CellPress

Optimal Control of Transient Dynamics in Balanced Networks Supports Generation of Complex Movements

Guillaume Hennequin,[1,2,*] Tim P. Vogels,[1,3,4] and Wulfram Gerstner[1,4]
[1]School of Computer and Communication Sciences and Brain Mind Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[2]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
[3]Centre for Neural Circuits and Behaviour, University of Oxford, Oxford OX1 3SR, UK
[4]Co-senior author
*Correspondence: gjeh2@cam.ac.uk

http://dx.doi.org/10.1016/j.neuron.2014.04.045

SUMMARY

Populations of neurons in motor cortex engage in complex transient dynamics of large amplitude during the execution of limb movements. Traditional network models with stochastically assigned synapses cannot reproduce this behavior. Here we introduce a class of cortical architectures with strong and random excitatory recurrence that is stabilized by intricate, fine-tuned inhibition, optimized from a control theory perspective. Such networks transiently amplify specific activity states and can be used to reliably execute multidimensional movement patterns. Similar to the experimental observations, these transients must be preceded by a steady-state initialization phase from which the network relaxes back into the background state by way of complex internal dynamics. In our networks, excitation and inhibition are as tightly balanced as recently reported in experiments across several brain areas, suggesting inhibitory control of complex excitatory recurrence as a generic organizational principle in cortex.

INTRODUCTION

The neural basis for movement generation has been the focus of several recent experimental studies (Churchland et al., 2010, 2012; Ames et al., 2014). In a typical experiment (Figure 1A), a monkey is trained to prepare a particular arm movement and execute it after the presentation of a go cue. Concurrent electrophysiological recordings in cortical motor and premotor areas show an activity transition from spontaneous firing into a movement-specific preparatory state with firing rates that remain stable until the go cue is presented (Figure 1B). Following the go cue, network dynamics begin to display quickly changing, multiphasic firing rate responses that form spatially and temporally complex patterns and eventually relax toward spontaneous activation levels (Churchland and Shenoy, 2007).

Recent studies (Afshar et al., 2011; Shenoy et al., 2011) have suggested a mechanism similar to a spring-loaded box, in which motor populations could act as a generic dynamical system that is driven into specific patterns of collective activity by preparatory stimuli (Figure 1). When released, intrinsic population dynamics would commandeer the network activity and orchestrate a sequence of motor commands leading to the correct movement. The requirements for a dynamical system of this sort are manifold. It must be highly malleable during the preparatory period, excitable and fast when movement is triggered, and stable enough to return to rest after an activity transient. Moreover, the dynamics must be sufficiently rich to support complex movement patterns (Maass et al., 2002; Sussillo and Abbott, 2009; Laje and Buonomano, 2013).

How the cortical networks at the heart of this black box (Figure 1C) could generate such complex transient amplification through recurrent interactions is still poorly understood. Randomly connected, globally balanced networks of leaky integrate-and-fire (LIF) neurons exhibit stable background states (van Vreeswijk and Sompolinsky, 1996; Tsodyks et al., 1997; Brunel, 2000; Vogels et al., 2005; Renart et al., 2010) but cannot autonomously produce the substantial yet reliable, spatially patterned departure from background activity observed in the experiments. Networks with strong recurrent pathways can exhibit ongoing, complex rate fluctuations beyond the population mean (Sompolinsky et al., 1988; Sussillo and Abbott, 2009; Rajan et al., 2010; Litwin-Kumar and Doiron, 2012; Ostojic, 2014) but do not capture the transient nature of movement-related activity. Moreover, such rate dynamics are chaotic, and sensitivity to noise seems improper in a situation in which the initial conditions dictate the subsequent evolution of the system. Chaos can be controlled either through continuous external feedback loops, or modifications of the recurrent connectivity itself (Sussillo and Abbott, 2009; Laje and Buonomano, 2013; Hoerzer et al., 2014). However, all of these models violate Dale's principle, according to which neurons can be either excitatory or inhibitory, but not of a mixed type. In other words, there is currently no biologically plausible network model to implement the spring-loaded box of Figure 1C, i.e., a system that

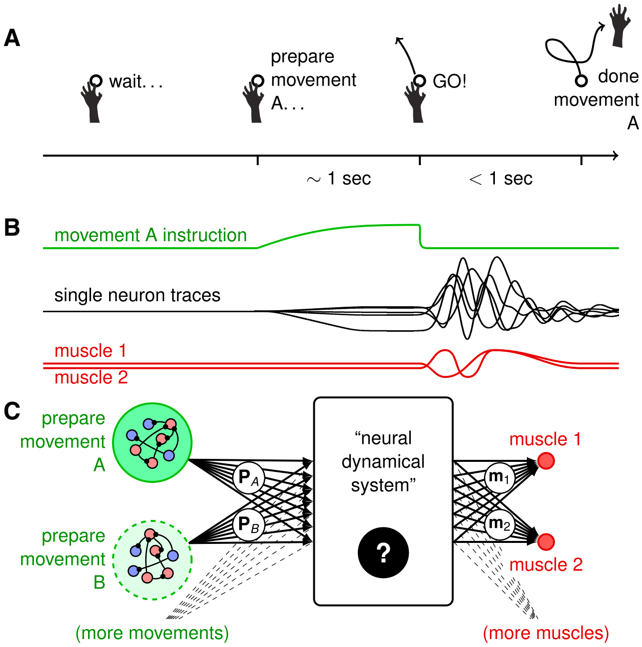

Figure 1. Dynamical Systems View of Movement Planning and Execution

(A) A typical delayed movement generation task starts with the instruction of what movement must be prepared. The arm must then be held still until the go cue is given, upon which the movement is performed.

(B) During the preparatory period, model neurons receive a ramp input (green). Following the go cue, that input is withdrawn, leaving the network activity free to evolve from the initial condition set up during the preparatory period. Model neurons then exhibit transient oscillations (black) that drive muscle activity (red).

(C) Black-box view on movement generation. Muscles (red, right) are thought to be activated by a population of motor cortical neurons ("neural dynamical system," middle). To prepare the movement, this network is initialized in a desired state by the slow activation of a movement-specific pool of neurons (green, left).

well-chosen inputs can prompt to autonomously generate multiphasic transients of large amplitude.

Here we introduce a class of neuronal networks composed of excitatory and inhibitory neurons that, similarly to chaotic networks, rely on strong and intricate excitatory synaptic pathways. Because traditional homogeneous inhibition is not enough to quench and balance chaotic firing rate fluctuations in these networks, we build a sophisticated inhibitory counterstructure that successfully dampens chaotic behavior but allows strong and fast break-out transients of activity. This inhibitory architecture is constructed with the help of an optimization algorithm that aims to stabilize the activity of each unit by adjusting the strength of existing inhibitory synapses, or by adding or pruning inhibitory connections. The result is a strongly connected, but *nonchaotic*, balanced network that otherwise looks random. We refer to such networks as "stability-optimized circuits," or SOCs. We study both a rate-based formulation of SOC dynamics and a more realistic spiking implementation. We show that external stimuli can force these networks into unique and stable activity states. When input is withdrawn, the subsequent free transient dynamics are in good qualitative agreement with the motor cortex data on single-cell and network-wide levels.

We show that SOCs connect unrelated aspects of balanced cortical dynamics. The mechanism that underlies the generation of large transients here is a more general form of "balanced amplification" (Murphy and Miller, 2009), which was previously discovered in the context of visual cortical dynamics. Additionally, during spontaneous activity in SOCs, a "detailed balance" (Vogels and Abbott, 2009) of excitatory and inhibitory inputs emerges, but it is much finer than expected from shared population fluctuations (Okun and Lampl, 2008; Cafaro and Rieke, 2010; Renart et al., 2010), beyond also what is possible with recently published inhibitory learning rules (Vogels et al., 2011; Luz and Shamir, 2012) that only alter the weights of inhibitory synapses, but not the structure of the network itself. Preparing such exquisitely balanced systems with an external stimulus into a desired initial state will then lead to momentary but dramatic departure from balance, demonstrating how realistically shaped cortical architectures can produce a large library of unique, transient activity patterns that can be decoded into motor commands.

RESULTS

We are interested in studying how neural systems (Figure 1C) can produce the large, autonomous, and stable "spring-box dynamics" as described above. We first investigate how to construct the architectures that display such behavior and show how their activity can be manipulated to produce motor-like activity. We then discuss the implications of the proposed architecture for the joint dynamics of excitation and inhibition. Finally, we confirm our results in a more realistic spiking network.

SOCs

We use $N = 200$ interconnected rate units (Dayan and Abbott, 2001; Gerstner and Kistler, 2002), of which 100 are excitatory and 100 are inhibitory. We describe the temporal evolution of their "potentials," gathered in a vector $\mathbf{x}(t)$, according to

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x}(t) + \mathbf{I}(t) + \mathbf{W}\Delta\mathbf{r}(\mathbf{x}, t) \qquad (1)$$

where $\tau = 200$ ms, the combined time constant of membrane and synaptic dynamics, is set to match the dominant timescale in the data of Churchland et al. (2012). $\mathbf{I}(t) = \xi(t) + \mathbf{S}(t)$ denotes all external inputs, i.e., an independent noise term $\xi(t)$ and a specific, patterned external stimulation $\mathbf{S}(t)$. The vector $\Delta\mathbf{r}(\mathbf{x},t)$ contains the instantaneous single-unit firing rates, measured relative to a low level of spontaneous activity ($r_0 = 5$ Hz). These rates are given by the nonlinear function $\Delta\mathbf{r}_i = g(x_i)$ of the potentials (Figure 3E and Experimental Procedures), although we also consider the linear case $\Delta\mathbf{r}_i \propto x_i$ in our analysis. The final term in Equation 1 accounts for the recurrent dynamics of the system due to its connectivity $\mathbf{W}$. We focus here on connectivities that obey Dale's principle, i.e., on weight matrices composed of separate positive and negative columns.

