

Article

Life on the Edge: Latching Dynamics in a Potts Neural Network

Chol Jun Kang ^{1,2} , Michelangelo Naim ^{1,3} , Vezha Boboeva ¹ and Alessandro Treves ^{1,4,*} 

¹ Cognitive Neuroscience, SISSA—International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy; ckang@sissa.it (C.J.K.); michelangelonaim@gmail.com (M.N.); vboboeva@sissa.it (V.B.)

² The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

³ Department of Physics, La Sapienza Università di Roma, Piazzale Aldo Moro, 5, 00185 Roma, Italy

⁴ Centre for Neural Computation, Norwegian University of Science and Technology, 7491 Trondheim, Norway

* Correspondence: ale@sissa.it; Tel.: +39-040-3787-623

Received: 2 August 2017; Accepted: 29 August 2017; Published: 3 September 2017

Abstract: We study latching dynamics in the adaptive Potts model network, through numerical simulations with randomly and also weakly correlated patterns, and we focus on comparing its slowly and fast adapting regimes. A measure, Q , is used to quantify the *quality of latching* in the phase space spanned by the number of Potts states S , the number of connections per Potts unit C and the number of stored memory patterns p . We find narrow regions, or *bands* in phase space, where distinct pattern retrieval and duration of latching combine to yield the highest values of Q . The bands are confined by the storage capacity curve, for large p , and by the onset of finite latching, for low p . Inside the band, in the slowly adapting regime, we observe complex structured dynamics, with transitions at high crossover between correlated memory patterns; while away from the band latching, transitions lose complexity in different ways: below, they are clear-cut but last such few steps as to span a transition matrix between states with few asymmetrical entries and limited entropy; while above, they tend to become random, with large entropy and bi-directional transition frequencies, but indistinguishable from noise. Extrapolating from the simulations, the band appears to scale almost quadratically in the p – S plane, and sublinearly in p – C . In the fast adapting regime, the band scales similarly, and it can be made even wider and more robust, but transitions between anti-correlated patterns dominate latching dynamics. This suggest that slow and fast adaptation have to be integrated in a scenario for viable latching in a cortical system. The results for the slowly adapting regime, obtained with randomly correlated patterns, remain valid also for the case with correlated patterns, with just a simple shift in phase space.

Keywords: neural network; Potts model; latching; recursion

1. Introduction

How can the human brain produce creative behaviour? Systems neuroscience has mainly focused on the states induced, in particular in the cortex, by external inputs, be these states simple distributions of neuronal activity or more complex dynamical trajectories. It has largely eschewed the question of how such states can be combined into novel sequences that express, rather than the reaction to an external drive, spontaneous cortical dynamics. However, the generation of novel sequences of states drawn from even a finite set has been characterized as the infinitely recursive process deemed to underlie language productivity, as well as other forms of creative cognition [1]. If the individual states, whether fixed points or stereotyped trajectories, are conceptualized as dynamical attractors [2], the cortex can be thought of as engaging in a kind of chaotic saltatory dynamics between such attractors [3]. Attractor dynamics has indeed fascinated theorists, and a major body of work has shown how to make relevant for neuroscience the concepts and analytical tools developed within

statistical physics, but the focus has been on compact, homogeneous neural networks [4–7]. These have been regarded as simplified models of local cortical networks—as well as, e.g., of the CA3 hippocampal field—and have not been analysed in their potential saltatory dynamics, given that it would make no sense to consider local cortical networks as isolated systems. Even in the case of a ground-breaking investigation of putative spatial trajectory planning [8], the hippocampal activity that expressed it was thought not to be entirely endogenous, but rather guided by external inputs, including those representing goals and path integration. Therefore, formal analyses of model networks endowed with attractor dynamics have been largely confined to the simple paradigm of cued retrieval from memory. Attempts have been made to explore methodologies to study mechanisms beyond simple cued retrieval [9,10], for example those involved in drawing, confabulation, thought processes in general, and language, which are all considered to be largely independent of external stimuli, at their core, and to combine generativity with recursion [11–16].

Potts neural networks, on the other hand, originally studied merely as a variant of mathematical or potentially applied interest [17–21], offer one approach to model spontaneous dynamics in extended cortical systems, in particular if simple mechanisms of temporal adaptation are taken into account [22]. They can be subject to rigorous analyses of e.g., their storage capacity [23], of the mechanics of saltatory transitions between states [24] and are amenable to a description in terms of distinct “thermodynamic” phases [25,26]. The dynamic modification of thresholds with timescales separate from that of retrieval, i.e., temporal adaptation, together with the correlation between cortical states, are key features characterizing cortical operations, and Potts network models may contribute to elucidate their roles. Adaptation and its role in semantic priming [27] have been linked to the instability manifested in schizophrenia [28].

The Potts description is admittedly an oversimplified effective model for an underlying two-level auto-associative memory network [29]. The even more drastically simplified model of latching dynamics considered by the Tsodyks group [30,31], however, has afforded spectacular success in explaining the scaling laws obtained for free recall in experiments performed 50 years ago. The Potts model may be relevant to a wide set of behaviours and to related experimental measures, once the correspondence between model parameters and the quantities characterizing the underlying two-level network are elucidated. On this correspondence, we elaborate in a separate study [32]. Here, we ask when does the Potts network latch?

2. The Model

We consider an attractor neural network model comprised of Potts units, as depicted in Figure 1. The rationale for the model is that each unit represents a local network of many neurons with its own attractor dynamics [4,6], but in a simplified/integrated manner, regardless of detailed local dynamics. Local attractor states are represented by $S + 1$ Potts states: S active ones and one quiescent state (intended to describe a situation of no retrieval in the local network), σ_i^k , $k = 0, 1, \dots, S$, with the constraint that $\sum_{k=0}^S \sigma_i^k \equiv 1$. We call this autoassociative network of Potts units a Potts network, and refer to our earlier studies of some of its properties [22–25,33].

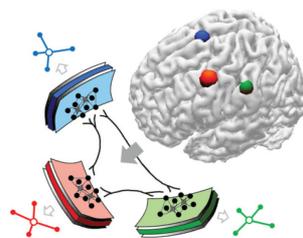


Figure 1. Global cortical model as a Potts neural network. Reprinted with permission from [25].

The “synaptic” connection between two Potts units is in fact a tensor summarizing the effect of very many actual connections between neurons in the two local networks, but still following the Hebbian learning rule [34], we write the connection weight between unit i in state k and unit j in state l as [23]

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

where c_{ij} is 1 if two units i and j have a connection and 0 otherwise, C is the average number of connections per unit, a is the sparsity parameter, i.e., the fraction of active units in every stored global activity pattern ($\{\xi_i^\mu\}$, $\mu = 1, 2, \dots, p$) and p is the number of stored patterns. The last two delta functions imply that the learned connection matrix does not affect the quiescent states. We will use the indices i, j for units, k, l for states and μ, ν for patterns. Units are updated in the following way:

$$\sigma_i^k = \frac{\exp(\beta r_i^k)}{\sum_{l=1}^S \exp(\beta r_i^l) + \exp[\beta(\theta_i^0 + U)]} \tag{2}$$

and

$$\sigma_i^0 = \frac{\exp[\beta(\theta_i^0 + U)]}{\sum_{l=1}^S \exp(\beta r_i^l) + \exp[\beta(\theta_i^0 + U)]} \tag{3}$$

where r_i^k is the input to (active) state k of unit i integrated over a time scale τ_1 , while U and θ_i^0 are, respectively, the constant and time-varying component of the effective overall threshold for unit i , which in practice act as inverse thresholds on its quiescent state. θ_i^0 varies with time constant τ_3 , to describe local network adaptation and inhibitory effects. The stiffness of the local dynamics is parametrized by the inverse “temperature” β (or T^{-1}), which is then distinct from the standard notion of thermodynamic noise. The input-output relations (2) and (3) ensure that

$$\sum_{k=0}^S \sigma_i^k = 1.$$

In addition to the overall threshold, θ_i^k is the threshold for unit i specific to state k , and it varies with time constant τ_2 , representing adaptation of the individual neurons active in that state, i.e., their neural or even synaptic fatigue. The time evolution of the network is then governed by equations that include three distinct time constants:

$$\tau_1 \frac{dr_i^k(t)}{dt} = h_i^k(t) - \theta_i^k(t) - r_i^k(t), \tag{4}$$

$$\tau_2 \frac{d\theta_i^k(t)}{dt} = \sigma_i^k(t) - \theta_i^k(t), \tag{5}$$

$$\tau_3 \frac{d\theta_i^0(t)}{dt} = \sum_{k=1}^S \sigma_i^k(t) - \theta_i^0(t), \tag{6}$$

where the field that the unit i in state k experiences reads

$$h_i^k = \sum_{j \neq i} \sum_{l=1}^S J_{ij}^{kl} \sigma_j^l + w \left(\sigma_i^k - \frac{1}{S} \sum_{l=1}^S \sigma_i^l \right). \tag{7}$$

The “local feedback term” w is a parameter, first introduced in [25], that modulates the inherent stability of Potts states, i.e., that of local attractors in the underlying network model. It helps the network converge to an attractor faster by giving positive feedback to the most active states and so it effectively deepens their basins of attraction. Note that, in this formulation, feedback is effectively spread over (at least) three time scales: w is positive feedback mediated by collective attractor effects

at the neural activity time scale τ_1 , θ_i^k is negative feedback mediated by fatigue at the slower time scale τ_2 , while θ_i^0 is also negative, and it can be used to model both fast and slow inhibition; for analytical clarity, we consider the two options separately, as the “slowly adapting regime”, with $\tau_3 > \tau_2$, and the “fast adapting regime”, with $\tau_3 < \tau_1$. It would be easy, of course, to introduce additional time scales, for example by distinguishing a component of θ_i^0 that varies rapidly from one that varies slowly, but it would greatly complicate the observations presented in the following.

