SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
COMPLEXITY
DYNAMICAL SYSTEMS
DYNAMIC NETWORKS
NETWORK MODELS

Received
6 March 2013

Accepted
30 May 2013

Published
20 June 2013

Correspondence and requests for materials should be addressed to C.G. (gros07@itp.uni-frankfurt.de)

Generating functionals for autonomous latching dynamics in attractor relict networks

Mathias Linkerhand & Claudius Gros

Institute for Theoretical Physics, Goethe University Frankfurt, Germany.

Coupling local, slowly adapting variables to an attractor network allows to destabilize all attractors, turning them into attractor ruins. The resulting attractor relict network may show ongoing autonomous latching dynamics. We propose to use two generating functionals for the construction of attractor relict networks, a Hopfield energy functional generating a neural attractor network and a functional based on information-theoretical principles, encoding the information content of the neural firing statistics, which induces latching transition from one transiently stable attractor ruin to the next. We investigate the influence of stress, in terms of conflicting optimization targets, on the resulting dynamics. Objective function stress is absent when the target level for the mean of neural activities is identical for the two generating functionals and the resulting latching dynamics is then found to be regular. Objective function stress is present when the respective target activity levels differ, inducing intermittent bursting latching dynamics.

The use of objective functions for the formulation of complex systems has seen a study surge of interest. Objective functions, in particular objective functions based on information theoretical principles[1–7], are used increasingly as generating functionals for the construction of complex dynamical and cognitive systems. There is then no need to formulate by hand equations of motion, just as it is possible, in analogy, to generate in classical mechanics Newton's equation of motion from an appropriate Lagrange function. When studying dynamical systems generated from objective functions encoding general principles, one may expect to obtain a deeper understanding of the resulting behavior. The kind of generating functional employed also serves, in addition, to characterize the class of dynamical systems for which the results obtained may be generically valid.

Here we study the interplay between two generating functionals. The first generating functional is a simple energy functional. Minimizing this objective function one generates a neural network with predefined point attractors, the Hopfield net[8,9]. The second generating functional describes the information content of the individual neural firing rates. Minimizing this functional results in maximizing the information entropy[6,7] and in the generation of adaption rules for the intrinsic neural parameters, the threshold and the gain. This principle has been denoted polyhomeostatic optimization[7,10], as it involves the optimization of an entire function, the distribution function of the time-averaged neural activities.

We show that polyhomeostatic optimization destabilizes all attractors of the Hopfield net, turning them into attractor ruins. The resulting dynamical network is an attractor relict network and the dynamics involves sequences of continuously latching transient states. This dynamical state is characterized by trajectories slowing down close to the succession of attractor ruins visited consecutively. The two generating functionals can have incompatible objectives. Each generating functional, on its own, would lead to dynamical states with certain average levels of activity. Stress is induced when these two target mean activity levels differ. We find that the system responds to objective function stress by resorting to intermittent bursting, with laminar flow interseeded by burst of transient state latching.

Dynamical systems functionally equivalent to attractor relict networks have been used widely to formulate dynamics involving on-going sequences of transient states. Latching dynamics has been studied in the context of the grammar generation with infinite recursion[11,12] and in the context of reliable sequence generation[13,14]. Transient state latching has also been observed in the brain[15,16] and may constitute an important component of the internal brain dynamics. This internal brain dynamics is autonomous, ongoing being modulated, but not driven, by the sensory input[17,18]. In this context an attractor relict network has been used to model autonomous

neural dynamics in terms of sequences of alternating neural firing patterns[19]. The modulation of this type of internal latching dynamics by a stream of sensory inputs results, via unsupervised learning, in a mapping of objects present in the sensory input stream to the pre-existing attractor ruins of the attractor relict network[20,21], the associative latching dynamics thus acquiring semantic content.

Results

As an introductory example we consider one of the simplest possible networks having two distinct attractors in terms of minima of the energy functional, as defined by Eq. (7). The dynamics of the 3-site network illustrated in Fig. 1 is given by

$$\begin{aligned}\dot{x}_1 &= -\Gamma x_1 + w^+ y_2 - w^- y_3 \\ \dot{x}_2 &= -\Gamma x_2 + w^+ (y_1 + y_3) \quad , \\ \dot{x}_3 &= -\Gamma x_3 + w^+ y_2 - w^- y_1\end{aligned} \tag{1}$$

with $w^+ > 0$ and $w^- > 0$ denoting the excitatory and inhibitory link strength respectively. Here the x_i/y_i correspond to neural membrane potentials and firing rates, respectively, related through a sigmoidal transfer function, see Eq. (5), which is parameterized by the gain a_i and the threshold b_i. For fixed intrinsic parameters $a \equiv a_i$ and $b \equiv b_i$, viz for vanishing adaption rates ϵ_a, $\epsilon_b = 0$, the network has two possible phases, as shown in Fig. 1. There is either a single global attractor, with activities y_i determined mostly by the value of the threshold b, or two stable attractors, with either y_1 or y_3 being small and the other two firing rates large.

The point attractors present for vanishing intrinsic adaption become unstable for finite adaption rates ϵ_a, $\epsilon_b > 0$. Point attractors, defined by $\dot{x}_i = 0 = \dot{y}_i$, have time-independent neural activities. The objective (9), to attain a minimal Kullback-Leibler divergence, can however not be achieved for constant firing rates y_i. The polyhomeostatic adaption (11) hence forces the system to become autonomously active, as seen in Fig. 2.

In Fig. 2 we present the results for the overlaps O_p and A_p, as defined by Eqs. (13) and (14), of the neural activity with the stored memories. One observes, that the original attractors become unstable in the presence of finite polyhomeostatic adaption, retaining however a prominent role in phase space. Unstable attractors, transiently attracting the phase space flow, can be considered to act as 'attractor relicts'. The resulting dynamical system is hence an attractor relict network, viz a network of coupled attractor ruins[19,20]. When coupling an attractor network to slow intrinsic variables, here the threshold b and the gain a, the attractors are destroyed but retain presence in the flow in terms of attractor ruins, with the dynamics slowing down close to the attractor relict.

The concept of an attractor ruin is quite general and attractor relict networks have implicitly been generated in the past using a range of functionally equivalent schemes. It is possible to introduce additional local slow variables, like a reservoir, which are slowly depleted when a unit is active[19,20]. Similarly, slowly accumulating spin variables have been coupled to local dynamic thresholds in order to destabilize attractors[13], and local, slowly adapting, zero state fields have been considered in the context of latching Potts attractor networks[11,12,22,23]. Alternatively, attractors have been destabilized in neural networks by introducing slowly adapting asymmetric components to the inter-neural synaptic weights[14].

