaptic swapping on the postsynaptic firing rate is. Now we generalize this intuition to a complete rewiring of E → E connections, which in a large network is asymptotically equivalent to the random permutation of all E → E connections. The larger the variance of the distribution of E → E synaptic weights and the larger the variance of the distribution of firing rates of excitatory presynaptic neurons, the more sensitive the firing rate of the postsynaptic neuron to E → E synaptic rewiring is. A similar argument can be made when considering I → E rewiring. A more precise analysis reveals that in the sparse network, the relevant parameters determining the effect of synaptic rewiring are the number of connections, the mean of the distribution of squared synaptic efficacies (sum of the squared mean and variance of the distribution), and the mean of the distribution of squared firing rates (see Methods).

Considering cortical parameters[21], the distributions of excitatory and inhibitory efficacies onto the excitatory neurons (E → E, I → E) are comparable (Supplementary Table 1). By contrast, the mean of the distribution of squared firing rates of excitatory neurons is substantially smaller than that of the inhibitory neurons (Fig. 1c and see Methods). As a result, the effect of excitatory rewiring is smaller than that of inhibitory rewiring. In Fig. 4a we plot the distribution of change in the time-averaged excitatory and inhibitory inputs to the postsynaptic excitatory neuron following the rewiring of excitatory and inhibitory synaptic connectivity. The fact that the

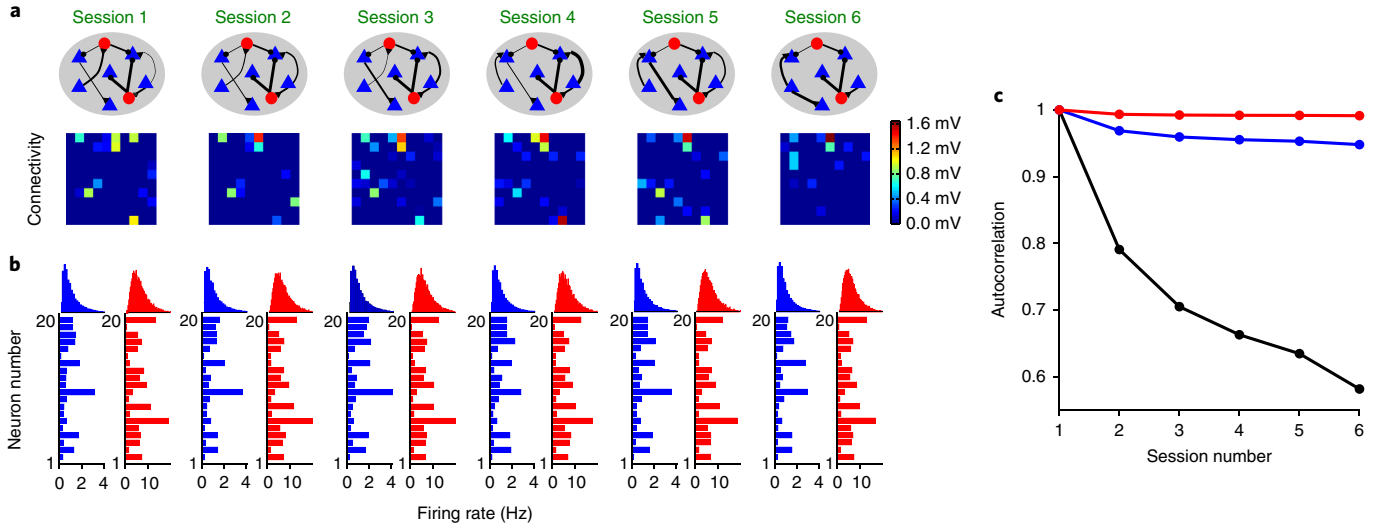

Fig. 2 | Volatility in the E → E connectivity has little effect on network activity. a, Top: schematic illustration of six networks, each corresponding to the cortical network in one of the imaging sessions. Bottom: the connectivity matrix between ten randomly selected excitatory neurons. Color denotes the synaptic efficacy (deep blue, unconnected pair). **b**, Top: distributions of firing rates of the excitatory (blue) and inhibitory (red) neurons (session 1 is same as Fig. 1c). Bottom: the firing rates of 20 excitatory (blue) and 20 inhibitory (red) randomly selected neurons. **c**, Autocorrelograms of the E → E connectivity matrices (black) and firing-rate vectors of the excitatory (blue) and inhibitory (red) neurons.

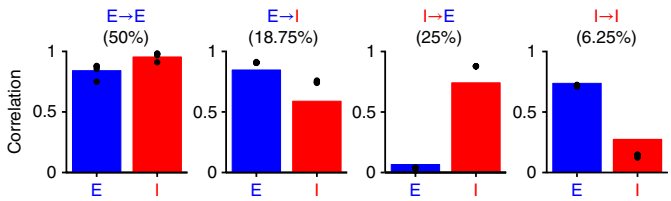

Fig. 3 | Rewiring different synaptic types differentially affects the patterns of network activity. Graphs present correlation coefficients of the firing-rate vectors before and after rewiring different synaptic types. Bars correspond to the theoretical prediction (see Methods); each of the five dots denotes the correlations computed from one numerical simulation (except E → E, the five dots overlap). Numbers in brackets denote the fraction of corresponding synapses in the network.

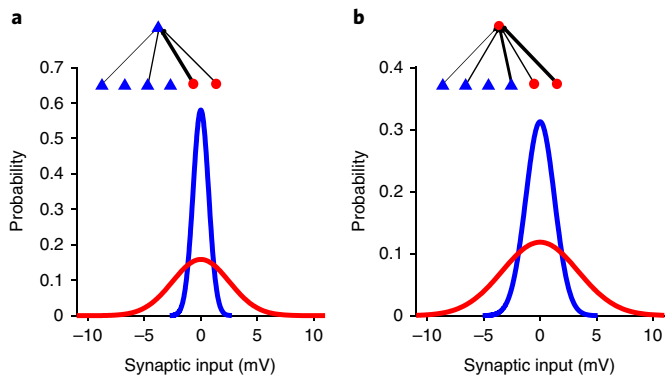

Fig. 4 | Synaptic rewiring in a feedforward network. a,**b**, The distributions of changes in time-averaged inputs to an excitatory (**a**) and an inhibitory (**b**) neuron in a feedforward configuration, following excitatory (blue) or inhibitory (red) rewiring. The wider the distribution, the larger the effect of rewiring is. For both the excitatory and inhibitory neurons, inhibitory rewiring has a larger effect than excitatory rewiring on the input to neuron.

excitatory distribution is narrower than the inhibitory distribution implies that E → E rewiring has a smaller effect on the postsynaptic activity than I → E rewiring. Thus, the seemingly innocuous difference in the distributions of the firing rates of the two populations of neurons (well documented in the cortex in vivo[17,22]) is the major determinant of the robustness of the network to excitatory rewiring and its sensitivity to inhibitory rewiring. We repeat this analysis for the inhibitory neurons (Fig. 4b). Because the distributions of excitatory and inhibitory efficacies onto the inhibitory neurons (E → I, I → I) are also comparable, the inhibitory neurons are also robust against excitatory rewiring and sensitive to inhibitory rewiring. The differential sensitivity of the firing rates of the feedforward network neurons to the excitatory and inhibitory inputs is depicted in Supplementary Fig. 1.

The intuitive explanation of Fig. 4 is not complete because of the recurrent connectivity. The E → E rewiring changes the inputs to all excitatory (postsynaptic) neurons in the network, which in turn are also the presynaptic neurons. In addition, the change in the firing rates of the excitatory neurons also changes the input to the inhibitory neurons, changing their firing rates and consequently, also changing the firing rates of the excitatory neurons that they innervate. To better understand the robustness of the recurrent network to excitatory rewiring and sensitivity to inhibitory rewiring, we approximate the spiking dynamics using a mean-field rate model. In this model, we can analytically compute the change in the pattern of activity following a change in the connectivity (see Methods). The predictions of this theory (Fig. 3) are qualitatively similar to the estimates from the numerical simulations.

To further test the role of the distributions of excitatory and inhibitory firing rates in the shaping of network activity, we study the dynamics of networks endowed with different levels of external inputs to the two populations. This manipulation maintains the distributions of synaptic connections while varying the firing rates distributions. The results of our mean field analysis are consistent with the intuitive explanation that is based on the feedforward network (Fig. 4), namely that sensitivity to the rewiring of excitatory synapses indeed increases with the relative value of the mean squared firing rates of the excitatory neurons (Supplementary Fig. 2).

Rewiring and memory. The results presented in Figs. 2 and 3 demonstrate that the pattern of activity in the random network is primarily determined by the inhibitory connections. This result motivated us to study the role of inhibition in memory storage. To address this question, we embed memory patterns in the network of Fig. 1. Each memory consists of a randomly chosen subset of neurons (sparseness = 0.1), designated to be more active during its recall. The excitatory and inhibitory connectivity are constructed using parameter-free Hebbian and anti-Hebbian learning rules on those patterns, respectively. The resultant probabilities of connections and the marginal distributions of synaptic efficacies are asymptotically equal (in the limit of a large number of memories) to that of the unstructured network (see Methods). Using simulations, we find that if the number of embedded memory patterns is not too large, the network is endowed with multiple attractors, corresponding to the different memory patterns. In Fig. 5a we plot the spike times of 400 excitatory and 100 inhibitory randomly chosen neurons in a network storing 2,000 memory patterns. The transient injection of current into the should-be-active neurons of a particular memory pattern sets the activity pattern of the network in an attractor state that is highly correlated with that memory pattern, which we refer to as recall state. Transitions to other recall states are induced by transient current injections.