The overlap or correlation of the activity state of the network with the global memory pattern μ can be measured as

$$m_\mu = \frac{1}{Na(1-a/S)} \sum_{j \neq i} \sum_{l \neq 0}^S \left(\delta_{\xi_j^\mu}^l - \frac{a}{S} \right) \sigma_j^l. \quad (8)$$

Randomly correlated memory patterns are generated according to the following probability distribution

$$P(\xi_i^\mu = k) = \frac{a}{S}, \quad (9)$$

$$P(\xi_i^\mu = 0) = 1 - a,$$

while correlated patterns are generated by the multi-parent algorithm sketched in [22], which will be discussed in a separate study [35].

3. Results

When does robust latching, as a model of spontaneous sequence generation, occur? We address this question with extensive computer simulations, mostly focused on latching between randomly correlated patterns. We consider first the slowly adapting regime ($\tau_1 \ll \tau_2 \ll \tau_3$) in which active states (τ_2) adapt slower than activity propagation to other units (τ_1), while inhibitory feedback is restricted to an even slower timescale, τ_3 . Next, we contrast with it the fast adapting regime ($\tau_3 \ll \tau_1 \ll \tau_2$) in which, instead, inhibitory feedback is immediate, relative to the other two time scales.

The critical parameters at play are the number of patterns, p , the number of active states, S , and the number of connections per unit, C , and we also look at the effect of the feedback term w . The other parameters, including T , τ_1 , τ_2 , and τ_3 , are kept fixed during simulations, after having chosen a priori values that can lead to robust latching dynamics in the two regimes.

3.1. Slowly Adapting Regime

In the slowly adapting regime, over a (short) time of order τ_1 the network, if suitably cued, may reach one of the global attractors, and stay there for a while; whereupon, after an adaptation time of order τ_2 , it may latch to another attractor, or else activity may die [25]. However, how distinct is the convergence to the new attractor? One may assess this as the difference between the two highest overlaps the network activity has, at time t , with any of the memory patterns, $m_1(t) - m_2(t)$: ideally, $m_1 \simeq 1$ and m_2 is small, so their difference approaches unity. A summary measure of memory pattern discrimination can be defined as $d_{12} \equiv \langle \int dt (m_1(t) - m_2(t)) \rangle_{\text{initial cue}}$, where, of course, the identity of patterns 1 and 2 changes over the sequence.

As discussed in [25], by looking at the latching length, how long a simulation runs before, if ever, the network falls into the global quiescent state, one can distinguish several “phases”. Depending on the parameters, the dynamics exhibit finite or infinite latching behaviour, or no latching at all. Typically, when increasing the storage load p , the latching sequence is prolonged and eventually extends indefinitely, but, at the same time, its distinctiveness decreases, since memory patterns cannot be individually retrieved beyond the storage capacity; and, even before, each acquires neighbouring patterns, in the finite and more crowded pattern space, with which it is too correlated to be well discriminated.

In Figure 2, we see that, for each $S = (2, 3, 4)$, as p is increased beyond a certain value, latching dynamics rapidly picks up and extends eventually through the whole simulation, but, in parallel, its

discriminative ability decreases and almost vanishes—the p -range where d_{12} is large is in fact when there is no latching, and d_{12} only measures the quality of the initial cued retrieval. For $S = 1$ no significant latching sequence is seen, whereas for higher values, at fixed p , its distinctiveness increases with S , but its length decreases from the peak value at $S = 2$.

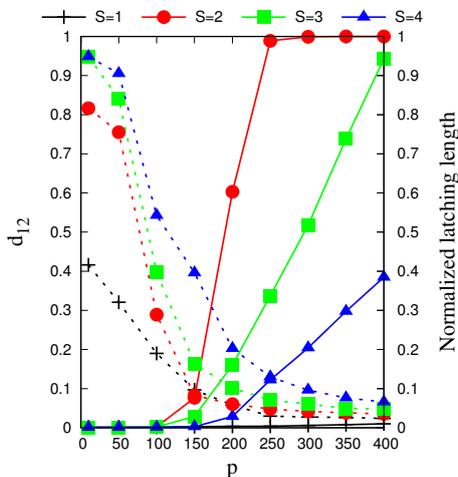


Figure 2. Trade-off between latching sequence length (solid lines) and retrieval discrimination (dashed lines). Different colors indicate different S values, while $C = 400$ throughout. The latching length l is in time steps (not in the number of transitions), normalized by the time of the simulation, $N_{update} = 6 \times 10^5$.

Since the latching length l is not itself sufficient to characterize latching and has to be complemented by discriminative ability, we find it convenient to quantify the overall *quality* of latching.

With a new quantity Q defined as

$$Q = d_{12} \cdot l \cdot \eta, \tag{10}$$

where η is introduced to exclude cases in which the network gets stuck in the initial cued pattern, so that no latching occurs; however, high d_{12} and l are:

$$\eta = \begin{cases} 1 & \text{if at least one transition to a second memory pattern occurs,} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Q is therefore a positive real number between 0 and 1, and we report its color-coded value to delineate the relevant phases in phase space.

Thus, *low quality* latching with small Q may result from either small d_{12} or short l , or both. The parameters that determine Q which we focus on are S , C and p , after having suitably chosen all the other parameters, which are kept fixed. Their default values in the slowly adapting regime are $N = 1000$, $a = 0.25$, $U = 0.1$, $T = 0.09$, $w = 0.8$, $\tau_1 = 3.3$, $\tau_2 = 100.0$, $\tau_3 = 10^6$, unless explicitly noted otherwise. If activity does not die out before, simulations are terminated after $N_{update} = 6 \times 10^5$ steps, the total number of updates of the entire Potts network, and are repeated with different cued patterns. Re the values of S , C and p , we use the following notation, for simplicity:

$$Q = Q(S, C, p) = \begin{cases} Q(S, p) & : C = 150 \text{ fixed,} \\ Q(C, p) & : S = 5 \text{ fixed.} \end{cases} \tag{12}$$

Figure 3 shows that there are narrow regions in the S - p and C - p planes, which we call *bands*, where relatively *high quality* latching occurs. The values of p with the “best” latching scale almost quadratically in S , and sublinearly in C . Moreover, one notices that, below certain values of S and C , no latching is seen, i.e., the band effectively ends at $S \sim 2$, $p \sim 90$ in Figure 3a and at $C \sim 50$,

$p \sim 70$ in Figure 3b. Importantly, the band in Figure 3a is confined in the area delimited by the cyan solid and dashed curves above and below it. The dashed curve is for the onset of latching, i.e., the phase transition to finite latching [25], while the solid curve above is the storage capacity curve in a diluted network, given by the approximate relation beyond which retrieval fails [25]. It should also be noted that overall Q values are not large, in fact well below 0.5 throughout both S - p and C - p planes. The reason is, again, in the conflicting requirements of persistent latching, favoured by dense storage, high p , and good retrieval, allowed instead only at low storage loads (in practice, relatively low p/S^2 and p/C values):

$$p_c \simeq \frac{CS^2}{4a \ln \frac{2S}{a\sqrt{\ln \frac{S}{a}}}} \tag{13}$$

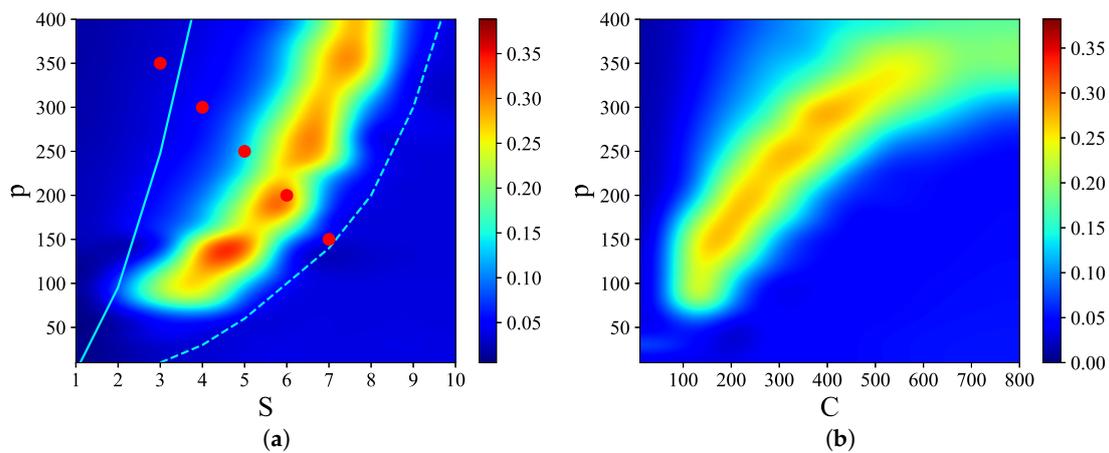


Figure 3. Phase space for $Q(S, p)$ in (a) and $Q(C, p)$ in (b) with randomly correlated patterns in the slowly adapting regime. The parameters are $C = 150$ and $S = 5$, if kept fixed, and $w = 0.8$. The red spots in (a) mark the parameter values used in the following analyses.