Random balanced networks can have qualitatively different types of dynamics depending on the overall magnitude of $\mathbf{W}$.

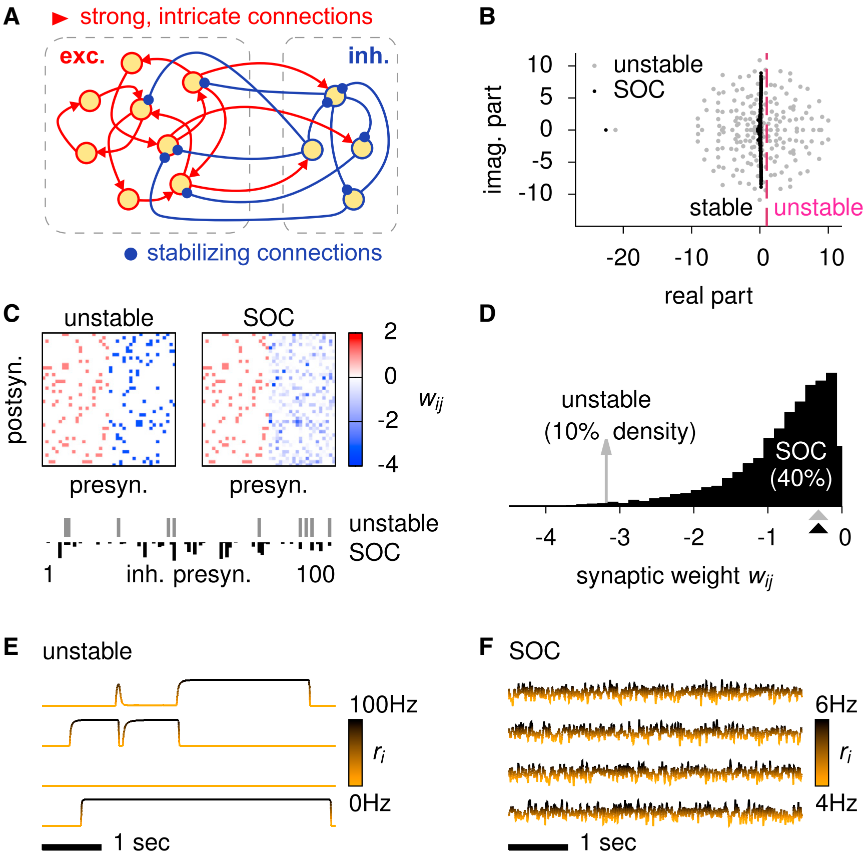

Figure 2. SOCs

(A) Schematic of a SOC. A population of rate units is recurrently connected, with strong and intricate excitatory pathways (red) that would normally produce unstable, chaotic activity. Stabilization is achieved through fine-tuned inhibitory feedback (blue).

(B) Eigenvalue spectrum of the connectivity matrix of a SOC (black) and that of the chaotic random network from which it is derived (gray). Stability requires all eigenvalues to lie to the left of the dashed vertical line. Note the large negative real eigenvalue, which corresponds to the spatially uniform activity pattern.

(C) Matrices of synaptic connectivity before (unstable) and after (SOC) stability optimization through inhibitory tuning. By design, the excitatory weights are the same in both matrices. Matrices were thinned out to 40 × 40 for visualization purposes. The bottom row shows the strengths of all the inhibitory input synapses to a single sample neuron, in the unstable network (gray) and in the corresponding SOC (black).

(D) Distribution of inhibitory weights in the unstable network (10% connection density, gray peak at $W_{ij} \sim -3.18$) and in the stabilized version (40% connection density, black). The mean inhibitory weight of all possible synapses is the same before and after optimization (~ -0.318, gray and black arrowheads).

(E and F) Spontaneous activity in the unstable network (E) and in the SOC (F), for four example units. Note the difference in firing rate scales.

With weak synapses, their activity decays rapidly against baseline when perturbed (not shown). To yield a more interesting, qualitatively different behavior, one can strengthen the existing connections (Figure 2C, left; Experimental Procedures), increasing the radius of the characteristically circular distribution of eigenvalues (Rajan and Abbott, 2006; Figure 2B). Small perturbations of the network dynamics can now propagate chaotically across the network (Sompolinsky et al., 1988; Rajan et al., 2010; Ostojic, 2014), generating uncontrollable, switch-like fluctuations in the neurons' firing rates even without external drive (Figure 2E).

Here we construct nonchaotic networks that exhibit stable background activity but retain interesting dynamical properties. Starting with the above-mentioned deeply chaotic network, we build a second network, a SOC (Figure 2A). The excitatory connections are kept identical to those in the reference network (Figure 2C), but the inhibitory connections are no longer drawn randomly. Instead, they are precisely matched against the excitatory connectivity. This "matching" is achieved by an algorithmic optimization procedure that modifies the inhibitory weights and wiring patterns of the reference network, aiming to pull the unstable eigenvalues of **W** toward stability (Experimental Procedures and Movie S1 available online). The total number of inhibitory connections is increased and the distribution of their strengths is wider, but the mean inhibitory weight is kept the same (Figure 2D). The resulting SOC network is as strongly connected as the reference chaotic network, *but it is no longer chaotic*, as indicated by the distribution of its eigenvalues in the complex plane, which all lie well within the stable side (Figure 2B, black dots). Accordingly, the background activity is now stable (Figure 2E), with small noisy fluctuations around the mean caused by $\xi(t)$. Shuffling the optimal inhibitory connectivity results in chaotic dynamics similar to the reference network (not shown), indicating that it is not the broad, sparse distribution of inhibitory weights but the precise inhibitory wiring pattern that stabilizes the dynamics.

SOCs Exhibit Complex Transient Amplification

To test whether SOCs can produce the type of complex transient behavior seen in experiments (Churchland and Shenoy, 2007; Churchland et al., 2012; cf. also Figure 1), we momentarily clamp each unit to a specific firing rate and then observe the network as it relaxes to the background state (later, we model the preparatory period explicitly). Depending on the spatial pattern of initial stimulation, the network activity exhibits a variety of transient behaviors. Some initial conditions result in fast monotonous decay toward rest, whereas others drive large transient deviations from baseline rate in most neurons.

To quantify this amplifying behavior of the network in response to a stimulus, we introduce the notion of "evoked energy" $E(\mathbf{a})$, measuring both the amplitude and duration of the collective transient evoked by initial condition **a** for a given

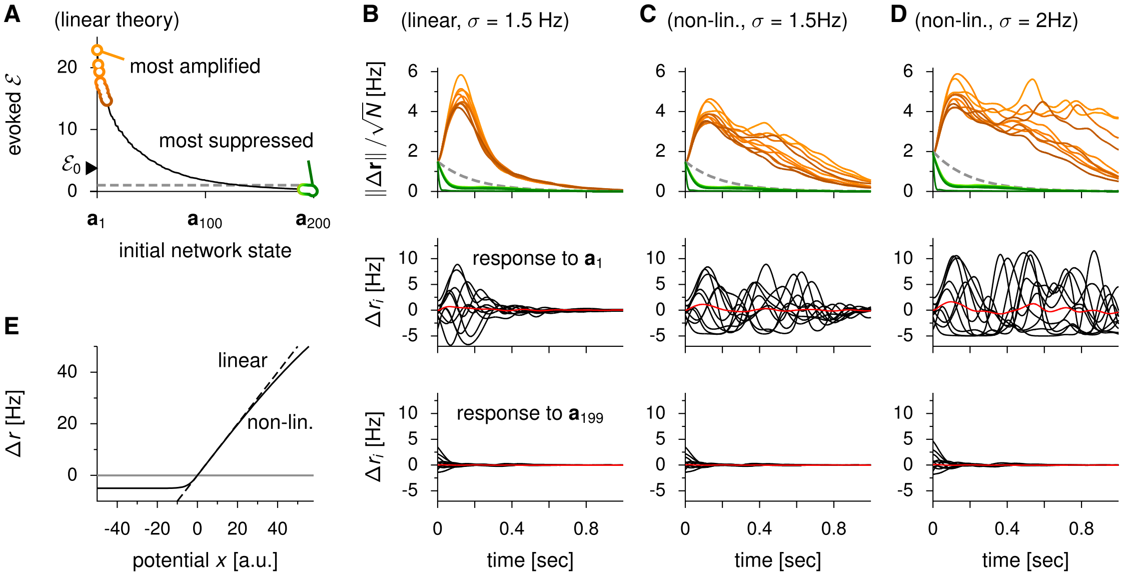

Figure 3. Transient Amplification in SOCs

(A) The energy E evoked by $N = 200$ orthogonal initial conditions ($\mathbf{a}_1, \ldots, \mathbf{a}_N$) as the network evolves linearly ($\Delta\mathbf{r}_i = x_i$) with no further input according to Equation 1. The energy (Equation 4) is normalized such that it equals 1 for an unconnected network ($\mathbf{W} = 0$) irrespective of the initial condition (dashed horizontal line). Each successive initial condition $\mathbf{a}_i$ is defined as the one that evokes maximum energy, within the subspace orthogonal to all previous input patterns $\mathbf{a}_{j<i}$ (Experimental Procedures). The black arrowhead indicates the mean, or the expected evoked energy E_0 when the neurons are initialized in a random activity state.