Phase boundary adaption. In Fig. 3 we have superimposed, for the three-site network, a typical set of trajectories, with finite adaption rates $\epsilon_a = 0.1$ and $\epsilon_b = 0.01$, onto the phase diagram evaluated for non-adapting neurons, with $\epsilon_a = 0 = \epsilon_b$, as shown in Fig. 1. The three trajectories (green: (a_2, b_2), red and blue: (a_1, b_1) and (a_3, b_3)) all start well in the region with a single global fixed point, at $a_i = 5$, $b_i = -0.5$ ($i = 1, 2, 3$). Two features of the flow are eye-catching.

- The intrinsic parameters of the active neurons settle, after a transient initial period, at the phase boundary and oscillate alternating between the phase with a single stable fixed point and the phase with two attractors.
- The trajectory (a_2, b_2) for the central neuron adapts to a large value for the threshold, taking care not to cross any phase boundary during the initial transient trajectory.

We first discuss the physics behind the second phenomenon. The Euclidean distance in phase space between the two attractors, when existing, can be used as an order parameter. In Fig. 4 the order parameter is given for vertical cuts through the phase diagram. The lower transition is evidently of second order, which can be confirmed by a stability analysis of (1), with the upper transition being of first order, in agreement with a graphical analysis of (1). A first order transition will not be crossed, in general, by an adaptive process, as there are no gradients for the adaption to follow. This is the reason

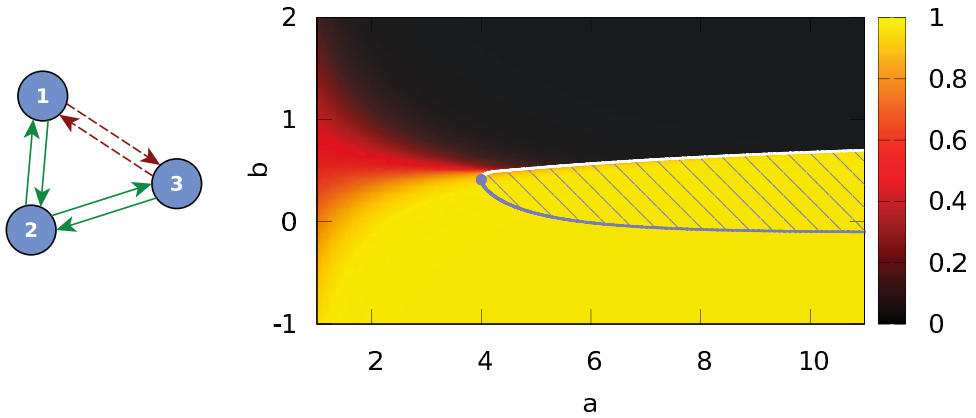

Figure 1 | Left: A three-site graph with symmetric excitatory (solid green lines) and inhibitory connections (red dashed lines). Right: The corresponding phase diagram, for fixed $\Gamma = 1$ and $w^+ = 1 = w^-$, as function of the gain $a \equiv a_i$ and threshold $b \equiv b_i$, for $i = 1, 2, 3$. For large/small thresholds the sites tend to be inactive/active. The color encodes the activity of the most active neuron. Inside the shaded area there are two stable attractors, outside a single, globally attracting state. The binary representations of the two non-trivial attractors are (1,1,0) and (0,1,1), where 1/0 denotes an active/inactive site. The lower and upper lines (blue/white) of the phase transition are of second and first order respectively, with the lower phase transition line determined by $1 = a\tilde{y}(1-\tilde{y})$ where $\tilde{y} \equiv y_1 = y_3$ is the activity of the single fixed point at site one and three. The two critical lines meet at $(a_c, b_c) = (4, 0.413)$ (blue dot).