To study the relative contribution of the different synapses to the ability of the network to store memories, we first rewire all E → E connections by resampling them from the same marginal distribution. This has no substantial effect on the pattern of activity in the network (Fig. 5b). By contrast, I → I rewiring results in a loss of the recall state (Fig. 5b).

Figure 5 demonstrates two things. First, in the balanced network, Hebbian and anti-Hebbian excitatory and inhibitory learning rules, respectively, can generate multiple attractors that have a substantial overlap with encoded memory patterns. Second, I → I rewiring is more detrimental than E → E rewiring for memory maintenance. To understand these results, we note that the co-existence of multiple attractors requires some form of patterned feedback. In our network, such feedback is mediated synergistically by three different synaptic mechanisms:

(1) Hebbian changes in the E → E connections result in stronger recurrent excitatory connectivity between neurons selective for the same memory pattern. Increasing the activity in any of these cell assemblies results in a selective positive feedback for that assembly, while inhibiting the other assemblies through the inhibitory connections.
(2) Positive feedback is also generated by anti-Hebbian changes in the I → I connections[26]: inhibitory neurons that belong to the same memory pattern inhibit each other less than they inhibit neurons that belong to different memory patterns.
(3) Finally, memories are maintained by the positive feedback generated by the E → I and I → E loop. Consider an excitatory and inhibitory neuron that belong to the same memory pattern: as a result of Hebbian plasticity, activation of the excitatory neuron will activate the inhibitory neuron, which will, in turn, preferentially inhibit excitatory neurons not belonging to this memory pattern due to anti-Hebbian learning.

All three mechanisms contribute to the initial memory maintenance in Fig. 5. Rewiring the E → E synapses (Fig. 5b) disables the first mechanism, exposing the contribution of the second and third mechanisms to memory maintenance. Similarly, rewiring the I → I synapses (Fig. 5b) disables the second mechanism. The robustness of the recall state to E → E rewiring, while being sensitive to I → I rewiring, indicates that inhibitory plasticity plays a more important role than excitatory plasticity in the maintenance of the memories in Fig. 5.

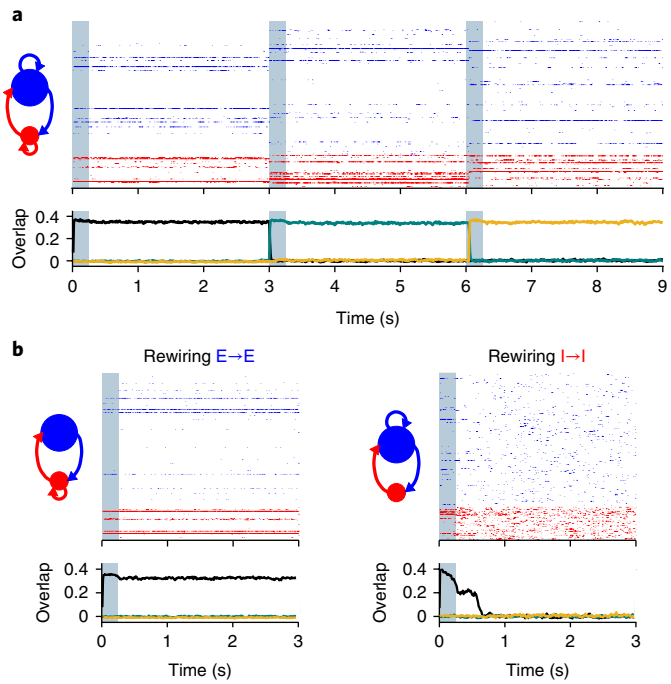

Fig. 5 | Synaptic rewiring and memory in a spiking network model. 2,000 memory patterns (sparseness: 0.1) were encoded in the integrate-and-fire network of Fig. 1a using Hebbian and anti-Hebbian learning rules on the excitatory and inhibitory synapses, respectively, preserving the probabilities of connections and the marginal distribution of synaptic efficacies of Fig. 1a (see text and Methods). **a**, Top: a raster plot of randomly selected 400 excitatory (blue) and 100 (red) inhibitory neurons. Patterned transient current injections (gray; see Methods) set the activity of the network in attractor states that are correlated with the memory patters. Bottom: black, teal, and orange traces depict the correlation coefficients of the firing rates of the neurons (averaged in bins of 10 ms) with the memory patterns corresponding to the current injected at times $t = 0$ s, $t = 3$ s, and $t = 6$ s, respectively. **b**, The same network with the same patterned current injection as in **a** after rewiring the E → E synapses (left) and I → I synapses (right). While E → E rewiring has almost no effect on the memory state of the network, following I → I rewiring, the network no longer maintains the memory pattern. Blue and red circles to the left of the raster plots denote excitatory and inhibitory populations, respectively. Arrows denote learned synaptic connections, whereas their absence denotes connections that were randomly drawn from the same marginal distribution.

To further dissect the relative contribution of each of the three mechanisms, we use the mean-field approximation (see Methods) to compute the memory capacity α_c of the network, the maximal number of patterns that the network can maintain, relative to the total number of neurons in the network. The results of this analysis are summarized in Fig. 6. When learning is restricted to the E → E connectivity (50% of the synapses), $\alpha_c = 0.007$. By contrast, $\alpha_c = 0.024$ when learning is restricted to the I → I connectivity, indicating that a much larger number of patterns can be maintained using I → I plasticity, despite the fact that it involves a substantially smaller number of synapses (only 6.25% of the synapses). Strikingly, the memory capacity per I → I synapse is more than 25 times larger than the capacity per E → E synapse. Finally, when learning is restricted to the E → I and I → E connections (approximately 44% of the synapses), the capacity is $\alpha_c = 0.021$. When all three mechanisms are used, $\alpha_c = 0.067$. Note that this number is larger than the sum of all three mechanisms operating individually ($0.007 + 0.024 + 0.021 = 0.052$), demonstrating synergy between the three mechanisms.

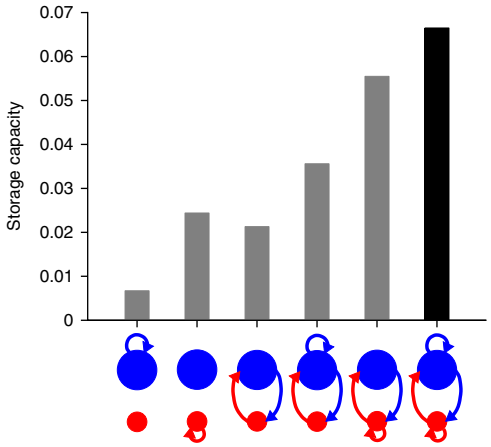

Fig. 6 | Storage capacity using different synaptic types. The memory capacity of a network is defined as the number of memories that the network can maintain relative to the total number of neurons in the network. We used mean-field approximation (see Methods) to estimate the memory capacity for random binary memory patterns (sparseness: 0.1). We used Hebbian and anti-Hebbian learning rules on the excitatory and inhibitory synapses, respectively, preserving the probabilities of connections and the marginal distributions of synaptic efficacies of Fig. 1a (see text and Methods). Bars depict the capacity when learning is restricted to different subsets of synapses. From left to right: E → E; I → I; E → I & I → E; E → E, E → I & I → E; I → I, E → I & I → E; all synapses (black). Note that memory capacity when learning is restricted to E → E synapses is less than one-third of the capacity when it is restricted to I → I synapses. This is despite the fact that there are 8 times more E → E synapses than I → I synapses in the network. When learning is restricted to E → E & I → I synapses, the memory states of the inhibitory and excitatory parts of the network are dissociated (not shown).

Rewiring in the structured network eliminates one of the mechanisms, leaving the other two mechanisms intact. As depicted in Fig. 6, E → E rewiring has only a small effect on the capacity, reducing it from $\alpha_c = 0.067$ to $\alpha_c = 0.056$. By contrast, I → I rewiring reduces the capacity to $\alpha_c = 0.036$. For comparison, the mean field approximation predicts that in the 40,000-neuron network of Fig. 5, the capacity after E → E rewiring is $0.056 \times 40{,}000 = 2{,}240$, a number larger than the 2,000 patterns embedded in the network. The capacity in this network after I → I rewiring is predicted to be smaller than 2,000: $0.036 \times 40{,}000 = 1{,}440$. Indeed, the recall states in Fig. 5 are lost following I → I rewiring but not following E → E rewiring. Finally, the rewiring of the E → I and I → E synapses splits the network into two functionally separate subnetworks, an excitatory network and an inhibitory network, in which the memory capacities of each subnetwork is limited by the corresponding mechanism (E → E for the excitatory subnetwork and I → I for the inhibitory network).