(B) Dynamics of the SOC in the linear regime. Top: time evolution of $\|\Delta\mathbf{r}\|/\sqrt{N}$, which measures the momentary spread of firing rates in the network above or below baseline, as the dynamics unfold from any of the ten best or ten worst initial states (same color code as in A). Initial states have a $\sigma = 1.5$ Hz across the population. The dashed gray line shows $\sigma \times \exp(-t/\tau)$, i.e., the behavior of an unconnected pool of neurons. Bottom: sample firing rate responses of ten randomly chosen neurons following initialization in state $\mathbf{a}_1$ or $\mathbf{a}_{199}$. The red line indicates the momentary population-averaged firing rate.

(C and D) Same as in (B), now with the nonlinear gain function shown in (E). Unlike in the linear case, the dynamics now depend on the spread σ of the initial firing rates across the network (1.5 Hz in C as in B, 2 Hz in D). The larger this spread, the longer the duration of the population transient. When $\sigma > 3$ Hz, the network initiates self-sustained chaotic activity (not shown).

(E) Single-unit input-output nonlinearity [solid line, $\Delta\mathbf{r}_i = g(x_i)$ given by Equation 2] and its linearization (dashed line, $\Delta\mathbf{r}_i = x_i$).

SOC compared to an unconnected network (Experimental Procedures). Of all initial conditions $\Delta\mathbf{r}$ with constant power $\sigma^2 = \sum_i \Delta\mathbf{r}_i^2/N$, we find the one that maximizes this energy and call it $\mathbf{a}_1$. We repeat this procedure among all patterns orthogonal to $\mathbf{a}_1$ to obtain the second best pattern $\mathbf{a}_2$, and iterate until we have filled a full basis of $N = 200$ orthogonal initial conditions $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N\}$ (an analytical solution exists for the linear case, $\Delta\mathbf{r}_i \propto x_i$; cf. Experimental Procedures). A large set of these orthogonal initial conditions are transiently amplified by the connectivity of the network, with the strongest states evoking energies ~25 times greater than expected from the exponential decay of activity in unconnected neurons (Figure 3A). For these strongly amplifying states, the population-averaged firing rate remains roughly constant during the transient (red line in Figure 3B, middle), but the average *absolute* deviation from baseline firing rate per unit can grow dramatically (Figure 3B, top), because some units become more active and others become less active than baseline. Amplifying behavior progressively attenuates but subsists for roughly the first half of the basis ($\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{100}$). Eventually, amplification disappears, and even turns into active dampening of the initial condition (Figure 3A, green dots). For $\mathbf{a}_{200}$, the least amplifying initial condition, return to rest occurs three times faster than it would in unconnected neurons (Figure 3B). Here, the least-preferred state $\mathbf{a}_{200}$ corresponds to the uniform spatial mode of activity (1, 1, ..., 1), i.e., the trivial case in which all neurons are initialized slightly above (or below) their baseline rate.

Finally, if we increase the firing rate standard deviation σ in the initial condition, such that a substantial number of (excitatory *and* inhibitory) neurons will reach lower saturation and stop firing during the transient, the duration of the response increases (Figures 3C and 3D). For $\sigma > 3$ Hz the network response begins to self-sustain in the chaotic regime (not shown). This behavior is beyond the scope of our study, and in the following we set $\sigma = 1.5$ Hz, which results in transients of ~1 s duration. Note also that we did not observe a return to chaotic behavior in the full spiking network, even though firing rates in the initial conditions deviated more dramatically from baseline.

SOC Dynamics Are Consistent with Experimental Data

In Churchland et al. (2012), monkeys were trained to perform 27 different cued and delayed arm movements (Figure 1A). The activity of the neurons recorded during this task (Figure 4A) displayed transient activity similar to the responses of appropriately initialized SOCs (Figure 3C). To model this behavior, we assume that each of the 27 instructed movements is associated with a pool of prefrontal cortical neurons (Figure 1C) feeding the motor network through sets of properly tuned input weights

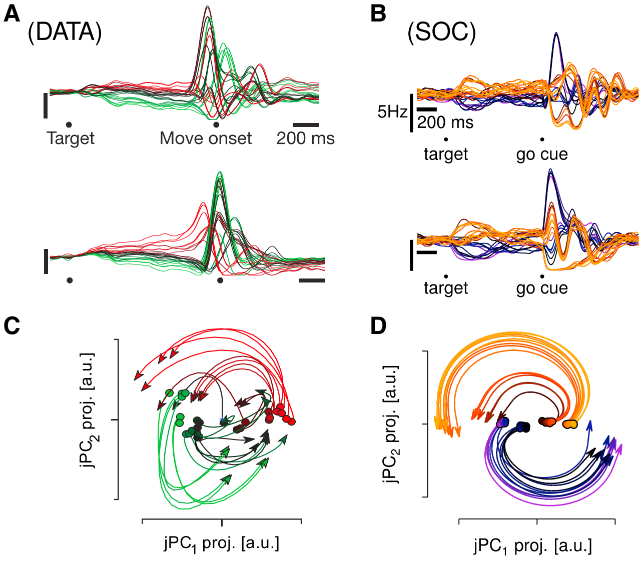

Figure 4. SOCs Agree with Experimental Data

(A) Experimental data, adapted with permission from Churchland et al. (2012). Each trace denotes the trial-averaged firing rate of a single cell (two sample cells are shown here) during a delayed reaching task. Each trace corresponds to one of 27 different movements. Vertical scale bars denote 20 spikes/s. The go cue is not explicitly marked here, but it occurs about 200 ms before movement onset.

(B) Time-varying firing rates of two neurons in the SOC, for 27 "conditions," each characterized by a different collective steady state of preparatory activity (see text).

(C) Experimental data adapted from Churchland et al. (2012), showing the first 200 ms of movement-related population activity projected onto the top jPC plane. Each trajectory corresponds to one of the 27 conditions mentioned in (A).

(D) Same analysis as in (C), for the SOC.

(Experimental Procedures). For a given movement, the corresponding command pool becomes progressively more active during the 1-s-long delay period (Amit and Brunel, 1997; Wang, 1999). Remarkably, this simple input drives the SOC into a stable steady state (Figure 4B). By adjusting the movement-specific input weights, we can manipulate this steady state and force the network into a specific spatial arrangement of activity. This is not possible in generic chaotic networks in which external inputs are overwhelmed by a strong and uncontrolled recurrent activity. We chose the input weights such that, by the end of the delay period, the network arrives at a state that is one of 27 different linear combinations of $\mathbf{a}_1$ and $\mathbf{a}_2$, i.e., the two orthogonal activity states that evoke the strongest collective responses. The go cue quickly silences the command pool, leaving the network free to depart from its preparatory state and to engage in transient amplification. The resulting recurrent dynamics produce strong, multiphasic, and movement-specific responses in single units (Figure 4B), qualitatively similar to the data.

In the data of Churchland et al. (2012), the complexity of the single-neuron multiphasic responses was in fact hiding orderly rotational dynamics on the population level. A plane of projection could be found in which the vector of population firing activity [$\Delta\mathbf{r}(t)$ in our model] would start rotating after the go cue, and consistently rotate in the same direction for all movements (Figure 4C). Our model, analyzed with the same dynamical variant of principal component analysis (jPCA, Churchland et al., 2012; Experimental Procedures) displays the same phenomenon (Figure 4D).

SOCs Can Generate Complex Movements

The complicated, multiphasic nature of the firing rate transients in SOCs suggests the possibility of reading out equally complex patterns of muscle activity. We illustrate this idea in a task in which the joint activation of two muscles must produce one of two target movements ("snake" or "butterfly" in Figure 5), within 500 ms following the go cue. Similarly to Figure 4, the preparatory input for the "snake" (respectively "butterfly") movement is chosen such that, by the arrival of the go cue, the network activity matches the network's preferred initial condition $\mathbf{a}_1$ (respectively $\mathbf{a}_2$). Two readout units ("muscles") compute a weighted sum of all neuronal activities in the network that we take to directly reflect the horizontal and vertical coordinates of the movement. Simple least-squares regression learning of the output weights (Experimental Procedures) can map the activity following each command onto the correct trajectory (compare the five test trials in Figure 5A).

We conclude that the SOC's single-neuron responses form a set of basis functions that is rich enough to allow readout of nontrivial movements. This is not possible in untuned, chaotic balanced networks without exquisite feedback loops or supervised learning of lateral connections (Sussillo and Abbott, 2009; Laje and Buonomano, 2013; Hoerzer et al., 2014) because of the high sensitivity to noise. Furthermore, in balanced networks with weak connections, each neuron's activity decays exponentially: this redundancy prevents the network to robustly learn the snake and butterfly trajectories (Figure 5B).