In Figure 4, we show representative latching dynamics at three selected points in the (S, p) plane, in terms of the time evolution of the overlap of the states with the stored activity patterns (see Equation (8)). The three points, marked in red, span across the band in Figure 3a, and we see that latching is indefinite but noisy in the example at $(5, 250)$, which is apparently too close to storage capacity, while memory retrieval is good at $(7, 150)$, but the sequence of states ends abruptly, as the network is in the phase of finite latching [25]. The two trends are representative of the two sides of the band, while in the middle, at $(6, 200)$, one finds a reasonable trade-off, with relatively good retrieval combined with protracted latching.

We use two statistical measures, the *asymmetry* of the transition probability matrix and Shannon’s *information entropy* [33,36,37] to characterize the essential features of the dynamics in different parameter regions. For that, we take all five red points from Figure 3a, such that they cut across the latching band in the S - p plane, and extend further upwards. We first compile a transition probability (or rather, frequency) matrix M from all distinct transitions observed along many latching sequences generated with the *same set* of stored patterns, as in [33]. The dimension of the matrix M is $(p + 1) \times (p + 1)$, as it includes all possible transitions between p patterns *plus* the global quiescent state. M is constructed from the transitions between states having both overlaps above a given threshold value, e.g., 0.5, in a data set of 1000 latching sequences, by accumulating their frequency between any two patterns into each element of the matrix and then normalizing to 1 row by row, so that $M_{\mu,\nu}$ reflects the probability of a transition from pattern μ to ν . A , the degree of asymmetry of M , is defined as

$$A = \frac{\|M - M^T\|}{\|M\|} \tag{14}$$

where M^T is the transpose matrix of M and $\|M\| = \sum_{\mu,\nu} |M_{\mu,\nu}|$. Note that A is small for unconstrained bi-directional dynamics and large for simpler stereotyped flows among global patterns, attaining its maximum value $A = 2$ for strictly uni-directional transitions. Note also that if the average had been taken over *different* realizations of the memory patterns, given sufficient statistics A would obviously vanish.

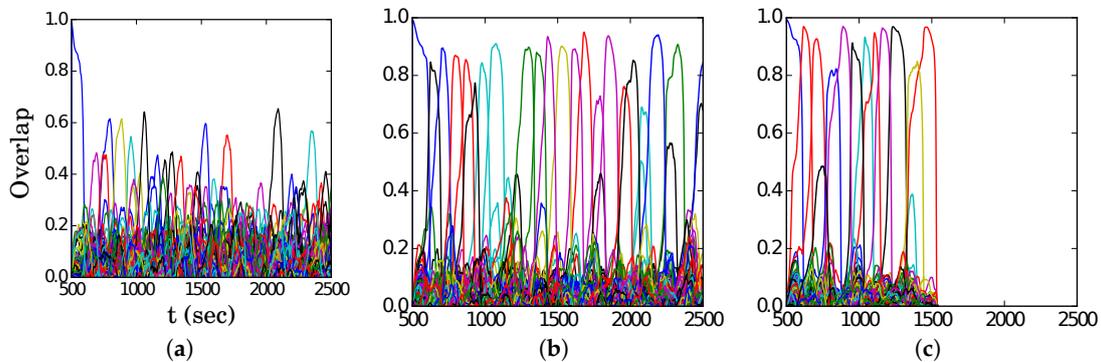


Figure 4. Latching behaviour for (S, p) equal to, respectively: (a) (5, 250); (b) (6, 200); and (c) (7, 150) in Figure 3a.

Another measure we apply to the transition matrix M is Shannon’s information entropy, defined as

$$I_\mu = \left\langle \frac{1}{\log_2(p + 1)} \sum_{\nu=1}^{p+1} M_{\mu,\nu} \log_2 \left(\frac{1}{M_{\mu,\nu}} \right) \right\rangle_\mu \tag{15}$$

I_μ takes positive real values from 0 (deterministic, all transitions from one state are to a single other state) to 1 (completely random), since it is normalized by $\log_2(p + 1)$, which corresponds to a completely random case.

We use these two measures, A and I_μ , on the points, marked red in Figure 3a.

$$(3, 350) - (4, 300) - (5, 250) - (6, 200) - (7, 150)$$

that lie on a segment going through the latching band observed in the slowly adapting regime. If we focus on transitions between states reaching at least a threshold overlap of 0.5, Figure 5 appears to show two complementary, almost opposite U-shaped curves as the two measures, asymmetry and entropy, are applied to the five points along the segment. One branch of each U shape extends over the range that includes the high- Q latching band: these are the right branches of the two curves, in which asymmetry decreases from a large value $A \simeq 1.6$ at (7, 150) to a smaller one $A \simeq 0.6$ at (5, 250), while concurrently the entropy increases from $I_\mu < 0.5$ at (7, 150) to $I_\mu > 0.8$ at (5, 250). As Figure 4 indicates, at (7, 150), latching sequences are distinct but very short, and few entries are filled in the transition matrix: generally either $M_{\mu\nu} = 0$ or $M_{\nu\mu} = 0$, so that asymmetry is high and entropy relatively low. This holds irrespective of the number of sequences that are averaged over. The opposite happens at (5, 250), where many transitions are observed, and in filling the transition matrix they approach the random limit. The point with the highest Q -value, (6, 200), is characterized by intermediate values of asymmetry and entropy which, we have previously observed, may be seen as a signature of complex dynamics [33]. Extending the range upwards, it seems as if the asymmetry, with threshold 0.5, were to eventually increase again, reaching its maximum $A = 2$ at (3, 350), with a decreasing entropy, vanishing at the same point (3, 350). These left branches are, however, dependent on the threshold values used, as Figure 5 shows, and do not imply that transitions become more deterministic because, in this region, there are simply fewer and fewer distinct transitions discernible above the noise (Figure 4). The left

branches merely reflect the increasing arbitrariness with which one can identify significant correlations with memory states in the rambling dynamics observed at higher storage loads.

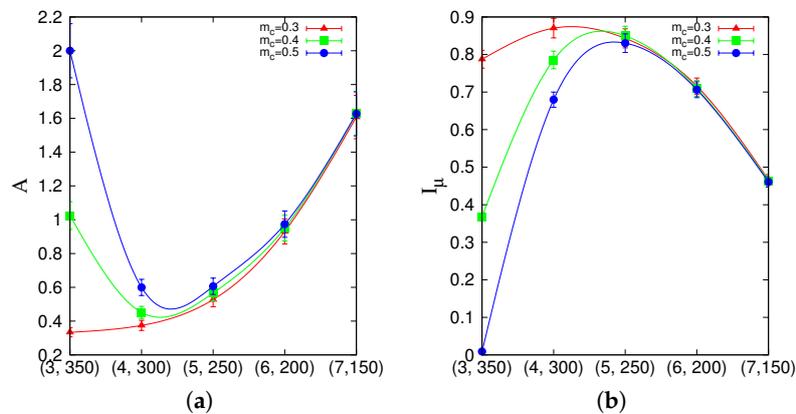


Figure 5. (a) asymmetry A of the transition matrix and (b) Shannon’s information entropy, I_μ along the (3, 350)–(4, 300)–(5, 250)–(6, 200)–(7, 150) parameter series from Figure 3. Different curves correspond to different thresholds for the overlap of the two states between which the network is defined to have a transition. The error bars report the standard deviation of either quantity for each of 1000 sequences.

In Figure 6, we see that the effect of the local feedback term, w , is first to enable latching sequences of reasonable quality, and then to also shift the latching band to higher values of S , effectively pushing this behaviour away from the storage capacity curve representing the retrieval capability of the Potts associative network. Hence, if one were to regard S as a structural parameter of the network, and w as a parameter that can be tuned, there is an optimal range of w values that allows good quality latching for higher storage. This argument has to be revised, however, by considering also the threshold U , since increasing w can be shown to be functionally equivalent, in terms of storage capacity, to decreasing U [32]. Also for U , in fact, one can find an optimal range for associative retrieval to occur, in the simple Potts network with no adaptation and with $w = 0$ [23]. This near equivalence between U and $-w$ does not hold anymore in the fast adapting regime, to which we turn next.

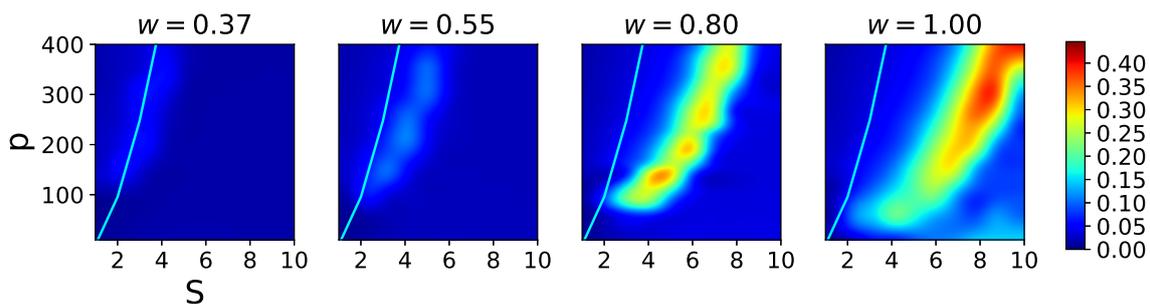


Figure 6. Latching quality $Q(S, p)$ with increasing local feedback, $w = 0.37, 0.55, 0.8,$ and 1.0 in the slowly adapting regime. Randomly correlated patterns are used, with $C = 150$ as in Figure 3a.

