Statistical analysis of a Potts glass model of cortical dynamics

Kwang Il Ryom, Alessandro Treves

SISSA - Cognitive Neuroscience, via Bonomea 265, 34136 Trieste, Italy

April 5, 2023

Abstract

We introduce, in a previously studied Potts model of long-range cortical interactions, a differentiation between a frontal and a posterior subnetwork. "Frontal" units, representing patches of anterior cortex, are endowed with a higher number S of attractor states, in keeping with the larger number of local synaptic contacts of neurons there, than in occipital cortex. A thermodynamic analysis and computer simulations confirm that disorder leads to glassy properties and slow dynamics but, surprisingly, the frontal network, which would be slower if isolated, becomes faster than the posterior network when interacting with it. From an abstract, drastically simplified model we take some steps towards approaching a neurally plausible one, and find that the speed inversion effect is basically preserved.

Contents

1 Local attractors in cortical dynamics **2**

2 Mean-field analysis of the long-time behaviour **4**

2.1 The random homogeneous Potts model, with a zero state 5

2.2 The hybrid Potts model without a zero state 10

2.3 The glassy phase of a Potts associative memory 13

3 Dynamics **15**
3.1 Dynamics close to steady states 16
3.2 Factors that accelerate the dynamics 17
3.3 Approaching a cortical scenario 21
3.4 Short-term dynamics for the associative memory model 23
4 Discussion **24**
5 Appendix **26**
5.1 The first-step RSB free energy of Potts glass model 26
5.2 Finite number of patterns in associative memory 27
5.3 Details on the computer simulations 28

1 Local attractors in cortical dynamics

Global oscillations in cortical state, as well as EEG and MEG response patterns, have been approached with linear decomposition analyses, such as spherical harmonics [1, 2, 3]. Yet, it is clear that in order to understand information processing and cognition one has to go beyond such macroscopic descriptions, and consider the local attractor dynamics which has been widely hypothesized to serve as the ubiquitous mechanism for expressing memory functionality [4, 5]. The compartment model by Valentino Braitenberg [6] proposes a separation between local and long range cortical interactions, both of which are envisaged to contribute to associative memory. The Potts associative network [7] can then be construed as an effective model of long-range interactions, that subsumes local ones into the *ad hoc* dynamics of Potts units, with S states each, which represent patches of cortex. We have analyzed the *latching* dynamics induced by adaptation and inhibition [8] and discussed how it could play out in free recall paradigms [9].

These studies, however, reduce the cortex to a *homogeneous* network of Potts units, each of which is characterized by the same number of states S, positive feedback w, time constants τ's for excitation, inhibition and adaptation. This is in contrast with prominent features of cortical organization, which for example point at much higher numbers of local synaptic contacts among pyramidal cells in temporal and frontal, compared to occipital cortex [10], suggestive of a capacity for more and/or stronger local attractor states in the former, or conversely at more

linear and prompt responses to afferent inputs in posterior visual cortices [11, 12], suggestive of reduced positive feedback relative to more anterior areas. Other features show gradients that roughly align with these, and all together have been proposed by Changeux and colleagues [13] to define, in particular in the human brain, a *natural cortical axis*. If one attempts to incorporate these features into a *non-homogeneous* Potts network, what are the implications for cortical dynamics? The indications that the dynamics in frontal cortex may be more affected by local attractors need not necessarily imply, it should be noted, an enhanced tendency for individual neurons to be routinely "stuck" in steady states, in which they keep firing at steady rates for a few hundred *msec*. This would be in apparent contrast with extensive evidence for more dynamical forms of coding in frontal cortex, e.g., for changing task contingencies rather than stable visual features [14]; or, moving up to entire populations of neurons and to the human brain, for the encoding of verbs rather than nouns [15, 16] (but see [17]) or of syntax rather than the lexicon [18].

Local attractor states may be composed, only transiently or more persistently, into global attractor states. Studying the dynamics of reactivating such global attractors requires assumptions about the nature and the statistics of the compositionality, and we have analyzed two distinct models in this respect, both for a homogeneous Potts network [19, 20]. Here, however, we want to focus on the dynamics unfolding away from previously acquired global attractors, for example as new attractors are being established, or *learned*. In a learning regime, we expect the lack of *a priori* relations between what has been already acquired and the new compositional representation to be established to turn the cortex, from the point of view of the latter, into basically a disordered system. Do long-range cortical interactions then result into "glassy" dynamics, with critical slowing down and persistent traces of initial conditions? If so, how does the glassy character express itself over the short time scales relevant to cognition? Is it affected by gross inhomogeneities, like the posterior-anterior gradients in cortical parameters mentioned above?

We consider here the most basic and mathematically well defined aspects of these issues, by analyzing a *hybrid* model that integrates in the Potts formulation a crude binary version of the gradient along the "natural" axis (Fig.1); and leaving for later reports more realistic models of cortical dynamics and applications to other domains. As we shall see, even the analysis of

what seems like a simple extension of a standard model for an infinite range spin glass reveals some surprising properties.

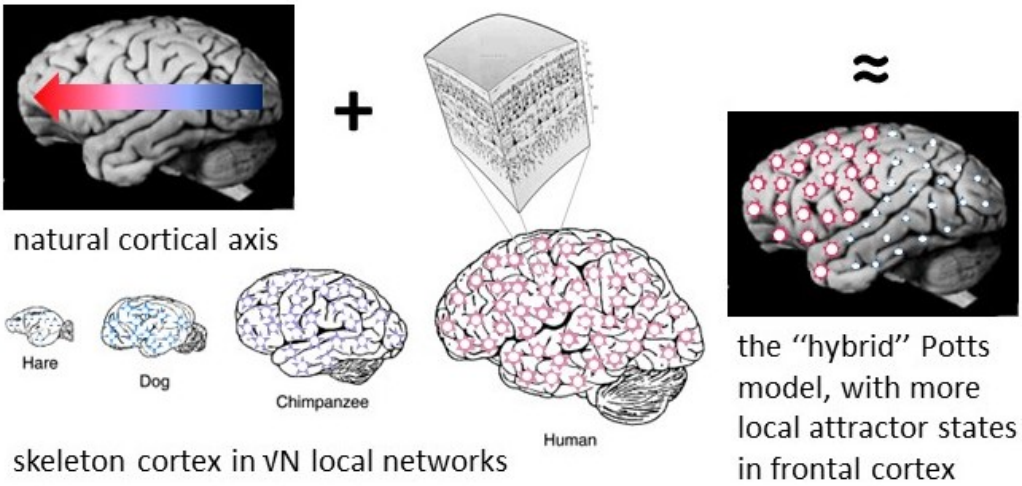

Figure 1: The hybrid Potts model combines the representation of local attractor dynamics in terms of units with S active states, inspired by Braitenberg's idea of an approximate $\sqrt{N}$ scaling [6], with a differentiation between frontal and posterior cortices, along the natural axis posited by Changeux and others [13] and expressed by a larger S value for frontal units. Note the assumption that the critical quantity that varies along the axis is S, the simplification of replacing a gradient with just two S values, and the ill-fitting temporal cortex areas, in which pyramidal cells have abundant recurrent collaterals [10] but are otherwise included among posterior regions.

2 Mean-field analysis of the long-time behaviour

As discussed in previous reports [21, 22], the analysis of the attractor states of associative Potts networks, in which each unit represents a patch of cortex, relies on the same assumption of symmetric interactions, proposed for the standard model [23] in which each unit represents a single neuron.

If we consider, as we do here, a local cortical network to behave effectively as a *discrete* Potts unit, $\sigma_i \in \{0, 1, \ldots, S\}$, which can take one of S active states (labelled by $k = 1, 2, \ldots, S$) as well as stay in the quiescent state (labelled by 0), it is convenient to introduce the Potts spin operator,

$$V_i^k \equiv (\delta_{\sigma_i k} - 1/S)(1 - \delta_{\sigma_i 0}). \tag{1}$$

2.1 The random homogeneous Potts model, with a zero state

First, we consider a network of Potts units all endowed with the same number of states S, that interact through random tensor connections. The Hamiltonian of the system reads

$$H = -\frac{1}{2}\sum_{i\neq j}^{N}\sum_{k,l>0} J_{ij}^{kl} V_i^k V_j^l + U\sum_i (1-\delta_{\sigma_i 0}), \tag{2}$$

where N is the number of Potts units and the $\{J_{ij}^{kl}\}$'s are sampled from Gaussian distributions with mean J_0/N and variance $\lambda^4 J^2/N$. We have introduced the normalisation factor λ,

$$\lambda^2 \equiv \frac{S}{\sqrt{S-1}}, \tag{3}$$

which makes the critical temperature for the transition to a glassy phase independent on S, in units of J (see below). The interactions satisfy

$$\begin{aligned} J_{ji}^{kl} &= J_{ij}^{lk}, \ i \neq j, \\ J_{ii}^{kl} &= 0. \end{aligned} \tag{4}$$

Note that in this model, although S is the same across all units, the states of one unit do not correspond to those of another unit, as they would if they represented, e.g., directions in physical space. This is in contrast to the Potts model considered by [24], in which such correspondence holds, and the interactions, albeit still random, are in the form $J_{ij}^{kl} \propto J_{ij}(\delta_{kl} - 1/S)$, with a single random variable J_{ij} per unit pair (and, in addition, there is no quiescent state). In that model, the symmetry among Potts states is global, whereas in our model it is local, as it must be in order to represent distinct codes by different patches of cortex.