Interaction between Excitation and Inhibition in SOCs

To understand the mechanism by which SOCs amplify their preferred inputs, we dissociated the excitatory (c_E) and inhibitory (c_I) synaptic inputs each unit received from other units in the network in the absence of specific external stimulation [$\mathbf{S}(t) = 0$]. We quantified the excitation/inhibition balance by $r_{EI}(t)$, the momentary Pearson correlation coefficient between c_E and c_I across the network population. The preferred initial states of the SOC momentarily produce substantially negative excitation/inhibition input correlations (Figure 6A), indicating an average mismatch between excitatory and inhibitory inputs. Balance is then quickly restored by internal network dynamics, with $r_{EI}(t)$ reaching ~0.8 at the peak of the transient triggered by initial condition $\mathbf{a}_1$. The effect subsists, although progressively attenuated, for roughly the first 100 preferred initial states ($\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_{100}$), which are also the initial states that trigger amplified responses.

Notably, the patterns of neuronal activity after 100 ms of recurrent processing have a larger amplitude than—but bear little spatial resemblance to—the initial condition. This is reflected by a rapid decay (within 100 ms) of the correlation coefficient between the momentary network activity and the initial state (Figure 6B, black). However, considering the excitatory and

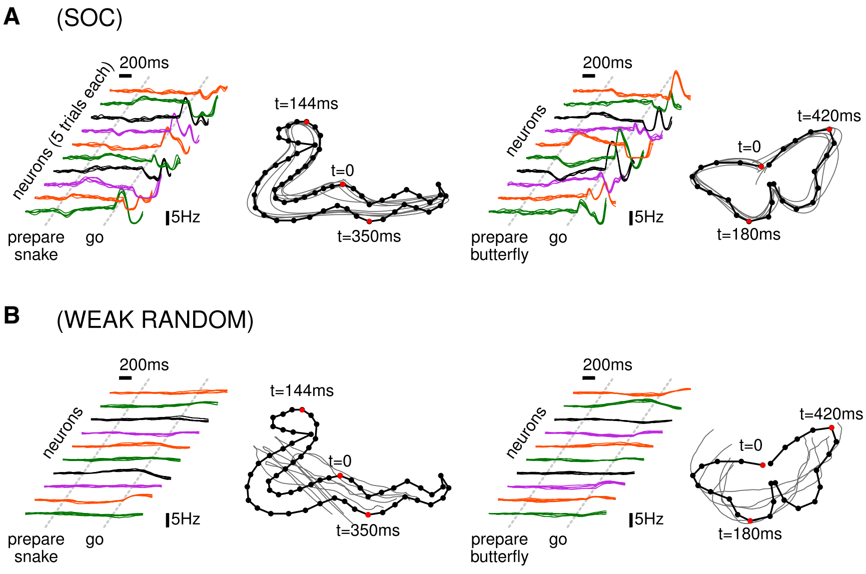

Figure 5. Generation of Complex Movements through SOC Dynamics

(A) Firing rates versus time for ten neurons of the SOC, as the system prepares and executes either of the two target movements (snake, left or butterfly, right). Five test trials are shown for each neuron. The corresponding muscle trajectories following the go cue are shown for the same five test trials (thin traces) and compared to the target movement (black trace and dots).

(B) Same as in (A), for a weakly connected (untuned) random balanced network (Experimental Procedures).

inhibitory populations separately shows that the excitatory subpopulation remains largely in the same spatial activity mode throughout the transient, i.e., units that were initially active (respectively inactive) tend to remain active (respectively inactive) throughout the relaxation (Figure 6B, red). In contrast, the inhibitory subpopulation becomes negatively correlated with its initial pattern after only 60 ms (Figure 6B, blue). In other words, it is mostly the swift reversal of inhibitory activity that quenches a growing excitatory transient and pulls the system back to rest.

The amplifying dynamics of excitation and inhibition seen on the level of transient responses to some initial conditions also shape the spontaneous background activity in SOCs (Figures 2F and 6D). In the absence of additional stimuli, the rate units are driven by private noise $\xi(t)$ (Experimental Procedures), such that firing rate fluctuations can be observed even in the unconnected case (**W** = 0) (Figure 6D, gray histogram). The recurrent SOC connectivity amplifies these unstructured fluctuations by one-third (Figure 6D, black histogram), because the noise stimulates each of the $\mathbf{a}_i$ modes evenly, and although some modes are suppressed by the recurrent dynamics and others are amplified, the net result is a mild amplification (Figure 3A, black arrowhead). Furthermore, because only a few activity modes experience very strong amplification, the resulting distribution of pairwise correlations among neurons is wide with a small positive mean (Figure 6E).

SOCs also exhibit an exquisite temporal match between excitatory and inhibitory inputs to single units during spontaneous activity (Figure 6F). The correlation between these two input streams averages to ~0.66 across units, because any substantial mismatch between recurrent excitatory and inhibitory inputs is instantly converted into a pattern of activity in which those inputs match again (cf. Figure 6A). The amplitude of such reactions is larger than the typical response to noise, so the network is constantly in a state of detailed excitation/inhibition balance (Vogels and Abbott, 2009). Furthermore, we have seen that it is mostly the spatial pattern of inhibitory activity that reverses during the course of amplification to restore the balance, whereas the excitatory activity is much less affected (Figure 6B). Thus, during spontaneous activity, inhibitory inputs are expected to lag behind excitatory inputs by a few milliseconds, which can indeed be seen in their average cross-correlogram (Figure 6G) and has also been observed experimentally (Okun and Lampl, 2008; Cafaro and Rieke, 2010).

The small temporal cofluctuations in the firing rates of the excitatory and inhibitory populations are known to translate into correlated excitatory and inhibitory inputs to single neurons, in densely connected circuits (Renart et al., 2010). Here, interestingly, excitatory and inhibitory inputs are correlated more strongly than expected from the magnitude of such shared population fluctuations. This can be seen by correlating the excitatory input stream taken in one unit with the inhibitory input stream taken in another unit (Figure 6F, bottom row). Such correlations average to ~0.26 only (to be compared with ~0.66 above; Figure 6G).

Spiking Implementation of a SOC

So far we have described neuronal activity on the level of firing rates. An important question is whether the dynamical features of rate-based SOCs are borne out in more realistic models of interconnected spiking neurons. To address this issue, we built a large-scale model of a SOC composed of 15,000 (12,000 excitatory + 3,000 inhibitory) LIF model neurons. The network was structured such that each neuron belonged to one of 200 excitatory or 200 inhibitory small neuron subgroups (of size 60 and 15, respectively), whose average momentary activities can be interpreted as the "rate variables" discussed until here.

In order to keep the network in the asynchronous and irregular firing regime, the whole network was, in part, randomly and sparsely connected, similar in this respect to traditional models (van Vreeswijk and Sompolinsky, 1996; Brunel, 2000; Vogels et al., 2005; Renart et al., 2010). In addition to those random, fast synapses, slower synapses were added that reflected the structured SOC connectivity between subgroups of neurons. The connectivity pattern between subgroups was given by a 400 × 400 SOC matrix obtained similarly to **W** in Figure 2. The value of a matrix element W_{ij} reflected the probability that a neuron in subgroup j be chosen as a presynaptic partner to another neuron in group i (Experimental Procedures). Overall, the average connection probability between spiking neurons was 0.2.

The spiking SOC operated in a balanced regime, with large subthreshold membrane potential fluctuations and occasional action potential firing (Figure 7A) with realistic rate and interspike interval statistics (Figure 7C). Spiking events were fully desynchronized on the level of the entire population, whose momentary activity was approximately constant at ~6 Hz.

Similar to our rate-based SOCs, the spiking network could be initialized in any desired activity state through the injection of specific ramping input currents into each neuron (Figure 7A). The go cue triggered sudden input withdrawal, resulting in large and rich transients in the trial-averaged spiking activities of single cells (Figure 7A, middle), which lasted for about 500 ms, and occurred reliably despite substantial trial-by-trial spiking variability in the preparation phase.

The trial-averaged firing rate responses to 27 different initial conditions, chosen in the same way as in Figure 4, as well as the diversity of single-cell responses, were qualitatively similar to the data in Churchland et al. (2012) (Figure 7; Figure S1). When projected onto the top jPC plane, the population activity also showed orderly rotations, as it did in our rate SOC (Figure 7E).

During spontaneous activity, subgroups of neurons in the SOC display large, slow and graded activity fluctuations (Figure 8A), which are absent from a control, traditional random network with equivalent synaptic input statistics (Figure S2; Experimental Procedures). Moreover, individual pairwise correlations between subgroup activities in the spiking SOC are accurately predicted by a linear rate model similar to Equation 1 (Figure 8B). Crucially, this rate model is nonchaotic, as the matrix that describes connectivity among subgroups has no eigenvalue larger than 1 (by construction of the SOC). We emphasize that our spiking network uses deterministic integrate-and-fire neurons without external noise, so that the spontaneous activity fluctuations seen in individual subgroups must have been intrinsically generated, similar to the voltage fluctuations seen in classical balanced networks (van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010). This is in contrast to the rate-based model where fluctuations arose from the amplification of an external source of noise (Equation 1).

