Supplementary Notes

1 Statistics of connectivity in networks storing fixed-point attractors

To compute the distribution of synaptic weights, we first note that the problem of learning p patterns in an attractor neural network is equivalent to N perceptron problems. Each neuron i has to learn p associations between the states of its inputs in pattern μ (η_j^μ, $j \neq i$), and its state in that pattern η_i^μ. Thus, the distribution of synaptic weights in an attractor network is identical to the distribution in a perceptron, in which the output coding level is equal to the input coding level. The distribution of synaptic weights in a perceptron with excitatory weights has been calculated by Kohler and Widmaier (1991, J. Phys. A 24, L495-L502) and Brunel et al (2004) [32] (see also refs[33,31,42] for generalizations of such calculations). Here, for the sake of completeness we describe below the calculation using the replica method. We also derive the distribution using another, more transparent, method: the cavity method[29] (Mézard (1989) J. Phys. A 22, 2181-2190), which gives a more intuitive understanding of the results. Furthermore, the cavity method allows us to compute joint distributions of sets of synaptic weights (section 1.2.3).

1.1 The replica method

The approach introduced by Gardner[26] (see also Shcherbina and Tirozzi (2003) Commun. Math. Phys. 234, 383-422) consists in computing the typical volume of the space of solutions to the learning problem. The capacity is then obtained as the value of p for which the volume vanishes. Distributions of relevant quantities (stabilities, weights) are averaged over the space of solutions.

We focus here on the space of weights of one particular neuron, k. The synaptic weights connecting other neurons in the network to neuron k are $w_i \equiv w_{ki}$. In the large N limit, $w_{ij} \sim O(1)$ leads to $T = O(N)$ and $K = O(\sqrt{N})$. We therefore rescale $T = N\theta$, $K = \sqrt{N}\kappa$ where θ and κ are of order 1 in the large N limit.

To compute the typical volume of the space of weights that satisfy learning of all p patterns for this particular neuron, one needs to compute the average of the logarithm of the volume over the distribution of patterns[26]. In practice, the calculation of this average is done using the replica method. One first computes the volume of n replicas of the system,

$$V^n = \int dw \prod_{\mu,a} \Theta(\Delta^{\mu a} - \kappa),$$

Nature Neuroscience: doi:10.1038/nn.4286

where $\Delta^{\mu a}$ is the *stability* of pattern μ for neuron k in replica a,

$$\Delta^{\mu a} = \frac{\xi^\mu}{\sqrt{N}} \left(\sum_i w_i^a \eta_i^\mu - N\theta \right)$$

where $\xi^\mu = \xi_k^\mu$, and w_i^a is the synaptic weight from neuron k to neuron i in replica a.

Averaging over patterns,

$$\langle V^n \rangle = \int dw \langle \prod_{\mu,a} \Theta(\Delta^{\mu a} - \kappa) \rangle,$$

and taking the limit $n \to 0$, one eventually obtains the desired quantity,

$$\langle \log V \rangle = \lim_{n\to 0} (\langle V^n \rangle - 1)/n. \tag{1}$$

The averaging over patterns is performed using integral representations for the Heaviside functions,

$$\Theta(\Delta^{\mu a} - \kappa) = \int dx^{\mu a} \int_\kappa^\infty dy^{\mu a} \exp(ix^{\mu a}(y^{\mu a} - \Delta^{\mu a})),$$

which gives

$$\begin{aligned} \langle V^n \rangle &= \int dw dx dy \exp \left[\sum_{\mu,a} ix^{\mu a} \left(y^{\mu a} + \sqrt{N}\theta\xi^\mu - \frac{f}{\sqrt{N}} \sum_j w_j^a \right) \right. \\ &\qquad \left. - \frac{f(1-f)}{2N} \sum_\mu \left(\sum_a w_j^a x^{\mu a} \right)^2 \right] \end{aligned} \tag{2}$$

The next step is to introduce order parameters,

$$\frac{1}{N} \sum_j w_j^a = \frac{\theta}{f} + \frac{M^a}{\sqrt{N}} \equiv \overline{w} + \frac{M^a}{\sqrt{N}} \tag{3}$$

$$\frac{1}{N} \sum_j \left(w_j^a \right)^2 = Q^a \tag{4}$$

$$\frac{1}{N} \sum_j w_j^a w_j^b = q^{ab}, \tag{5}$$

together with conjugate parameters $\hat{M}^a$, $\hat{Q}^a$ and $\hat{q}^{ab}$. The following steps are to use a replica-symmetric ansatz, and perform the limit $n \to 0$. This leads to

$$\langle V^n \rangle = \int d \ldots \exp(NnF) \tag{6}$$

$$\begin{aligned} F &= -\hat{Q}Q + \frac{1}{2}\hat{q}q + \overline{w}\hat{M} \\ &\quad + \int_{-\infty}^{+\infty} Du \log \int_0^\infty dw \exp \left[(\hat{Q} - \frac{\hat{q}}{2}) w^2 + w(u\sqrt{\hat{q}} - \hat{M}) \right] \\ &\quad + \alpha \sum_{\xi=\pm 1} p_\xi \int_{-\infty}^{+\infty} Du \log H \left[\frac{\kappa - \xi f M + u\sqrt{qf(1-f)}}{\sqrt{f(1-f)(Q-q)}} \right] \end{aligned} \tag{7}$$

Nature Neuroscience: doi:10.1038/nn.4286

where $Du = du \exp(-u^2/2)/\sqrt{2\pi}$ and $H(x) = \int_x^\infty Du$.