Despite the larger number of random variables the thermodynamic analysis proceeds along similar lines to that in [24] and it is in some respects simpler. Using the replica method [25],

the free energy of the system is written as

$$
\begin{aligned}
f &= \lim_{n\to 0} \frac{1}{n} f_n, \\
\beta n f_n[\{q_{ab}\}] &= \frac{(\beta J)^2 \lambda^4}{2} \sum_{a<b} q_{ab}^2 + (\beta J)^2 \lambda^4 \sum_{a=1}^{n} q_{aa}^2 - \ln \sum_{\sigma^1=0}^{S} \sum_{\sigma^2=0}^{S} \cdots \sum_{\sigma^n=0}^{S} \exp(K), \\
K &\equiv (\beta J)^2 \lambda^4 \sum_{a<b} q_{ab} \sum_{k=1}^{S} V_a^k V_b^k + (\beta J)^2 S \sum_{a=1}^{n} q_{aa}(1-\delta_{\sigma^a 0}) - \beta U \sum_{a=1}^{n} (1-\delta_{\sigma^a 0}).
\end{aligned}
\tag{5}
$$

where q_{ab} is the Edward-Anderson order parameter [25], $\beta = 1/T$ is the inverse temperature and replica indices a and b run from 1 to n. The saddle-point equations of Eqs. (5) are

$$
\begin{aligned}
q_{ab} &= \frac{\sum_{\sigma^1=0}^{S} \sum_{\sigma^2=0}^{S} \cdots \sum_{\sigma^n=0}^{S} [\sum_{k=1}^{S} V_a^k V_b^k \exp(K)]}{\sum_{\sigma^1=0}^{S} \sum_{\sigma^2=0}^{S} \cdots \sum_{\sigma^n=0}^{S} \exp(K)}, \quad a \neq b, \\
q_{aa} &= \frac{S-1}{2S} \frac{\sum_{\sigma^1=0}^{S} \sum_{\sigma^2=0}^{S} \cdots \sum_{\sigma^n=0}^{S} [(1-\delta_{\sigma^a 0}) \exp(K)]}{\sum_{\sigma^1=0}^{S} \sum_{\sigma^2=0}^{S} \cdots \sum_{\sigma^n=0}^{S} \exp(K)}.
\end{aligned}
\tag{6}
$$

The physical meaning of q_{ab} ($a \neq b$) is the same as in the Sherrington-Kirkpatrick (SK) model [26] (see also [27]), while $2q_{aa}S/(S-1)$ is the fraction of active units in replica a of the Potts network. Note that the free energy in Eq. (5) does not depend on J_0, the mean of the normal distribution from which the J_{ij}^{kl}'s are sampled. This is in contrast with the Elderfield-Sherrington (ES) model [24], where low enough values of J_0 should be chosen to avoid ferromagnetic ordering at low temperatures ([24, 28]). Since the symmetry in our model is local – a sort of *gauge* invariance – there is no meaning to ferromagnetic alignment.

Properties near the critical temperature

The paramagnetic solution ($q_{ab} = 0$, $a \neq b$) is the ground state of the system at high enough temperatures. Lowering the temperature, a phase transition from the paramagnetic to the spin glass phase occurs at $T = T_c$. To determine T_c, one can (Landau-) expand the free energy close

to it, assuming q_{ab} $(a < b)$ to be small, to find

$$\begin{aligned}
\beta n f_n &\approx \frac{A}{2}\sum_{(ab)} q_{ab}^2 - \frac{B}{3}\sum_{(ab)} q_{ab}^3 - \frac{C}{3}\sum_{(abc)} q_{ab}q_{bc}q_{ca} - \frac{D}{12}\sum_{(ab)} q_{ab}^4, \\
A &= \frac{(\beta J)^2}{2}\frac{S^2}{S-1}[1-(\beta J)^2 S^2\psi^2], \\
B &= \frac{(\beta J)^6}{4}\frac{S-2}{\sqrt{S-1}}S^2\psi^2, \\
C &= \frac{(\beta J)^6}{2}\frac{S^3\psi^3}{\sqrt{S-1}}, \\
D &= (\beta J)^8\Big[\frac{3(3S-1)}{4(S-1)}S^4\psi^4 - 3S^3\psi^3 + \frac{S^2-3S+3}{4(S-1)}S^2\psi^2\Big].
\end{aligned} \tag{7}$$

Here (abc) means that replica indices $a,\ b,\ c$ are all distinct in the summation. Following [29], we have retained only the quartic term that is relevant for replica-symmetry breaking (RSB) in Eqs. (7). We have also assumed that the order parameter q_{aa} does not depend on the replica index a near T_c and thus have introduced a symbol $\psi \equiv 2q_{aa}/(S-1)$ to reduce the burden of heavy notation. Then, ψ should satisfy

$$\psi = \frac{\exp\left[(\beta J)^2\frac{S(S-1)}{2}\psi - \beta U\right]}{1 + S\exp\left[(\beta J)^2\frac{S(S-1)}{2}\psi - \beta U\right]} \tag{8}$$

and the quantity $S\psi$ gives the fraction of active units in the network.

Under the replica symmetric (RS) assumption, $q_{ab} = q$ $(a \neq b)$, the critical temperature is determined by numerically solving

$$(\beta J)^2 S^2\psi^2 - 1 = 0 \tag{9}$$

together with Eq. (8), since ψ contains T. If $U \to -\infty$, then $S\psi \to 1$ and we get a simple formula, $T_c = J$. The phase transition is a continuous one if

$$0 < 4S\psi - (S-2). \tag{10}$$

For $S < 6$, there exists a critical value of U, U_c, above (below) which the transition is discontinuous (continuous). For $S > 6$, the transition is discontinuous for all values of U. This RS solution is however unstable against RSB in the whole glassy phase. Thus, replica symmetry

should be broken.

To probe replica symmetry breaking, following Parisi's hierarchical scheme [30] we write the free energy

$$
\begin{aligned}
-\beta f \approx & \int_0^1 dx \left[\frac{A}{2} q^2(x) - \frac{B}{3} q^3(x) - \frac{D}{12} q^4(x) \right] \\
& + \frac{C}{3} \int_0^1 dx \left[x q^3(x) + 3 q(x) \int_0^x q^2(y) dy \right],
\end{aligned}
\tag{11}
$$

which is to be maximised with respect to Parisi's function $q(x)$ [29]. We note that Eq. (11) has the same form as in the ES model, except for the coefficients. Thus, we can envisage that the nature of RSB is similar to that in the ES model (see also [28, 31] for its detailed properties). The non-trivial solution of Eq. (11) is

$$
q(x) = \begin{cases} 0, & x \leq x_0 \\ \dfrac{Cx - B}{D}, & x_0 \leq x \leq x_1 \\ q_1 = \dfrac{A}{B - C} + O(q_1^2), & x_1 \leq x, \end{cases}
\tag{12}
$$

where

$$
\begin{aligned}
x_0 &= \frac{B}{C} = \frac{1}{2\psi} \frac{S-2}{S}, \\
x_1 &= \frac{B}{C} + \frac{AD}{C(B-C)}
\end{aligned}
\tag{13}
$$

From Eq. (12), we can see how replica symmetry is broken, for a given value of U. The scheme

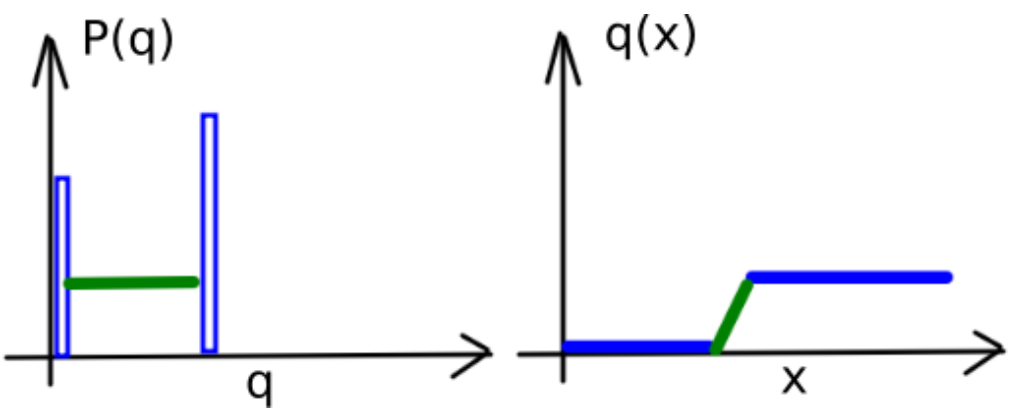

Figure 2: Schematic description of replica symmetry breaking, from Eq. (12). Left: the probability density $P(q)$, with blue rectangles denoting Dirac delta functions. Right: Parisi's function. Colour coding is used to facilitate a visual comparison.

in Fig. 2 is similar to the one for the ES model. Note that x_0 is always zero for $S = 2$, regardless

of U, whereas it remains positive for $S > 2$. This means that $P(q)$ has a Dirac delta at $q = 0$ for $S > 2$, whereas there is no Dirac delta at $q = 0$ for $S = 2$, as in the SK model. The phase transition to the glassy phase is continuous if

$$\begin{aligned} &0 < 2S\psi - (S-2), \\ &0 < 3S^2(3S-1)\psi^2 - 12S(S-1)\psi + S^2 - 3S + 3. \end{aligned} \tag{14}$$

In general, these two conditions are numerically probed together with Eqs. (8) and (9) for

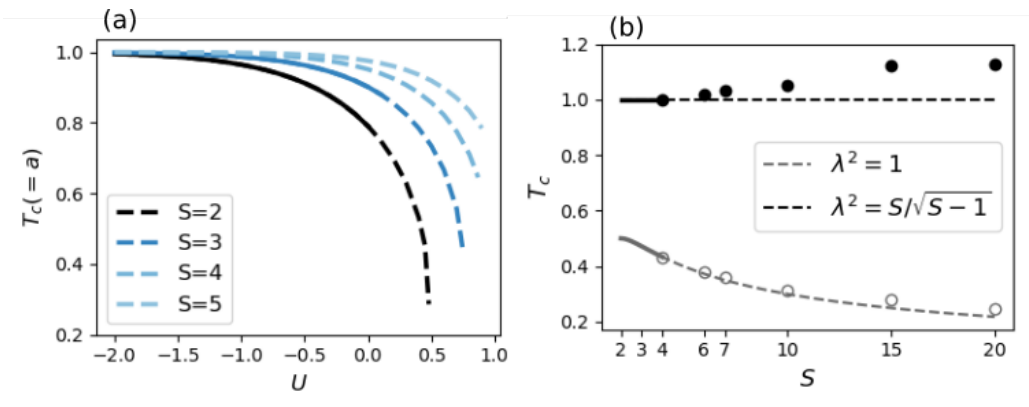

Figure 3: **The critical temperature** (T_c) for the onset of the glassy phase of a homogeneous Potts network. **(a)**: T_c as a function of the threshold U for a model with a zero state. With the normalisation set as in Eq. (3), the mean activity a of the network at $T = T_c$ is equal to T_c itself (that is, to T_c/J). Dashed curves are predicted by RS theory and solid curves are from RSB theory. All transitions shown here are continuous. **(b)** T_c as a function of S for a model without a zero state ($U \to -\infty$): colour encodes the normalisation used (as indicated in the legend). Solid curves are obtained analytically from the Landau expansion of the free energy (a continuous phase transition) and dashed curves are their mere extensions, to guide the eye. Circles are obtained by numerically maximising the 1-step RSB free energy (a discontinuous transition), Eq. (16). We set $J = 1$.

a given value of U. As a special case, when $U \to -\infty$, the second condition is guaranteed. However, unlike the RS Eq. (10), the first of RSB Eqs. (14) ceases to hold for $S > 4$. Thus, the transition can be continuous only for $S \leq 4$. We can compute the range of U where Eqs. (14) hold by solving them together with Eq. (8) and Eq. (9). The result is shown in Fig. 3(a).