3.2. Fast Adapting Regime

We characterize the fast adapting regime by the alternative ordering of time scales $\tau_3 < \tau_1 \ll \tau_2$, such that the mean activity in each Potts unit is rapidly regulated by fast inhibition, at the time scale τ_3 . Equation (6) stipulates that $\sum_{k=1}^S \sigma_i^k(t)$, the total activity of each unit, is followed almost immediately, or more precisely at speed τ_3^{-1} , by the generic threshold $\theta_i^0(t)$. Extensive simulations, with the same parameters as for the slowly adapting regime, except for $w = 1.37, \tau_1 = 20, \tau_2 = 200$ and $\tau_3 = 10$, show that, similarly to the slowly adapting regime, there are latching bands in the $Q(S, p)$ and $Q(C, p)$

planes (see Figure 7). With these parameters, in particular, the larger value chosen for the feedback term w , the bands occupy a similar position as in the slowly adapting regime. Again, they appear to vanish below certain values of S and C , more precisely around $S \sim 3, p \sim 120$ in Figure 7a and around $C \sim 50, p \sim 90$ in Figure 7b, and to scale subquadratically in S and sublinearly in C . The band in the S - p plane is again confined by the storage capacity (solid cyan curve) and by the onset of (finite) latching (dashed curve). The storage capacity curve, which is independent of threshold adaptation, follows the same Equation (13).

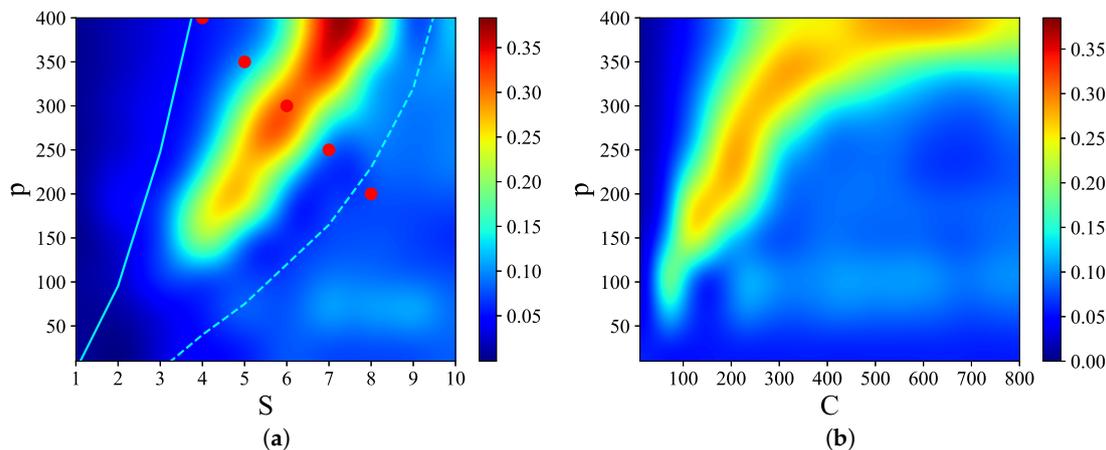


Figure 7. Phase space for $Q(S, p)$ in (a) and $Q(C, p)$ in (b) with randomly correlated patterns in the fast adapting regime. The parameters are identical to those in the slowly adapting regime, with the exception of $w = 1.37, \tau_1 = 20, \tau_2 = 200, \tau_3 = 10$. The red spots in (a) mark, again, the parameter values used in the Figures below.

Examples of latching behaviour outside and inside the band are presented in Figure 8, at the same values for S but shifted by $\Delta p = 100$, i.e., at the “red” points (5, 350), (6, 300), and (7, 250) in the S - p plane. Again, we see from Figure 7a that (5, 350) lies just above the band, while (6, 300) is right on the centre. To the right of the band, e.g., at (7, 250), the transitions are distinct but latching dies out very soon, while on the left, e.g., at (5, 350), the progressively reduced overlaps are a manifestation of increasingly noisy retrieval dynamics. In all three examples, we observe that latching steps proceed slowly, even slower than the doubled time scale $\tau_2 = 200$ would have led to predict. This appears to be because often a significant time elapses between the decay of the overlap of the network with one pattern and the emergence of a new one.

Figure 9 shows the asymmetry and entropy measures, A and I_μ , along the points

$$(4, 400) - (5, 350) - (6, 300) - (7, 250) - (8, 200),$$

in Figure 7a, where, again, we have chosen a series shifted by $\Delta p = 100$ upwards in order to centre it better on the high quality latching band. Only an overlap threshold of 0.5 is considered. What one can see, in contrast with the slowly adapting regime, is that now the two measures are not quite complementary. The point (6, 300) that lies inside the band, very much at its quality peak, shows again an intermediate value for the asymmetry, but the highest value, given the overlap threshold, for the entropy. The discrepancy may be ascribed to the different prevailing type of latching transition observed in the fast adapting regime, Figure 8. As discussed in [24], in a Potts network latching transitions with a high cross-over, which can only occur between memory patterns with a certain degree of correlation, can be distinguished from those with a vanishing cross-over, which are much more random. In the fast adapting regime, as indicated by the examples in Figure 8, all transitions tend to be of the latter type. A more careful analysis indicates, in fact, that they are quasi-random, in that they avoid a memory pattern in which largely the same Potts units are active as in the preceding

pattern. In fact, the value of the entropy at (6, 300) implies that on average from each of the 300 memory patterns there are transitions to at least 190 other patterns (to 190 if they were equiprobable, in practice many more); therefore, only the few patterns that happen to be more (spatially) correlated are avoided.

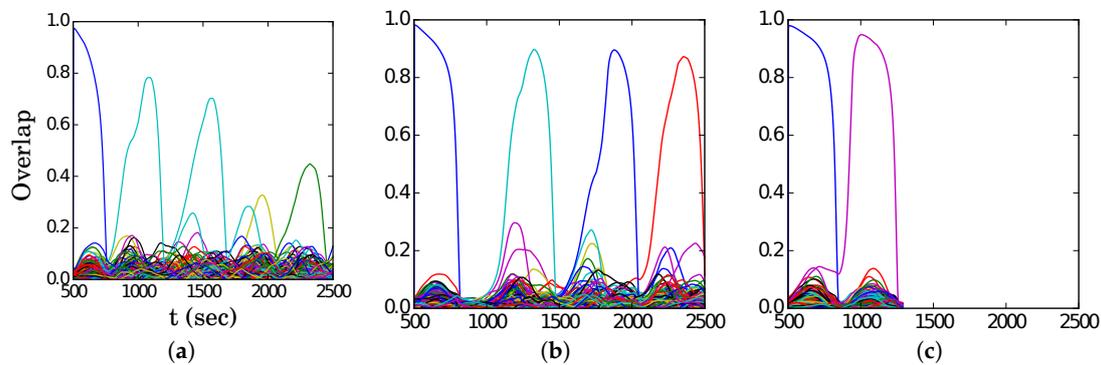


Figure 8. Latching behaviour for (S, p) equal to, respectively: (a) (5, 350); (b) (6, 300); and (c) (7, 250) in Figure 7a.

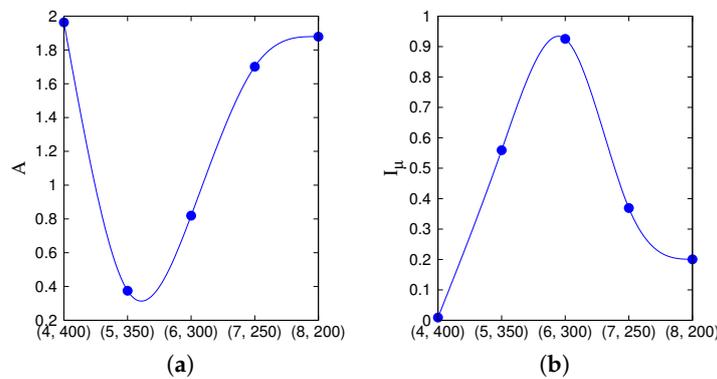


Figure 9. (a) asymmetry A of the transition matrix and (b) Shannon’s information entropy, I_μ along the (4, 400)–(5, 350)–(6, 300)–(7, 250)–(8, 200) parameter series from Figure 7a, using only a threshold 0.5 for the overlaps before and after each transition.

Towards the left, the curves do not vary much depending on the threshold chosen for the overlaps, but the asymmetry eventually becomes maximal and the entropy vanishes simply because sequences of robustly retrieved patterns do not last long, so, in this particular case, it would take more than 1000 sequences to accumulate sufficient statistics.