In the large N limit, the integral in Eq. (6) is dominated by the saddle point, given by the equations

$$\overline{w} = \int_{-\infty}^{+\infty} Du \frac{\int_0^\infty dw w \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]}{\int_0^\infty dw \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]} \tag{8}$$

$$Q = \int_{-\infty}^{+\infty} Du \frac{\int_0^\infty dw w^2 \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]}{\int_0^\infty dw \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]} \tag{9}$$

$$q = \int_{-\infty}^{+\infty} Du \frac{\int_0^\infty dw (w^2 - wu/\sqrt{\hat{q}}) \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]}{\int_0^\infty dw \exp\left[(\hat{Q} - \frac{\hat{q}}{2})w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]} \tag{10}$$

$$0 = \sum_\xi p_\xi \xi \int Du \frac{G(a_\xi(u))}{H(a_\xi(u))} \tag{11}$$

$$\hat{Q} = \frac{\alpha}{2} \sum_\xi p_\xi \int Du \frac{a_\xi(u)}{(Q-q)} \frac{G(a_\xi(u))}{H(a_\xi(u))} \tag{12}$$

$$\hat{q} = \alpha \sum_\xi p_\xi \int Du \left(\frac{u}{\sqrt{q(Q-q)}} + \frac{a_\xi(u)}{(Q-q)} \right) \frac{G(a_\xi(u))}{H(a_\xi(u))} \tag{13}$$

$$a_\xi(u) = \sqrt{\frac{q}{Q-q}}(u - \tau_\xi) \tag{14}$$

$$\tau_\xi = -\frac{\kappa - \xi f M}{\sqrt{qf(1-f)}} \tag{15}$$

where $G(x) = \exp(-x^2/2)/\sqrt{2\pi}$.

At maximal capacity, $\alpha = \alpha_c$, $q \to Q$. In that limit, we rewrite

$$2\hat{Q} \sim \hat{q} \sim \frac{C}{(Q-q)^2} \tag{16}$$

$$\hat{q} - 2\hat{Q} \sim \frac{A}{Q-q} \tag{17}$$

$$\hat{M} \sim \frac{B\sqrt{C}}{Q-q}. \tag{18}$$

Saddle point equations give in that limit

$$\overline{w} = \frac{\sqrt{C}}{A} (G(B) - BH(B)) \tag{19}$$

$$Q = \frac{C}{A^2} \Big((1+B^2)H(B) - BG(B)\Big) \tag{20}$$

$$A = H(B) \tag{21}$$

Nature Neuroscience: doi:10.1038/nn.4286

$$0 = \sum_\xi p_\xi \xi \left(G(\tau_\xi) - \tau_\xi H(\tau_\xi)\right) \tag{22}$$

$$C = \alpha_c Q \sum_\xi p_\xi \left((1+\tau_\xi^2)H(\tau_\xi) - \tau_\xi G(\tau_\xi)\right) \tag{23}$$

$$A = \alpha_c \sum_\xi p_\xi H(\tau_\xi) \tag{24}$$

These equations can be solved as follows. For a given

$$\tilde{\kappa} = \kappa/\sqrt{Qf(1-f)},$$

we obtain $\tilde{M} = M/\sqrt{Qf(1-f)}$ by solving Eq. (22). This gives us the τ_ξs. Then, we can obtain B through the relationship

$$\frac{(1+B^2)H(B) - BG(B)}{H(B)} = \frac{\sum_\xi p_\xi H(\tau_\xi)}{\sum_\xi p_\xi \left((1+\tau_\xi^2)H(\tau_\xi) - \tau_\xi G(\tau_\xi)\right)} \tag{25}$$

obtained from combining Eqs. (20,21,23). Since we already know $\overline{w} = \theta/f$, we can now compute all other parameters Q, A, C, and α_c. In particular, for $\kappa = 0$, $f = 0.5$, we get $M = 0$, $B = 0$, $A = 0.5$, and recover the well-known result $\alpha_c = 1$ (Amit et al (1989), J. Phys. A 22, 4687-4693).

The replica method can also be used to compute the distribution of stabilities (Kepler and Abbott (1988), J. Phys. France 49 1657-1662) and the distribution of weights (Kohler and Widmaier (1991), J. Phys. A, 24, L495-L502)[32]. The distribution of stabilities is given at maximal capacity by

$$P(\Delta) = \sum_\xi p_\xi \left[\frac{G\left(\frac{\Delta - \xi f M}{\sqrt{f(1-f)Q}}\right)}{\sqrt{f(1-f)Q}}\Theta(\Delta - K) + H(\tau_\xi)\delta(\Delta - K)\right] \tag{26}$$

Note that the fraction of saturated constraints is $\sum_\xi p_\xi H(\tau_\xi)$.