Properties at all temperatures

At temperatures well below T_c and when the transition is discontinuous, one should directly deal with the free energy, Eqs. (5), in the full RSB formalism. Even for the SK model, solving Parisi's equations requires sophisticated numerical techniques (see [32]). However, Potts spins seem to have a distinguishing property from Ising spins, at least when we compare the ES

model with the SK model: while any finite-step RSB solution is unstable in the SK model ([30]), the first-step RSB (1RSB) solution is locally stable in the ES model ($S > 2$) below T_c, down to a certain temperature, where another phase transition occurs [28, 31]. So, one can study discontinuous transitions for $S > 4$, where the Landau expansion does not apply, by using a 1RSB formalism ([33]). Below we use this method to study the discontinuous transition of our random Potts model, Eq. (2).

While the expression of the free energy Eq. (5) within a 1RSB ansatz is given in the Appendix (see Eq. (40)), its numerical solution is computationally hard, especially for large values of S. Thus, we restrict ourselves to a special case: the threshold U goes to $-\infty$ (the zero state then drops out of the equations). Inspired from the shape of Eq. (12), we seek solutions of the form

$$P(q) = m\delta(0) + (1-m)\delta(q). \tag{15}$$

Then, the 1RSB free energy becomes,

$$\beta f \approx \frac{(\beta J)^2\lambda^4}{4}(m-1)q^2 + \frac{(\beta J)^2\lambda^4}{2}\frac{m+S-1}{S}q - \frac{1}{m}\ln\int D\overrightarrow{y}\,\Big[\sum_{l=1}^{S}\exp(\beta J\lambda^2\sqrt{q}y_l)\Big]^m, \tag{16}$$

where

$$D\overrightarrow{y} \equiv \prod_{k=1}^{S}\Big[\frac{dy_k}{\sqrt{2\pi}}\exp\Big(-\frac{y_k^2}{2}\Big)\Big].$$

We can numerically maximise Eq. (16) by using the same numerical trick as in [33], up to $S = 20$. Critical temperatures obtained that way are reported in Fig. 3(b), while the order parameters are shown in Fig. 4.

2.2 The hybrid Potts model without a zero state

We now consider a network of Potts units that have different values for S: a unit i has its own number S_i of Potts states. For the sake of simplicity, we consider Potts units without the quiet state. We group units according to their number of states: there are N_l units in group l ($l = 1, 2, \ldots, L$) and they have S_l Potts states each. If the total number of Potts units in the

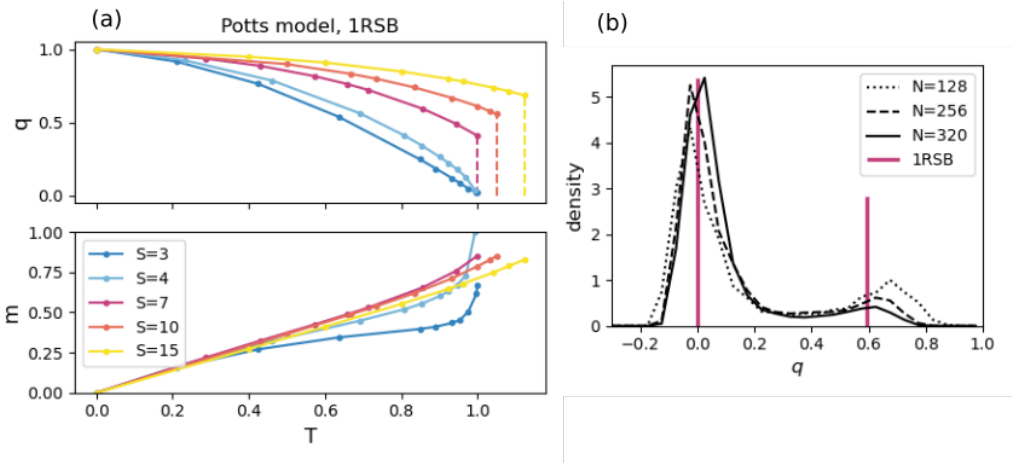

Figure 4: Order parameters of a homogeneous Potts network without a zero state ($U \to -\infty$), predicted by 1RSB theory. **(a)**: solutions of the 1RSB free energy as a function of T. Note the discontinuous jumps in q at $T = T_c$ for $S > 4$. **(b)**: Probability density, $P(q)$, obtained from Monte Carlo simulations, for $S = 7$ and $T \approx \frac{6}{7}T_c$. Red vertical lines indicate Dirac delta functions, estimated from (a). The peak at higher q seems to be lower with increasing values of N, but this is due to the insufficient relaxation time in our simulations. Since the relaxation time grows exponentially with N ([34]), we did not attempt to obtain the exact ground states.

network is N,

$$\eta_l \equiv \frac{N_l}{N},$$
$$1 = \sum_{l=1}^{L} \eta_l.$$

We write

$$H = -\frac{1}{2}\sum_{i\neq j}^{N}\sum_{k=1}^{S_i}\sum_{m=1}^{S_j} \lambda_i J_{ij}^{km} \lambda_j (\delta_{\sigma_i k} - 1/S_i)(\delta_{\sigma_j m} - 1/S_j), \tag{17}$$

where $\lambda_j \equiv \sqrt{\frac{S_j}{\sqrt{S_j - 1}}}$ normalizes the interactions with both a pre- and a post-synaptic factor, and the $\{J_{ij}^{km}\}$'s are sampled from a Gaussian distribution of mean J_0/N and variance J^2/N and satisfy Eqs. (4).

The replica method gives the free energy,

$$\beta n f_n \approx \frac{(\beta J)^2}{2}\sum_{a<b} q_{ab}^2 - \sum_{l=1}^{L} \eta_l \ln \mathrm{Tr}^l \exp[(\beta J \lambda_l)^2 \sum_{a<b} q_{ab}(\delta_{\sigma^a\sigma^b} - 1/S_l)], \tag{18}$$

where q_{ab} is the Edward-Anderson order parameter,

$$q_{ab} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^2 \langle \delta_{\sigma_i^a \sigma_i^b} - 1/S_i \rangle, \tag{19}$$

and

$$\mathrm{Tr}^l \equiv \sum_{\sigma^1=1}^{S_l} \cdots \sum_{\sigma^n=1}^{S_l} . \tag{20}$$

As in the previous subsection, we expand Eq. (18) around $q_{ab} = 0$ and apply the Parisi algebra [35] to probe the nature of the equilibrium state. The corresponding free energy functional and the Parisi function that maximises it have the same form as for the homogeneous network (see Eqs. (7), (11), (12)), after a redefinition of the coefficients A, B, C and D

$$\begin{aligned} A &= \frac{(\beta J)^2}{2}[1-(\beta J)^2], \\ B &= \frac{(\beta J)^6}{4}\sum_{l=1}^{L}\eta_l \frac{S_l-2}{\sqrt{S_l-1}}, \\ C &= \frac{(\beta J)^6}{2}\sum_{l=1}^{L}\eta_l \frac{1}{\sqrt{S_l-1}}, \\ D &= \frac{(\beta J)^8}{4}\sum_{l=1}^{L}\eta_l \frac{S_l^2-6S_l+12}{S_l-1}. \end{aligned} \tag{21}$$

The critical temperature for the onset of the glassy phase is again given by

$$T_c = J, \tag{22}$$

where the phase transition is continuous in terms of q whenever

$$\sum_{l=1}^{L}\eta_l \frac{S_l-4}{\sqrt{S_l-1}} < 0. \tag{23}$$

As an example, consider a hybrid network with two types of Potts units: half with S_1 and half with S_2 states. The phase transition is continuous if

$$\frac{S_1-4}{\sqrt{S_1-1}} + \frac{S_2-4}{\sqrt{S_2-1}} \leq 0. \tag{24}$$

Several cases are interesting (we set $1 < S_1 \leq S_2$):

- $S_1 = 2$. The transition is continuous for $S_2 \leq 10$ and discontinuous otherwise.
- $S_1 = 3$. The transition is continuous for $S_2 \lesssim 5.5$ and discontinuous otherwise.
- $S_1 \geq 4$. The transition is always discontinuous (except for $S_1 = S_2 = 4$, but then the

network is again homogeneous, as in the previous section).