Consistent with the effective rate picture, the distribution of spike correlations in the SOC (Figure 8C) is wide with a very small positive mean (ρ ~ 0.0027), indicating that cells fire asynchronously. The same is true in the control random network (ρ ~ 0.0005; Renart et al., 2010). However, within SOC subgroups, spiking was substantially correlated (Figure 8C, blue; ρ ~ 0.17), and particularly so on the 100 ms timescale, suggesting that the correlations can be attributed to joint activity fluctuations of all neurons in a given subgroup. Interestingly thus, in situations in which the subgroup partitioning would be unknown a priori

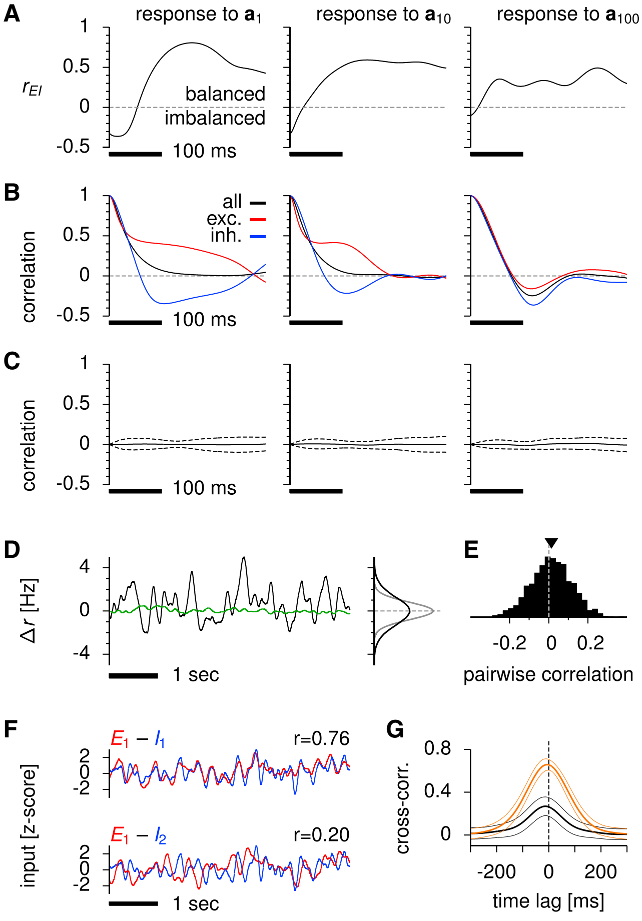

Figure 6. Precise Balance of Excitation and Inhibition in SOCs

The network is initialized in state $\mathbf{a}_1$ (left), $\mathbf{a}_{10}$ (middle), or $\mathbf{a}_{100}$ (right) and runs freely thereafter. The amplitude of the initial condition is chosen weak enough for the dynamics of amplification to remain linear (cf. Figure 3).

(A) Temporal evolution of the Pearson correlation coefficient r_{EI} between the momentary excitatory and inhibitory recurrent inputs across the population.

(B) Corresponding time course of the correlation coefficients between the network activity and the initial state, calculated from the activity of the entire population (black), the excitatory subpopulation (red), and the inhibitory subpopulation (blue).

(C) Temporal evolution of the correlation coefficient between the network activity when initialized in state $\mathbf{a}_i$, where $i = 1$ (left), 10 (middle), or 100 (right), and when initialized in a different state $\mathbf{a}_j$ ($j \neq i, j < 100$). Solid lines denote the average across j, and the dashed flanking lines indicate 1 σ. Small values indicate that the responses to the various initial conditions $\mathbf{a}_i$ are roughly decorrelated.

(D) Black: spontaneous fluctuations around baseline rate of a sample unit in the network. The corresponding rate distribution is shown on the right (black) and compared to the distribution obtained if the unit were not connected to the rest of the network (gray). Green: denotes the momentary population average rate, which fluctuates much less.

(E) Histogram of pairwise correlations between neuronal firing rates estimated from 100 s of spontaneous activity. The black triangular mark indicates the mean (~0.014).

(F) Excitatory (red) and inhibitory (blue) inputs taken in the same sample unit (top) or in a pair of different units (bottom), and normalized to Z scores. The corresponding Pearson correlation coefficients are indicated above each combination and computed from 100 s of spontaneous dynamics.

(G) Brown: lagged cross-correlogram of excitatory and inhibitory inputs to single units, each normalized to Z score (cf. F, top row). The solid line is an average across all neurons; flanking lines denote ± 1 σ. Inhibition lags behind excitation by a few milliseconds. Cross-correlating the E input into one unit with the inhibition input into another unit (cf. F, bottom row) yields the black curve, which is an average over 1,000 randomly chosen such pairs in the SOC.

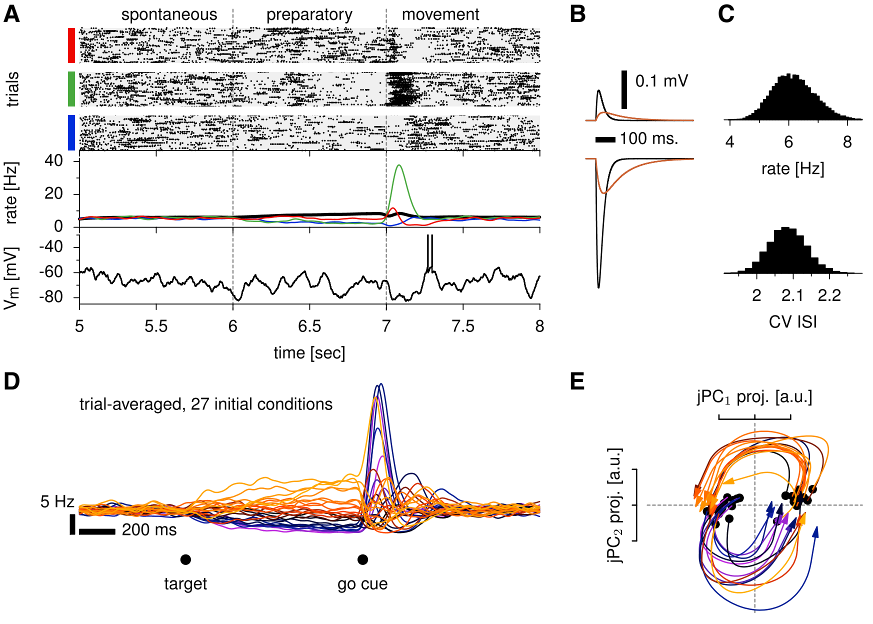

Figure 7. Transient Dynamics in a Spiking SOC

(A) The network is initialized in a mixture of its top two preferred initial states during the preparatory period. Top: raster plot of spiking activity over 200 trials for three cells (red, green, blue). Middle: temporal evolution of the trial-averaged activity of those cells (same color code) and that of the overall population activity (black). Rate traces were computed over 1,000 trials and smoothed with a Gaussian kernel (20 ms width), to reproduce the analysis of Churchland et al. (2012). Bottom: sample voltage trace of a randomly chosen neuron.

(B) Fast (black) and slow (brown) synaptic PSPs, corresponding to random and structured connections in the spiking circuit, respectively.

(C) Distribution of average firing rates (top) and interspike interval (ISI) coefficients of variation (bottom) during spontaneous activity.

(D) Trial-averaged firing rate traces for a single sample cell, when the preparatory input drives the SOC into one of 27 random mixtures of its first and second preferred initial conditions. Averages were computed over 1,000 trials, and smoothed as described in (A).

(E) First 200 ms of movement-related population activity, projected onto the top jPC plane. Each trajectory corresponds to a different initial condition in (D), using the same color code.

See also Figure S1.

(e.g., in actual experiments), clustering could potentially be performed on the basis of those large correlations (though admittedly they would be measured only rarely) to achieve subgroup identification. Not surprisingly, membrane potentials followed a similar pattern of correlations (Figure 8D).

Importantly, the detailed balance prediction made above for the rate-based scenario (Figures 6F and 6G) remains true on the level of single cells in the spiking network. Slow excitatory and inhibitory inputs (corresponding to the structured SOC recurrent synapses) to single neurons are substantially more correlated ($r \sim 0.24$) than pairs of excitation and inhibition currents taken from different neurons ($r \sim 0.12$; compare red and black in Figures 8E and 8F). This is not true in the control random network, in which the balance is merely a reflection of the synchronized fluctuations of the excitatory and inhibitory populations as a whole.

DISCUSSION

The motor cortex data of Churchland et al. (2012) showcase two seemingly conflicting characteristics. On the one hand, motor cortical areas appear to be precisely controllable during movement preparation, and dynamically stable with firing rates evolving well below saturation during movement execution. In most network models, such stability arises from weak recurrent interactions. On the other hand, the data show rich transient amplification of specific initial conditions, a phenomenon that requires strong recurrent excitation. To reconcile these opposing aspects, we introduced and studied the concept of SOCs, broadly defined as precisely balanced networks with strong and complex recurrent excitatory pathways. In SOCs, strong excitation mediates fast activity breakouts following appropriate input, whereas inhibition keeps track of the activity and acts as a retracting spring force. In the presence of intricate excitatory recurrence, inhibition cannot instantaneously quench such activity growth, leading to transient oscillations as excitation and inhibition waltz their way back to a stable background state. This results in spatially and temporally rich firing rate responses, qualitatively similar to those recorded by Churchland et al. (2012).

To build SOCs, we used progressive optimal refinement of the inhibitory synaptic connectivity within a normative, control-theoretic framework. Our method makes use of recent techniques for stability optimization (Vanbiervliet et al., 2009) and can in principle produce SOCs from any given excitatory connectivity. In simple terms, we iteratively refined both the absence/presence and the strengths of the inhibitory connections to pull all the unstable eigenvalues of the network's connectivity matrix back into the stable regime (Figure 2B). Even though we constrained the procedure to yield plausible network connectivity, notably one that respects Dale's law (Dayan and Abbott, 2001, chapter 7), it does not constitute—and is not meant to be—a synaptic plasticity rule. However, the phenomenology achieved by recent models of inhibitory synaptic plasticity (Vogels et al., 2011; Luz and Shamir, 2012; Kullmann et al., 2012) is similar to, although more crude than, that of our SOCs. It raises the possibility that nature solves the problem of network stabilization through a form of inhibitory plasticity, potentially aided by appropriate pre- and rewiring during development (Terauchi and Umemori, 2012).