The distribution of weights is

$$Q(w) = \Theta(w)\int_{-\infty}^{+\infty} Du \frac{\exp\left[(\hat{Q} - \frac{\hat{q}}{2})\, w^2 + w(u\sqrt{\hat{q}} - \hat{M})\right]}{\int_0^{+\infty} dw' \exp\left[(\hat{Q} - \frac{\hat{q}}{2})\, w'^2 + w'(u\sqrt{\hat{q}} - \hat{M})\right]} \tag{27}$$

At maximal capacity the distribution becomes

$$Q_c(w) = H(-B)\delta(w) + \frac{1}{\sqrt{2\pi}w_s}\exp\left[-\frac{1}{2w_s^2}\left(w + Bw_s\right)^2\right]\Theta(w) \tag{28}$$

where

$$w_s = \frac{\sqrt{C}}{A} = \frac{\overline{w}}{G(B) - BH(B)} \tag{29}$$

Nature Neuroscience: doi:10.1038/nn.4286

Note that the fraction of zero weight synapses is $H(-B)$; while the fraction of strictly positive synapses (in other words, the connection probability) is $H(B)$. Interestingly, the fraction of strictly positive synapses is related to the fraction of saturated constraints through the relationship

$$H(B) = \alpha \sum_{\xi} p_\xi H(\tau_\xi)$$

In other words, the number of strictly positive synapses is equal to the number of saturated constraints[1].

Note that the distribution of weights can be computed both below capacity (for $\alpha < \alpha_c$), using Eq. (27) where $\hat{Q}$, $\hat{q}$, $\hat{M}$ are obtained solving numerically Eqs. (8-13), and at maximal capacity (for $\alpha = \alpha_c$), using Eq. (28) where B and w_s are obtained as outlined above [32]. How the distribution changes as a function of storage capacity is illustrated in Supplementary Figure 1.

1.2 The cavity method

The cavity method was introduced by physicists working on disordered systems as an alternative, more transparent method than the replica method[29]. It has been applied to a wide range of problems, including combinatorial optimization problems. In the following we apply the cavity method to our network, along the lines of Mézard's calculation for the unconstrained perceptron (Mézard 1989, J. Phys. A 22, 2181-2190). The calculation proceeds in two steps: we first introduce a new pattern, and compute the distribution of stabilities for this new pattern; then, we introduce a new weight, and compute the distribution of this weight. The calculation will then be generalized to joint distributions of pairs of weights.

1.2.1 Distribution of stabilities

We assume that a network of N neurons has already learned p patterns. We add a new randomly drawn pattern ($\vec{\eta}$) to the set of patterns to be learned. We focus on one particular neuron k and define for convenience $\xi = 2\eta_k - 1$, $p_\xi = f\xi + (1-\xi)/2$. Its associated stability is

$$\Delta = \frac{\xi}{\sqrt{N}} \left(\sum_i w_i \eta_i - N\theta \right) \tag{30}$$

where w_i is the weight from neuron i to k.

We rewrite Eq. (30) as

$$\Delta = \frac{\xi}{\sqrt{N}} \left(\sum_i w_i(\eta_i - f) - \left[N\theta - f \sum_i w_i \right] \right)$$

[1] I am indebted to Peter Dayan for this observation.

Nature Neuroscience: doi:10.1038/nn.4286

We first consider the distribution of Δ over the space of weights satisfying the previous p constraints. This distribution is a Gaussian, with mean

$$h = \frac{\xi}{\sqrt{N}} \sum_i \langle w_i \rangle (\eta_i - f) + \xi f M \tag{31}$$

where $\langle . \rangle$ denotes 'thermal' averages, i.e. averages over the allowed space of weights, and we have introduced $f\sqrt{N}M = f \sum_i \langle w_i \rangle - N\theta$.

Its variance is

$$\sigma_h^2 = \frac{1}{N} \left[\sum_i \left(\left\langle w_i^2 \right\rangle - \langle w_i \rangle^2 \right) \left(\eta_i (1 - 2f) + f^2 \right) \right]$$

Learning this pattern consists in removing from the space of weights those weights leading to a stability lower than κ. Therefore, the distribution of stabilities after learning is a truncated Gaussian,

$$P(\Delta, h) = \frac{1}{\sigma_h} \frac{G\left(\frac{\Delta - h}{\sigma_h}\right)}{H\left(\frac{\kappa - h}{\sigma_h}\right)} \Theta(\Delta - \kappa). \tag{32}$$

where $G(x) = \exp(-x^2/2)/\sqrt{2\pi}$, $H(x) = \int_x^\infty G(u)du$.