2.3 The glassy phase of a Potts associative memory

We consider now an attractor neural network comprised of Potts units. The Hamiltonian is the same as in Eq. (2), with the connection J_{ij}^{kl} now given by the Hebbian-learning rule,

$$J_{ij}^{kl} = \frac{1-\delta_{ij}}{Na(1-\tilde{a})} \sum_{\mu=1}^{p} \left(\delta_{\xi_i^\mu k} - \tilde{a}\right)\left(\delta_{\xi_j^\mu l} - \tilde{a}\right)(1-\delta_{k0})(1-\delta_{l0}), \tag{25}$$

where $\tilde{a} = a/S$ and $\{\xi_i^\mu\}$ are p randomly correlated memory patterns. The free energy is obtained by the replica trick (see [36] for the Hopfield model and [21, 22] for the Potts model):

$$\begin{aligned} n f_n[\overrightarrow{m}, \mathbf{q}, \mathbf{r}] = & \frac{1}{2}\sum_\gamma (m_\gamma)^2 + \frac{\alpha}{2\beta}\mathrm{Tr}\ln\left[\mathbf{I} - \beta\tilde{a}\mathbf{q}\right] + \sum_{\gamma\lambda} r_{\gamma\lambda} q_{\gamma\lambda} + \\ & + \left[\frac{\alpha\tilde{a}}{2} + \frac{SU}{S-1}\right]\sum_\gamma q_{\gamma\gamma} - \frac{1}{\beta}\left\langle \ln \mathrm{Tr}_{\{\sigma^\gamma\}} \exp[\beta L_\sigma^\xi]\right\rangle_\xi, \end{aligned} \tag{26}$$

where

$$L_\sigma^\xi \equiv \sum_\gamma m_\gamma \sum_{k>0} (\delta_{\xi k} - \tilde{a}) V_\gamma^k + \sum_{\gamma\lambda} r_{\gamma\lambda} \sum_{k>0} V_\gamma^k V_\lambda^k \tag{27}$$

and $\alpha \equiv p/N$. The saddle-point equations read

$$\begin{aligned} m_\gamma &= \sum_{k>0} \left\langle (\delta_{\xi k} - \tilde{a}) \left\langle V_\gamma^k \right\rangle_{L_\sigma^\xi} \right\rangle_\xi, \\ q_{\gamma\delta} &= \left\langle \sum_{k>0} \left\langle V_\gamma^k V_\delta^k \right\rangle_{L_\sigma^\xi} \right\rangle_\xi, \\ r_{\gamma\delta} &= \frac{\alpha\tilde{a}}{2}\left[\mathbf{I} - \beta\tilde{a}\mathbf{q}\right]_{\gamma\delta}^{-1} - \delta_{\gamma\delta}\left[\frac{\alpha\tilde{a}}{2} + \frac{SU}{S-1}\right] \end{aligned} \tag{28}$$

where

$$\langle X(\sigma,\xi)\rangle_{L_\sigma^\xi} \equiv \frac{\mathrm{Tr}_\sigma\left[X(\sigma,\xi)\exp(\beta L_\sigma^\xi)\right]}{\mathrm{Tr}_\sigma \exp(\beta L_\sigma^\xi)}. \tag{29}$$

One can solve Eqs. (28) by using either RS or RSB assumptions to compute, *inter alia*, the storage capacity of network. We refer to Refs. [21, 22] for a discussion of the storage capacity (see also [7, 37] for related but different models). Here, we are interested in the phases prevailing at higher temperature: the paramagnetic and the glassy phase.

At high enough values of T, in fact, retrieval solutions do not exist. So, we set $m_\gamma = 0$

and the terms including ξ and m_γ drop out of the equations. We can easily see that q_{ab} and r_{ab} converge to zero in the high temperature limit, if $a \neq b$. We expand the free energy with respect to these two variables around zero

$$
\begin{aligned}
n\beta f \approx\ & \frac{n\alpha}{2}\ln(1-\beta\tilde{a}\tilde{q}) + \beta\sum_{(ab)} r_{ab}q_{ab} + n\beta\tilde{q}\big(\frac{\alpha\tilde{a}}{2} + \frac{SU}{S-1} + \tilde{r}\big) \\
& - \frac{\alpha\Lambda^2}{2}\Big(\frac{1}{2}\sum_{(ab)} q_{ab}^2 + \frac{\Lambda}{3}\sum_{(abc)} q_{ab}q_{bc}q_{ca} + \frac{\Lambda^2}{4}\sum_{(abcd)} q_{ab}q_{bc}q_{cd}q_{da}\Big) \\
& - (S-1)\beta^2\psi^2\Big\{\sum_{(ab)} r_{ab}^2 + \frac{4\beta\psi}{3}\sum_{(abc)} r_{ab}r_{bc}r_{ca} + \frac{2\beta(S-2)}{3S}\sum_{(ab)} r_{ab}^3 \\
& + \beta^2\Big[\psi^2(3S-1) - 4\psi\frac{S-1}{S} + \frac{S^2-3S+3}{4S^2}\Big]\sum_{(ab)} r_{ab}^4\Big\},
\end{aligned}
\tag{30}
$$

where

$$
\Lambda \equiv \Lambda(T) = \frac{\beta\tilde{a}}{1-\beta\tilde{a}\tilde{q}} = \frac{\tilde{a}}{T-\tilde{a}\tilde{q}}.
$$

For the sake of simplicity, let us consider a RS ansatz. Then, the free energy reads, up the third order in q and r,

$$
\begin{aligned}
\beta f_{\mathrm{RS}} \approx\ & \frac{\alpha}{2}\ln(1-\beta\tilde{a}\tilde{q}) - \beta r q + \beta\tilde{q}\big(\frac{\alpha\tilde{a}}{2} + \frac{SU}{S-1} + \tilde{r}\big) \\
& + \frac{\alpha\Lambda^2}{4}q^2\Big[1 - \frac{4}{3}\Lambda q\Big] + (S-1)\beta^2\psi^2 r^2\Big[1 - \frac{8}{3}\beta\psi r + \frac{2(S-2)}{3S}\beta r\Big].
\end{aligned}
\tag{31}
$$

This free energy is maximised with respect to r and q, while $\tilde{q}$ and $\tilde{r}$ satisfy

$$
\begin{aligned}
\tilde{q} &= (S-1)\psi, \\
\tilde{r} &= \frac{\alpha\tilde{a}}{2}\big(\frac{1}{1-\beta\tilde{a}\tilde{q}} - 1\big) - \frac{S}{S-1}U, \\
\psi &= \frac{\exp\big(\beta\tilde{r}\frac{S-1}{S}\big)}{1 + S\exp\big(\beta\tilde{r}\frac{S-1}{S}\big)}.
\end{aligned}
\tag{32}
$$

In addition to the trivial (paramagnetic) solution of $q = 0$, we have

$$
q = 2\frac{\alpha\Lambda^2\psi^2(S-1) - 1}{\alpha\psi^2\Lambda^3(S-1)\big(4 + \alpha\Lambda\frac{4S\psi+2-S}{S}\big)}.
\tag{33}
$$

A phase transition from the paramagnetic to the glassy phase occurs at

$$T_c = \tilde{q}\tilde{a} + \psi\tilde{a}\sqrt{\alpha(S-1)} \to \frac{(S-1)a}{S^2} + \frac{a}{S^2}\sqrt{\alpha(S-1)}, \tag{34}$$

where the last expression is for the limit of $U \to -\infty$. It is a continuous transition if $S \leq 6$. For $S > 6$, the transition is continuous if $\alpha < \alpha_0$ and discontinuous otherwise, with

$$\alpha_0 = \frac{16S^2\psi^2(S-1)}{(S-4\psi S-2)^2} \to \frac{16(S-1)}{(S-6)^2},$$

where the last expression is again for $U \to -\infty$.

As in the the random Potts model considered above, there is a value of U above which the phase transition cannot be treated by the Landau expansion (either the phase transition does not occur or it cannot be described by the Landau expansion). This result is shown in Appendix.

3 Dynamics

Although dynamics can be studied within mean-field theory to a certain extent ([38]), here we stick to Monte Caro (MC) simulations. Throughout this work, we use the heat bath algorithm to simulate the dynamics of Potts networks. Specifically, the local field of each Potts unit is computed as

$$h_i^k = \sum_{j=1(j\neq i)}^{N} \sum_{l>0} (J_{ij}^{kl} - \frac{1}{S}\sum_{k'} J_{ij}^{k'l})V_j^l, \ k > 0. \tag{35}$$

At each MC step, one Potts unit is randomly chosen to be updated based on the following equations

$$\begin{aligned} \text{Prob}[\sigma_i = k] &= \frac{\exp[\beta h_i^k]}{\sum_{l=1}^{S}\exp[\beta h_i^l] + \exp[\beta U]}, \ k > 0, \\ \text{Prob}[\sigma_i = 0] &= \frac{\exp[\beta U]}{\sum_{l=1}^{S}\exp[\beta h_i^l] + \exp[\beta U]}. \end{aligned} \tag{36}$$

For models without a zero state, the second of Eqs. (36) is not used.

For most of the simulations presented here, we run two systems ([34]) with the same quenched disorder (i.e., the set of interactions between Potts units) and measure the overlap

between the two configuration at time t:

$$q_{ab}(t) = \frac{S}{N(S-1)} \sum_{i=1}^{N} (\delta_{\sigma_i^a \sigma_i^b} - 1/S). \tag{37}$$

3.1 Dynamics close to steady states

Fig. 5 shows sample trajectories at temperatures $T \ll T_c$, to illustrate their glassy nature: after an initial transient the system is trapped in metastable states for a while before finding a way out, along which it can further lower its energy. The opportunities to escape a metastable state however become rarer and rarer, and the time spent near it longer and longer, a process called thermalization.