The results depicted in Figs. 5 and 6 demonstrate that recall states in the structured network are robust to E → E rewiring and sensitive to I → I rewiring. This is qualitatively similar to the robustness and sensitivity of the ongoing pattern of activity to these two perturbations, depicted in Fig. 3. As explained in Fig. 4, the differential effect of E → E and I → I rewirings stems from the difference in the distributions of firing rates of inhibitory and excitatory neurons. A similar mechanism underlies the differential effects of E → E and I → I rewirings on recall states in the structured network. To understand why, we initialize the network in one of the memory patterns. For this recall state to be maintained over time, the synaptic inputs to the active neurons should be larger than those to the quiescent neurons. Both excitatory and inhibitory synapses can contribute to a differential input to the active and quiescent neurons. The excitatory connections can contribute because by construction, the excitatory connections amongst active neurons in a memory pattern are stronger than between active and quiescent neurons. Similarly, the connections from active inhibitory neurons to other active neurons in a pattern are weaker than the connections from active inhibitory neurons to quiescent neurons. In both cases, the higher the firing rates of the active neurons, the larger the difference is between the inputs to the active and quiescent neurons. Excitatory rewiring impairs the former contribution, whereas inhibitory rewiring impairs the latter. The firing rates of active inhibitory neurons are higher than those of active excitatory neurons. The reason for this difference in firing rates is that the population-averaged firing rates of the excitatory and inhibitory neurons are fixed (determined by balance conditions) in both unstructured and structured networks. A recall state is associated with a higher firing rate of the 10% of neurons associated with that memory pattern and a lower firing rate of all other neurons. These firing rates are bounded from below. Therefore, the average firing rates of the active neurons is bounded by the population average divided by the sparseness (0.1). Because the average firing rate of inhibitory neurons is six times higher than that of the excitatory neurons, memory capacity is dominated by inhibition.

Excitatory plasticity reconsidered. Our findings that changes in excitatory connections have only little effect on network activity seem at odds with the fact that plasticity of glutamatergic synapses is essential for learning and memory[13,27]. A possible solution to this puzzle is the fact that so far we have only considered manipulations to the connectivity that preserve the marginal distribution of connections. By contrast, several studies have demonstrated that learning is accompanied by a transient increase in the total number of spines, specifically in regions that are associated with the task[2–5]. To emulate such learning-induced plasticity in a network of spiking neurons (Fig. 7a), we randomly increase the number of E → E connections by 20% onto a subset of the excitatory population, which we refer to as 'targeted neurons'. When the fraction of targeted neurons is 20% (Fig. 7a), this manipulation results in a 4% increase in the total number of E → E connections, a modest change to the connectivity, compared with the complete E → E rewiring of Fig. 3. Naively, this manipulation should have only a minute effect on the pattern of firing rates. We observe, instead, a substantial change in firing rates of the individual excitatory neurons (Fig. 7b). The average firing rate of the targeted neurons was 2.05 ± 0.02 Hz, substantially larger than that of the nontargeted neurons (0.79 ± 0.01 Hz; distributions in Fig. 7c).

Next, we considered the effect of an even larger change to the connectivity: a 20% increase in the number of E → E connections onto all excitatory neurons (all excitatory neurons being targeted neurons; Fig. 7a). One may expect that a larger change in the number of connections should result in a larger change to the pattern of firing rates. Contrary to this expectation, this manipulation has only a small effect on the pattern of neural activity (Fig. 7b). Thus, the change in the pattern of firing rates in response to the addition of E → E connections to a subset of the excitatory population is not a monotonically increasing function of the fraction of targeted neurons (Fig. 7d and Supplementary Fig. 3). It is maximal when the fraction of targeted neurons is approximately 20%. In that case, the magnitude of change in the pattern of activity is many times larger than that following a complete rewiring of the E → E connections (Fig. 3). The reason for this is that adding connections to some but not all of the neurons disrupts the balance in the network[28]. Because synapses are strong, even a relatively small change in their number has a large effect on the excitatory input to the targeted neurons. The increased excitatory input is immediately balanced by increased inhibitory activity. This increased inhibition, however, is not restricted to the targeted neurons. As a result of the increased inhibitory input, the firing rates of the nontargeted

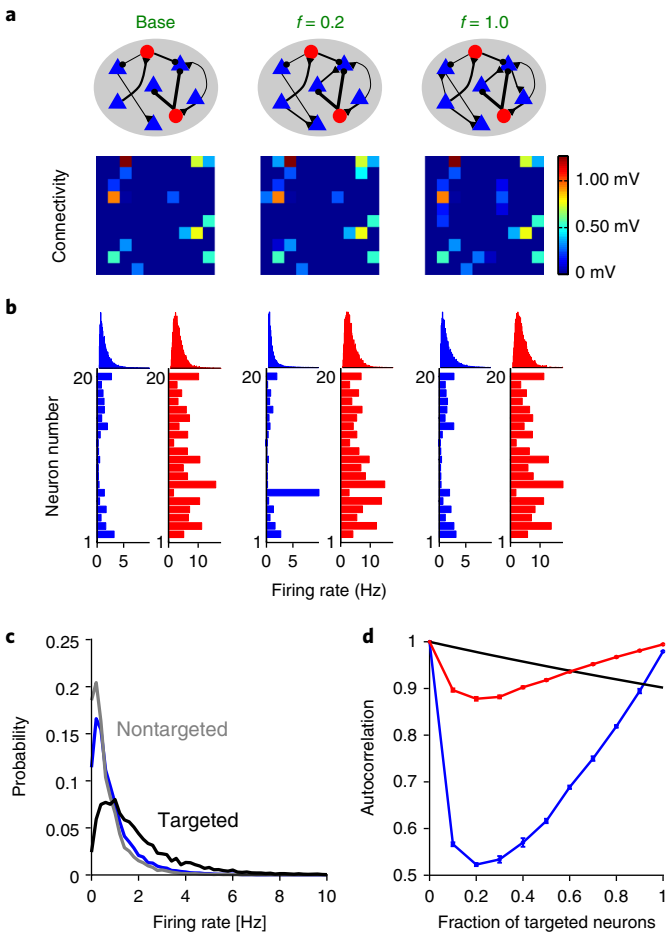

Fig. 7 | The effect of heterogeneous addition of E → E connections. a, Top: schematic illustration of three sparsely connected networks of spiking neurons (see Methods). Left: baseline network, in which the probability of E → E connection between each pair of neurons (*f*) is 0.2. Middle: the same network as in the left panel, in which randomly chosen 20% of the excitatory neurons are targeted: each targeted neuron is now connected to the 80% previously unconnected excitatory neurons with a probability of 5%, such that the overall probability of incoming E → E connection to the targeted neurons is 24%. Connectivity onto nontargeted neurons is unchanged and their probability of incoming E → E connection remains at 20%. Right: the same network when all excitatory neurons are targeted. Bottom: connectivity matrices between 10 randomly selected excitatory neurons given the conditions in the top row. **b**, Top: distributions of firing rates of excitatory (blue) and inhibitory (red) neurons given the conditions in **a**. Bottom: firing rates of randomly selected 20 excitatory (blue) and 20 inhibitory (red) neurons. **c**, Distributions of firing rates of the targeted (black) and nontargeted (gray) excitatory neurons when 20% of the neurons are targeted (middle panels in **a** and **b**. For comparison, the blue curve denotes the distribution of firing rates of the excitatory neurons in the random network (left panel in **a**. **d**, Autocorrelogram of the E → E connectivity matrices (black) and of the vectors of firing rates of excitatory (blue) and inhibitory (red) neurons as a function of the fraction of targeted neurons. Points denote average over five simulations; error bars are s.e.m.

neurons decrease substantially. Thus, the pattern of excitatory network activity is dominated by the targeted neurons (Fig. 7b). To understand the nonmonotonic effect of the fraction of targeted neurons on the change in the pattern of activity (Fig. 7d), we note that the smaller the number of targeted neurons, the larger the number of 'silenced' nontargeted neurons is. Hence, we expect that the smaller the fraction of targeted neurons is, the larger the change in network activity should be. This argument, however, is accurate only in the mathematical limit of balanced networks. In finite networks, such as the one used in the simulation shown in Fig. 7, if the fraction of targeted neurons is too small, their overall activity is not sufficient to substantially affect the inhibitory neurons or to silence the remaining nontargeted neurons. In this regime of a small fraction of targeted neurons, we expect a larger change in the pattern of activity as the fraction of targeted neurons increases. The combined effect of these two mechanisms results in the U-shaped function of Fig. 7d and Supplementary Fig. 3.

This cell-targeted excitatory plasticity, which tags specific neurons, generates a cell assembly of neurons whose firing rate is, on average, higher than that of the nontargeted neurons (Fig. 7c). In learning experiments, the increase in the number of spines is transient, and the density of spines reverts to basal levels within days[2–5]. We hypothesize that during this window of disrupted balance between excitation and inhibition, inhibitory plasticity can form a new memory by encoding this specific pattern of activity as an attractor state[29,30].