The next step is to average over the distribution of patterns. These averages are denoted by $\overline{X}$. The first two moments of h and the average of σ_h^2 are:

$$\begin{aligned} \overline{h} &= \xi f M & (33) \\ \overline{\left(h - \overline{h}\right)^2} &= \frac{f(1-f)}{N} \sum_i \overline{\langle w_i \rangle^2} & (34) \\ \overline{\sigma_h^2} &= \frac{f(1-f)}{N} \sum_i \overline{\langle w_i^2 \rangle} - \overline{\langle w_i \rangle^2} & (35) \end{aligned}$$

To simplify these expressions we introduce the order parameters

$$\begin{aligned} \frac{1}{N} \sum_i \overline{\langle w_i^2 \rangle} &= Q & (36) \\ \frac{1}{N} \sum_i \overline{\langle w_i \rangle^2} &= q & (37) \end{aligned}$$

where for obvious reasons we use the same notations as in the replica method.

Using these order parameters, we obtain

$$\begin{aligned} \overline{(h - \xi f M)^2} &= q f (1 - f) & (38) \\ \overline{\sigma_h^2} &= f(1-f)(Q - q) & (39) \end{aligned}$$

h is distributed as a Gaussian with mean $\xi f M$ and variance $qf(1-f)$, while σ_h^2 has a mean $f(1-f)(Q-q)$ with a variance that goes to zero in the thermodynamical limit.

Nature Neuroscience: doi:10.1038/nn.4286

Using the above results, we can write the sample-averaged distribution of stabilities as

$$\overline{P(\Delta)} = \sum_{\xi=\pm 1} p_\xi \int \frac{dh}{\sqrt{2\pi q f(1-f)}} G\left(\frac{h - \xi f M}{\sqrt{q f(1-f)}}\right) P(\Delta, h) \tag{40}$$

which coincides with the expression obtained using the replica method. At maximal capacity, the space of available synaptic connectivities shrinks to a single point, and consequently $q \to Q$. To obtain the distribution at maximal capacity, we thus take the limit $q \to Q$. In this limit, σ_h goes to zero. Therefore, $P(\Delta, h)$ becomes a delta function which is either peaked on h if $h > \kappa$, or on κ if $h < \kappa$. This leads to the distribution of stabilities obtained using the replica method, Eq. (26). Intuitively, the truncated Gaussian corresponds to those patterns that do not need any synaptic change to be learned, while the delta function in κ corresponds to those patterns that needed synaptic change to be learned.

1.2.2 Distribution of synaptic weights

We next turn to the distribution of synaptic weights. We add a single neuron, $i = 0$, with its associated weight w_0, to the set of neurons that send inputs to neuron k. The values of the patterns for this new neuron are η_0^μ, $\mu = 1, \ldots, p$. This changes slightly the stability of each of the patterns: $\Delta^\mu \to \Delta^\mu + \epsilon^\mu$ where

$$\epsilon^\mu = \frac{\xi^\mu}{\sqrt{N}} w_0 (\eta_0^\mu - f)$$

Assuming the joint distribution of stabilities approximately factorizes in the product of the distributions of individual stabilities (Mézard (1989) J. Phys. A 22 2191-2190), the distribution of weights w_0 satisfying all the p constraints, averaged over the space of all the other couplings, is

$$Q(w_0) \propto \int \prod_\mu P(\Delta^\mu, h^\mu) d\Delta^\mu \Theta\left(\Delta^\mu + \epsilon^\mu - \kappa\right) \Theta(w_0) \exp(-c w_0)$$

where $P(\Delta^\mu, h^\mu)$ is given by Eq. (32), in which h^μ is given by Eq. (31), replacing (ξ, η_i) by (ξ^μ, η_i^μ), and $\Theta(w_0)\exp(-cw_0)$ is a prior distribution which enforces the constraint $w > 0$, and c is a constant that will be determined self-consistently in the following. Expanding the r.h.s. in ϵ gives

$$Q(w_0) \propto \exp\left((a - c) w_0 - \frac{b}{2} w_0^2\right) \Theta(w_0)$$

where

$$a = \frac{1}{\sqrt{N}} \sum_\mu \xi^\mu (\eta_0^\mu - f) P(\kappa, h^\mu) \tag{41}$$

$$b = \frac{1}{N} \sum_\mu \left(\eta_0^\mu (1 - 2f) + f^2\right) \left(P(\kappa, h^\mu)^2 + \frac{\partial P}{\partial \Delta}(\kappa, h^\mu)\right) \tag{42}$$

Nature Neuroscience: doi:10.1038/nn.4286

Hence, we get again a truncated Gaussian. a and b depend in particular on the properties of the distribution of stabilities at κ, which makes sense: only the patterns whose stability is close to κ constrain the choice of w_0. The sign of a also tells us whether the weight would tend to be positive or negative in absence of the constraint: inspecting the r.h.s. of Eq. (41), we see that the more associations leading to positive output in which this synapse is activated, the more positive it will tend to be, as one would expect.