The effects of increasing the w term in the fast adapting regime are shown in Figure 10, where one notices two main features. First, there is heightened sensitivity to the exact value of w , so that relatively close data points at $w = 1.33, 1.37, 1.41,$ and 1.45 yield rather different pictures. Second, although again increasing w shifts the latching band rightward, by far the main effect is a widening of the band itself. This is because in the presence of rapid feedback inhibition a larger w term ceases to be functionally similar to a lower threshold, which in the slowly adapting regime was leading in turn to noisier dynamics and eventually indiscernible transitions. In the fast adapting regime, the increased positive feedback can be rapidly compensated by inhibitory feedback, so that in the high-storage region overlaps remain large, until they are suppressed by storage capacity constraints (the cyan curve, which remains at approximately the same distance from the larger and larger latching band).

We now turn to more explicit comparison of the transition dynamics in two regimes.

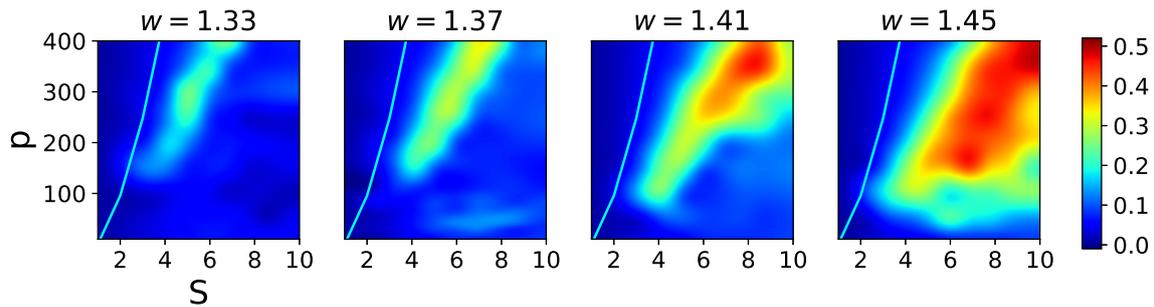


Figure 10. Latching quality $Q(S, p)$ with increasing local feedback, $w = 1.33, 1.37, 1.41,$ and 1.45 in the fast adapting regime. Randomly correlated patterns are used, with $C = 150$ as in Figure 7a.

3.3. Comparison of Two Regimes

To look more closely at latching dynamics in the slowly and fast adapting regimes, we take the following points from Figures 3a and 7a, which allow us to cut through the bands at two different storage levels

$$\begin{cases} p = 200, & S = (4, 5, 6, 7), \\ p = 400, & S = (6, 7, 8, 9). \end{cases} \quad (16)$$

Figure 11 shows in different colors the overlaps of the state of the network with the global patterns, for sample sequences along the points (16), in the slowly adapting regime. For both $p = 200$ and 400 , latching length is observed to decrease with S , unlike the discrimination between patterns, as measured by d_{12} , in agreement with Figure 2. Note that the two rows in the figure are similar, indicating that the shift $\Delta p = 200$ is approximately compensated by the rightward shift $\Delta S = 2$.

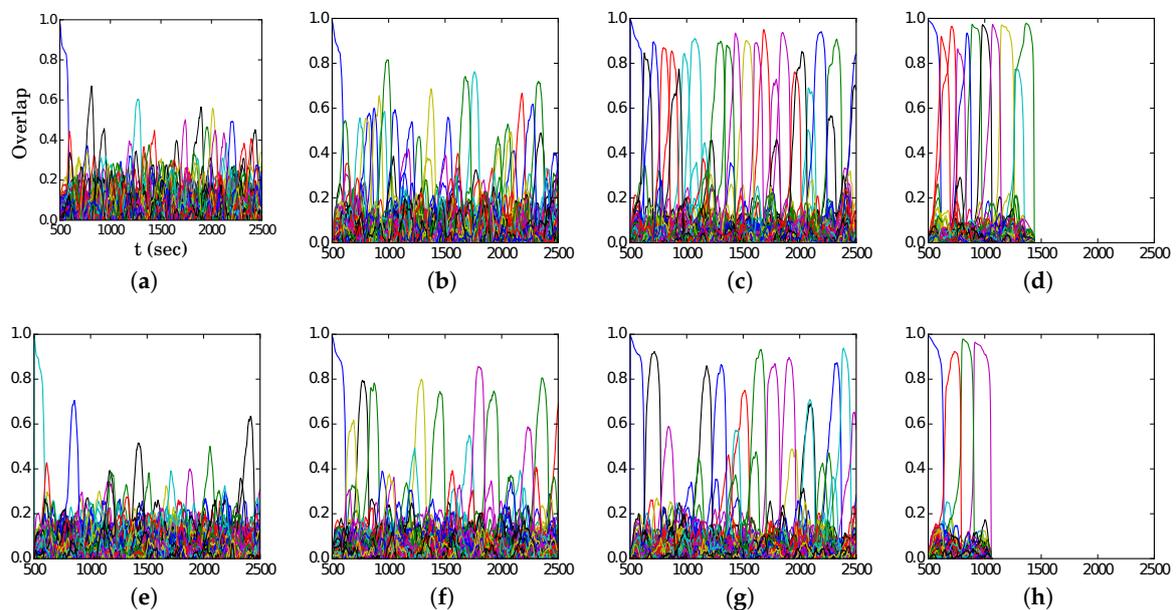


Figure 11. Latching behaviour in the slowly adapting regime. A sample of points (16) from Figure 3a. (a) $p = 200, S = 4$; (b) $p = 200, S = 5$; (c) $p = 200, S = 6$; (d) $p = 200, S = 7$; (e) $p = 400, S = 6$; (f) $p = 400, S = 7$; (g) $p = 400, S = 8$; (h) $p = 400, S = 9$.

The fast adapting regime shows the same trends, again one sees in Figure 12 the approximate compensation between the two shifts $\Delta p = 200$ and $\Delta S = 2$, but latching appears in general less noisy.

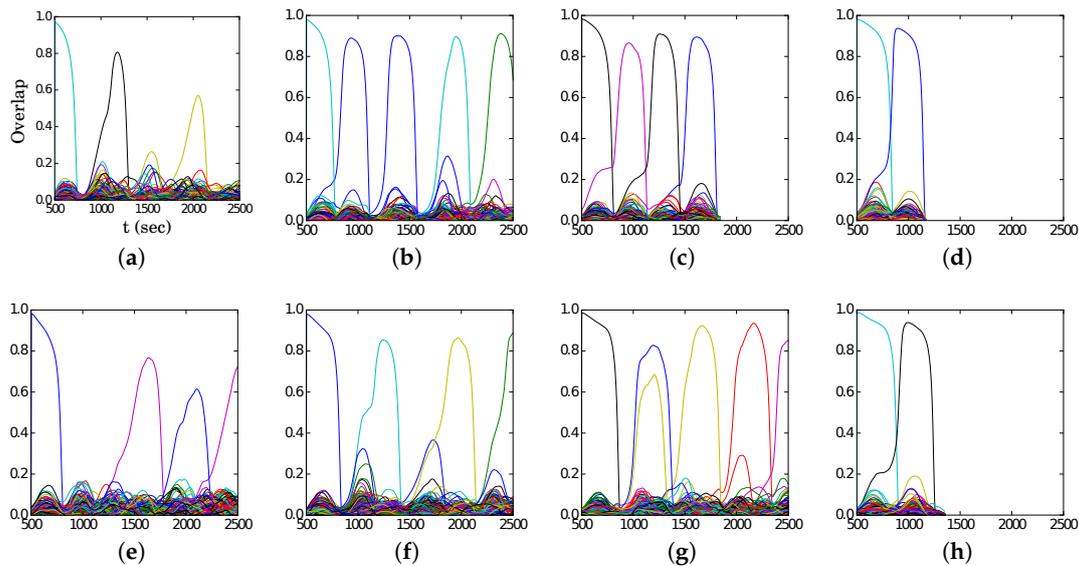


Figure 12. Latching behaviour in the fast adapting regime. A sample of points (16) from Figure 7a. (a) $p = 200, S = 4$; (b) $p = 200, S = 5$; (c) $p = 200, S = 6$; (d) $p = 200, S = 7$; (e) $p = 400, S = 6$; (f) $p = 400, S = 7$; (g) $p = 400, S = 8$; (h) $p = 400, S = 9$.

The main difference between the two regimes, however, is in the distribution of crossover values, those when the network has equal overlap with the preceding and the following pattern: their distribution (PDF, or probability density function) is shown in Figures 13 and 14.