Discussion

Studying a model of a cortical circuit, we find that a higher and more heterogeneous firing rate of the inhibitory neurons compared to that of excitatory neurons results in a pattern of cortical activity that is dominated by the inhibitory synapses. Consequently, the cortex is robust to the experimentally observed volatility of E → E synapses. Similarly, learning that is mediated by inhibitory plasticity is more effective than excitatory plasticity for memory storage. Finally, we show that excitatory plasticity can still effectively shape the pattern of activity in the network if the pattern of connectivity is structured. These results imply that inhibitory synapses play a dominant role in computation, despite their smaller number.

To understand the relevance of this theoretical analysis for cortical function, we discuss the assumptions that, in the model, underlie the dominance of inhibitory connectivity. In our model, all cells are identical and only differ in their input. Inhibition dominates over excitation because most of the heterogeneity between cells is due to heterogeneity in their inhibitory input. This result is quantitative rather than qualitative and is a consequence of the parameters that were used in the model, as they were measured in the mouse barrel cortex[17,21], as well as our assumptions about the pattern of connectivity. The structure of the mathematical equations themselves does not dictate this result. Rather, it is a combination of the fact that the distributions of excitatory and inhibitory synaptic efficacies are comparable, while the firing rates of the inhibitory neurons are higher and more diverse than those of the excitatory neurons. If these assumptions do not hold in a different brain area or a different animal type, then the conclusions may change. Our framework can nevertheless be used to compute the expected effect of synaptic rewiring for any set of parameters. Another important assumption of our model is that rewiring is random. As a result, in the process of rewiring, a high-firing-rate presynaptic neuron can occasionally be replaced by a low-firing-rate one, or vice versa. If rewiring is nonrandom, for example, such that a high-firing-rate presynaptic neuron is necessarily replaced by another high-firing-rate neuron, then the effect of rewiring can be very different than the one predicted by our theory.

We show in Fig. 3 that in random unstructured networks with cortical parameters, the heterogeneity in inputs to the neurons is dominated by inhibition, and therefore, E → E rewiring has a limited effect on network activity. However, E → E rewiring can substantially affect network activity when connectivity is structured[31,32], for example, when a subset of neurons receive substantially more excitatory input (Fig. 7) or when memory patterns are embedded within the excitatory network (Fig. 6). Nevertheless,

even when considering atypical connectivity, inhibitory synapses are more flexible in their ability to control network activity. This is demonstrated by the larger memory capacity of learning that is based on inhibitory plasticity, compared with that based on excitatory plasticity (Fig. 6).

Inhibition is a key determinant of firing patterns, both during the critical period and during adulthood[33–41]. Moreover, the power of disinhibitory connectivity (I→I) in cortical computation has been recently unveiled[42]. Our theory provides a mechanistic explanation for conditions in which inhibition can dominate computation. Specifically, we show how a stable inhibitory scaffold can stabilize cortical dynamics even in the presence of substantial volatility of excitatory synapses. However, there is some evidence that cortical activity changes over days and weeks[10,43,44]. In our framework, this is an indication of inhibitory volatility.

Indeed, not only excitatory but also inhibitory synapses are plastic and change over time[30,37,38,45–47]. Plasticity of inhibitory synapses is likely co-orchestrated with excitatory plasticity in the process of memory formation[29,48]. Consistent with the observation that inhibitory plasticity lags excitatory plasticity[29], we suggest that the formation of memory is a two-stage process, in which excitatory and inhibitory plasticity play qualitatively different roles. Heterogeneous transient changes in the E→E connectivity disrupt the balance between excitation and inhibition by changing the overall excitatory input to a task-specific subset of excitatory neurons (as in Fig. 7), allowing for changes in the network activity. Activity-dependent inhibitory plasticity incorporates these changes in the inhibitory connectome[29,49,50]. Rebalancing the network occurs over time, when spontaneous E→E synaptic rewiring eliminates the excess excitatory connections on the subset of neurons, but the memory trace is maintained in the inhibitory connections. Thus, inhibition sets the realm for excitatory plasticity and controls functional stability of the network.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-018-0226-x.

Received: 27 April 2018; Accepted: 30 July 2018;
Published online: 17 September 2018

References

1. Lai, C. S., Franke, T. F. & Gan, W. B. Opposite effects of fear conditioning and extinction on dendritic spine remodelling. *Nature* **483**, 87–91 (2012).
2. Moczulska, K. E. et al. Dynamics of dendritic spines in the mouse auditory cortex during memory formation and memory recall. *Proc. Natl. Acad. Sci. USA* **110**, 18315–18320 (2013).
3. Xu, T. et al. Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462**, 915–919 (2009).
4. Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
5. Hayashi-Takagi, A. et al. Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* **525**, 333–338 (2015).
6. Caroni, P., Donato, F. & Muller, D. Structural plasticity upon learning: regulation and functions. *Nat. Rev. Neurosci.* **13**, 478–490 (2012).
7. Zuo, Y., Lin, A., Chang, P. & Gan, W. B. Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron* **46**, 181–189 (2005).
8. Holtmaat, A. J. et al. Transient and persistent dendritic spines in the neocortex in vivo. *Neuron* **45**, 279–291 (2005).
9. Loewenstein, Y., Yanover, U. & Rumpel, S. Predicting the dynamics of network connectivity in the neocortex. *J. Neurosci.* **35**, 12535–12544 (2015).
10. Chambers, A. R. & Rumpel, S. A stable brain from unstable components: Emerging concepts and implications for neural computation. *Neuroscience* **357**, 172–184 (2017).
11. Mongillo, G., Rumpel, S. & Loewenstein, Y. Intrinsic volatility of synaptic connections–a challenge to the synaptic trace theory of memory. *Curr. Opin. Neurobiol.* **46**, 7–13 (2017).
12. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* **10**, 647–658 (2009).
13. Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A. & Noguchi, J. Structural dynamics of dendritic spines in memory and cognition. *Trends Neurosci.* **33**, 121–129 (2010).
14. Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex in vivo. *Nat. Neurosci.* **9**, 1117–1124 (2006).
15. Loewenstein, Y., Kuras, A. & Rumpel, S. Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. *J. Neurosci.* **31**, 9481–9488 (2011).
16. DeFelipe, J. & Jones, E. G. in *Handbook of Brain Microcircuits* (eds. Shepherd, G. M. & Grillner, S.). Ch. 1,5–14 (Oxford University Press, 2010).
17. Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. & Petersen, C. C. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65**, 422–435 (2010).
18. Karnani, M. M., Agetsuma, M. & Yuste, R. A blanket of inhibition: functional inferences from dense inhibitory connectivity. *Curr. Opin. Neurobiol.* **26**, 96–102 (2014).
19. Bhatt, D. H., Zhang, S. & Gan, W. B. Dendritic spine dynamics. *Annu. Rev. Physiol.* **71**, 261–282 (2009).
20. Holtmaat, A. et al. Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window. *Nat. Protoc.* **4**, 1128–1144 (2009).
21. Avermann, M., Tomm, C., Mateo, C., Gerstner, W. & Petersen, C. C. Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J. Neurophysiol.* **107**, 3116–3134 (2012).
22. Buzsáki, G. & Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* **15**, 264–278 (2014).
23. Renart, A. et al. The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
24. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
25. Roxin, A., Brunel, N., Hansel, D., Mongillo, G. & van Vreeswijk, C. On the distribution of firing rates in networks of cortical neurons. *J. Neurosci.* **31**, 16217–16226 (2011).
26. Hendin, O., Horn, D. & Tsodyks, M. V. The role of inhibition in an associative memory model of the olfactory bulb. *J. Comput. Neurosci.* **4**, 173–182 (1997).
27. Rumpel, S., LeDoux, J., Zador, A. & Malinow, R. Postsynaptic receptor trafficking underlying a form of associative learning. *Science* **308**, 83–88 (2005).
28. Landau, I. D., Egger, R., Dercksen, V. J., Oberlaender, M. & Sompolinsky, H. The impact of structural heterogeneity on excitation-inhibition balance in cortical networks. *Neuron* **92**, 1106–1121 (2016).
29. Froemke, R. C., Merzenich, M. M. & Schreiner, C. E. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**, 425–429 (2007).
30. Rubinski, A. & Ziv, N. E. Remodeling and tenacity of inhibitory synapses: relationships with network activity and neighboring excitatory synapses. *PLoS Comput. Biol.* **11**, e1004632 (2015).
31. Denève, S. & Machens, C. K. Efficient codes and balanced networks. *Nat. Neurosci.* **19**, 375–382 (2016).
32. Hennequin, G., Vogels, T. P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
33. Griffen, T. C. & Maffei, A. GABAergic synapses: their plasticity and role in sensory cortex. *Front. Cell. Neurosci.* **8**, 91 (2014).
34. Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
35. Doron, G., von Heimendahl, M., Schlattmann, P., Houweling, A. R. & Brecht, M. Spiking irregularity and frequency modulate the behavioral report of single-neuron stimulation. *Neuron* **81**, 653–663 (2014).
36. Liberti, W. A. et al. Unstable neurons underlie a stable learned behavior. *Nat. Neurosci.* **19**, 1665–1671 (2016).
37. Chen, J. L. et al. Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron* **74**, 361–373 (2012).
38. van Versendaal, D. et al. Elimination of inhibitory synapses is a major component of adult ocular dominance plasticity. *Neuron* **74**, 374–383 (2012).
39. Hensch, T. K. et al. Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science* **282**, 1504–1508 (1998).
40. Huang, Z. J. et al. BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell* **98**, 739–755 (1999).
41. Levelt, C. N. & Hübener, M. Critical-period plasticity in the visual cortex. *Annu. Rev. Neurosci.* **35**, 309–330 (2012).
42. Letzkus, J. J., Wolff, S. B. & Lüthi, A. Disinhibition, a circuit mechanism for associative learning and memory. *Neuron* **88**, 264–276 (2015).
43. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell* **170**, 986–999.e16 (2017).