Let us now compute the statistics of a and b over the distribution of patterns. a is a Gaussian random variable with zero mean and variance σ_a^2, while b becomes in the thermodynamical limit equal to its average value $\bar{b}$. Using Eq. (32), writing $h^\mu = \xi^\mu f M - u\sqrt{qf(1-f)}$ where u is a Gaussian variable with zero mean and unit variance, and replacing $\sum_\mu$ by $\alpha N \sum_\xi p_\xi \int Du$ leads to

$$\sigma_a^2 = \alpha f(1-f)\sum_\xi p_\xi \int Du \frac{1}{\sigma_h^2}\frac{G\left(a_\xi(u)\right)^2}{H\left(a_\xi(u)\right)^2} \tag{43}$$

$$\bar{b} = \alpha f(1-f)\sum_\xi p_\xi \int Du \frac{1}{\sigma_h^2}\frac{G\left(a_\xi(u)\right)^2}{H\left(a_\xi(u)\right)^2} - \frac{a_\xi(u)}{\sigma_h^2}\frac{G\left(a_\xi(u)\right)}{H\left(a_\xi(u)\right)} \tag{44}$$

$$a_\xi(u) = \frac{\kappa - \xi f M + u\sqrt{qf(1-f)}}{\sqrt{f(1-f)(Q-q)}} \tag{45}$$

The averaged distribution of weights is

$$\overline{Q(w_0)} = \int Du \frac{\exp\left(-\frac{\bar{b}}{2}w_0^2 + w_0\left(-c + u\sigma_a\right)\right)\Theta(w_0)}{\int_0^{+\infty} dw \exp\left(-\frac{\bar{b}}{2}w^2 + w\left(-c+u\sigma_a\right)\right)} \tag{46}$$

i.e. the same equation as Eq. (27), provided we have

$$\bar{b} = \hat{q} - 2\hat{Q} \sim \frac{A}{Q-q} \tag{47}$$

$$c = \hat{M} \sim \frac{B\sqrt{C}}{Q-q} \tag{48}$$

$$\sigma_a^2 = \hat{q} \sim \frac{C}{(Q-q)^2} \tag{49}$$

In the $q \to Q$ limit, Eqs. (43,44) become

$$C = \alpha Q \sum_\xi p_\xi \left((1+\tau_\xi^2)H(\tau_\xi) - \tau_\xi G(\tau_\xi)\right) \tag{50}$$

$$A = \alpha \sum_\xi p_\xi H(\tau_\xi) \tag{51}$$

where τ_ξ is given by Eq. (15). These are the same equations as those obtained using the replica method, Eqs. (24,23).

Nature Neuroscience: doi:10.1038/nn.4286

1.2.3 Joint distributions of synaptic weights

The cavity method can be used to compute the joint distribution of arbitrary n-tuples of weights. We focus here on the simplest case of the distribution of pairs of weights.

We consider a pair of neurons, i and j, and add a new pair of weights w_{ij} and w_{ji} to the network. These weights will change slightly the stability of each of the patterns at the corresponding sites, $\Delta_i^\mu \to \Delta_i^\mu + \epsilon_{ij}^\mu$, $\Delta_j^\mu \to \Delta_j^\mu + \epsilon_{ji}^\mu$ where

$$\epsilon_{ij}^\mu = \frac{\xi_i^\mu}{\sqrt{N}} w_{ij}(\eta_j^\mu - f) \tag{52}$$

$$\epsilon_{ji}^\mu = \frac{\xi_j^\mu}{\sqrt{N}} w_{ji}(\eta_i^\mu - f) \tag{53}$$

Assuming again that the joint distribution of stabilities approximately factorizes in the product of the distributions of individual stabilities, the joint distribution of weights satisfying all the p constraints, averaged over the space of all the other couplings, is

$$Q(w_{ij}, w_{ji}) \propto \int \prod_{\mu,k=i,j} P(\Delta_k^\mu, h_k^\mu) d\Delta_k^\mu \Theta\left(\Delta_k^\mu + \epsilon_{kk'} - \kappa\right) \Theta(w_{kk'}) \exp(-c w_{kk'})$$

where $k' = j$ if $k = i$, while $k' = i$ if $k = j$, $P(\Delta_k^\mu, h_k^\mu)$ is given by Eq. (32), in which h_k^μ is given by Eq. (31), replacing (ξ, η_i) by $(\xi_k^\mu, \eta_{k'}^\mu)$. Expanding the r.h.s. in ϵ gives

$$Q(w_{ij}, w_{ji}) \propto \exp\left((a_{ij} - c)w_{ij} + (a_{ji} - c)w_{ji} - \frac{b_{ij}}{2} w_{ij}^2 - \frac{b_{ji}}{2} w_{ji}^2\right) \Theta(w_{ij})\Theta(w_{ji})$$

where

$$a_{ij} = \frac{1}{\sqrt{N}} \sum_\mu \xi_i^\mu (\eta_j^\mu - f) P(\kappa, h_i^\mu) \tag{54}$$

$$a_{ji} = \frac{1}{\sqrt{N}} \sum_\mu \xi_j^\mu (\eta_i^\mu - f) P(\kappa, h_j^\mu) \tag{55}$$

$$b_{ij} = \frac{1}{N} \sum_\mu \left(\eta_j^\mu (1 - 2f) + f^2\right) \left(P(\kappa, h_i^\mu)^2 + \frac{\partial P}{\partial \Delta}(\kappa, h_i^\mu)\right) \tag{56}$$

$$b_{ji} = \frac{1}{N} \sum_\mu \left(\eta_i^\mu (1 - 2f) + f^2\right) \left(P(\kappa, h_j^\mu)^2 + \frac{\partial P}{\partial \Delta}(\kappa, h_j^\mu)\right) \tag{57}$$