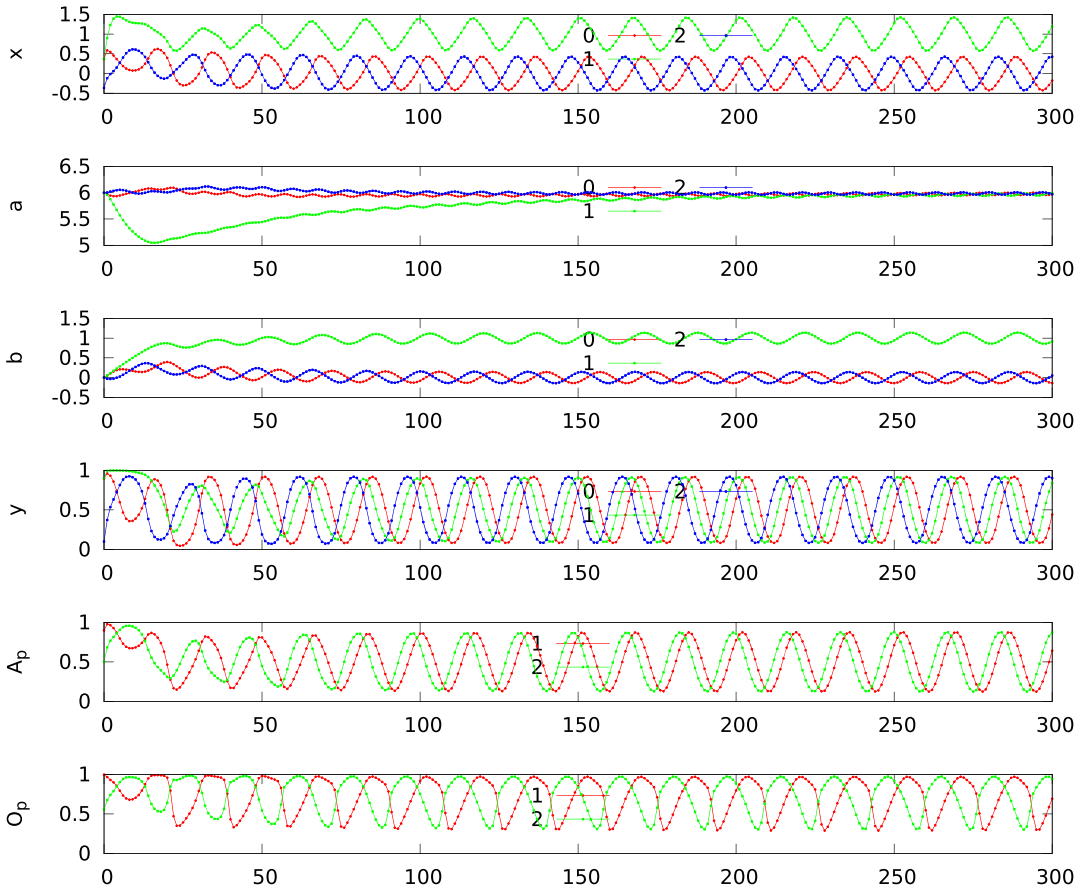

Figure 2 | Time series of the three-site network shown in Fig. 1, for $\Gamma = 1$, $\epsilon_a = 0.1$, $\epsilon_b = 0.01$, $\lambda_1 = 0$, $\lambda_2 = 0$ and $w^{\pm} = 1$. The dynamics is given by (6) and (11). The dynamics retraces periodically the original attractor states, as one can see from the oscillation of the overlap O_p and the A_p, compare Eqs. (13) and (14), between the patterns of neural and attractor activities.

that the trajectory for (a_2, b_2) shown in Fig. 3 needs to take a large detour in order to arrive to its target spot in phase space. This behavior is independent of the choice of initial conditions.

Using an elementary stability analysis one can show that the condition for the second-order line is $1 = a\tilde{y}(1-\tilde{y})$, where $\tilde{y} \equiv y_1 = y_3$ (for the single global fixed point). Only a single global attractor is consequently stable for $a \leq 4$ (since $\tilde{y} \in [0, 1]$ and $\tilde{y}(1-\tilde{y}) \leq 0.25$) and the second and the first order lines meet at $(a_c, b_c) = (4, 0.413)$, where the critical threshold b_c is determined by the self-consistent solution of $b_c = 1/[1 + \exp(4b_c - 4)] - 1/2$. The trajectories (a_1, b_1) and

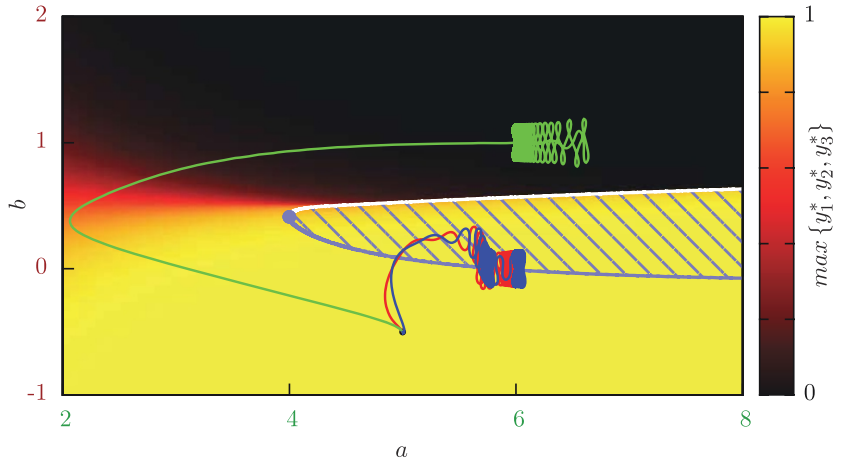

Figure 3 | The trajectories of $(a_i(t), b_i(t))$, for finite adaption rates, $\epsilon_a = 0.1$ and $\epsilon_b = 0.01$, superimposed onto the phase diagram of the three-site network shown in Fig. 1. The color encodes the activity of the most active site. Inside the shaded blue area two attractors are stable, outside there is only a single global attractor. The parameters are the same as for the simulation presented in Fig. 2. The thresholds $b_1(t)$ and $b_3(t)$ (red/blue trajectories) oscillate across the second-order phase boundary (blue line), the threshold $b_2(t)$ (green trajectory) acquires a large value, receiving two excitatory inputs, and avoids the first-order phase boundary (white line) during the initial transient. All three trajectories start in the center, lower-half of the phase diagram, at $a = 5$, $b = -0.5$ (black dot). Note that the underlying phase diagram is for identical thresholds b_i and gains a_i.

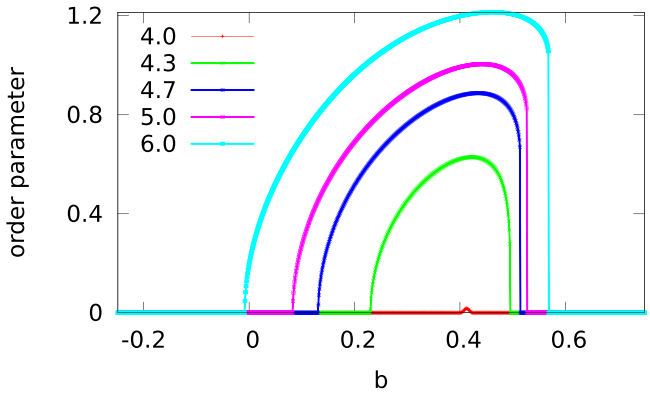

Figure 4 | Vertical cuts (for constant gains $a = 4, 0, 4.3, 4.7, 5.0, 6.0$, bottom to top curves) through the phase diagram of the 3-site graph shown in Fig. 1. Plotted is the order parameter $\sqrt{\sum_{i=1}^{3}\left(y_i^{(1)}-y_i^{(2)}\right)^2}$, where $y^{(\gamma)}=\left(y_1^{(\gamma)},y_2^{(\gamma)},y_3^{(\gamma)}\right)$ are the two fixed point solutions ($\gamma = 1, 2$) in the coexistent region. The lower/upper transitions are second/first order.

(a_3, b_3) are hence oscillating across the locus in phase space where there would be a second-order phase transition, compare Fig. 3, for the case of identical internal parameters a_i and b_i. With adaption, however, the respective internal parameters take distinct values, $(a, b \pm \delta b)$ for the first/third neuron and (a, b_2) for the second neuron, with $a \approx 6$, $b \approx 0$ and $b_2 \approx 1$. In fact the system adapts the intrinsic parameters dynamical in a way that a non-stopping sequence of states

$$\ldots \to \xi^1 \to \xi^0 \to \xi^2 \to \xi^0 \to \xi^1 \to \ldots \tag{2}$$

is visited, where $\xi^1 = (1, 1, 0)$, $\xi^2 = (0, 1, 1)$ denote the two stable cliques and $\xi^0 = (1, 1, 1)$ the fixed point in the phase having only a single global attractor. It is hence not a coincidence that the system adapts autonomously, as shown in Fig. 3, to a region in phase space where there would be a second order phase transition for identical internal parameters. At this point, small adaptive variations make the sequence (2) achievable. The adaption process hence shares phenomenological some aspects with what one calls 'self-organized quasi-criticality' in the context of absorbing phase transitions[27,28], which denotes the circumstance that a system without energy conservation may become critical only when tuning an external recharging rate. For imperfect tuning the trajectories would linger close to second-order phase transition point, without actually reaching it, a phenomenology close to one seen in Fig. 3.