44. Maass, W. Searching for principles of brain computation. *Curr. Opin. Behav. Sci.* **11**, 81–92 (2016).
45. Gaiarsa, J. L., Caillard, O. & Ben-Ari, Y. Long-term plasticity at GABAergic and glycinergic synapses: mechanisms and functional significance. *Trends Neurosci.* **25**, 564–570 (2002).
46. Kullmann, D. M., Moreau, A. W., Bakiri, Y. & Nicholson, E. Plasticity of inhibition. *Neuron* **75**, 951–962 (2012).
47. Woodin, M. A., Ganguly, K. & Poo, M. M. Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. *Neuron* **39**, 807–820 (2003).
48. Donato, F., Rompani, S. B. & Caroni, P. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).
49. Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
50. Luz, Y. & Shamir, M. Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. *PLoS Comput. Biol.* **8**, e1002334 (2012).

Acknowledgements

We thank L. Abbott and D. Hansel for their careful reading of our manuscript and their insightful comments. This work was performed in the framework of the the France-Israel Center for Neural Computation and was supported by the Israel Science Foundation (Grant No. 757/16, Y.L.), the DFG (CRC 1080, Y.L. and S.R.), the Gatsby Charitable Foundation (Y.L.), and by ANR (14-NEUC-0001-01 and 13-BSV4-0014-02, G.M.).

Author contributions

S.R. performed the experiments, G.M., S.R. and Y.L. analyzed the data, G.M. and Y.L. developed the theory, G.M. performed the numerical simulations, G.M., S.R. and Y.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0226-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to G.M. or Y.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Experimental methods. In this study, we used the data described in[9,15]. Animal procedures were approved by the Cold Spring Harbor Laboratory Animal Care and Use Committee and carried out in accordance with National Institutes of Health standards. The experimental procedures are described in detail in refs [9,15]. In short, we implanted glass windows in the crania of 6 male adult (~6 months old) inhouse-bred mice of the GFP-M transgenic line (Tg(Thy1-EGFP)MJrs/J) selected for sparse GFP expression in the cortex[51] and characterized the dynamics of spines using in vivo two-photon imaging. For image analysis, best projections of all dendrites for all timepoints were constructed that allowed identification of spines at a given timepoint and indexing of identical spines across time[8]. The volume of the spine was estimated as the spine's integrated intensity, normalized by the intensity of the adjacent dendrite. It has previously been shown that this normalized integrated spine intensity in in vivo two-photon imaging is tightly correlated with the volume of the spine as subsequently estimated by ssEM reconstruction[7,8,52–54]. Furthermore, there is good evidence that the volume of a spine can serve as a proxy for the functional strength of the corresponding synaptic connection[5,12,19,20]. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in a previous publication[8].

Numerical methods. *The spiking network.* The network is composed of N_E excitatory and N_I inhibitory current-based integrate-and-fire neurons. The depolarization of neuron i in population a $(=E;I)$, $\nu_a^i(t)$ evolves according to

$$\dot{\nu}_a^i(t) = -\frac{\nu_a^i(t)}{\tau_m} + h_a^i(t) + H_a^{(ext)} \tag{1}$$

where $i=1,\ldots,N_a$; τ_m denotes the membrane time constant, $h_a^i(t)$ denotes the afferent input originating from the recurrent synaptic connectivity, and $H_a^{(ext)}$ denotes a constant input that originates from other brain regions and is homogeneous across neurons within the population a. The neuron fires a spike when reaching a fixed threshold θ (that is, $\nu_a^i(t) \geq \theta$), and becomes refractory for a period τ_{arp} after which equation (1) resumes from a subthreshold rest potential v_R.

The recurrent input to neuron i in population a, $h_a^i(t)$, is given by

$$h_a^i(t) = \sum_{j=1}^{N_E} c_{aE}^{ij} W_{aE}^{ij} \sum_k \delta(t - t_{E,k}^j) - \sum_{j=1}^{N_I} c_{aI}^{ij} W_{aI}^{ij} \sum_k \delta(t - t_{I,k}^j) \tag{2}$$

where $c_{ab}^{ij}=1$ if there exists a synaptic connection from neuron j in population b to neuron i in population a, and $c_{ab}^{ij}=0$ otherwise; W_{ab}^{ij} is the corresponding synaptic efficacy; the sums over j are over all neurons in the corresponding populations, while the sums on k are over all the emission times of the action potentials, $t_{b,k}^j$, of the presynaptic neuron j in population b. For simplicity, we neglect the temporal dynamics of the synapses.

All parameters are reported in Supplementary Table 1.

The spiking network was simulated in two different configurations: the random network configuration (used to generate Figs. 2, 3 and 7) and the structured network configuration (used to generate Fig. 5).

The random network. In the random network configuration, connectivities for the synaptic populations E→I, I→E and I→I were generated as follows. The c_{ab}^{ij} values were randomly and independently set to 1 with probability c_{ab}, while the W_{ab}^{ij} values were independently drawn from a log-normal distribution, with $\langle W_{ab} \rangle$ and $\langle W_{ab}^2 \rangle$ estimated from experimental data. The E→E connectivity was extracted from the spines' data by a procedure described below.

The structured network. In the structured network configuration, we first generated P memories. Each memory μ consists of a binary vector, $\{\xi_E^i(\mu);\ \xi_I^i(\mu)\}$, where ξ_a^i is set to 1 with probability f independently for each μ and each $i=1,\ldots,N_a$, where $(a=E,I)$ and 0 otherwise. Next, for each pair of neurons we computed the corresponding Hebbian terms

$$z_{ab}^{ij} = \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \sum_{\mu=1}^{P} \epsilon_{ab}^{ij}(\mu)(\xi_a^i(\mu) - f)(\xi_b^j(\mu) - f) \tag{3}$$

$$z_{ba}^{ji} = \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \sum_{\mu=1}^{P} \epsilon_{ba}^{ji}(\mu)(\xi_b^j(\mu) - f)(\xi_a^i(\mu) - f) \tag{4}$$

where $\epsilon_{ab}^{ij}(\mu)$ is a binary variable which takes the value 1 with probability 1/2 and 0 otherwise, and $\epsilon_{ba}^{ji}(\mu) = 1 - \epsilon_{ab}^{ij}(\mu)$. The variables $\epsilon_{ab}^{ij}(\mu)$ are introduced to prevent correlations between z_{ab}^{ij} and z_{ba}^{ji}, which in turn manifest as correlations between W_{ab}^{ij} and W_{ba}^{ji}. Finally, we set $c_{aE}^{ij}=0$ and we set $c_{aI}^{ij}=0$ if and only if (iff)

$$z_{aE}^{ij} \leq \zeta_{aE} \equiv \Phi^{-1}(1-c_{aE});\ -z_{aI}^{ij} \leq \zeta_{aI} \equiv \Phi^{-1}(1-c_{aI}); \tag{5}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the Gaussian cumulative distribution function. Otherwise, we set $c_{ab}^{ij}=1$.

The idea behind this scheme is that the only nonzero excitatory synapses are the ones for which the Hebbian term is largest. Similarly, the only nonzero inhibitory synapses are the ones for which the Hebbian term is smallest (most negative, that is, most anti-Hebbian).

In addition,

$$\ln W_{ab}^{ij} = \ln\langle W_{ab} \rangle - \frac{1}{2} \ln \frac{\langle W_{ab}^2 \rangle}{\langle W_{ab} \rangle^2} + \Phi^{-1}(y_{ab}^{ij}) \cdot \sqrt{\ln \frac{\langle W_{ab}^2 \rangle}{\langle W_{ab} \rangle^2}} \tag{6}$$

where <...> denote parameters, as in the random network, and

$$y_{aE}^{ij} = \frac{1}{c_{aE}} \int_{\zeta_{aE}}^{z_{aE}^{ij}} Dz;\ y_{aI}^{ij} = \frac{1}{c_{aI}} \int_{z_{aI}^{ij}}^{\zeta_{aI}} Dz. \tag{7}$$

where Dz is the standard Gaussian measure, that is, $e^{-z^2/2}dz/\sqrt{2\pi}$. For large P, z_{ab}^{ij} is asymptotically normally distributed, with zero mean and unitary variance, and thus the nonzero synaptic efficacies W_{ab}^{ij} are log-normally distributed, with mean $\langle W_{ab} \rangle$ and variance $\langle W_{ab}^2 \rangle - \langle W_{ab} \rangle^2$. As a result of this procedure, the excitatory synapses associated with the largest Hebbian term and the inhibitory synapses associated with the smallest Hebbian term are largest in their magnitude.