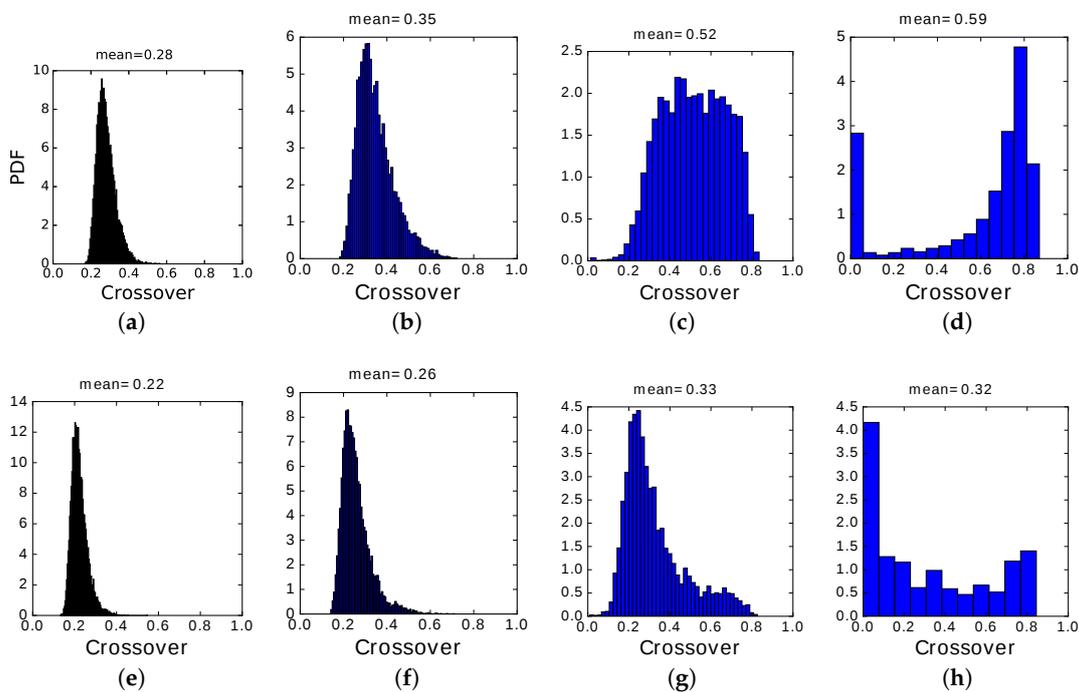


Figure 13. Probability density function (PDF) of crossover values in the slowly adapting regime. (a) $p = 200, S = 4$; (b) $p = 200, S = 5$; (c) $p = 200, S = 6$; (d) $p = 200, S = 7$; (e) $p = 400, S = 6$; (f) $p = 400, S = 7$; (g) $p = 400, S = 8$; (h) $p = 400, S = 9$.

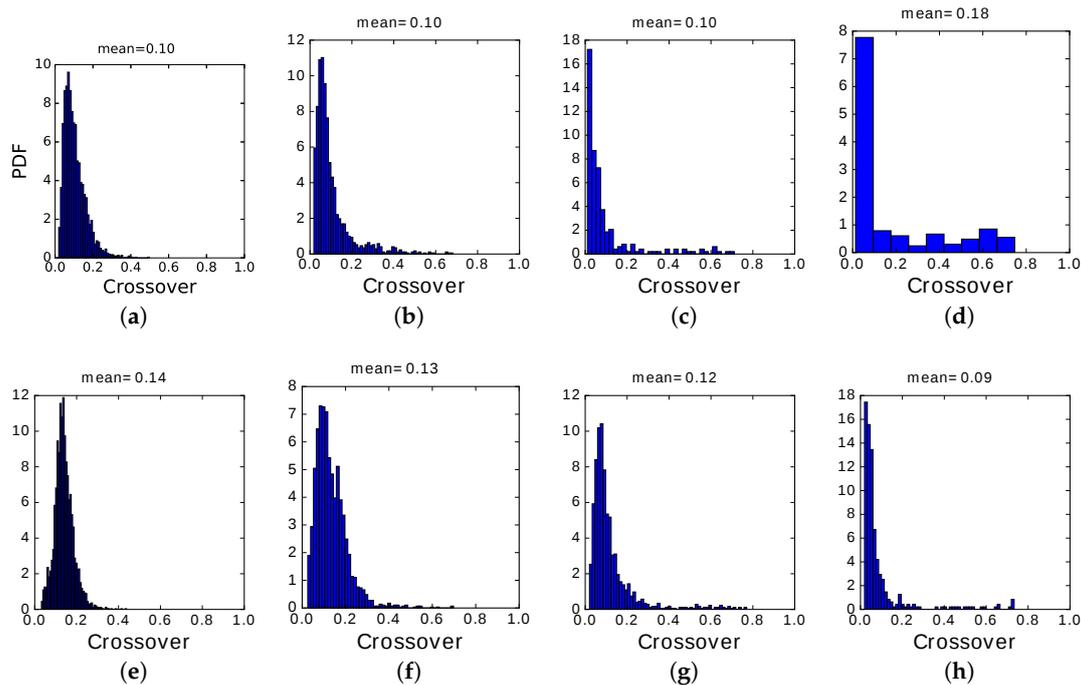


Figure 14. Probability density function (PDF) of crossover values in the fast adapting regime. (a) $p = 200$, $S = 4$; (b) $p = 200$, $S = 5$; (c) $p = 200$, $S = 6$; (d) $p = 200$, $S = 7$; (e) $p = 400$, $S = 6$; (f) $p = 400$, $S = 7$; (g) $p = 400$, $S = 8$; (h) $p = 400$, $S = 9$.

We see that, in the fast adapting regime, most transitions occur at very low crossover, i.e., the correlation with the preceding memory has to decay almost to zero before the next memory pattern can be activated. Only in regions of the (S, p) plane where latching sequences are very short, a few transitions only, we begin to see a small fraction of them with crossover values above 0.2. In most cases, the inhibitory feedback conveyed by the variable θ_i^0 is so fast as not to allow transitions to be carried through by positive correlations, i.e., by the subset of Potts units which are in the same active state in the preceding and successive pattern. The choice of the next pattern is not completely random, as indicated by the relative entropy values still below unity, but is determined essentially by negative selection, as mentioned above: the next pattern tends to have few active Potts units that coincide with those active in the preceding pattern.

In the slowly adapting regime, instead, due to the slow variation of the non-specific threshold, active Potts units can remain active, but they are *encouraged* by the variables θ_i^k to switch between active states if they have been in the same for too long. This can produce, particularly in the center of the latching band, sequences of patterns succeeding each other at high crossover, as shown by the distribution in Figure 13c. Even when latching is very noisy and approaches randomness, as in panels Figure 13a,e, crossover values are consistently above 0.2, indicating a preference for patterns insisting on the same set of active Potts units, unlike the fast adapting regime. Finally, when the number of states S is too large or, equivalently, that of patterns p too low, we observe some transitions with minimal crossover and a majority with very large crossover, as if occurring only with those patterns that were already partially retrieved when the network had still the largest overlap with the preceding pattern, but the main observation is that there are very few transitions at all, so that to plot a probability density distribution we need to use wide bins, in panels Figure 13d,h (and in Figure 14d).

This difference between the two regimes is confirmed by an analysis of the correlations between successive patterns in latching sequences. In the Potts network, at least two types of spatial correlation between patterns are relevant: how many active Potts units the two patterns share, and how many of these units are active and in the same state. We quantify them with C_1 , the fraction of the units active

in one pattern that are active also in the other, *and* in the same state; and with C_2 , the fraction that are also active, but in a different state. In a large set of randomly determined patterns, the mean values are $\langle C_1 \rangle = a/S$ and $\langle C_2 \rangle = a(S-1)/S$. The full distribution, among all pairs, is scattered around these mean values. However, do transitions occur between any pair of patterns?

Figure 15 shows that relative to the full distribution, in blue, transitions tend to occur, in the slowly adapting regime on the left, only between patterns with C_1 above and C_2 below (or at most around) their average values. Thus, when the network has retrieved a memory representation, it looks for correlated ones, as it were, where to jump. In the fast adapting regime, this is not the case: transitions are almost random, except there appears to be a slight tendency to avoid those with C_1 well above its mean value. Note that the values of p and w are different in the two panels, and are chosen so as to be in roughly equivalent positions within the respective latching bands.

The analysis of the crossover points, therefore, affords insight into the rather different transition dynamics prevailing in the fast and slowly adapting regimes, in particular in the center of their latching bands, suggesting that in a more realistic cortical model, which combines both types of activity regulation, there should still be a significant component of “slow adaptation” for interesting sequences of correlated patterns to emerge. The preceding simulations, however, were all carried out with randomly correlated patterns, in which the occasional high or low correlation of a pair is merely the result of a statistical fluctuation. Does the insight carry over to a more structured model of the correlations among memory patterns? This is what we ask next.