We can now compute the statistics of a_{ij}, a_{ji}, b_{ij} and b_{ji} over the distribution of patterns. The means and variances have already been computed in the previous section, Eqs. (43,44). The covariance of a_{ij} and a_{ji} is

$$\overline{a_{ij} a_{ji}} = \frac{\alpha}{4} \left(\sum_\xi p_\xi \int Du \frac{1}{\sigma_h} \frac{G\left(a_\xi(u)\right)}{H\left(a_\xi(u)\right)}\right)^2 \tag{58}$$

Nature Neuroscience: doi:10.1038/nn.4286

The joint distribution of w_{ij} and w_{ji} is

$$\overline{Q(w_{ij},w_{ji})} = \int \prod_{a=\{ij\},\{ji\}} du_a P(u_{ij},u_{ji}) \frac{\exp\left(\sum_a -\frac{\bar{b}}{2}w_a^2 + w_a\left(-c+u_a\sigma_a\right)\right)\prod_a \Theta(w_a)}{\prod_a \int_0^\infty dw_a \exp\left(-\frac{\bar{b}}{2}w_a^2 + w_a\left(-c+u_a\sigma_a\right)\right)} \quad (59)$$

$$P(u_{ij},u_{ji}) = \frac{1}{2\pi\sqrt{1-\lambda^2}}\exp\left(-\frac{1}{2(1-\lambda^2)}\left(u_{ij}^2+u_{ji}^2-2\lambda u_{ij}u_{ji}\right)\right) \quad (60)$$

$$\lambda = \frac{\left(\sum_\xi p_\xi \int Du \frac{G\left(a_\xi(u)\right)}{H\left(a_\xi(u)\right)}\right)^2}{4f(1-f)\sum_\xi p_\xi \int Du \frac{G\left(a_\xi(u)\right)^2}{H\left(a_\xi(u)\right)^2}} \quad (61)$$

In the limit $\alpha \to \alpha_c$, $q \to Q$, we get

$$\begin{aligned}
\overline{Q(w_{ij},w_{ji})} = & \int DzH\left(-\frac{B+z\sqrt{\lambda}}{\sqrt{1-\lambda}}\right)\delta(w_{ij})\delta(w_{ji}) \\
& +\frac{1}{\sqrt{2\pi}w_s}\exp\left(-\frac{1}{2}\left(B+\frac{w_{ji}}{w_s}\right)\right)H\left(-\sqrt{\frac{1-\lambda}{1+\lambda}}B\right. \\
& \left. +\frac{\lambda}{\sqrt{1-\lambda^2}}\frac{w_{ji}}{w_s}\right)\delta(w_{ij})\Theta(w_{ji}) \\
& +(i\leftrightarrow j) \\
& +\frac{1}{2\pi\sqrt{1-\lambda^2}}\exp\left(-\frac{1}{2(1-\lambda^2)}\left(\left[B+\frac{w_{ij}}{w_s}\right]^2+\left[B+\frac{w_{ji}}{w_s}\right]^2\right.\right. \\
& \left.\left. -2\lambda\left[B+\frac{w_{ij}}{w_s}\right]\left[B+\frac{w_{ji}}{w_s}\right]\right)\right)\Theta(w_{ij})\Theta(w_{ji})
\end{aligned} \quad (62)$$

$$\lambda = \frac{\left(\sum_\xi p_\xi \xi\left(G(\tau_\xi)-\tau_\xi H(\tau_\xi)\right)\right)^2}{4f(1-f)\sum_\xi p_\xi\left((1+\tau_\xi^2)H(\tau_\xi)-\tau_\xi G(\tau_\xi)\right)} \quad (63)$$

From the above equation one can compute the probabilities of the three 2-neuron motifs,

$$p_{00} = \int DzH\left(-\frac{B+z\sqrt{\lambda}}{\sqrt{1-\lambda}}\right)^2 \quad (64)$$

$$p_{10} = 2\int DzH\left(-\frac{B+z\sqrt{\lambda}}{\sqrt{1-\lambda}}\right)H\left(\frac{B-z\sqrt{\lambda}}{\sqrt{1-\lambda}}\right) \quad (65)$$

$$p_{11} = \int DzH\left(\frac{B-z\sqrt{\lambda}}{\sqrt{1-\lambda}}\right)^2 \quad (66)$$

as well as the distribution of weights for bidirectionnally or unidirectionnally coupled pairs,

$$Q_{10}(w) = \frac{H\left(-\sqrt{\frac{1-\lambda}{1+\lambda}}B+\frac{\lambda}{\sqrt{1-\lambda^2}}\frac{w}{w_s}\right)}{H(-B)}Q(w) \quad (67)$$