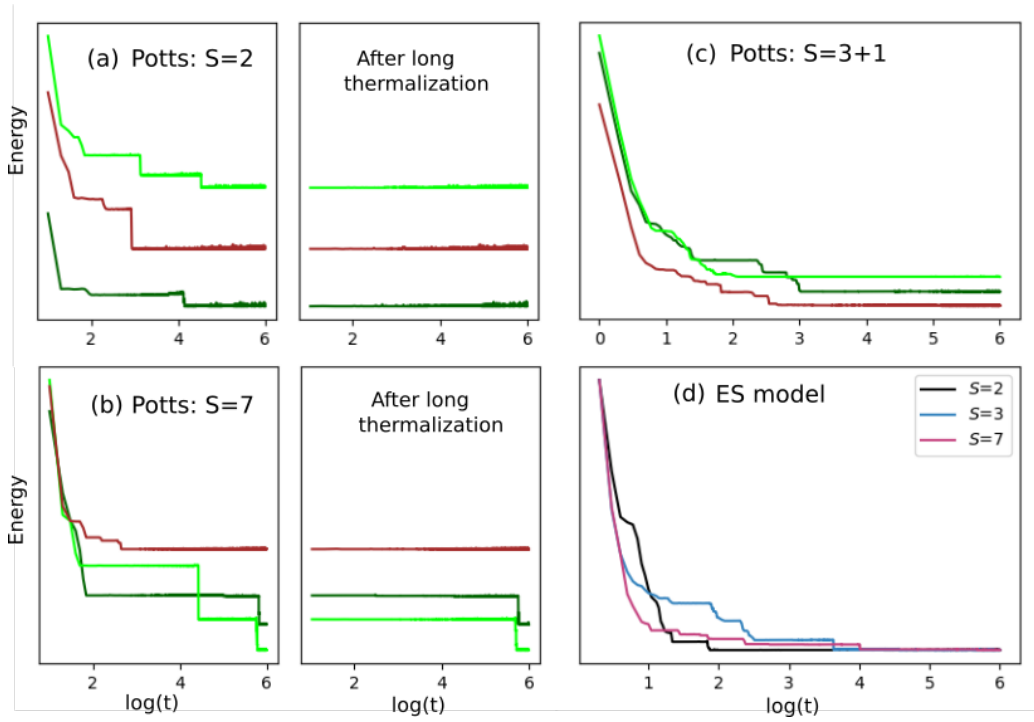

Figure 5: **Energy as a function of MC sweeps per unit** for sample MC trajectories. Note the log scale of the abscissa. **(a)** and **(b)**: three example trajectories are shown for a homogeneous Potts network without a zero state. In the right panels, t restarts after $t_0 \simeq 10^5$, to focus on long-time glassy dynamics. **(c)**: example trajectories of a homogeneous Potts network with a zero state. **(d)**: example trajectories of the ES model ([24]). The three curves are rescaled and shifted for better visibility (only in panel (d)). Note that the ES model reduces to the SK model if $S = 2$. The number of units is $N = 256$ for all three panels, and each data point is averaged over 10 MC sweeps, except for the first 100 points.

To measure how fast the dynamics unfolds on the glassy free energy landscape, we first "thermalize" a configuration by letting it evolve for $t_0 = 10^3$ time steps, and then start from it two simulations with identical interactions, until at τ their overlap reaches half its initial value.

Since the times τ are quite scattered depending on the realization of the interaction – their logarithms are approximately normally distributed – we consider their cumulative distribution, for a given network, and in particular the thermal *half-life* scale $\zeta_g(T)$, defined as the median $\mu_{1/2}[\log(\tau)]$ when the cumulative distribution, at a temperature T, reaches the value 0.5.

With this procedure, we find that a homogeneous random Potts network "moves" faster the smaller are its Potts units, i.e., the lower their S. This is shown in Fig. 6a, which indicates that $\zeta_g(T) \equiv \mu_{1/2}[\log(\tau)]$ increases approximately with $\log(S)$, with the parameters we use.

If we measure τ (and ζ_g) separately for the units with a given S in a hybrid network, we find that the small units get slower and the large units get faster, due to the hybridization. Surprisingly, however, the effect is not simply an interpolation or averaging of the temporal scale between the two subnetworks, that would come to share a common speed, because in many cases the large units get markedly faster than the small units. This is shown in Fig. 6b for $T = 0.8$ and large units with $S_2 = 7$, that interact in a hybrid network with small spins, $S_1 = 2$. Fig. 6c shows that the slowing down of these spins scales roughly with the log of S_2, the number of states of the units that "bog them down". Simultaneously, the large units "speed up" after the hybridization, Fig. 6d and, particularly when the interactions are not renormalized as in Eq. (17), can get to be faster, on average, than the small spins (Fig. 6f).

The speed inversion phenomenon indicates that the same free-energy landscape is "perceived" as rougher, near the metastable states, by Potts units with fewer degrees of freedom. Does the same effect occur away from the metastable states, e.g. in the initially rapid dynamics to the left of the panels in Fig. 5, or when *asymmetric* connection weaken the very stability of such disordered states?

3.2 Factors that accelerate the dynamics

In the Hopfield model, imposing symmetry in the interactions, which established the connection with Hamiltonian systems, thus enabling the analytical approach [23], entailed gross disregard for Dale's law – stating that excitatory and inhibitory neurotransmitters are released by distinct types of cortical neurons – and also of plausible statistical models of connectivity among excitatory neurons alone. Interestingly, it was argued early on that spin-glass-like metastability would still characterize networks with asymmetrically "diluted" connectivity, whereas it was

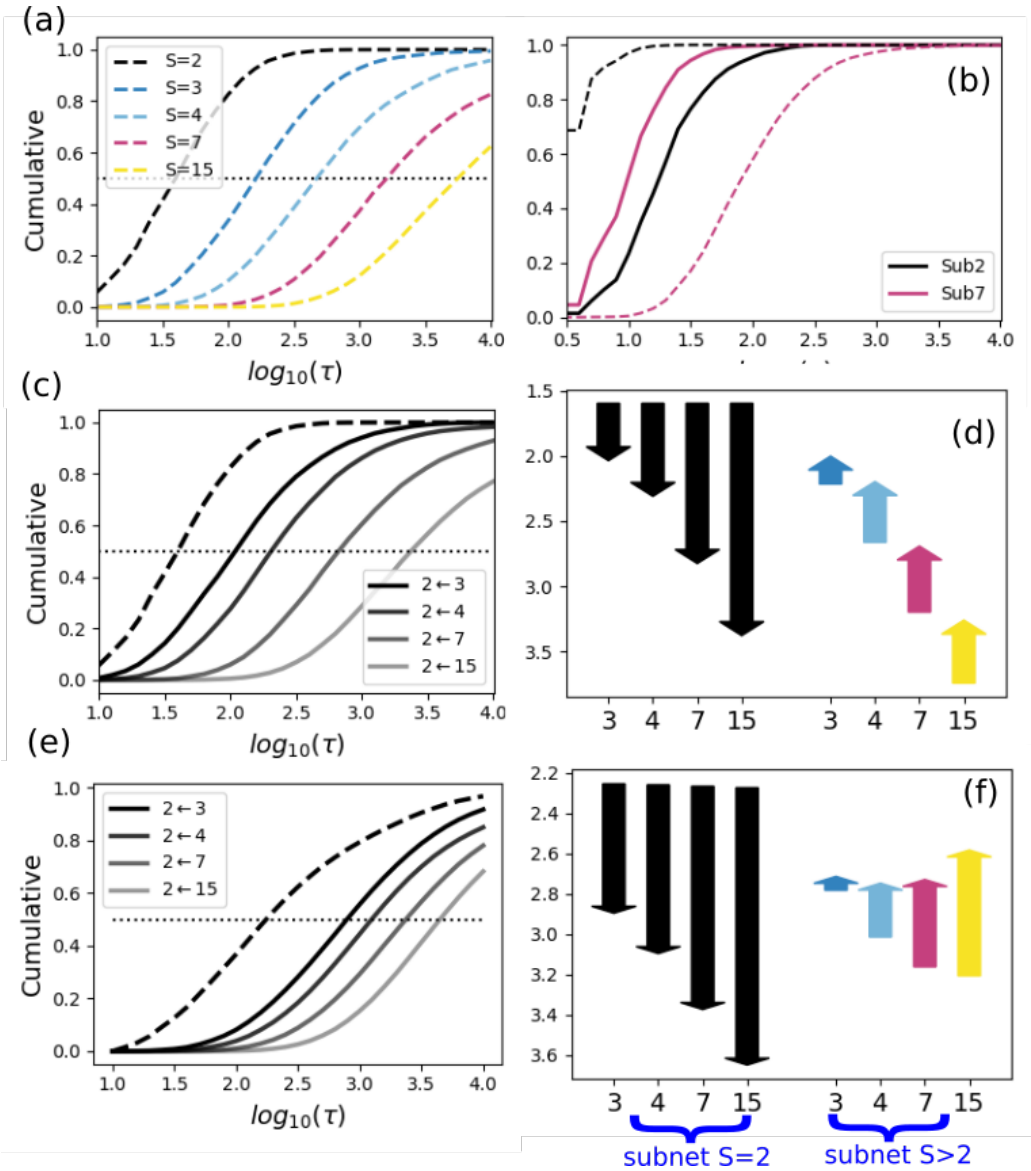

Figure 6: **Speed-up and slow-down in a hybrid Potts model**. **(a)**: Cumulative distribution of τ, computed for homogeneous networks of $N = 256$ units, as a function of S. **(b)**: the inversion of speed due to hybridization between spins ($S_1 = 2$) and large units with $S_2 = 7$. **(c)**:a sub-network of $S_1 = 2$ that interacts with another sub-network with $S_2 > 2$ is more slowed down the higher is S_2. Note that the case with $S_2 = 2$ is the homogeneous network of panel (a). **(d)**: The speed-up and slow-down of sub-networks are shown by the arrows, which head up for units that accelerate. The color of bars stands for S as in panel (a), while the height measures the difference $\Delta\zeta_g$ in the median of the cumulative distribution of τ, between hybrid and homogeneous networks. $T = 0.5$, $t_0 = 10^3$. **(e)** and **(f)**: same as (c) and (d), but without the normalisation constants λ_i in Eq. (17) and $T = 0.2$, $t_0 = 5 \times 10^3$.

suggested that more profound changes due to asymmetry might be observed in the dynamics [39]. In the Potts network, inspired by Braitenberg's model [6], Dale's law is not relevant as long-range connectivity (the component modelled by the tensor interactions among Potts units) is only excitatory; and there is no urgency to consider diluted connectivity either, as the tensor connections themselves are considered to recapitulate thousands of individual synaptic connections [22]. Still, it makes sense to consider the effect of asymmetric non-zero values in the random interactions, by writing them in the form

$$J_{ij}^{kl} = \gamma J_{\rm asym} + (1-\gamma) J_{\rm sym}, \tag{38}$$

where $J_{ij,\rm sym}^{kl} = J_{ji,\rm sym}^{lk}$ and $J_{ij,\rm asym}^{kl}$ is unrelated to $J_{ji,\rm asym}^{lk}$; thus the former are *symmetric* and the latter *a-symmetric* random components, drawn from the usual distribution with zero mean and variance J.

Fig. 7b,c shows that introducing asymmetry does have a major effect in speeding up the dynamics, across the board, while maintaining the slowing down of small units and speeding up of large units due to hybridization. With $\gamma = 0.5$, symmetric and asymmetric components have equal average weight, and dynamics in the hybrid network is extremely fast.