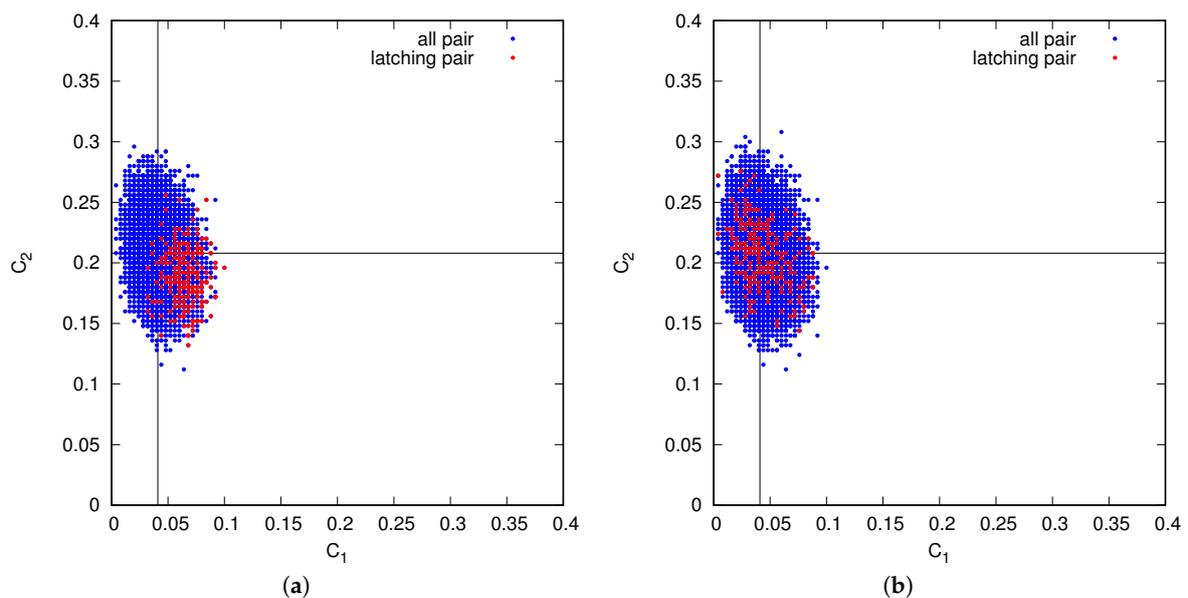


Figure 15. Scatterplots of the fractions C_1 and C_2 of Potts units active in one pattern that are active also in another, and in the same state or, respectively, in another active state. The panels show the full distribution between any pattern pair, in the slowly (a) and fast adapting (b) regimes, in blue; and the distribution between successive patterns in latching transitions, in red. The blue distribution for the fast adapting regime (for which $a = 0.25$, $S = 6$, $p = 300$ and $w = 1.32$) is similar to the one for the slowly adapting regime (for which again $a = 0.25$, $S = 6$, but $p = 200$ and $w = 0.65$), except that it is slightly wider, because of the higher storage load, while the red distributions are markedly different. Vertical lines indicate mean values: (a) slowly adapting regime; (b) fast adapting regime.

3.4. Analysis with Correlated Patterns

Correlated patterns were generated according to the algorithm mentioned by [22] and discussed in detail in [35]. The multi-parent pattern generation algorithm works in three stages. In the first step, a total set of Π random patterns are generated to act as parents. In the second step, each of the total set of parents are assigned to p_{par} randomly chosen children. Then, a “child” pattern is generated:

each pattern, receiving the influence of its parents with a probability a_p , aligns itself, unit by unit, in the direction of the largest field. In the third and final step, a fraction a of the units with the highest fields is set to become active. In this way, child patterns with a sparsity a are generated. In addition, another parameter ζ can be defined, according to which the field received by a child pattern is weighted with a factor $\exp(-\zeta k)$ where the index k runs through all parents. This is meant to express a non-homogeneous input from parents.

It is clear that such patterns, however, cannot be considered as independent and identically distributed, as in Equation (9), because their activity is drawn from a common pool of parents. In fact, they are correlated, in the sense that those children receiving congruent input from a larger number of common parents will tend to be more similar. All of these observations are studied in more detail in [35], and here we only focus on how correlations affect the phase diagrams. In the following simulations, the parameters pertaining to the patterns are $a_p = 0.4$, $\Pi = 100$, $\zeta = 0.1$ while p_{par}/p , the probability that a pattern be influenced by a parent is kept constant at 0.277.

Simulations with correlated patterns were carried out across the same S - p and C - p planes in phase space, in the slowly adapting regime, as shown in Figure 16. We focused on the slowly adapting regime based on the results of the crossover analysis. All other simulation parameters were kept at the values used with randomly correlated patterns.

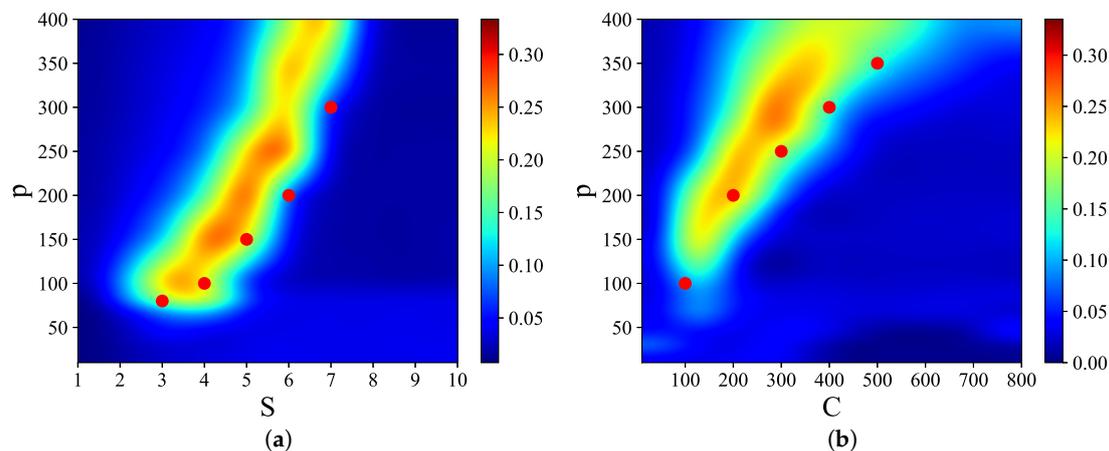


Figure 16. Phase space, cut across the $Q(S, p)$ plane in (a) and $Q(C, p)$ in (b), with correlated patterns in the slowly adapting regime. Red dots represent the quality peaks in the the same planes, with randomly correlated patterns. The parameters are $C = 150$ and $S = 5$, if kept fixed, and $w = 0.8$.

We see from the figure that the presence of non-random correlations among the memory patterns, albeit weak, shifts the bands to the left and upward in phase space, keeping approximately the dependence of the viable storage load p on S and C , but at somewhat higher values. It is as if more memories could “fit”, if correlated, into the same latching dynamics.

Figure 17 shows the S - p plane cut along $p = 200$, to better compare the cases with correlated (blue) and random (red) patterns. It is apparent that there is a leftward shift, in the case of correlated patterns, from the red curve applying to the random case, but the dependence on S remains very similar.

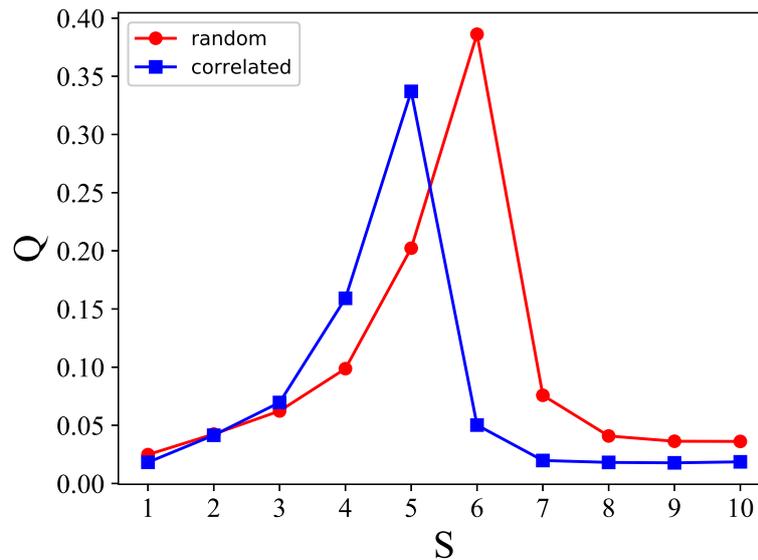


Figure 17. Comparison of S - p phase spaces along $p = 200$ with random (red dotted line) and correlated (blue dotted) patterns in the slow adapting regime.

4. Conclusions

In this paper, we have found the region in the Potts network phase space spanned by the number of Potts states S , the number of connections per unit C and the storage load p , where latching dynamics occur, and we have described their character, comparing and contrasting the slowly and fast adapting regimes. In relation to our earlier paper [22], where the possibility of such a latching region was pointed out on the basis of limited simulations, we have now a firmer basis to extrapolate to regions of parameter space of relevance to the human cortex, possibly a step toward quantitatively studying human specific capacities, including creative behaviour. A common hallmark in both regimes is that good quality latching occupies a band which scales almost quadratically in the p - S plane, while it is sublinear in the p - C plane. These bands are bounded by the storage capacity line, above, and by the boundary between no latching and finite latching, below. If, as discussed elsewhere [32], we were to take $C \approx 10^2$ and $S \approx 10^2$ as the orders of magnitude of interest for the human brain, we would conclude that the relevant storage load, or semantic depth, is in the region $p \approx 10^5$, in both regimes. At the center of the band in the slowly adapting regime, asymmetry and entropy take intermediate values, pointing at maximally complex and potentially useful dynamics, intermediate between the deterministic and the random extremes. High crossover values indicate that many transitions occur between highly correlated patterns. Using correlated patterns shifts the position of the band in phase space, but preserving the features observed with random patterns, still in the slowly adapting regime.

In the fast adapting regime, instead, in the center of the band, which can be made wider and more robust, the entropy is higher, and correspondingly only low crossover transitions are observed, indicating that the network latches most of the time from one pattern to any other among the many with which it is weakly or anti-correlated, avoiding only those few with which it is highly correlated.