Nature Neuroscience: doi:10.1038/nn.4286

$$Q_{11}(w) = \frac{H\left(\sqrt{\frac{1-\lambda}{1+\lambda}}B - \frac{\lambda}{\sqrt{1-\lambda^2}}\frac{w}{w_s}\right)}{H(-B)}Q(w) \quad (68)$$

2 Statistics of connectivity in networks storing sequences

In networks storing sequences, we now define Δ_i^μ as the stability of the transition from pattern μ to pattern $\mu+1$ at neuron i,

$$\Delta_i^\mu = \frac{\xi_i^{\mu+1}}{\sqrt{N}}\left(\sum_i w_i\eta_i^\mu - N\theta\right)$$

where for convenience $\xi_i^{\mu+1} = 2\eta_i^{\mu+1} - 1$. The full sequence is learned if and only if

$$\Delta_i^\mu \geq \kappa \quad (69)$$

for all i, μ.

The calculation of the statistics of connectivity proceeds as in the case of fixed point attractors. The distribution of synaptic weights ends up being identical to the case of fixed point attractors. On the other hand, the joint distribution of weights of a pair of neurons is different in the case of sequences. It is given by

$$Q(w_{ij}, w_{ji}) \propto \exp\left((a_{ij}-c)w_{ij} + (a_{ji}-c)w_{ji} - \frac{b_{ij}}{2}w_{ij}^2 - \frac{b_{ji}}{2}w_{ji}^2\right)\Theta(w_{ij})\Theta(w_{ji})$$

where

$$a_{ij} = \frac{1}{\sqrt{N}}\sum_\mu \xi_i^{\mu+1}(\eta_j^\mu - f)P(\kappa, h_i^\mu) \quad (70)$$

$$a_{ji} = \frac{1}{\sqrt{N}}\sum_\mu \xi_j^{\mu+1}(\eta_i^\mu - f)P(\kappa, h_j^\mu) \quad (71)$$

$$b_{ij} = \frac{1}{N}\sum_\mu \left(\eta_j^\mu(1-2f) + f^2\right)\left(P(\kappa, h_i^\mu)^2 + \frac{\partial P}{\partial \Delta}(\kappa, h_i^\mu)\right) \quad (72)$$

$$b_{ji} = \frac{1}{N}\sum_\mu \left(\eta_i^\mu(1-2f) + f^2\right)\left(P(\kappa, h_j^\mu)^2 + \frac{\partial P}{\partial \Delta}(\kappa, h_j^\mu)\right) \quad (73)$$

The difference with the fixed point scenario is that $\xi_i^{\mu+1}$, $\xi_j^{\mu+1}$ enters in the sums over patterns in the r.h.s. of Eqs. (70,71), instead of ξ_i^μ, ξ_j^μ in Eqs. (54,55). As a consequence, $\overline{a_{ij}a_{ji}} = 0$, the distribution of weights factorizes, and as a result there is no overrepresentation of bidirectionnally connected pairs.

Nature Neuroscience: doi:10.1038/nn.4286

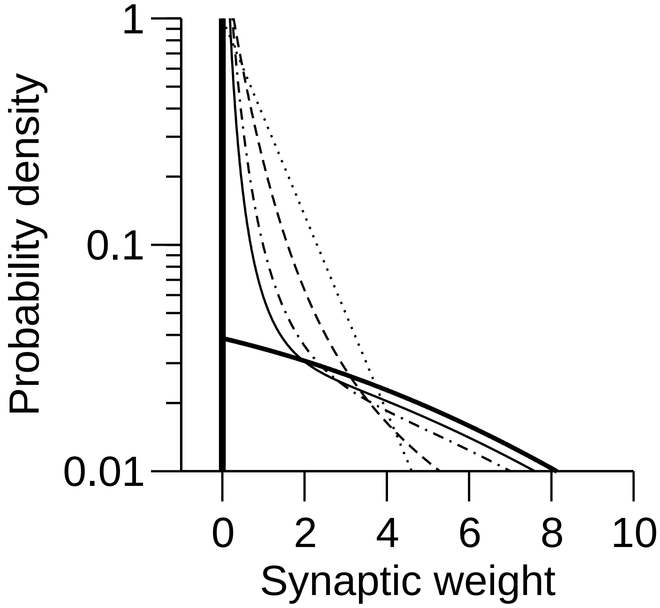

Figure S1: Distribution of weights vs storage capacity. The distribution is shown in semi-log plot for $f = 0.25$ and $\rho = 2.1$, for various values of $\alpha = 0$ (dotted - in this limit, the distribution is exponential), $0.5\alpha_c$ (dashed), $0.8\alpha_c$ (dot-dashed) $0.9\alpha_c$ (thin solid), α_c (thick solid).