To probe the dynamics away from the vicinity of the metastable states, without touching the symmetry of the interactions, we use a variant of the simulation paradigm above, that mimics the arrival of an external input to the Potts network. That is, after a configuration has been thermalized as in previous simulations, a fraction of the units are randomly reset in a new state (different from the thermalized one), and then two independent trajectories evolve with the heat bath procedure from this common starting configuration, until the time τ when their overlap has been halved. Fig. 7 shows that the basic inversion effect, and in particular the selective slowing down of the "small" units, persists over wider regions of activity space. With respect to the standard thermalization paradigm in Fig. 7a, panel d shows that resetting a quarter of the units does indeed accelerate the dynamics of the $S = 3$ network, when it is homogeneous; whereas when it is hybridized with $S = 7$ units, these latter get faster, and slightly faster than the $S = 3$ ones.

In Fig. 7d, the external stimulus or perturbation is applied to a quarter of the units distributed in both sub-networks; when they are concentrated among the small $S = 3$ units (panel

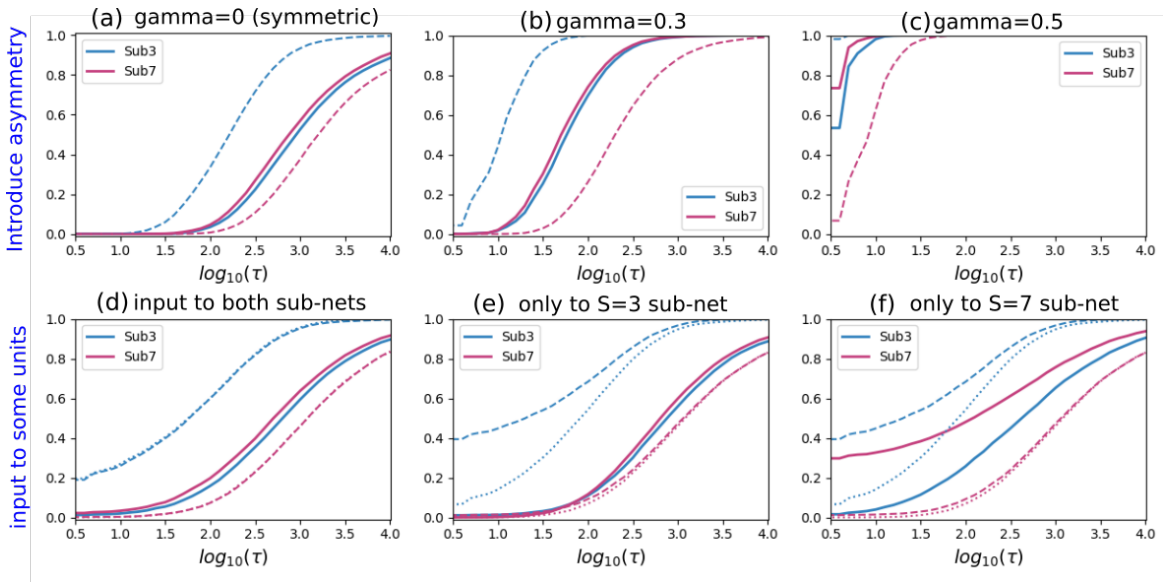

Figure 7: **Speeding up the dynamics with asymmetric connections and external inputs**. **(a)**: The speed-up and slow-down of sub-networks (relative to their homogeneous counterparts) are shown without asymmetry or perturbation, to serve as the "control" case. $T = 0.5$, $t_0 = 10^3$. **(b)**: The effect of asymmetry, where $\gamma = 0.3$, is to speed up the dynamics across both subnetworks, homogeneous or hybrid. **(c)**: With more asymmetry, $\gamma = 0.5$, the same general speed-up is seen as in (b), but more extreme. **(d)**: $N/4$ units are perturbed or reset, after thermalization, mimicking an external input; they are selected uniformly across the whole network. **(e)**: Those perturbed by the input are all in the smaller unit sub-network. **(f)**: They are all in the larger unit sub-network. In both (e) and (f) the dotted curves refer to unperturbed halves of homogeneous networks, and the dashed ones to the halves including the units receiving the input.

e, solid curves), the already minimal acceleration effect is reduced even further. When they are concentrated among the large $S = 7$ units, instead, their sub-network activity is markedly accelerated (panel f, solid curves), but only if it is part of a hybrid network with $S = 3$ units, with only minimal acceleration if they are part of a homogeneous network.

The results of the simulated external input procedure are therefore rather counter-intuitive: if affecting one fourth of the Potts units, the input effectively distances the network from its slow-evolving glassy state in two situations: when it is applied to a homogeneous network of small, but not large units, *or*, in a hybrid network, only when it is applied to the large units, but then it accelerates essentially their dynamics alone. These complex effects are observed still within the domain of networks with symmetric interactions, and they beg the question of what happens when an external input is combined with relaxing the symmetry constraint in a cortically plausible manner.

3.3 Approaching a cortical scenario

An interesting model of how cortical dynamics might influence cortical connectivity might be expressed by setting $\gamma = 0$ only for the interactions among the small units. This leads to a remarkable inversion effect, illustrated in Fig. 8a. One can see a self-consistent pattern potentially at play: the hybridization makes the large-S subnetwork fast, which upon synaptic plasticity would tend to result in more asymmetric tensorial couplings connecting those units.

To combine a putative external sensory input and the same type of asymmetry of Fig. 8a, in a cortically plausible scenario, we show in Fig. 8b what happens when resetting a fraction η of the small-S units (thus simulating an input to posterior cortex) after thermalization. The result is a moderate general speed-up, for both sub-networks, and very fast dynamics in about 30% of the runs, for the posterior network. It appears that in those runs the input has brought the small-S units close to the boundary between deep basins of attraction, so that fast noise leads to the immediate divergence of trajectories with the same starting point. For most of the other runs, instead, presumably well inside each large basin, the posterior network remains slower than the frontal one.

Finally, in Fig. 8c we take a major step towards cortical plausibility, by re-introducing the quiet state until now considered only in the thermodynamic analysis. The quiet state implies

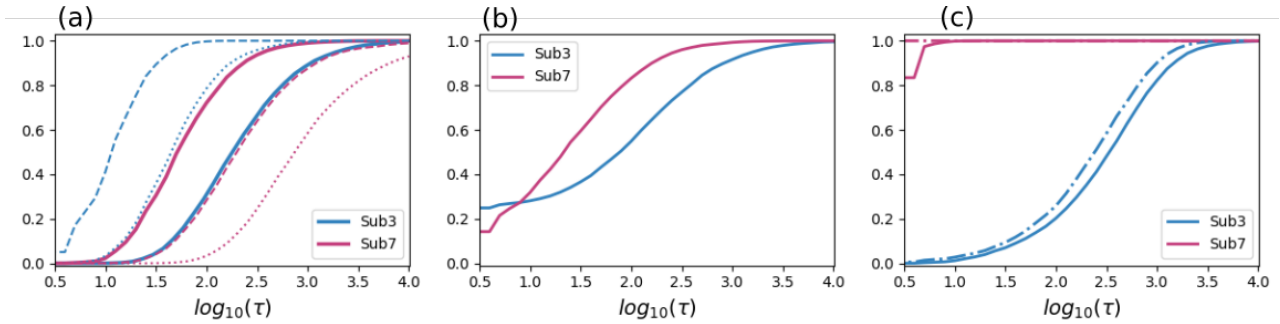

Figure 8: **The speed inversion effect likely applies to the cortex.** **(a)** Distribution of divergence times when the asymmetric component is zero only within the $S = 3$ sub-network and $\gamma = 0.3$ otherwise. For homogeneous networks, dotted curves are for the sub-networks that have zero asymmetric component. **(b)** Same as in (a), but half the $S = 3$ sub-network units are perturbed after thermalization. **(c)**. Potts glass model with a quiet state and with regulated mean-activity. After thermalization, a persistent external input is applied to the $S = 3$ sub-network, by flipping to a different active state a proportion $a\eta$ of its active units, inactivating a proportion $(1 - a)\eta$, and activating (in a random active state) the exact same number as those that get inactivated (which is close to $Na(1-a)\eta$, but varies somewhat in the course of each thermalization). The newly activated units are clamped. Broken curves show results when reintroducing asymmetry, $\gamma = 0.3$, in the connections involving the $S = 7$ units.

sparse activity (only a fraction a of the Potts units in one of their active states) and this overall level of sparsity must be conceived as being regulated by inhibition (in the analysis, this amounts to considering the activity level rather than the threshold U as a parameter, whereas for the implementation in the simulations see the Appendix). We first consider in this case purely symmetric random connections, and an input applied to some of the posterior units. To maintain the sparsity level, the input is applied after thermalization both to units already in an active state (which are then flipped to a different state) and to units in their quiescent state – in this case the input is *clamped* to keep them in the new state, simulating the strong effect of thalamic inputs impinging on an inactive local network. Again, we refer to the Appendix for a full description of the procedure. The result, in Fig. 8c is a strong differentiation between slow dynamics in the posterior network and an immediate divergence of nearly all trajectories in the frontal one. While this outcome stems to a large extent from clamping a few critical units only in the posterior network, it suggests that the main speed inversion phenomenon is not necessarily reversed back again when moving towards actual cortical dynamics. Reintroducing the asymmetry in the connections involving the $S = 7$ units only makes their network diverge immediately in *all* trajectories (the broken curves in Fig. 8c).

3.4 Short-term dynamics for the associative memory model

In this last Results subsection we consider the associative memory model, in which the interactions are not random but rather tend to align the network along one of a number p of pre-acquired memory states. Here there is no hypothesis about the overarching structure of memory representations in the cortex (we have reported elsewhere on the problems in applying to the cortex the simplest autoassociative retrieval scenario [20]) but rather we aim to to assess the effects on glassy dynamics of the presence of the large attractors associated with the memories. The logic is that we are probing the establishment of new representations, driven by either external inputs or internal dynamics, and if the network gets stuck into a previously acquired memory, no new configuration can be learned.

First, Fig. 9a shows that hybridization, i.e., the differentiation between large- and small-S units, in this case speeds up both subnetworks. In a homogeneous network, the $S = 7$ units are extremely slow, as nearly all trajectories are trapped in one of the large basins of attraction of the memories encoded in the connections, reflecting the very extensive storage capacity of the Potts network, quadratic in S [7, 21]. Also the trajectories of the $S = 3$ homogeneous network are slower than in the random network, which does not have the memory attractors, but faster than the $S = 7$ ones. The effect of hybridization is then much stronger on the $S = 7$ units.