Objective function stress. We will now investigate networks of larger size N. In principle we could generate random synaptic link matrices $\{w_{ij}\}$ and find their eigenstates $\xi^p = (\xi_i^p, \ldots, \xi_N^p)$ using a diagonalization routine. When using the Hopfield encoding (15) for the synaptic weights w_{ij} the attractors are known, corresponding, as long as the number N_p of patterns is not too large[8,9], to the to the stored patterns ξ^p. We here make use of the Hopfield encoding for convenience, no claim is made that memories in the brain are actually stored and defined by (15).

We consider in the following random binary patterns, as illustrated in Fig. 5,

$$\xi_i^p = \begin{cases} 1 & \text{with probability} \quad \alpha \\ 0 & \text{with probability} \quad 1-\alpha \end{cases} \tag{3}$$

where α is the mean activity level or sparseness. The patterns ξ^p have in general a finite, albeit small, overlap, as illustrated in Fig. 5. The target distribution $q(y)$ for the intrinsic adaption has, for $\lambda_2 = 0$, an expected mean

$$\mu = \int_0^1 \mathrm{d}y\; y\; q(y) = 1 - \frac{1}{\lambda_1} + \frac{1}{e^{\lambda_1}-1}, \tag{4}$$

which can be evaluated noting that the support of $q(y)$ is [0, 1]. The target mean activity μ can now differ from the average activity of an attractor relict, the mean pattern activity α, compare (3). The difference between the two quantities, viz between the two objectives, energy minimization vs. polyhomeostatic optimization, induces stress into the latching dynamics.

In Fig. 6 we present the time evolution of the overlaps A_p and O_p for $\alpha = 0.3 = \mu$ and adaption rates $\epsilon_a = 0.1$, $\epsilon_b = 0.01$. In this case the target mean activity, μ, of the intrinsic adaption rules (11) is consistent with the mean activity α of the stored patterns ξ^p, viz with the mean activity level of the attractor relicts. One observes that the system settles into a limiting cycle with all seven stored patterns becoming successively transiently active, a near-to-perfect latching dynamics. The dynamics is very stable and independent of initial conditions, which were selected randomly.

In Fig. 7 we present the time evolution, for 20 out of the $N = 100$ neurons, of the respective individual variables. The simulation parameters are identical for Figs. 6 and 7. Shown in Fig. 7 are the individual membrane potentials $x_i(t)$, the firing rates $y_i(t)$, the gains $a_i(t)$ and the thresholds $b_i(t)$. The latching activation of the attractor relicts seen in Fig. 6 reflects in corresponding transient activations of the respective membrane potentials and firing rates. The oscillations in the thresholds $b_i(t)$ drive the latching dynamics, interestingly, even though the adaption rate is larger for the gain.

The synaptic weights w_{ij} are symmetric and consequently also the overlap matrix presented in Fig. 5. The latching transitions evident in Fig. 6 are hence spontaneous in the sense that they are not induced by asymmetries in the weight matrix[14]. We have selected uncorrelated patterns ξ^p and the chronological order of the transient states is hence determined by small stochastic differences in the pattern overlaps. It would however be possible to consider correlated activity patters ξ^p incorporating a rudimental grammatical structure[22,23], which is however beyond the scope of the present study.

Stress-induced intermittency. In Fig. 8 we present the time evolution of the overlaps A_p and O_p for the case when the two objective

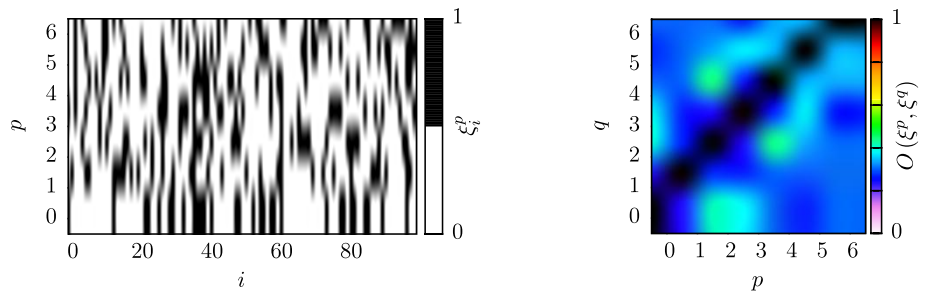

Figure 5 | Left: Example of $N_P = 7$ random binary patterns ξ^p, compare (3), with sparseness $\alpha = 0.3$ for a network of $N = 100$ neurons. Right: Respective pairwise mutual pattern overlaps $O(\xi^p, \xi^q)$. Patterns maximally overlap with themselves, the inter-pattern overlap is random and small.