In a protocol qualitatively similar to the experimental design of Churchland et al. (2012) (Figure 1), we could generate complex activity transients by forcing the SOC into one of a few specific

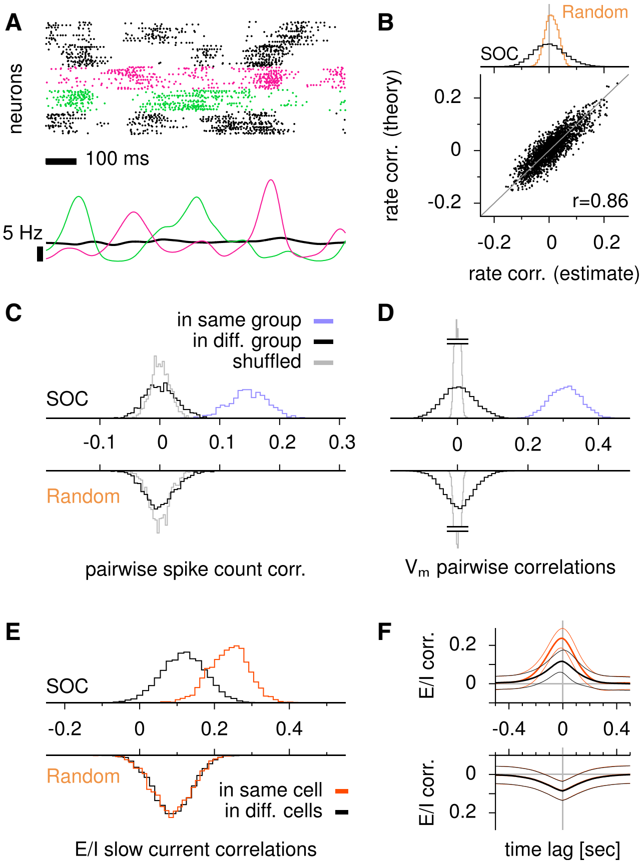

Figure 8. Spontaneous Activity in Spiking SOCs

(A) Top: raster plot of spontaneous spiking activity in the SOC. Only the neurons in the first five subgroups (300 neurons) are shown. Bottom: momentary activity of the whole population (black) and of the second (green) and third (magenta) subgroups. Traces were smoothed using a Gaussian kernel of 20 ms width.

(B) Pairwise correlations between instantaneous subgroup firing rates in the SOC, as empirically measured from a 1,000-s-long simulation (x axis) versus theoretically predicted from a linear stochastic model (y axis). Rate traces were first smoothed using a Gaussian kernel (20 ms width) as in (A). Distributions of pairwise correlations are shown at the top, for the SOC (black) and for a control random network with equivalent synaptic input statistics (brown; Experimental Procedures).

(C) Distributions of pairwise spike correlations in the SOC (top) and in the control random network the random network (bottom), between pairs of neurons belonging to the same subgroup (blue), or to different subgroups.

(black). Spike trains were first convolved with a Gaussian kernel of 100 ms width. Gray curves were obtained by shuffling the ISIs, thus destroying correlations while preserving the ISI distribution.

(D) Distributions of subthreshold membrane potential correlations. Colors have the same meaning as in (C). Voltage traces were cut off at the spike threshold. Gray curves were obtained by shuffling the time bins independently for each voltage trace.

(E) Distributions of pairwise correlations between the slow excitatory and inhibitory currents, taken in the same cells (red) or in pairs of different cells (black).

(F) Full lagged cross-correlograms between the slow excitatory and inhibitory currents, taken in the same cells (red) or in pairs of different cells. Thick lines denote averages over such excitation/inhibition current pairs across the network, and thin flanking lines denotes ± 1 σ. The peak at negative time lag corresponds to excitation currents leading inhibition currents.

See also Figure S2.

preparatory states through the delivery of appropriate inputs, which were then withdrawn to release the network into free dynamics (Figure 4). Those “engine dynamics” (Shenoy et al., 2011) could easily be converted into actual muscle trajectories. Simple linear readouts, with weights optimized through least-squares regression, were sufficient to produce fast and elaborate two-dimensional movements (Figure 5). Three aspects of the SOC dynamics make this possible. First, the firing rates strongly deviate from baseline during the movement period, effectively increasing the signal-to-noise ratio in the network response. Second, the transients are multiphasic (Figure 4B), as opposed to simple rise-and-decay, allowing the readouts not to overfit on multicurved movements. Third, the preferred initial conditions of the SOC are converted into activity modes that are largely nonoverlapping (Figure 6C). Thus, not only is the system highly excitable from a large set of states, but also those states produce responses that are distinguishable from one another, ensuring that different motor commands can be mapped onto distinct muscle trajectories (Figure 5).

Relation to Balanced Amplification and Relevance to Sensory Circuits

Transient amplification in SOCs is an extended, more intricate form of “balanced amplification,” first described by Murphy and Miller (2009) in a model of V1 synaptic organization. In their model, small patterns of spatial imbalance between excitation and inhibition, or “difference modes,” drive large activity transients in which neighboring excitation and inhibition neurons fire in unison (“sum modes”). Due to the absence of a topology in SOCs, it is impossible to tell which neuron is a neighbor to which, making sum and difference modes difficult to define. Nevertheless, they can be understood more broadly as patterns of average balance and imbalance in the excitatory and inhibitory synaptic *inputs* to single cells. With this definition, we showed here (Figure 6A) that the phenomenology of amplification in SOCs is similar to balanced amplification, i.e., small stimulations of difference modes drive large activations of sum modes. This accounts for the large transient firing rate deflections of individual neurons that follow appropriate initialization. A key difference between SOCs and Murphy and Miller’s model of V1 is the complexity of lateral excitatory connections in SOCs, which gives rise to temporally rich transients (Figures 3 and 4). Furthermore, although the “spring-box” analogy may not apply directly to sensory cortices, SOCs (as inhibition-stabilized networks) could still provide an appropriate conceptual framework for such cortical areas, as suggested by Ozeki et al. (2009). Likewise, the method we have used here to build such circuits could prove useful in finding conditions for inhibitory stabilization of known and nontrivial excitatory connectivities (see, e.g., Ahmadian et al., 2013). Finally, although we were able to calculate and rank the most (or least) amplified initial states analytically only in the linear regime, we found this ranking was preserved in the more realistic, nonlinear model in which neurons can saturate at zero and maximum firing rates (Figure 3). This is not surprising, as the onset of amplification after a weak perturbation relies on the connectivity matrix of SOCs being mathematically “nonnormal,” which is a linear property (Ganguli et al., 2008; Murphy and Miller, 2009; Goldman, 2009; Hennequin et al., 2012).

Relation to Detailed Excitation/Inhibition Balance

SOCs make a strong prediction regarding how excitation and inhibition interact in cortical networks: excitatory and inhibitory synaptic inputs in single neurons should be temporally correlated in a way that cannot be explained by the activity cofluctuations that occur on the level of the entire population.

During spontaneous activity in SOCs, balanced amplification of external noise (or intrinsically generated stochasticity, as in our spiking SOC) results in strongly correlated excitatory/inhibitory inputs in single units. This phenomenon is a recurrent equivalent to what has been referred to as "detailed balance" in feedforward network models (Vogels and Abbott, 2009; Vogels et al., 2011; Luz and Shamir, 2012), and it cannot be attributed here to mere cofluctuations of the overall activity of excitation and inhibition neurons. Such covariations can be substantial in balanced networks (Vogels et al., 2005; Kriener et al., 2008; Murphy and Miller, 2009), but they have been quenched here by requiring inhibitory synaptic connections to be three times stronger than excitatory connections on average (Renart et al., 2010; Hennequin et al., 2012). The residual shared population fluctuations accounted for only one-third of the total excitation/inhibition input correlation (Figures 6F and 6G). Thus, the excess correlation can only be explained by the comparatively large fluctuations of balanced, zero-mean activity modes (the responses to the preferred initial conditions of the SOC; Figure 6A).

A certain degree of such excitation/inhibition balance has been observed in several brain areas, and on levels as different as trial-averaged excitatory and inhibitory synaptic input conductances in response to sensory stimuli (Wehr and Zador, 2003; Mariño et al., 2005; Froemke et al., 2007; Dorrn et al., 2010; but see Haider et al., 2013), single-trial synaptic responses in which the trial-average has been removed ("residuals," Cafaro and Rieke, 2010), and spontaneous activity (Okun and Lampl, 2008; Cafaro and Rieke, 2010). However, the latter spontaneous excitatory/inhibitory input fluctuations have been simultaneously recorded either in the same cell or in different cells, making it impossible to estimate the contribution of global population activity fluctuations to the overall excitation/inhibition balance.