Note that the excitatory synapses (that is, E→E and E→I) follow a Hebbian rule, that is, the larger the Hebbian term the larger the corresponding synaptic efficacy is, while the inhibitory synapses (that is, I→E and I→I) follow an anti-Hebbian rule, that is, the larger the Hebbian term the smaller the corresponding synaptic efficacy is. Small values of z_{aE}^{ij} and large values of z_{aI}^{ij} result in the lack of the corresponding synaptic connection.

To activate memory μ during the simulations (Fig. 5), we increase $H_a^{(ext)}$ by 75% to neurons for which $\xi_a^i(\mu)=1$ and decrease it by the same proportion to neurons for which $\xi_a^i(\mu)=0$. External inputs to all neurons are then restored to their baseline levels linearly over 250 ms.

Extracting the E→E connectivity from spine imaging data. Our spine imaging data consisted of 3,688 spines, 1,420 of which were present in the first imaging session. The E→E connectivity used for the simulations of Fig. 1 is constructed in the following way. First, the probability of connection between any two excitatory neurons is set to $c_{EE}=0.2$. For those connected pairs, we randomly associate (without replacement) a spine from the 1,420 spines imaged in the first session to each of these connections. The spine sizes are converted into the corresponding putative synaptic efficacies in the following procedure: we assume that synaptic efficacy, W, is proportional to spine size, S, (that is, $W=g \cdot S$), and computed the proportionality factor by requiring that the average synaptic efficacy (over all spines in all sessions) was the same as the one reported in Supplementary Table 1 (that is, $g=\langle W_{EE} \rangle / \langle S \rangle$).

To simulate the effects of the spontaneous synaptic reorganization observed in the experiment, Fig. 2, we construct six synaptic matrices that are identical apart from the E→E segment. The E→E segment is generated as follows. First, the probability of connection between any two excitatory neurons in any of the 6 matrices is set to $c_{EE}=0.51$. For those potentially connected pairs, we randomly associate (without replacement) a spine from the 3,688 imaged spines to each of these connections. The size of that spine in the six imaging sessions was used to generate the connection between the corresponding excitatory neurons in all six networks. Note that, as most of the spines were transient, the effective probability of connection between any two excitatory neurons is approximately constant ($c_{EE} \cong 0.2$) across the six synaptic matrices. Specifically, the first synaptic matrix in Fig. 2 is identical to that of Fig. 1.

To simulate the effects of E→E rewiring in Fig. 3, we use a log-normal distribution of synaptic efficacies, where the probability of connection between any two excitatory neurons is set to $c_{EE}=0.2$ and the mean and variance of the distribution is taken from the mean and variance of the distribution of 3,688 spines, with the same conversion of size to efficacy as in Figs. 1 and 2.

In our analysis, we assumed that a spine is equivalent to a connection. However, there is some evidence that cortical neurons can be connected through more than one synapse[55,56]. The implications of this possibility for the results presented in Fig. 2 depend on the way in which spine volatility is correlated within a single connection. However, this will not affect the results presented in Figs. 3–7.

Parameter setting. The parameters used for the numerical simulations are reported in Supplementary Table 1. In all cases for which experimental estimates are available, the parameters are selected from within the corresponding range.

For simplicity, we choose the single-cell parameters to be the same for the excitatory and inhibitory neurons. The spike generation threshold is set at $\theta=33$ mV, within the range 17–43.8 mV estimated in ref. [21]. The membrane time constant is set at $\tau_m=10$ ms, within the range 9.3–28.4 ms estimated in ref. [21].

The means of the four types of synaptic weights are extracted from ref. [21] in the following way. $\langle W_{EE}\rangle$ is taken from ref. [21]. With respect to the inhibitory synapses, because the parameters in the literature are estimated separately for fast-spiking and non-fast-spiking inhibitory neurons, we estimate $\langle W_{EI}\rangle$, $\langle W_{IE}\rangle$, and $\langle W_{II}\rangle$ as weighted averages of the values reported for each population of inhibitory neurons, where the weighting is equal to the number of cells measured in each subpopulation. For the variances, we first estimate the variances in the synaptic efficacies for each subpopulation of inhibitory neurons, based on the reported s.e.m. values and numbers of synapses in ref. [21]. As with the estimated means, we use a weighted average to lump together the numbers for fast-spiking and non-fast-spiking inhibitory neurons. $\langle W_{EE}^2\rangle = 0.26$ mV^2 is estimated from the spines' size data after linear scaling (see Sec. 2.2). Note that estimating $\langle W_{EE}^2\rangle$ from ref. [21] using the s.e.m. of $\langle W_{EE}\rangle$ as we do for the other synaptic connections' types would have yielded a comparable parameter, $\langle W_{EE}^2\rangle = 0.29$ mV^2. The choice of connection probabilities is less constrained by the literature, in particular because connection probabilities depend on the distance between neurons, which is not modeled here. Similarly, there are no hard constraints on the values of $H_E^{(ext)}$ and $H_I^{(ext)}$. These parameters are all chosen to have the average firing rate within commonly reported ranges for cortical networks[17], $\langle \nu_E\rangle \cong 1$ Hz and $\langle \nu_I\rangle \cong 7$ Hz. For these parameters, $H_E^{(ext)}$ is approximately 75% of the total average excitatory input onto the excitatory neurons, and $H_I^{(ext)}$ is approximately 46% of the total average excitatory input onto the inhibitory neurons.

Analytical methods. *Mean-field analysis of the random network.* In the steady state, the network operates in an asynchronous, low-rate state of activity[23,24]. In such a state, spikes are the result of fast temporal fluctuations in the afferent synaptic input, whose average level is typically below the threshold. In these conditions, spiking is akin to a noise-driven escape process and the $f-I$ curve, $\phi(\cdot)$ is well approximated by[57]

$$\nu_a^i = \phi_a(h_a^i) = \left(\tau_{arp} + \tau_m \int_{\frac{v_R - \tau_m h_a^i}{\sigma_a\sqrt{\tau_m}}}^{\frac{\theta - \tau_m h_a^i}{\sigma_a\sqrt{\tau_m}}} dy\, e^{y^2}(1+\mathrm{erf}(y)) \right)^{-1} \tag{8}$$

where ν_a^i is the average firing rate of neuron i in population a, h_a^i is the corresponding time-averaged input, τ_{arp} is the absolute refractory period, τ_m is the membrane time constant, θ is the spike emission threshold, v_R is the postspike reset potential, and σ_a^2 is the variance per unit time of the synaptic input. In other words, in these conditions we can approximate the spiking dynamics of the network using a mean-field rate model, in which the fast-noise (due to spiking) is integrated into the shape of the single-neuron $f-I$ curve. The time-averaged input to neuron i in population a is given by

$$h_a^i = \sqrt{N_E} h_a^{(ext)} + \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} w_{aE}^{ij} \nu_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} w_{aI}^{ij} \nu_I^j \tag{9}$$

where, for later convenience, we have made explicit the scaling of the external inputs and synaptic efficacies with the size of the network, that is,

$$c_{ab}^{ij} W_{ab}^{ij} \to \frac{w_{ab}^{ij}}{\sqrt{N_b}}\,;\; H_a^{(ext)} \to \sqrt{N_E} h_a^{(ext)} \tag{10}$$

The activities of the different neurons are weakly correlated, and each neuron receives a large number of randomly distributed connections. Thus, the distribution of inputs over the neurons within each neuronal population is well approximated by a Gaussian distribution. The corresponding mean, u_a, and variance, s_a^2, are given by

$$u_a = \sqrt{N_E} \cdot \left(h_a^{(ext)} + \langle w_{aE}\rangle \langle \nu_E\rangle - \frac{\sqrt{N_I}}{\sqrt{N_E}} \langle w_{aI}\rangle \langle \nu_I\rangle \right) \tag{11}$$

$$s_a^2 = \sum_{b=E,I} \left(\langle w_{ab}^2\rangle \langle \nu_b^2\rangle - \langle w_{ab}\rangle^2 \langle \nu_b\rangle^2 \right) \tag{12}$$

and the variance per unit time of the synaptic input σ_a^2 (see equation (8)) is given by

$$\sigma_a^2 = \sum_{b=E,I} \langle w_{ab}^2\rangle \langle \nu_b\rangle \tag{13}$$

The steady state properties of the network activity can now be computed by using a self-consistency argument. Knowing the statistics of the afferent inputs, one can compute the statistics of firing rates in the network. In particular, one can compute the first two moments, which are given by

$$\langle \nu_a\rangle = \int D\eta \phi_a(u_a + \eta \cdot s_a) \tag{14}$$

$$\langle \nu_a^2\rangle = \int D\eta (\phi_a(u_a + \eta \cdot s_a))^2 \tag{15}$$

Incorporating equations (14) and (15), into equations (11)–(13), one obtains a set of self-consistency equations whose solution determines the statistics of the inputs in the steady state.