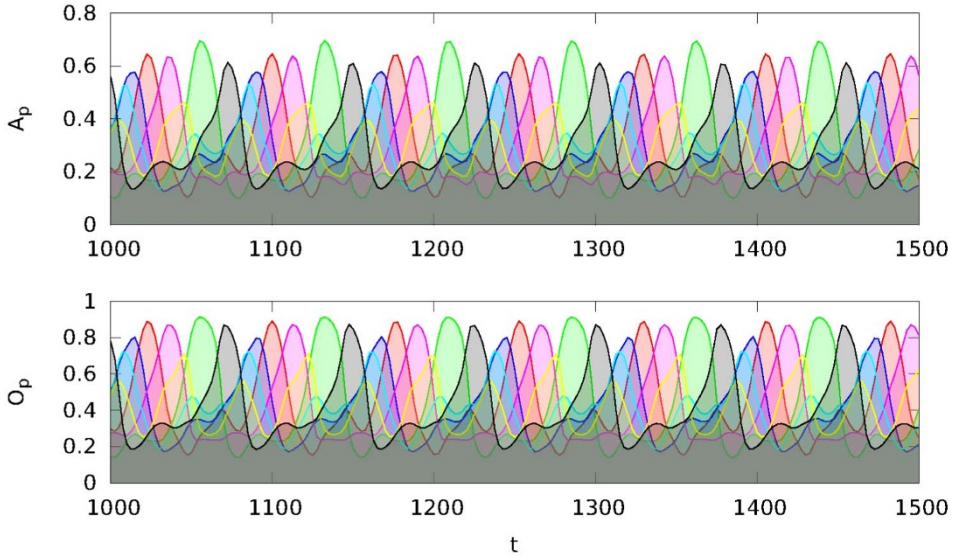

Figure 6 | Overlaps O_p and A_p (color encoded, see Eqs. (13) and (14)), for a $N = 100$ network, of the neural activities with the $N_p = 7$ attractor ruins, as a function of time t. The sparseness of the binary patterns defining the attractor ruins is $\alpha = 0.3$. Here the average activity α of the attractor relicts and the target mean activity level $\mu = 0.3$ have been selected to be equal, resulting in a limiting cycle with clean latching dynamics.

functions, the energy functional and the polyhomeostatic optimization, incorporate conflicting targets. We retain the average sparseness $\alpha = 0.3$ for the stored patterns, as for Fig. 6, but reduced the target mean firing rate to $\mu = 0.15$. This discrepancy between α and μ induces stress into the dynamics. The pure latching dynamics, as previously observed in Fig. 6, corresponds to a mean activity of about 0.3, in conflict with the target value $\mu = 0.15$.

Phases of laminar flow are induced by the objective function stress, and the latching dynamics occurs now in the form of intermittent bursts. The neural activity, see A_p in Fig. 8, is substantially reduced during the laminar flow and the time averaged mean firing rate such reduced towards the target activity level of $\mu = 0.15$. The trajectory does not come close to any particular attractor relict during the laminar flow, the overall O_p being close to 0.5 for all $N_p = 7$ stored pattern.

The time evolution is hence segmented into two distinct phases, a slow laminar phase far from any attractor and intermittent phases of bursting activity in which the trajectories linger transiently close to attractor relicts with comparatively fast (relative to the time scale of the laminar phase) latching transitions. Intermittent bursting

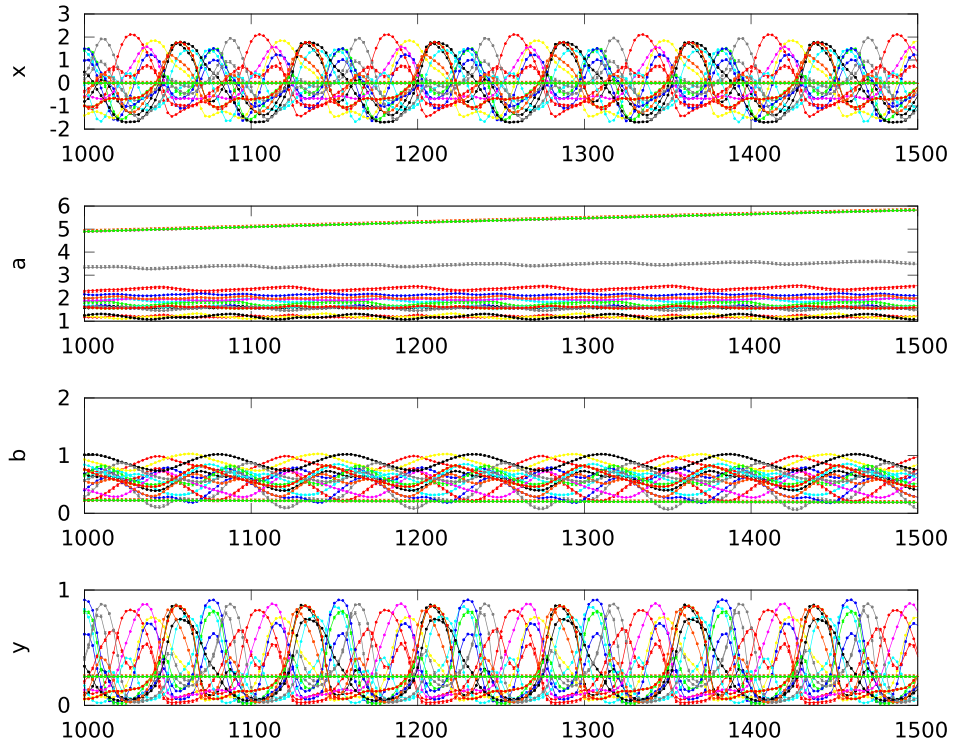

Figure 7 | Time evolution, for a selection of 20 out of $N = 100$ neurons, of the membrane potential x_i, gain a_i, threshold b_i and firing rate y_i for the time series of overlaps presented in Fig. 6.

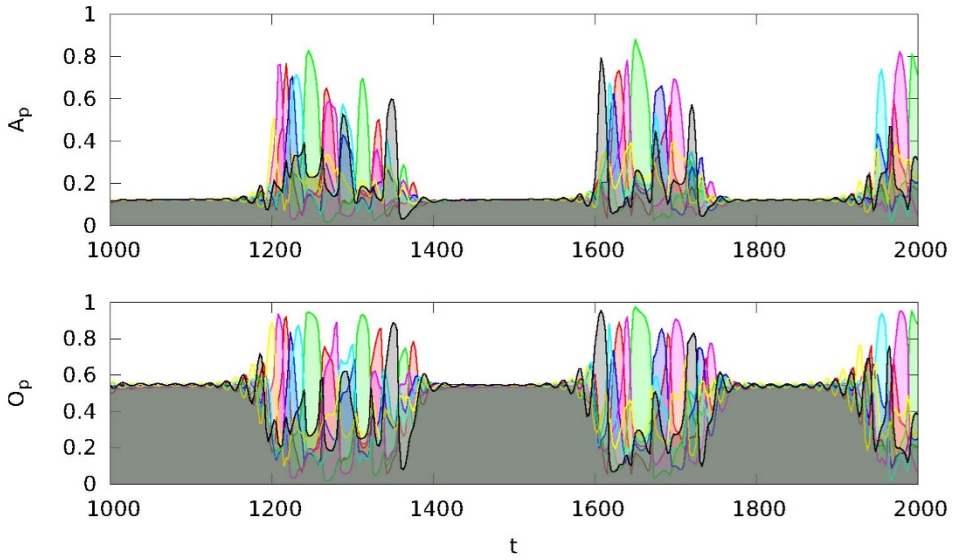

Figure 8 | Time evolution of overlaps O_p and A_p (color coded, compare Eqs. (13) and (14)) for all $N_P = 7$ binary patterns, with sparseness $\alpha = 0.3$, for a network of $N = 100$ neurons. The target mean neural activity is $\mu = 0.15$, the difference between the two objective functions, viz between μ and α induces stress. The intermittent latching dynamics has a mean activity of about 0.3, which is too large. The phases of laminar flows between the burst of latching has a reduced average activity, thus reducing the time-averaged mean activity level toward the target $\mu = 0.15$.

dynamics has been observed previously in polyhomeostatically adapting neural networks with randomly selected synaptic weights[7,10], the underlying causes had however not been clear. The results presented above show that the concept of competing generating functionals can be used to define objective function stress and that a given difference in target mean activities is the underlying force driving the system into the intermittent bursting regime.