Spiking Models of SOCs

The simplicity and analytical tractability of rate models make them appealing to theoretical studies such as ours. One may worry, however, that some fundamental aspects of collective dynamics are being overseen when spiking events are reduced to their probabilities of occurrence, i.e., to rate variables. To verify our results, we embedded a SOC in a standard balanced spiking network, in which millions of randomly assigned synapses connect two populations of excitatory and inhibitory neurons. The SOC structure was embodied by additional connections between subgroups of these neurons, each containing on the order of tens of spiking cells. The resulting network displayed simultaneous firing rate and spiking variability (Churchland and Abbott, 2012), thus phenomenologically similar to the networks of Litwin-Kumar and Doiron (2012) and Ostojic (2014). However, slow rate fluctuations in SOCs arise from a completely different mechanism. The sea of random synapses in our network induces strong excitatory and inhibitory inputs to single cells that cancel each other on average, leaving large subthreshold fluctuations in membrane potential and therefore irregular spiking whose variability is mostly "private" to each neuron. This feature is common to all traditional balanced network models (van Vreeswijk and Sompolinsky, 1996; Brunel, 2000; Vogels et al., 2005; Renart et al., 2010). On the level of subgroups of neurons, this source of variability is not entirely lost to averaging: although all the cells in a given subgroup n fire at the same rate r_n at any given time, receiver neurons in another subgroup m will only "sense" a noisy sample estimate r_n of this rate, because n connects onto m through a finite number of synapses. Now, because the connectivity between subgroups is strong, but stabilized, this intrinsic source of noise (the "residual" $\xi_n = r_n - r_n$) is continuously amplified into large, structured firing rate fluctuations on the level of subgroups. The underlying mechanism is the same as for the rate model, i.e., balanced amplification of noise (Murphy and Miller, 2009), with the notable difference that the noise in the spiking network is intrinsically generated (the external excitatory drive that each neuron receives was chosen constant here to make this point).

In order to match the timescale of the rate transients in our spiking SOC to those in the data of Churchland et al. (2012), we assumed that the structured SOC synapses had slower time constants than the random synapses. Functional segregation of fast/slow synapses in the cortex has been reported in the visual cortex (Self et al., 2012) and could also be motivated by recent experiments in which the distance from soma along the dendritic arbor was shown to predict the magnitude of the NMDA component in the corresponding somatic postsynaptic potentials (PSPs) (Branco and Häusser, 2011). Thus, distal synapses tend to evoke slower PSPs than proximal synapses. It is in fact an interesting and testable prediction of our model that distal synapses are actively recruited in the motor cortex during movement preparation and generation. Finally, pilot simulations suggest that this separation of timescales, although necessary to obtain realistically long movement-related activity, is not a requirement for the emergence of large transients, which could indeed be obtained with a single synaptic time constant of ~10 ms (not shown).

Conclusions

In summary, we have shown that specific, recurrent inhibition is a powerful means of stabilizing otherwise unstable, complex circuits. The resulting networks are collectively excitable and display rich transient responses to appropriate stimuli that resemble the activity recorded in the motor cortex (Churchland et al., 2012) on both single-neuron and populations levels. We found that SOCs can be used as "spring-loaded motor engines" to generate complicated and reliable movements. The intriguing parallels to the detailed balance of excitatory and inhibitory inputs in cortical neurons, as well as to recent theories that apply specifically to the visual cortex (Ozeki et al., 2009; Murphy and Miller, 2009), suggest cortical-wide relevance for this class of neuronal architectures.

EXPERIMENTAL PROCEDURES

Network Setup and Dynamics

Single-neuron dynamics followed Equation 1, which we integrated using a standard fourth-order Runge-Kutta method. Following Rajan et al. (2010), we used the gain function

$$g(x) = \begin{cases} r_0 \tanh[x/r_0] & \text{if} \quad x<0 \\ (r_{\max}-r_0)\tanh[x/(r_{\max}-r_0)] & \text{if} \quad x \geq 0 \end{cases} \tag{2}$$

with baseline firing rate r_0 = 5 Hz and maximum rate $r_{\max}$ = 100 Hz (Figure 3E). Unless indicated otherwise, the input $\mathbf{I}(t) = \xi(t) + \mathbf{S}(t)$ included a noise term $\xi(t)$, which we modeled as an independent Ornstein-Uhlenbeck process for each neuron, with time constant τ_ξ = 50 ms. We set the variance of these processes to $\sigma_0^2(\tau + \tau_\xi)/\tau_\xi$, such that, in the limit of very weak synaptic connectivity, the firing rate of each cell in the network fluctuated around baseline with a standard deviation σ_0^2 = 0.2 Hz.

In order to "prepare" the network and drive its activity **x** into a specific steady-state pattern $\mathbf{a}_k$ (Figures 4 and 5), we delivered a slow ramping input to each cell during ongoing activity. This input was delivered as vector $\mathbf{S}(t) = R(t)\,\mathbf{P}_k$, where $R(t)$ denotes the ramp activation of the input pool k and P_k are the projection weights from pool k onto the motor network (Figures 1B and 1C). The ramp $R(t)$ had a slow exponential rise with time constant 400 ms beginning with the target cue at $t = -1$ s., followed by a fast exponential decay with time constant 2 ms after the go cue. The projection weights were set to

$$\mathbf{P}_k = \mathbf{a}_k - \mathbf{W}\,g(\mathbf{a}_k) \tag{3}$$

in order to guarantee $\mathbf{x}(t = 0) \sim \mathbf{a}_k$.

In Figure 4B, the 27 arm reaching movements in Churchland et al. (2012) were modeled as 27 different initial conditions ($\mathbf{b}_1$, …, $\mathbf{b}_{27}$) for the SOC. We chose each vector $\mathbf{b}_k$ as a random linear combination of the SOC's first and second preferred initial conditions $\mathbf{a}_1$ and $\mathbf{a}_2$ (see below). More precisely, $\mathbf{b}_k = \sum_{c=\{1,2\}} \mathrm{s}_{kc}\, \mathrm{z}_{kc}\, \mathbf{a}_c$ where the s_{kc}'s were random signs and the z_{kc}'s were drawn uniformly between 0.5 and 1.

Preferred Initial States

To find the preferred initial conditions of the SOC, we restricted ourselves to the linear regime in which $\Delta\mathbf{r}_i \sim x_i$. To quantify the response evoked by some unit-norm initial condition $\Delta\mathbf{r}(t = 0) \equiv \mathbf{a}$, we defined the "energy" $E(\mathbf{a})$ of the response as

$$E(\mathbf{a}) = \frac{2}{\tau}\int_0^\infty \|\Delta\mathbf{r}(t)\|^2 \mathrm{d}t \tag{4}$$

also assuming that the network dynamics run freely without noise [$\xi(t)$ = 0]. Here $2/\tau$ is a normalizing factor such that E = 1 for an unconnected network ($\mathbf{W}$ = 0), irrespective of the (unit-norm) initial condition **a** (in which case $\|\Delta\mathbf{r}(t)\|^2 = \exp(-2t/\tau)$). Because the SOC is linearly stable, E is finite, in the sense that any initial condition is bound to decay (exponentially) after a sufficiently long period of time.

The "best" input direction is then defined as the initial condition $\mathbf{a}_1$ that maximizes $E(\mathbf{a})$. By iterating, we can define a collection $\mathbf{a}_1$, $\mathbf{a}_2$, …, $\mathbf{a}_N$ of N orthogonal input states that each maximize the evoked energy within the subspace orthogonal to all previous best input directions. In the linear regime, this maximization can be performed analytically (Supplemental Information). Note that in the linear regime, $E(\mathbf{a}_k) = E(-\mathbf{a}_k)$. In the nonlinear network, this needs not be the case, and in Figures 3C and 3D we resolved this sign ambiguity by picking the sign that evoked most energy.

Construction of the SOC Architecture

Random connectivity matrices of size $N = 2M$, with M positive (excitatory) columns and M negative (inhibitory) columns, were generated as in Hennequin et al. (2012) with connectivity density p = 0.1. Non-zero excitatory (respectively inhibitory) weights were set to $w_0/\sqrt{N}$ (respectively $-\gamma w_0/\sqrt{N}$), where $w_0^2 = 2R^2/(p(1-p)(1+\gamma^2))$ and R is the desired spectral radius *before* stability optimization (Rajan and Abbott, 2006).

To generate a SOC, we generated such a random connectivity matrix with R = 10, producing unstable, deeply chaotic network behavior. After the creation of the initial **W**, all excitatory connections remained fixed. To achieve robust linear stability of the dynamics, we refined the inhibitory synapses to minimize the "smoothed spectral abscissa" (SSA) of **W**, a relaxation of the spectral abscissa (the largest real part in the eigenvalues of **W**) that—among other advantages—leads to tractable optimization (Vanbiervliet et al., 2009). In short, inhibitory weights followed a gradient descent on the SSA subject to three constraints. First, we kept the inhibitory weights inhibitory, i.e., negative. Second, we enforced a constant ratio between the average magnitude of the inhibitory weights, and its excitatory counterpart (γ = 3, cf. Discussion). Third, the density of inhibitory connections was restricted to less than 40%, to yield realistically sparse connectivity structures. This constrained gradient descent usually converged within a few hundred iterations. All details can be found in Supplemental Information.