What happens when applying, after thermalization, an external input to some of the $S = 3$ units? Not much, Fig. 9b shows, if the simulated input is applied to half of them (following the procedure used for Fig. 8c, with $\eta = 0.5$ and no clamp). If $\eta = 1.0$, instead, i.e. the input is applied to the entire $S = 3$ sub-network, the there is a major effect, particularly in producing immediate or very early divergence of many of the trajectories, but the speed inversion remains more or less unaltered (broken lines).

Finally, Fig. 9c shows that introducing moderate levels of asymmetry by diluting or cutting 30% of the connections *between* the two sub-networks does not have much of an effect either – unless one also clamps some of the units in the posterior network, in the manner already described; then, the posterior network slows down, almost to a standstill, which is intuitive, while surprisingly the anterior network speeds up further, as if unable to find any single satisfactory accomodation to the configuration imposed posteriorly.

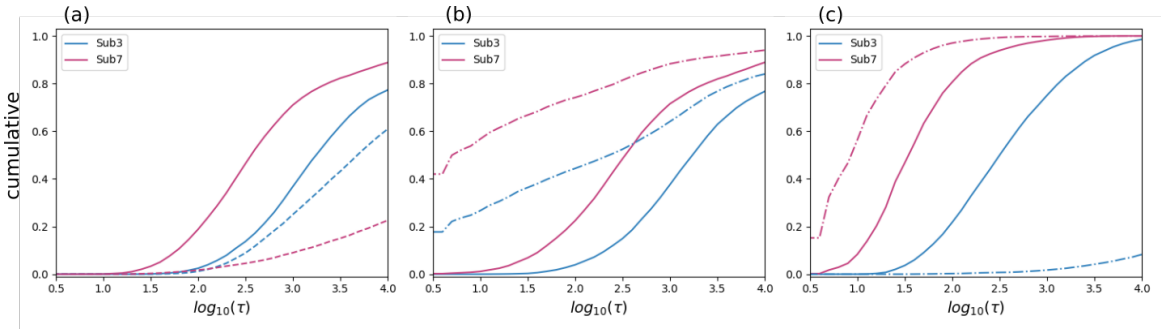

Figure 9: **Speed inversion occurs also in the associative memory model**. **(a)** Cumulative distribution for τ (on a log scale) without external input. Dashed curves are for the homogeneous network. **(b)** The input-driven divergence times, i.e., when half of the $S = 3$ active units are perturbed ($\eta = 0.5$, solid curves) and all of the $S = 3$ active units are perturbed ($\eta = 1.0$, broken curves). **(c)** Asymmetric connections between the two sub-networks, obtained by removing/pruning 30% of them, results in only quantitative changes. The slow-down and also the speed-up are dramatic, instead, when in addition, like in Fig. 8c, the newly activated units are clamped by persistent external inputs (broken curves).

4 Discussion

Our study is premised on the hypothesis that some of the characteristics of cortical dynamics have their roots in the statistical physics of disordered systems. Prior to attempting to validate the connection between two levels of analysis so distant from each other, we wanted to explore what the statistical properties might be, that might find – or not – their expression at the neural systems level. We have considered the reduction of Braitenberg's model of cortical connectivity to a Potts network, and reviewed the thermodynamic analysis that predicts different types of transition from paramagnetic to glassy phase, as a function of the number S of local states. Surprisingly, when combining in a "hybrid" network two halves with "low" and "high" S units, the former are slowed down by the interaction, and the latter are sped up, to the point of overtaking the former. This effect might be related to the different order of the phase transition to the glassy phase, but remarkably it is a reversal of the difference presented by homogeneous networks. Although one can construct seemingly intuitive explanations *a posteriori*, those did not enable us to *predict* it, in the least.

The speed inversion effect appears to survive largely unaltered the introduction of additional elements and details, and, importantly, the replacement of the random network with an associative memory with connections structured by learning.

What are the implications for cortical processing? First, one should note that such impli-

cations should be taken with more than a grain of salt, if anything because the key concept of a single cortical axis is rather ill-defined, at best. Perusing the many parameters of cortical circuitry that have been observed to vary across cortical areas, and the many more likely to be reported in the future, describing their variation as aligned to an axis, let alone whether it is the *same* axis across parameters, is a wishful simplification. The sensory-motor hierarchies conceptualized e.g. by Fuster [40] have their final station in motor cortex *after* passing through the more anterior prefrontal cortices, while the termination layers of intracortical fibers, used to distinguish between feed-forward and feedback projections, define a cortical hierarchy with the hippocampus at the top, the limbic cortices next to it just below, then the association cortices of both temporal and frontal lobes, going down all the way to primary sensory *and* motor cortices [41]. In terms of the number of largely local inputs to the basal dendrites of pyramidal cells, instead, Elston [42] gives estimates for areas V1, 7a, TE and 12, in macaque monkey, that are roughly in the ratios 1:4:11:16, more or less along a posterior-to-anterior axis – but then measures in other areas do not necessarily align, for example area 10 at the frontal pole is anterior to 12, but its pyramidal cells are estimated to have on average 17% fewer spines.

Our hybrid Potts network discards such complexity anyway to favor simplicity, and the speeding up of the large-S units that it reveals may have to be factored in, as an underlying phenomenon, in any complex scenario that envisages an imbalance between the effective numerosity of local attractor states across the cortical mantle.

Interestingly, machine learning has pointed out the usefulness of combining "processing units" with memory properties at different time scales (*LSTM* units), e.g. to tackle syntax in language production and understanding. In particular, it has been predicted that long- and short-range units, which are taken to correspond to patches of cortex of perhaps 10^6 neurons, similar to our Potts units, should reside in different cortical regions [43]. Our findings should prove useful to research in this natural language processing framework, by at least contributing a warning that the properties of the units in a homogeneous network, or even in isolation, may differ, to the point of being the opposite, from those of the same units in a hybrid one.

A rather different linguistic domain in which the effective speed or slowness of glassy dynamics may be important is language evolution. There, it has long been hypothesized that the syntactic parameters that determine the internal structure of language and that evolve or even

"mutate", like units of a genetic code, on a scale of hundreds or thousands of years [44], may all be *binary*. Notably, many other features which are needed to describe natural languages and to implement them in artificial systems are obviously far from binary and appear to evolve, largely, on faster time scales. Our study suggests that in a network of parameters with effectively random interactions, those that emerge in evolution as more resistant to change, and therefore describe the most stable internal structure or set of *motifs* of a natural language are precisely the binary ones, whether or not they possess a default value [45].

Yet other seemingly distant domains are those of protein folding and evolution, which have been approached with simplified Potts models [46, 47]. The possible application of our results to these different fields is left for future work.

5 Appendix

5.1 The first-step RSB free energy of Potts glass model

The Edward-Anderson order parameter is set as

$$q_{ab} = \begin{cases} \tilde{q}, \ a = b \\ q_1, \ a \neq b, \ \lfloor \frac{a}{m} \rfloor = \lfloor \frac{b}{m} \rfloor \\ q_0, \ a \neq b, \ \lfloor \frac{a}{m} \rfloor \neq \lfloor \frac{b}{m} \rfloor. \end{cases} \tag{39}$$

Then the free energy reads,

$$f[q_1, q_0, m] = \frac{\beta^2 J^2}{4} \lambda^4 [m(q_1^2 - q_0^2) - q_1^2] + \beta^2 J^2 \lambda^4 \tilde{q} - \frac{1}{m} \int \big(\prod_{l>0} Dz_l \big) \ln \int \big(\prod_{k>0} Dy_k \big) L^m, \tag{40}$$

where

$$L \equiv \sum_{\sigma=0}^{S} \exp \Big\{ \big[\beta^2 J^2 S \big(\tilde{q} - \frac{q_1 - q_0}{2} \big) - \beta U \big] (1 - \delta_{\sigma 0}) + \beta J \lambda^2 \sum_{l>0} \big(\sqrt{q_0} z_l + \sqrt{q_1 - q_0} y_l \big) V_\sigma^l \Big\} \tag{41}$$

and

$$Dy \equiv \frac{dy}{\sqrt{2\pi}} \exp \Big(- \frac{y^2}{2} \Big). \tag{42}$$

5.2 Finite number of patterns in associative memory

We compute the maximum value of U, below which the onset of glassy phase is continuous in the limit of $\alpha = p/N \to 0$. The trivial solution of Eqs. (28), $\overrightarrow{m} = 0$, is stable as long as the corresponding eigenvalue

$$\lambda = 1 - \frac{\tilde{a}(S-1)}{T} \frac{\exp(-U/T)}{1 + S\exp(-U/T)} \tag{43}$$

remains positive. This is always the case for $T > T_c$, the critical temperature where the trivial solution becomes unstable and Mattis states appear. Mattis states satisfy

$$m = a \frac{\sum_{\sigma>0}(\delta_{\sigma 1} - 1/S)\exp[\beta m(\delta_{\sigma 1} - 1/S) - \beta U]}{1 + \sum_{\sigma>0}\exp[\beta m(\delta_{\sigma 1} - 1/S) - \beta U]}. \tag{44}$$

and the critical temperature for a given value of U is determined by solving

$$T_c = \tilde{a}(S-1)\frac{\exp(-U/T_c)}{1 + S\exp(-U/T_c)}. \tag{45}$$

We can see that if $U \to +\infty$, $T_c \to 0$ and the trivial solution $\overrightarrow{m} = 0$ is stable for all temperature. In Fig. 10, we show values of U_c and the critical temperature at $U = U_c$.