Robustness of transient state dynamics. The latching dynamics presented in Figs. 6 and 8 is robust with respect to system size N. We did run the simulation for different sizes of networks with up to $N = 10^5$ neurons, a series of sparseness parameters α and number of stored patterns N_P. As an example we present in Fig. 9 the overlap O_p for $N = 1000$ neurons and $N_P = 20$ binary patterns with a sparseness of $\alpha = 0.2$. No stress is present, the target mean activity level is $\mu = 0.2$. The latching dynamics is regular, no intermittent bursting is observed. There is no constraint, generically, to force the limiting cycle to incorporate all attractor relicts. Indeed, for the transient state dynamics presented in Fig. 9, some of the stored patterns are never activated.

The data presented in Figs. 6, 8 and 9 is for small numbers of attractor relicts N_p, relative to the systems size N, a systematic study for large values of N_p is beyond the scope of the present study. Latching dynamics tends to break down, generically speaking, when the overlap between distinct attractors, as shown in Fig. 5, becomes large. The autonomous dynamics then becomes irregular.

Discussion

The use of generation functionals has a long tradition in physics in general and in classical mechanics in particular. Here we point out that using several multivariate objective functions may lead to novel dynamical behaviors and an improved understanding of complex systems in general. We propose in particular to employ generating functionals which are multivariate in the sense that they are used to derive the equations of motion for distinct, non-overlapping subsets of dynamical variables. In the present work we have studied a neural network with fast primary variables $x_i(t)$, the membrane potentials, and slow secondary variables $a_i(t)$ and $b_i(t)$, characterizing the internal behavior of individual neurons, here the gains a_i and the thresholds b_i. The time evolution of these sets of interdependent variables is determined respectively by two generating functionals.

The equations of motion (leaky integrator) for the primary dynamical variables $\dot{x}_i(t)$, the individual membrane potentials. are generated minimizing an *energy functional*. The Kullback-Leibler divergence is, on the other side, an example of an *information-theoretical fucntional*. Minimizing the Kullback-Leibler divergence between the distribution $p_i(y)$ of the time-average neural firing rate y_i and a target distribution function $q(y)$ maximizing information entropy generates intrinsic adaption rules $\dot{a}$ and $\dot{b}$ for the gain a and the threshold b (polyhomeostatic optimization).

Generating functionals may incorporate certain targets or constraints, either explicitly or implicitly. We denote the interplay between distinct objectives incorporated by competing generating functionals *"objective functions stress"*. For the two generating functionals considered in this study there are two types of objective functions stress.

The minima of the energy functional are time-independent point attractors leading to firing-rate distributions $p_i(y)$ which are sharply peaked. The target firing-rated distribution $q(y)$ for the information-theoretical functional is however smooth (polyhomeostasis). This *functional stress* leads to the formation of an attractor relict network. The mean target neural firing rate μ is, on the other side, a scalar parameter for the target firing-rate distribution $q(y)$, and hence encoded explicitly within the information theoretical functional. The local minima of the energy functional, determined by the synaptic weights w_{ij}, determine implicitly the mean activity levels α of the corresponding point attractors. *Scalar objective function stress* is present for $\alpha \neq \mu$.

For the two generating functionals considered, we find that the scalar objective function stress induces a novel dynamical state, characterized by periods of slow laminar flow interseeded by bursts of rapid latching transitions. We propose that objective function stress is a powerful tool, in general, for controlling the behavior of complex dynamical systems. The interplay between distinct objective functions may hence serve as a mechanism for guiding self organization[30–32], a process also denoted targeted self organization[25].

The main focus of this study has been to to investigate and to quantify, for larger size networks, the competition between an objective function based on energy minimization and a generating functional

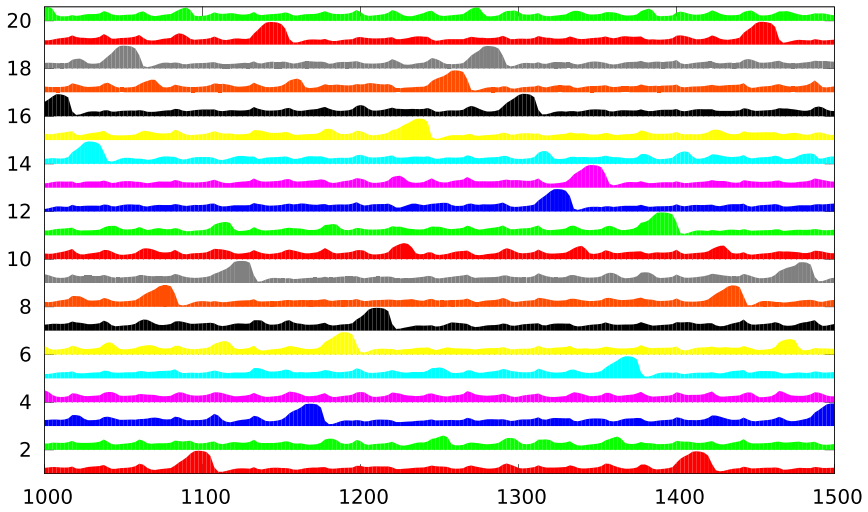

Figure 9 | Time evolution, for $N = 1000$, of the overlaps O_p for all $N_P = 20$ binary patterns (vertically displaced) with sparseness $\alpha = 0.2$ and a target mean activity of $\mu = 0.2$. There is no objective-function stress and the latching is regular, the adaption rates are $\epsilon_a = 0.1$, $\epsilon_b = 0.01$.

based on the principle of homeostasis. Within this investigation we also detailed out the phase diagram of a model 3-site network having either one or two transiently stable attractors. This application to a simple model network allowed to investigate the adaptive flow within the underlying phase diagram, as generated in the adiabatic limit by the energy functional, compare Fig. 3. The results indicate that poly-homeostatic adaption has the tendency to drive the system toward regions of phase space close to a second order phase transition, as this offers to generate the desired variability in neural activity. We plan to investigate further this interesting phenomenon in future studies.