Effect of rewiring in the feedforward network. To gain insight, we commence by considering the effect of synaptic rewiring in the feedforward network depicted in Fig. 4. The time-averaged input to the postsynaptic neuron i in population a is given by (see equations (11) and (12)):

$$h_a^i = u_a + \eta_E^i s_{aE} + \eta_I^i s_{aI} \tag{16}$$

where η_E^i and η_I^i are two uncorrelated Gaussian variables with zero mean and unitary variance; and s_{ab}^2 is the variance in the time-averaged input due to the heterogeneity in the inputs from population b. The rewiring of synaptic connections from population b to neuron i in population a is equivalent to the resampling of the corresponding Gaussian variable η_b^i. The larger the corresponding variance, that is, s_{ab}^2, the larger the resulting change is in the time-averaged input to the postsynaptic neuron and, thus, the larger the effect of rewiring.

The variance s_{ab}^2 is given by

$$s_{ab}^2 = \langle w_{ab}^2\rangle \langle \nu_b^2\rangle - \langle w_{ab}\rangle^2 \langle \nu_b\rangle^2 = N_b \left(c_{ab} \langle W_{ab}^2\rangle \langle \nu_b^2\rangle - c_{ab}^2 \langle W_{ab}\rangle^2 \langle \nu_b\rangle^2 \right) \tag{17}$$

$$\cong c_{ab} N_b \langle W_{ab}^2\rangle \langle \nu_b^2\rangle \tag{18}$$

where we used the fact that $c_{ab}^2 \ll c_{ab}$ when $c_{ab} \ll 1$ (sparse network). Thus, as explained in the 'Results' section, the effect of synaptic rewiring can be predicted from the number of connections, $c_{ab}N_b$, the mean of the distribution of squared synaptic efficacies, $\langle W_{ab}^2\rangle$ and the mean of the distribution of squared firing rates, $\langle \nu_b^2\rangle$.

In the recurrent network, the synaptic rewiring also affects the firing rates of the presynaptic neurons. The effect of such rewiring is discussed in the next section.

Effect of rewiring in the random network. We investigate the impact of changes in the network synaptic structure on the firing rates in the recurrent network. Hereafter, we denote with a tilde (~) the quantities in the perturbed network, while those without the tilde denote quantities in the original network. We shall consider only changes that do not affect the statistical properties of the synaptic structure: the distribution of synaptic efficacies, as well as the probability of connection between any two neurons in the perturbed network, are the same as in the original. As a result, the marginal distributions of inputs in the two networks are the same (that is, $u_a = \tilde{u}_a$, $s_a^2 = \tilde{s}_a^2$, $\sigma_a^2 = \tilde{\sigma}_a^2$ for $a = E, I$). The inputs to the same neuron in the two networks are Gaussian and correlated, that is,

$$h_a^i = u_a + \eta_a^i s_a \tag{19}$$

$$\tilde{h}_a^i = u_a + (\rho_a \eta_a^i + \sqrt{1-\rho_a^2}\,\tilde{\eta}_a^i) s_a \tag{20}$$

where ρ_a denotes the correlation coefficient, and η_a^i and $\tilde{\eta}_a^i$ are uncorrelated Gaussian variables with zero mean and unitary variance. The correlation coefficient can be computed by evaluating $\langle h_a^i \tilde{h}_a^i\rangle$. From equations (19) and (20)

$$\rho_a s_a^2 = h_a^i \tilde{h}_a^i - u_a^2 \tag{21}$$

where u_a is defined in equation (11). We now evaluate $\langle h_a^i \tilde{h}_a^i\rangle$

$$\langle h_a^i \tilde{h}_a^i\rangle = \left(\sqrt{N_E} h_a^{(ext)}\right)^2 + \sqrt{N_E} h_a^{(ext)} \cdot \left\langle \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} w_{aE}^{ij} \nu_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} w_{aI}^{ij} \nu_I^j \right\rangle \tag{22}$$

$$+ \sqrt{N_E} h_a^{(ext)} \cdot \left\langle \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} \tilde{w}_{aE}^{ij} \tilde{\nu}_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} \tilde{w}_{aI}^{ij} \tilde{\nu}_I^j \right\rangle \tag{23}$$

$$-\left\langle \frac{1}{\sqrt{N_E N_I}} \sum_{j=1}^{N_E} w_{aE}^{ij} v_E^j \cdot \sum_{k=1}^{N_I} \tilde{w}_{aI}^{ik} \tilde{v}_I^k \right\rangle \tag{24}$$

$$-\left\langle \frac{1}{\sqrt{N_E N_I}} \sum_{j=1}^{N_E} \tilde{w}_{aE}^{ij} \tilde{v}_E^j \cdot \sum_{k=1}^{N_I} w_{aI}^{ik} v_I^k \right\rangle \tag{25}$$

$$+\left\langle \frac{1}{N_E} \sum_{j,k=1}^{N_E} w_{aE}^{ij} \tilde{w}_{aE}^{ik} v_E^j \tilde{v}_E^k \right\rangle + \left\langle \frac{1}{N_I} \sum_{j,k=1}^{N_E} w_{aI}^{ij} \tilde{w}_{aI}^{ik} v_I^j \tilde{v}_I^k \right\rangle \tag{26}$$

All terms except the last two can be straightforwardly evaluated by noticing that (i) the w and the v values are independent because of the random connectivity; (ii) all variables with the tilde and without the tilde are uncorrelated at different sites (that is, for $i \neq j$ and/or for the different populations, E and I); (iii) all moments of the variables with the tilde are the same as those of the variables without the tilde (for example, $\langle w_{ab}^{ij} \rangle = \langle \tilde{w}_{ab}^{ij} \rangle$ and $\langle v_a^j \rangle = \langle \tilde{v}_a^j \rangle$). For the last two terms, we separate in the sum terms with $j=k$ from terms with $j \neq k$

$$\left\langle \frac{1}{N_E} \sum_{j,k=1}^{N_E} w_{aE}^{ij} \tilde{w}_{aE}^{ik} v_E^j \tilde{v}_E^k \right\rangle = \tag{27}$$

$$\left\langle \frac{1}{N_E} \sum_{j=1}^{N_E} w_{aE}^{ij} \tilde{w}_{aE}^{ij} v_E^j \tilde{v}_E^j \right\rangle + \left\langle \frac{1}{N_E} \sum_{j \neq k=1}^{N_E} w_{aE}^{ij} \tilde{w}_{aE}^{ik} v_E^j \tilde{v}_E^k \right\rangle = \tag{28}$$

$$\langle w_{aE} \tilde{w}_{aE} \rangle \langle v_E \tilde{v}_E \rangle + (N_E - 1) \langle w_{aE} \rangle^2 \langle v_E \rangle^2 \tag{29}$$

where

$$\langle w_{aE} \tilde{w}_{aE} \rangle = \frac{1}{N_E} \sum_{j=1}^{N_E} w_{aE}^{ij} \tilde{w}_{aE}^{ij} \tag{30}$$

Similarly,

$$\left\langle \frac{1}{N_I} \sum_{j,k=1}^{N_I} w_{aI}^{ij} \tilde{w}_{aI}^{ik} v_I^j \tilde{v}_I^k \right\rangle = \langle w_{aI} \tilde{w}_{aI} \rangle \langle v_I \tilde{v}_I \rangle + (N_I - 1) \langle w_{aI} \rangle^2 \langle v_I \rangle^2 \tag{31}$$

Putting all together, one obtains

$$\rho_a s_a^2 = \sum_{b=E,I} \left(-\langle w_{ab} \rangle^2 \langle v_b \rangle^2 + \langle w_{ab} \tilde{w}_{ab} \rangle \langle v_b \tilde{v}_b \rangle \right) \tag{32}$$

Knowing $\rho_a s_a^2$, one can compute $v_a \tilde{\mathbf{v}}_a$ in the following way

$$\langle v_a \tilde{v}_a \rangle = \int D\eta D\tilde{\eta} \phi_a (u_a + \eta \cdot s_a) \phi_a \left(u_a + (\rho_a \eta + \sqrt{1 - \rho_a^2}\, \tilde{\eta}) s_a \right) \tag{33}$$

Incorporating equation (33) into equation (32), one obtains a set of self-consistent equations whose solution determines ρ_a ($a = E, I$) as a function of the network parameters and of the synaptic perturbation.

Mean-field analysis of the storage capacity. We are interested in the existence of steady retrieval states (to be defined shortly) as a function of the loading level, that is, the number of memories per neuron. A memory is defined by a binary vector $\{\xi_E^i(\mu); \xi_I^i(\mu)\}$, defined in the section 'The structured network,' where we also describe the embedding of memories into the synaptic structure.