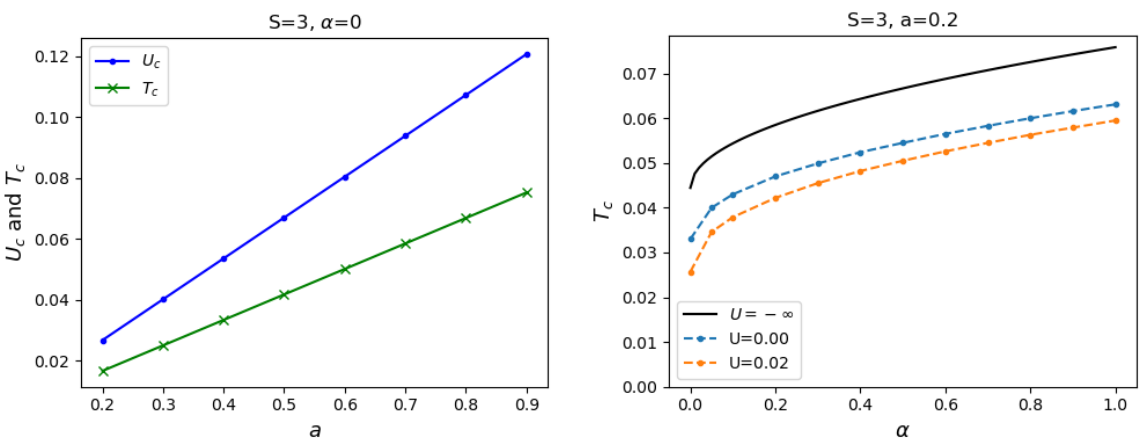

Figure 10: **High temperature phase of associative memory** for $S = 3$. **Left**: maximum value of U (blue) above which the transition is no longer continuous, and the corresponding critical temperature (green) are plotted against sparsity of patterns for $\alpha \to 0$. **Right**: Critical temperature as a function of α for $a = 0.2$. Note that $U = 0.02$ is just below $U_c \approx 0.026$ given on the left panel.

5.3 Details on the computer simulations

For models with a quiet state, the Edward-Anderson order parameter is computed as, instead of Eq. (37),

$$q_{ab} = \frac{S}{S-1} \frac{\sum_i (\delta_{\sigma^a \sigma^b} - 1/S)(1 - \delta_{\sigma^a 0})(1 - \delta_{\sigma^b 0})}{\sum_i (1 - \delta_{\sigma^a 0})(1 - \delta_{\sigma^b 0})}. \tag{46}$$

The mean activity of the network is controlled by time-dependent threshold

$$U(t) = U_0 + k\Big[\frac{1}{N}\sum_i (1 - \delta_{\sigma_i 0}) - a\Big]^3, \tag{47}$$

where a is the sparsity of patterns in associative memory model and k is set as 1000. For the Potts glass model with a quiet state, we have used the same activity level a.

The external input to the posterior sub-network is modelled by persistent external fields applied (after thermalization) to a fraction η of its units, which will maintain its states during dynamics (clamping in the main text). Specifically, we randomly select a fraction η of all active units in the $S = 3$ sub-network. Among the selected units, a fraction a of them is flipped into a different active state, while the remaining fraction $1 - a$ of them is set into a quiet state. The same number of units among quiet units is activated to maintain the same level of activity.

If not specified explicitly, parameters are set as in Table 1.

Table 1: Parameters of the network

Symbol	Meaning	Default value
N	number of Potts units	256
S	number of states per unit	7 (3)
T	Noise level (temperature)	0.5
γ	degree of asymmetry	0.3
η	fraction of units with external inputs	0.5
p	number of memory patterns	1024
t_0	number of thermalization updates	1000
a	mean activity	0.25

References

[1] Paul L Nunez. The brain wave equation: a model for the eeg. *Mathematical Biosciences*, 21(3-4):279–297, 1974.

[2] Pedro Valdés, J Bosch, R Grave, J Hernandez, J Riera, R Pascual, and R Biscay. Frequency domain models of the eeg. *Brain topography*, 4:309–319, 1992.

[3] Daniele Daini, Giacomo Ceccarelli, Enrico Cataldo, and Viktor Jirsa. Spherical-harmonics mode decomposition of neural field equations. *Physical Review E*, 101(1):012202, 2020.

[4] Daniel J Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1989.

[5] V Braitenberg and A Schütz. Anatomy of the cortex: Statistics and geometry, 1991.

[6] Valentino Braitenberg. Cortical architectonics: general and areal. In *Architectonics of the cerebral cortex*, pages 443–465. Raven Press, 1978.

[7] Ido Kanter. Potts-glass models of neural networks. *Physical Review A*, 37(7):2739, 1988.

[8] Eleonora Russo and Alessandro Treves. Cortical free-association dynamics: Distinct phases of a latching network. *Physical Review E*, 85(5):51920, 2012.

[9] Kwang Il Ryom, Vezha Boboeva, Oleksandra Soldatkina, and Alessandro Treves. Latching dynamics as a basis for short-term recall. *PLoS computational biology*, 17(9):e1008809, 2021.

[10] Guy N Elston, Ruth Benavides-Piccione, and Javier DeFelipe. The pyramidal cell in cognition: a comparative study in human and monkey. *Journal of Neuroscience*, 21(17):RC163–RC163, 2001.

[11] Earl K Miller, Cynthia A Erickson, and Robert Desimone. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of neuroscience*, 16(16):5154–5167, 1996.

[12] Pia Rotshtein, Richard NA Henson, Alessandro Treves, Jon Driver, and Raymond J Dolan. Morphing marilyn into maggie dissociates physical and identity face representations in the brain. *Nature neuroscience*, 8(1):107–113, 2005.

[13] Claus C Hilgetag, Alexandros Goulas, and Jean-Pierre Changeux. A natural cortical axis connecting the outside and inside of the human brain. *Network Neuroscience*, 6(4):950–959, 2022.

[14] Ethan M Meyers, David J Freedman, Gabriel Kreiman, Earl K Miller, and Tomaso Poggio. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of neurophysiology*, 100(3):1407–1419, 2008.

[15] Antonio R Damasio and Daniel Tranel. Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences*, 90(11):4957–4960, 1993.

[16] Antonio Daniele, Laura Giustolisi, M Caterina Silveri, Cesare Colosimo, and Guido Gainotti. Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. *Neuropsychologia*, 32(11):1325–1341, 1994.

[17] Lorraine K Tyler, Richard Russell, Jalal Fadili, and Helen E Moss. The neural representation of nouns and verbs: Pet studies. *Brain*, 124(8):1619–1634, 2001.

[18] Michael T Ullman. Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2):231–270, 2004.

[19] Vezha Boboeva, Romain Brasselet, and Alessandro Treves. The capacity for correlated semantic memories in the cortex. *Entropy*, 20(11):824, 2018.

[20] Kwang Il Ryom, Debora Stendardi, Elisa Ciaramelli, and Alessandro Treves. Computational constraints on the associative recall of spatial scenes. *bioRxiv*, pages 2022–10, 2022.

[21] Emilio Kropff and Alessandro Treves. The storage capacity of potts models for semantic memory retrieval. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(08):P08010, 2005.

[22] Michelangelo Naim, Vezha Boboeva, Chol Jun Kang, and Alessandro Treves. Reducing a cortical network to a potts model yields storage capacity estimates. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(4):043304, 2018.

[23] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[24] David Elderfield and David Sherrington. The curious case of the potts spin glass. *Journal of Physics C: Solid State Physics*, 16(15):L497, 1983.

[25] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[26] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[27] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[28] David J Gross, Ido Kanter, and Haim Sompolinsky. Mean-field theory of the potts glass. *Physical review letters*, 55(3):304, 1985.

[29] Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13(3):1101, 1980.

[30] Giorgio Parisi. A sequence of approximated solutions to the sk model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[31] V Janiš and A Klíč. Mean-field solution of the potts glass near the transition temperature to the ordered phase. *Physical Review B*, 84(6):064446, 2011.

[32] Andrea Crisanti and Tommaso Rizzo. Analysis of the ∞-replica symmetry breaking solution of the sherrington-kirkpatrick model. *Physical Review E*, 65(4):046137, 2002.

[33] Emilio De Santis, Giorgio Parisi, and Felix Ritort. On the static and dynamical transition in the mean-field potts glass. *Journal of Physics A: Mathematical and General*, 28(11):3025, 1995.

[34] AP Young. Direct determination of the probability distribution for the spin-glass order parameter. *Physical review letters*, 51(13):1206, 1983.

[35] Giorgio Parisi. Magnetic properties of spin glasses in a new mean field theory. *Journal of Physics A: Mathematical and General*, 13(5):1887, 1980.

[36] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Statistical mechanics of neural networks near saturation. *Annals of physics*, 173(1):30–67, 1987.

[37] Désiré Bollé, Roland Cools, Patrick Dupont, and J Huyghebaert. Mean-field theory for the q-state potts-glass neural network with biased patterns. *Journal of Physics A: Mathematical and General*, 26(3):549, 1993.

[38] Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47(5):359, 1981.

[39] Alessandro Treves and Daniel J Amit. Metastable states in asymmetrically diluted hopfield networks. *Journal of Physics A: Mathematical and General*, 21(14):3155, 1988.

[40] Joaquin Fuster. *The prefrontal cortex*. Academic press, 2015.

[41] Helen Barbas and Nancy Rempel-Clower. Cortical structure predicts the pattern of corticocortical connections. *Cerebral cortex (New York, NY: 1991)*, 7(7):635–646, 1997.

[42] Guy N Elston. Pyramidal cells of the frontal lobe: all the more spinous to think with. *Journal of Neuroscience*, 20(18):RC95–RC95, 2000.

[43] Yair Lakretz, Stanislas Dehaene, and Jean-Rémi King. What limits our capacity to process nested long-range dependencies in sentence comprehension? *Entropy*, 22(4):446, 2020.

[44] Giuseppe Longobardi and Cristina Guardiano. Evidence for syntax as a signal of historical relatedness. *Lingua*, 119(11):1679–1706, 2009.

[45] Giuseppe Longobardi and Alessandro Treves. Grammatical parameters from a gene-like code to self-organizing attractors: a research program. page 2023, 2023.

[46] Simona Cocco, Remi Monasson, and Martin Weigt. From principal component to direct coupling analysis of coevolution in proteins: Low-eigenvalue modes are needed for structure prediction. *PLoS computational biology*, 9(8):e1003176, 2013.

[47] Simona Cocco, Christoph Feinauer, Matteo Figliuzzi, Rémi Monasson, and Martin Weigt. Inverse statistical physics of protein sequences: a key issues review. *Reports on Progress in Physics*, 81(3):032601, 2018.