Methods

We use N rate encoding neurons in continuous time, with a non-linear transfer function,

$$y_i = \frac{1}{1 + e^{a_i(b_i - x_i)}}, \tag{5}$$

where $x_i(t) \in \mathcal{R}$ is the membrane potential and $y_i(t) \in [0, 1]$ the firing rate of the ith neuron. The dynamics of the neural activity is

$$\dot{x}_i = -\Gamma x_i + \sum_{j=1}^{N} w_{ij} y_j, \tag{6}$$

which describes a leaky integrator[24], with Γ being the leak rate and the w_{ij} the inter-neural synaptic weights. This kind of dynamics can be derived minimizing a simple energy functional

$$\frac{\Gamma}{2} \sum_i x_i^2 - \frac{1}{2} \sum_{ij} y_i w_{ij} y_j \tag{7}$$

with respect to the membrane potential x_i[8,9]. When using an energy functional to derive (6), the resulting synaptic weights are necessarily symmetrized, viz $w_{ij} = w_{ji}$. Here we do indeed consider only symmetric synaptic links, the dynamics of a network of leaky integrators would however remain well behaved when generalizing to non-symmetric weight matrices.

The transfer function (5) contains two intrinsic parameters, a and b, which can be either set by hand or autonomously adapted, an approach considered here. The intrinsic parameters determine the shape of the individual firing-rated distributions $p_i(y)$, given generically by

$$p_i(y) = \frac{1}{T} \int_0^T \delta(y - y_i(t - \tau)) d\tau, \tag{8}$$

where $\delta(\cdot)$ is the Dirac δ-function, and $T \to \infty$ the observation time-interval. The Kullback-Leibler divergence[25]

$$D_{\mathrm{KL}}(p_i, q) = \int_0^1 \mathrm{d}y \; p_i(y) \log \frac{p_i(y)}{q(y)} \tag{9}$$

measures the distance between the neural firing rate distribution $p_i(y)$ and a normalized distribution $q(y)$, with $D_{\mathrm{KL}} \geq 0$ generically and $D_{\mathrm{KL}} = 0$ only for $p_i(y) = q(y)$. Gaussian distributions

$$q(y) \propto \exp\left(\lambda_{1y} + \lambda_{2y^2}\right) \tag{10}$$

maximize the information entropy whenever mean and standard deviation are given[25]. Minimizing the Kullback-Leibler divergence (9) with respect to the individual intrinsic parameters a_i and b_i is hence equivalent to optimizing the information content of the neural activity[6]. The optimization can be performed using variational calculus[7,10,26], one obtains

$$\begin{aligned} \dot{a}_i &= \epsilon_a (1/a_i + (x_i - b_i)\theta) \\ \dot{b}_i &= \epsilon_b (-a_i)\theta, \end{aligned} \tag{11}$$

with $\theta = 1 - 2y_i + (\lambda_1 + 2\lambda_2 y_i)(1 - y_i) y_i$ and where the ϵ_a and ϵ_b are adaption rates. When these adaption rates are small, with respect to the time scale of the cognitive dynamics (6), an implicit time average with respect to the distribution of input signals is performed[10,26], compare (8), and the respective firing rate distribution $p_i(y)$ approaches the target distribution function $q(y)$. The rules (11) are also denoted *stochastic* adaption rules, since they correspond[7,10,26], for small adaption rates ϵ_a and ϵ_b, to an implicit (stochastic) average over all input signals x_i. The adaption rules (11) implement the optimization of an entire distribution function, and not of a single scalar quantity, and are hence equivalent to a polyhomeostatic optimization process[7].

The original network (6) has point attractors, as given by $\dot{x}_i = 0, \forall i$. These attractors are destabilized for any $\epsilon_a, \epsilon_b > 0$, the coupled system (6) and (11) has no stationary points, as defined by $\dot{x} = 0 = \dot{a} = \dot{b}$. Stationary point attractors lead to neural firing statistics far from the target distribution (10) and to a large Kullback-Leibler divergence D_{KL}, and hence to finite stochastic gradients (11). The principle of polyhomeostatic adaption is hence intrinsically destabilizing.

In deriving the stochastic adaption rules (11) one rewrites the Kullback-Leibler divergence D_{KL} as an integral over the distribution $p_i(x)$ of the respective membrane potential x_i, namely as $D_{KL} = \int \mathrm{d}x \; p_i(x) d(x)$, with an appropriate kernel $d(x)$. The optimal overall adaption rules depend on the specific shape of $p_i(x)$. The optimal adaption rates not dependent on the distribution of the membrane potential are, on the other side, given by minimizing the kernel, $\dot{a} \propto \partial d(x)/\partial a$ and $\dot{b} \propto \partial d(x)/\partial b$ respectively, which leads to the adaption rates (11), which are instantaneous in time[7,10,26].

For the analysis of the simulation results we use with

$$\xi^p = (x_1^p, \ldots, x_N^p), \quad \|\xi^p\| = \sqrt{\sum_{i=1}^{N} \left(\xi_i^p\right)^2} \tag{12}$$

the binary activity patterns of the attractors of (6). The are $\xi^1 = (1, 1, 0)$ and $\xi^2 = (0, 1, 1)$ for the case of the $N = 3$ network presented in Fig. 1. We use two criteria in order to

Table 1 | Relation between the parameter λ_1, for $\lambda_2 = 0$, and the mean value μ, as given by Eq. 4, for the target distribution $q(y)$

μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
λ_1	−9.995	−4.801	−2.672	−1.229	0	1.229	2.672	4.801	9.995

investigate to which degree the original binary attractors are retraced. The first criterion is the overlap $O_p \in [0, 1]$,

$$O_p = \frac{\langle \xi^p, y \rangle}{\|\xi^p\| \|y\|}, \quad \langle \xi^p, y \rangle = \sum_{i=1}^{N} \xi_i^p y_i, \tag{13}$$

which corresponds to the cosine of the angle between the actual neural activity $(y_1, \ldots, y_N)$ and the attractor state ξ^p. As a second measure, of how close the actual activity pattern and the original attractors are, we consider the reweighted scalar product $A_p \in [0, 1]$

$$A_p = \langle \xi^p, y \rangle \Big/ \sum_{i=1}^{N} \xi_i^p. \tag{14}$$

Note that the ξ_i^p are positive, representing the neural firing rate in attractor states. For the general network analysis we use the Hopfield encoding[8,9]

$$w_{ij} \propto \frac{1}{\alpha(N-1)} \sum_{p=1}^{N_p} \left(\xi_i^p - \overline{\xi_i}\right)\left(\xi_j^p - \overline{\xi_j}\right) \tag{15}$$

for the synaptic weights w_{ij}, where the $\overline{\xi_i} = \frac{1}{N_p} \sum_{p=1}^{N_p} \xi_i^p$ are the arithmetic means of all pattern activities, for the respective sites. Here N_p is the number of encoded attractors and $\alpha \in [0, 1]$ the mean pattern activity,

$$\alpha = \frac{1}{N} \sum_i x_i^p. \tag{16}$$

When using the Hopfield encoding (15), the attractors are known to correspond, to a close degree, to the stored patterns ξ^p, as long as the number N_p of patterns is not too large[8,9].

We simulated Eqs. (6) and (11) using 4th order classical Runge-Kutta[29] and a timestep of $\Delta t = 0.1$. The resulting dynamics is dependent on the magnitude of the adaption rates for the gain a and for the threshold b. In general latching dynamics is favored for small $\epsilon_b \sim 0.01$ and somewhat larger $\epsilon_a \sim 0.01 \ldots 1$. We kept a constant leak rate $\Gamma = 1$. The results presented in Figs. 6 and 7 have been obtained setting $\lambda_2 = 0$ and using Table 1 for λ_1.

1. Ay, N., Bertschinger, N., Der, R., Güttler, F. & Olbrich, E. Predictive information and explorative behavior of autonomous robots. *Eur. Phys. J. B* **63**, 329–339 (2008).
2. Ay, N., Bernigau, H., Der, R. & Prokopenko, M. Information-driven self-organization: the dynamical system approach to autonomous robot behavior. *Theory in Biosciences* **131**, 161–179 (2012).
3. Sporns, O. & Lungarella, M. Evolving coordinated behavior by maximizing information structure. *Artificial life X: proceedings of the tenth international conference on the simulation and synthesis of living systems*, 323–329 (2006).
4. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* **7**, 1129–1159 (1995).
5. Becker, S. Mutual information maximization: models of cortical self-organization. *Network: Computation in Neural Systems* **7**, 7–31 (1996).
6. Triesch, J. A gradient rule for the plasticity of a neuron's intrinsic excitability. In *Proceedings of ICANN 2005*, Duch, W. *et al.* (Eds.), LNCS **3696**, 65–70 (2005).
7. Marković, D. & Gros, C. Self-organized chaos through polyhomeostatic optimization. *Physical Review Letters* **105**, 068702 (2010).
8. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences* **79**, 2554–2558 (1982).
9. Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the national academy of sciences* **81**, 3088–3092 (1984).
10. Marković, D. & Gros, C. Intrinsic Adaption in Autonomous Recurrent Neural Networks. *Neural Computation* **24**, 523–540 (2012).
11. Russo, E., Namboodiri, V. M., Treves, A., & Kropff, E. Free association transitions in models of cortical latching dynamics. *New Journal of Physics* **10**, 015008 (2008).
12. Akrami, A., Russo, E. & Treves, A. Lateral thinking, from the hopfield model to cortical dynamics. *Brain Research* **1434**, 4–16 (2012).
13. Horn, D. & Usher, M. Neural networks with dynamical thresholds. *Phys. Rev. A* **40**, 1036 (1989).
14. Sompolinsky, H. & Kanter, I. Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* **57**, 2861–2864 (1986).
15. Abeles, M. *et al.* Cortical activity flips among quasi-stationary states. *Proceedings of the National Academy of Sciences* **92**, 8616–8620 (1995).
16. Ringach, D. L. States of mind. *Nature* **425**, 912–913 (2003).
17. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **421**, 573–578 (2004).
18. MacLean, J. N., Watson, B. O., Aaron, G. B. & Yuste, R. Internal Dynamics Determine the Cortical Response to Thalamic Stimulation. *Neuron* **48**, 811–823 (2005).
19. Gros, C. Neural networks with transient state dynamics. *New Journal of Physics* **9**, 109 (2007).
20. Gros, C. Cognitive computation with autonomously active neural networks: An emerging field. *Cognitive Computation* **1**, 77–90 (2009).
21. Gros, C. & Kaczor, G. Semantic learning in autonomously active recurrent neural networks. *J. Algorithms in Cognition, Informatics and Logic* **81**, 686–704 (2010).
22. Treves, A. Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsychology* **22**, 276–291 (2005).
23. Kropff, E. & Treves, A. The complexity of latching transitions in large scale cortical networks. *Natural Computing* **6**, 169–185 (2007).
24. Beer, R. D. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior* **11**, 209–243 (2003).
25. Gros, C. Complex and Adaptive Dynamical Systems. A Primer. Springer (2008); third edition 2013.
26. Linkerhand, M. & Gros, C. Self-organized stochastic tipping in slow-fast dynamical systems. *Mathematics and Mechanics of Complex Systems* **1–2**, 129 (2013).
27. Bonachela, J. A. & Muñoz, M. A. Self organized quasi criticality Self-organization without conservation: true or just apparent scale-invariance? *Journal of Statistical Mechanics: Theory and Experiment* **2009**, P09009 (2009).
28. Marković, D. & Gros, C. Powerlaws and Self-Organized Criticality in Theory and Nature. to be published.
29. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press, September 2007.
30. Martius, G., Herrmann, J. & Der, R. Guided self-organisation for autonomous robot development. *Advances in Artificial Life* **46–48**, 766–775 (2007).
31. Prokopenko, M. Guided self-organization. *HFSP Journal* **3**, 287 (2009).
32. Haken, H. Information and self-organization: A macroscopic approach to complex systems. *Springer* (2006).

Author contributions

Both authors, C.G. and M.L., contributed equally to the research and to the preperation of the manuscript, M.L. prepared the figures. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Linkerhand, M. & Gros, C. Generating functionals for autonomous latching dynamics in attractor relict networks. *Sci. Rep.* **3**, 2042; DOI:10.1038/srep02042 (2013).