The memory μ is being successfully retrieved if there exists a steady state of activity of the network such that the average activity level of the neurons belonging to the memory (that is, the neurons for which $\xi_a^i(\mu) = 1$) is larger than the global average activity level of the network. More precisely, we define

$$m_a(\mu) = \frac{1}{f N_a} \sum_{j=1}^{N_a} \xi_a^j(\mu) v_a^j - \frac{1}{N_a} \sum_{j=1}^{N_a} v_a^j \tag{34}$$

so that memory μ is being successfully retrieved by population a if $m_a(\mu) > 0$. Following a time-honored tradition, we choose memory 1 and study under which conditions the corresponding retrieval state exists. The time-averaged input to neuron i in population a is given by

$$h_a^i = \sqrt{N_E} h_a^{(ext)} + \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} F_{aE}(z_{aE}^{ij}) v_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} F_{aI}(-z_{aI}^{ij}) v_I^j \tag{35}$$

for $a = E, I$. To extract the specific contribution of memory 1 to synaptic structuring, we Taylor expand the function $F_{aE}(\cdot)$ as follows

$$z_{ab}^{ij} = \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \epsilon_{ab}^{ij}(1) (\xi_a^i(1) - f) (\xi_b^j(1) - f) + \tag{36}$$

$$\frac{\sqrt{2}}{f(1-f)\sqrt{P}} \sum_{\mu>1}^{P} \epsilon_{ab}^{ij}(\mu) (\xi_a^i(\mu) - f) (\xi_b^j(\mu) - f) = \tag{37}$$

$$= \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \epsilon_{ab}^{ij}(1) (\xi_a^i(1) - f) (\xi_b^j(1) - f) + \tilde{z}_{ab}^{ij} + O\left(\frac{1}{P}\right) \tag{38}$$

where

$$\tilde{z}_{ab}^{ij} \equiv \frac{\sqrt{2}}{f(1-f)\sqrt{P-1}} \sum_{i>1}^{P} \epsilon_{ab}^{ij}(\mu) (\xi_a^i(\mu) - f) (\xi_b^j(\mu) - f) \tag{39}$$

and noting that

$$\sum_{\mu>1}^{P} \epsilon_{ab}^{ij}(\mu) (\xi_a^i(\mu) - f) (\xi_b^j(\mu) - f) = O(\sqrt{P-1}) \tag{40}$$

Hence,

$$F_{ab}(z_{ab}^{ij}) = F_{ab}(\tilde{z}_{ab}^{ij}) + F'_{ab}(\tilde{z}_{ab}^{ij}) \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \epsilon_{ab}^{ij} (\xi_a^i - f) (\xi_b^j - f) + O\left(\frac{1}{P}\right) \tag{41}$$

where we have dropped the dependence on μ, being understood that we are considering memory 1. Incorporating equation (41) in equation (35) yields

$$h_a^i = \sqrt{N_E} h_a^{(ext)} + \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} F_{aE}(\tilde{z}_{aE}^{ij}) v_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} F_{aI}(-\tilde{z}_{aI}^{ij}) v_I^j + \tag{42}$$

$$\frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} \sqrt{\frac{2N_E}{N_E + N_I}} \frac{1}{N_E} \sum_{j=1}^{N_E} \epsilon_{aE}^{ij} (\xi_E^j - f) F'_{aE}(\tilde{z}_{aE}^{ij}) v_E^j + \tag{43}$$

$$\frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} \sqrt{\frac{2N_I}{N_E + N_I}} \frac{1}{N_I} \sum_{j=1}^{N_I} \epsilon_{aI}^{ij} (\xi_I^j - f) F'_{aI}(-\tilde{z}_{aI}^{ij}) v_I^j \tag{44}$$

where we defined $\alpha = P/(N_E + N_I)$. We can use the central limit theorem to estimate the different sums over j in the large N limit. Note that, in this limit, the statistics of the $\tilde{z}_{aE}^{ij}$ values become independent of the specific realization of the memory patterns $\xi_a^i(\mu)$, $\mu > 1$, at neuron i (that is, it is self-averaging[58]). Thus,

$$\sqrt{N_E} h_a^{(ext)} + \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} F_{aE}(\tilde{z}_{aE}^{ij}) v_E^j - \frac{1}{\sqrt{N_I}} \sum_{j=1}^{N_I} F_{aI}(-\tilde{z}_{aI}^{ij}) v_I^j \to u_a + \eta_a^i s_a \tag{45}$$

where, as in equations (11) and (12),

$$u_a = \sqrt{N_E} \cdot \left(h_a^{(ext)} + \langle w_{aE} \rangle \langle v_E \rangle - \frac{\sqrt{N_I}}{\sqrt{N_E}} \langle w_{aI} \rangle \langle v_I \rangle \right) \tag{46}$$

$$s_a^2 = \sum_{b=E,I} \left(\langle w_{ab}^2 \rangle \langle v_b^2 \rangle - \langle w_{ab} \rangle^2 \langle v_b \rangle^2 \right) \tag{47}$$

Equations (43) and (44) become

$$\frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} \sqrt{\frac{2N_E}{N_E+N_I}} \frac{1}{N_E} \sum_{j=1}^{N_E} \epsilon_{aE}^{ij} (\xi_E^j - f) F'_{aE}(\tilde{z}_{aE}^{ij}) v_E^j \rightarrow \frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} A_{aE} m_E \quad (48)$$

$$\frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} \sqrt{\frac{2N_I}{N_E+N_I}} \frac{1}{N_I} \sum_{j=1}^{N_I} \epsilon_{aI}^{ij} (\xi_I^j - f) F'_{aI}(-\tilde{z}_{aI}^{ij}) v_I^j \rightarrow \frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} A_{aI} m_I \quad (49)$$

with

$$A_{aE} = \sqrt{\frac{N_E}{2(N_E+N_I)}} \int Dz F'_{aE}(z); \; A_{aI} = \sqrt{\frac{N_I}{2(N_E+N_I)}} \int Dz F'_{aI}(-z) \quad (50)$$

Putting all together we obtain

$$h_a^i = u_a + \eta_a^i s_a + \frac{\xi_a^i - f}{f(1-f)\sqrt{\alpha}} \sum_b A_{ab} m_b \quad (51)$$

This mean field theory obtained is similar to the one obtained in the section "Mean-field analysis of the random network" but with two additional order parameters, m_E and m_I, which describe the retrieval states ($m_a > 0$).The equations that determine the first two moments of the firing rates $\langle v_a^n \rangle$ ($n = 1, 2$) and m_a are, similar to equations (14) and (15)

$$\langle v_a^n \rangle = \left\langle \int D\eta \left(\phi_a \left(u_a + \eta \cdot s_a + \frac{\xi - f}{f(1-f)\sqrt{\alpha}} \sum_b A_{ab} m_b \right) \right)^n \right\rangle_\xi \quad (52)$$

and

$$\langle m_a \rangle = \frac{1}{f} \left\langle \int D\eta \xi \phi_a \left(u_a + \eta \cdot s_a + \frac{\xi - f}{f(1-f)\sqrt{\alpha}} \sum_b A_{ab} m_b \right) \right\rangle_\xi - \langle v_a \rangle \quad (53)$$

It is easy to see that the mean-field equations always have a solution with $m_a = 0$, while retrieval solutions exist only for suitably small α. The critical capacity, α_c, is defined as the largest α for which the retrieval solution still exists.

Effect of rewiring on memory storage. The rewiring of a synaptic population (for example, E → E) removes the correlation between the synaptic efficacies in that population and the memories stored. In the mean-field theory, this is equivalent to setting the corresponding signal term (that is, the A_{ab} values) to zero. Thus, to investigate the effect of E → E rewiring on memory capacity, we solve the mean-field equations with $A_{EE} = 0$. To investigate the effect of I → I rewiring, we solve the mean-field equations with $A_{II} = 0$. Similarly, to compute the storage capacity of the E → E synapses, we solve the mean-field equations with all signal terms set to zero but A_{EE}.

It is important to note that, for the parameters used, the signal terms associated with excitatory synapses (that is, A_{aE}) are larger than those associated with the inhibitory synapses ($A_{EE} = 1.44$, $A_{IE} = 3.28$, $A_{EI} = 0.75$ and $A_{II} = 0.84$). Thus, the larger memory capacity associated with the inhibitory synapses is not simply due to stronger dependence of the efficacies on the memory patterns.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The code for numerical simulations is available from the authors upon reasonable request.

Data availability

The dataset analyzed in the current study is available in http://bio.huji.ac.il/yonatanlab/spines/.

References

51. Feng, G. et al. Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
52. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
53. Trachtenberg, J. T. et al. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
54. Keck, T. et al. Massive restructuring of neuronal circuits during functional reorganization of adult visual cortex. *Nat. Neurosci.* **11**, 1162–1167 (2008).
55. Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
56. Markram, H. et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
57. Renart, A., Brunel, N. & Wang, X. *Computational Neuroscience: A Comprehensive Approach*. (CRC Press, Boca Raton, FL, USA,, 2003). Chapter 15.
58. Amit, D. J., Gutfreund, H. & Sompolinsky, H. Spin-glass models of neural networks. *Phys. Rev. A. Gen. Phys.* **32**, 1007–1018 (1985).

nature research

Corresponding author(s): Yonatan Loewenstein

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No new data was collected
Data analysis	MATLAB (The MathWorks)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

In the Methods section we state that the full dataset is publicly available at http://bio.huji.ac.il/yonatanlab/spines/

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A - This study is based on previously- published data
Data exclusions	No data were excluded
Replication	No replication was attempted
Randomization	Not relevant - there were no experimental groups
Blinding	Not relevant - there were no experimental groups

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Male adult (~6 months old) in-house bred mice of the GFP-M transgenic line (Tg(Thy1-EGFP)MJrs/J)
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve samples collected from the field