Nature Neuroscience: doi:10.1038/nn.4286

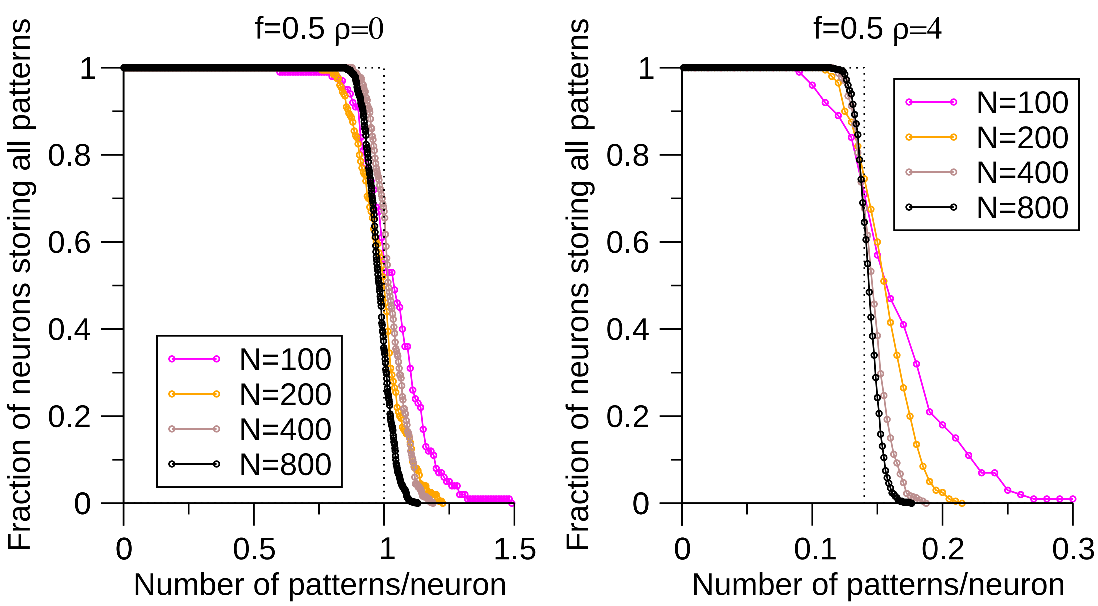

Figure S2: Fraction of neurons storing all patterns correctly, as a function of $\alpha = p/N$, for different network sizes, and two values of the robustness parameters. The analytical prediction for the storage capacity is shown by the dotted line.

Nature Neuroscience: doi:10.1038/nn.4286

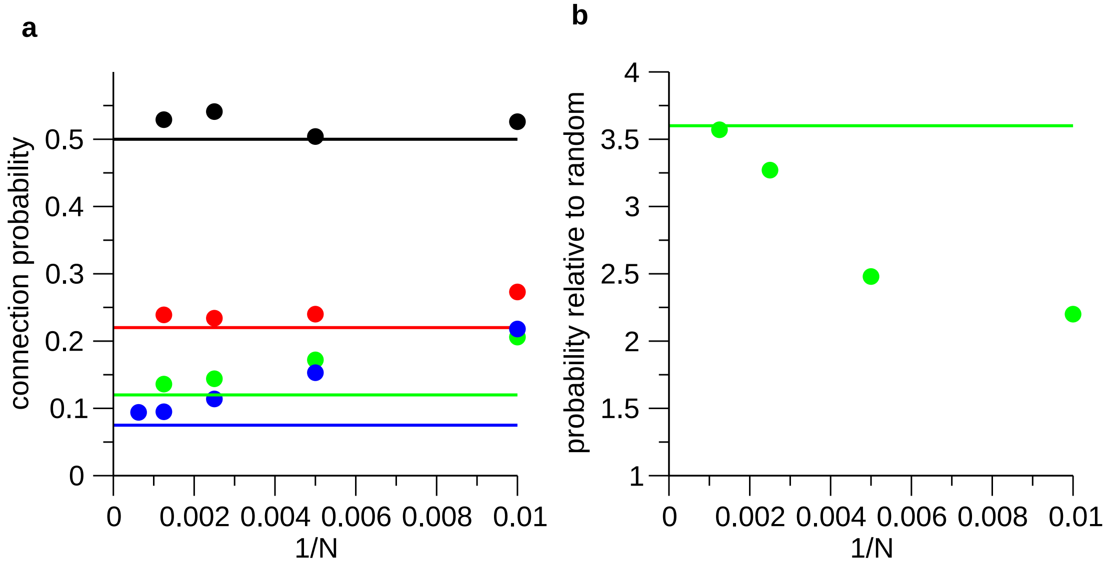

Figure S3: Finite-size effects. (**a**). Connection probability as a function of $1/N$ for $\rho = 0$ (black), 2 (red), 4 (green), 6 (blue), and $f = 0.5$. horizontal lines are analytical predictions for the large N limit, filled circles are results of simulations. (**b**). Probability of observing bidirectionnally connected pairs relative to random networks, as a function of $1/N$, for $f = 0.5$, $\rho = 4$.

Nature Neuroscience: doi:10.1038/nn.4286