Therefore, we can conclude that the fast adapting regime, modelling rapid inhibitory feedback, offers a robust framework for latching dynamics, but of an essentially random, not very useful nature; whereas in the slowly adapting regime, modelling slow inhibition or local fatigue, correlations can drive latching transitions, potentially enabling semantic content in a stream of thoughts or linguistic productions, but with fragile dynamics, living at the very edge between memory overload and sequence termination because of the inability of the network to jump forward. This suggests the opportunity of considering models that integrate both fast and slowly adapting dynamics in their

non-specific thresholds, so as to combine the useful features of both regimes. It will be the object of future work.

We would like to note, in the end, the inherent limitation of considering a simple homogeneous Potts network, with no differentiation among its units and no internal structure. In order to make contact with cognitive processes, of any kind, this limitation has to be overcome, as perhaps attempted, with one first step among many possible ones, by arranging Potts units on a ring [38]. Nevertheless, even in its crudest form, the Potts network with its latching dynamics can be used to explore e.g., novel theories as to the evolutionary origin of complex cognition [39]. It establishes a quantitative framework to understand phase transitions [25], complementary to the perspective offered by other modelling approaches to sequence generation in cortical networks [40]. At the most abstract level, it can be considered an implementation of a fuzzy logic system [41,42], but with the critical advantage that its parameters can eventually be related to cortical parameters, as we begin to describe in a related study [32].

Acknowledgments: We are grateful to Leonardo Romor who optimized the Potts network code, and to the Human Frontier Science Program RGP0057/2016 collaboration, that has also supported open access publication.

Author Contributions: All authors conceived and designed the simulations, which were performed primarily by Chol Jun Kang, with major contributions by Michelangelo Naim and Vezha Boboeva. Alessandro Treves and Chol Jun Kang wrote the paper, with input from the other two authors. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hauser, M.D.; Chomsky, N.; Fitch, W.T. The Faculty of language: What is it, who has it, and how did it evolve? *Science* **2002**, *298*, doi:10.1126/science.298.5598.1569.
2. Amit, D.J. The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behav. Brain. Sci.* **1995**, *18*, doi:10.1017/S0140525X00040164.
3. Kaneko, K.; Tsuda, I. Dynamic link of memory—Chaotic memory map in nonequilibrium neural networks. *Chaos* **2003**, *13*, doi:10.1016/S0893-6080(05)80029-2.
4. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558.
5. Amit, D.J.; Gutfreund, H.; Sompolinsky, H. Statistical mechanics of neural networks near saturation. *Ann. Phys.* **1987**, *173*, doi:10.1016/0003-4916(87)90092-3.
6. Amit, D.J. *Modeling Brain Function*; Cambridge University Press: New York, NY, USA, 1992.
7. Rolls, E.T.; Treves, A. *Neural Networks and Brain Function*; Oxford University Press: New York, NY, USA, 1998.
8. Pfeiffer, B.E.; Foster, D.J. Hippocampal place cell sequences depict future paths to remembered goals. *Nature* **2013**, *497*, 74–79.
9. Abeles, M. *Local Cortical Circuits: An Electrophysiological Study*; Springer Science & Business Media: Berlin, Germany, 2012.
10. Chossat, P.; Krupa, M.; Lavigne, F. Latching dynamics in neural networks with synaptic depression. *arXiv* **2016**, arXiv:1611.03645v2 .
11. Burgess, P.W.; Shallice, T. Confabulation and the control of recollection. *Memory* **1996**, *4*, 359–411.
12. Epstein, R. The neural-cognitive basis of the Jamesian stream of thought. *Conscious. Cognit.* **2000**, *9*, 550–575.
13. Abeles, M. Time is precious. *Science* **2004**, *304*, 523–524.
14. Pulvermüller, F. Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* **2005**, *6*, 576–582.
15. Shmiel, T.; Drori, R.; Shmiel, O.; Ben-Shaul, Y.; Nadasdy, Z.; Shemesh, M.; Teicher, M.; Abeles, M. Temporally precise cortical firing patterns are associated with distinct action segments. *J. Neurophysiol.* **2006**, *96*, 2645–2652.
16. Sosnik, R.; Shemesh, M.; Abeles, M. The point of no return in planar hand movements: An indication of the existence of high level motion primitives. *Cognit. Neurodyn.* **2007**, *1*, 341–358.
17. Kanter, I. Potts-glass models of neural networks. *Phys. Rev. A* **1988**, *37*, 2739–2742.

18. Bollé, D.; Dupont, P.; Mourik, J.V. Stability properties of potts neural networks with biased patterns and low loading. *J. Phys. A Math. Gen.* **1991**, *24*, doi:10.1088/0305-4470/24/5/021.
19. Bollé, D.; Dupont, P.; Huyghebaert, J. Thermodynamic properties of the Q-state potts-glass neural network. *Phys. Rev. A* **1992**, *45*, doi:10.1103/PhysRevA.45.4194.
20. Bollé, D.; Inck, B.; Zagrebnoy, V.A. On the parallel dynamics of the Q-state potts and Q-ising neural networks. *J. Stat. Phys.* **1993**, *70*, 1099–1119.
21. Bollé, D.; Cools, R.; Dupont, P.; Huyghebaert, J. Mean-field theory for the Q-state potts-glass neural network with biased patterns. *J. Phys. A Math. Gen.* **1993**, *26*, doi:10.1088/0305-4470/26/3/017.
22. Treves, A. Frontal latching networks: A possible neural basis for infinite recursion. *Cogn. Neuropsychol.* **2005**, *22*, 276–291.
23. Kropff, E.; Treves, A. The storage capacity of potts models for semantic memory retrieval. *J. Stat. Mech. Theor. Exp.* **2005**, doi:10.1088/1742-5468/2005/08/P08010.
24. Russo, E.; Namboodiri, V.M.K.; Treves, A.; Kropff, E. Free association transitions in models of cortical latching dynamics. *New J. Phys.* **2008**, *10*, doi:10.1088/1367-2630/10/1/015008.
25. Russo, E.; Treves, A. Cortical free-association dynamics: Distinct phases of a latching network. *Phys. Rev. E* **2012**, *85*, doi:10.1103/PhysRevE.85.051920.
26. Abdollah-nia, M.F.; Saeedghalati, M.; Abbassian, A. Optimal region of latching activity in an adaptive Potts model for networks of neurons. *J. Stat. Mech. Theory Exp.* **2012**, doi:10.1088/1742-5468/2012/02/P02018.
27. Lerner, I.; Bentin, S.; Shriki, O. Spreading activation in an attractor network with latching dynamics: Automatic semantic priming revisited. *Cogn. Sci.* **2012**, *36*, 1339–1382.
28. Lerner, I.; Bentin, S.; Shriki, O. Excessive attractor instability accounts for semantic priming in schizophrenia. *PLoS ONE* **2012**, *7*, doi:10.1371/journal.pone.0040663.
29. O’Kane, D.; Treves, A. Short-and long-range connections in autoassociative memory. *J. Phys. A Math. Gen.* **1992**, *25*, doi:10.1088/0305-4470/25/19/018.
30. Romani, S.; Pinkoviezky, I.; Rubin, A.; Tsodyks, M. Scaling laws of associative memory retrieval. *Neural Comput.* **2013**, *25*, 2523–2544.
31. Recanatesi, S.; Katkov, M.; Romani, S.; Tsodyks, M. Neural network model of memory retrieval. *Front. Comput. Neurosci.* **2015**, *9*, doi:10.3389/fncom.2015.00149.
32. Naim, M.; Boboeva, V.; Kang, C.J.; Treves, A. From multi-modular Hopfield networks to the Potts network and its storage capacity. Unpublished work, 2017.
33. Kropff, E.; Treves, A. The complexity of latching transitions in large scale cortical networks. *Nat. Comput.* **2007**, *6*, 169–185.
34. Hebb, D.O. *The Organization of Behavior: A Neuropsychological Theory*; John Wiley & Sons: New York, NY, USA, 2005.
35. Boboeva, V.; Treves, A. The storage capacity of the Potts network with correlated patterns. Unpublished work, 2017.
36. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2006.
37. Russo, E.; Pirmoradian, S.; Treves, A. Associative latching dynamics vs. syntax. In *Advances in Cognitive Neurodynamics (II)*; Wang, R., Gu, F., Eds.; Springer: Dordrecht, The Netherlands, 2011; pp. 111–115.
38. Song, S.; Yao, H.; Treves, A. A modular latching chain. *Cogn. Neurodyn.* **2014**, *8*, 37–46.
39. Amati, D.; Shallice, T. On the emergence of modern humans. *Cognition* **2007**, *103*, 358–385.
40. Rajan, K.; Harvey, C.D.; Tank, D.W. Recurrent network models of sequence generation and memory. *Neuron* **2016**, *9*, 128–142.
41. Jiang, Y.; Chung, F.L.; Deng, Z.; Wang, S. Multitask TSK fuzzy system modeling by mining intertask common hidden structure. *IEEE Trans. Cybern.* **2015**, *45*, 548–561.
42. Tu, C.C.; Juang, C.F. Recurrent type-2 fuzzy neural network using Haar wavelet energy and entropy features for speech detection in noisy environments. *Expert Syst. Appl.* **2012**, *39*, 2479–2488.