Analysis of Rotational Dynamics

The plane of projection of Figure 4D was found with jPCA, a dynamical variant of principal component analysis used to extract low-dimensional rotations from multidimensional time series (Churchland et al., 2012). Given data of the form ($\mathbf{y}(t)$,$\mathrm{d}\mathbf{y}(t)/\mathrm{d}t$), jPCA fits (through standard least-squares regression) a linear oscillatory model of the form $\mathrm{d}\mathbf{y}/\mathrm{d}t = \mathbf{M}_{\text{skew}}\,\mathbf{y}(t)$, where $\mathbf{M}_{\text{skew}}$ is a skew-symmetric matrix, therefore one with purely imaginary eigenvalues. The two leading eigenvectors of the best-fitting $\mathbf{M}_{\text{skew}}$ (associated with the largest conjugate pair of imaginary eigenvalues) define the plane in which the trajectory rotates most strongly.

Here we computed the jPC projection exactly as prescribed in Churchland et al. (2012). Our model data consisted of the population responses $\Delta\mathbf{r}(t)$ during the first 200 ms following the go cue for each of our 27 initial conditions, sampled in 1 ms time steps. Note that the temporal derivatives are directly given by Equation 1, except in the spiking network (see below) where we estimated those derivatives using a finite-difference approximation. To make sure that the jPC projection captures enough of the data variance, that is, that the observed rotational dynamics (if any) are significant, the data were first projected down to the top six standard principal components (as in Churchland et al., 2012).

Muscle Activation through Linear Readouts

In Figure 5, a single pair of muscle readouts was learned from 200 training trials (100 trials for each of the "snake" and "butterfly" movements). We assumed the following linear model:

$$\mathbf{z}_t = (\mathbf{m}_1;\mathbf{m}_2)^{\mathrm{T}}\,\Delta\mathbf{r}_t + \mathbf{b} + \varepsilon_t \tag{5}$$

where $\mathbf{z}_t$ (size 2) denotes the vector of target muscle activations at discrete time t, $\Delta\mathbf{r}_t$ is the vector of momentary deviation from baseline firing rate in the network (size N), and ε_t is the vector of residual errors (size 2). The readout weights (column vectors $\mathbf{m}_1$ and $\mathbf{m}_2$) are parameters that we optimized through simple least-squares regression, together with a pair of biases **b**. The snake (respectively butterfly) target trajectory was made of 58 points (respectively 26 points), equally spaced in time over 500 ms following the go cue. Those points defined the discrete time variable t in Equation 5, and the activity vector $\Delta\mathbf{r}_t$ was sampled accordingly for each movement.

Spiking Network Simulations

We simulated a network of 15,000 neurons, composed of 12,000 excitatory and 3,000 inhibitory neurons, divided into 200 subgroups of excitatory neurons and 200 subgroups of inhibitory neurons, which can be interpreted as the "rate units" we have focused on until here.

Single-Neuron Model

Single cells were modeled as LIF neurons (e.g., Gerstner and Kistler, 2002, chapter 4) according to

$$\tau_{\mathrm{m}}\frac{\mathrm{d}V_{\mathrm{m}}^{(i)}}{\mathrm{d}t} = -V_{\mathrm{m}}^{(i)} + V_{\text{rest}} + h_{\text{exc.}}^{(i)} + h_{\text{inh.}}^{(i)} + h_{\text{ext.}} \tag{6}$$

with τ_{m} = 20 ms and $V_{\text{rest}} = -70$ mV. Neuron i emitted a spike whenever $V_{\mathrm{m}}^{(i)}(t)$ crossed -55 mV from below. Following a spike, the voltage was reset to -60 mV and held constant for an absolute refractory period of 2 ms. The excitatory and inhibitory synaptic inputs, $h_{\text{exc.}}^{(i)}$ and $h_{\text{inh.}}^{(i)}$, were sums of alpha-shaped postsynaptic currents (PSCs) of the form $c[\exp(-t/\tau^{\text{decay}}) - \exp(-t/\tau^{\text{rise}})]$, where c is a synapse-type-specific scaling factor that regulates peak excitatory and inhibitory postsynaptic potential (PSP) amplitudes after further membrane integration through Equation (6) (Figure 7B). τ^{rise} was set to 1 ms, and τ^{decay} depended on the synapse type (see below).

Recurrent Synapses

Each neuron received input from 1,500 excitatory and 1,500 inhibitory network neurons. For 50% of those recurrent connections (750 excitatory and 750 inhibitory synapses), the presynaptic partner was drawn randomly and uniformly from the corresponding population (excitatory or inhibitory), providing a sea of unspecific, random synapses that was instrumental in maintaining the network in a regime of asynchronous and irregular firing. These connections were thought to target proximal dendritic zones and therefore to evoke fast PSCs (τ^{decay} = 10 ms). The other half of the network synapses were used to mirror the structure of the network of rate units described throughout the article and were therefore drawn according to probabilities jointly determined by (1) the subgroups that the pre- and postsynaptic neurons belonged to and (2) an optimized SOC matrix **W** of size 400 × 400 that described the connectivity between subgroups.

We first normalized the excitatory and inhibitory parts for each row of **W**, obtaining a matrix $\hat{\mathbf{w}}$ of connection *probabilities*. Then, for any cell i in group m ($1 \leq m \leq 400$, excitatory or inhibitory), each of 750 excitatory partners were chosen in two steps: first, a particular group n was picked with probability $\hat{\mathbf{w}}_{mn}$; and second, a presynaptic neuron was picked at random from this group n. We applied the same procedure to generate the second half of the inhibitory synapses (750 per neuron). These structured SOC connections were given a slower PSC decay time constant (τ^{decay} = 100 ms), and can be interpreted as targeting more distal dendritic parts.

Sample PSPs are shown in Figure 7B for all four types of synapses. The ratio between excitatory and inhibitory synaptic efficacies was set to achieve a stable background firing state of 5 Hz. Note that because of the amplifying behavior of SOCs and the superlinear nature of the input-output function of LIF neurons, the network ended up with a mean of 6 Hz instead (Figure 7C).

Each neuron also received a constant positive external input current $h_{ext.}$ that was set to the mean current a cell would receive from 5,000 independent Poisson sources at 5 Hz with fast synapses. We boiled this input down to its mean to motivate that the slow, seemingly stochastic rate fluctuations we observed in the spiking SOC (Figure 8A) did not require any external source of noise.

Generation of W

SOC matrices for spiking networks were generated in a similar manner as described above for rate-based networks, except for a few simple variations to account for the effective gains of the excitatory and inhibitory synaptic pathways between subgroups. These details are described in Supplemental Information.

Control Random Network

The random network used for comparison in Figure 8 was identical in every respect to the SOC, except that presynaptic partners for slow synapses were drawn completely randomly (there was no notion of neuronal subgroups).

Simulations were custom-written in OCaml and parallelized onto eight cores following the strategy developed in Morrison et al. (2005), taking advantage of a finite axonal propagation delay which we set to 0.5 ms. We used simple Euler integration of Equation (6) with a time step of 0.1 ms.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2014.04.045.

ACKNOWLEDGMENTS

This research was supported by the Swiss National Science Foundation grant 200020 13287 (Coding Characteristics), ERC grant 268689 (MultiRules), and European FP7 grants 215910 (BIOTACT) and 269921 (BrainScaleS). Additionally, G.H. was supported by a Swiss National Science Foundation Postdoctoral fellowship PBELP3.146554, and T.P.V. was supported by the European Community's Seventh Framework Marie Curie International Reintegration grant 268436 and by a Sir Henry Dale Wellcome Trust and Royal Society research fellowship. We thank D. Rezende, A. Seeholzer, and F. Zenke for useful discussions and M. Lengyel for pointing us to Churchland et al. (2012).

Accepted: April 14, 2014
Published: June 18, 2014

REFERENCES

Afshar, A., Santhanam, G., Yu, B.M., Ryu, S.I., Sahani, M., and Shenoy, K.V. (2011). Single-trial neural correlates of arm movement preparation. Neuron *71*, 555–564.

Ahmadian, Y., Rubin, D.B., and Miller, K.D. (2013). Analysis of the stabilized supralinear network. Neural Comput. *25*, 1994–2037.

Ames, K.C., Ryu, S.I., and Shenoy, K.V. (2014). Neural dynamics of reaching following incorrect or absent motor preparation. Neuron *81*, 438–451.

Amit, D.J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb. Cortex *7*, 237–252.

Branco, T., and Häusser, M. (2011). Synaptic integration gradients in single cortical pyramidal cell dendrites. Neuron *69*, 885–892.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J. Comput. Neurosci. *8*, 183–208.

Cafaro, J., and Rieke, F. (2010). Noise correlations improve response fidelity and stimulus encoding. Nature *468*, 964–967.

Churchland, M.M., and Abbott, L.F. (2012). Two layers of neural variability. Nat. Neurosci. *15*, 1472–1474.

Churchland, M.M., and Shenoy, K.V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. J. Neurophysiol. *97*, 4235–4257.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Ryu, S.I., and Shenoy, K.V. (2010). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? Neuron *68*, 387–400.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Dayan, P., and Abbott, L.F. (2001). Theoretical neuroscience. (Cambridge: MIT Press).

Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E., and Froemke, R.C. (2010). Developmental sensory experience balances cortical excitation and inhibition. Nature *465*, 932–936.

Froemke, R.C., Merzenich, M.M., and Schreiner, C.E. (2007). A synaptic memory trace for cortical receptive field plasticity. Nature *450*, 425–429.

Ganguli, S., Huh, D., and Sompolinsky, H. (2008). Memory traces in dynamical systems. Proc. Natl. Acad. Sci. USA *105*, 18970–18975.

Gerstner, W., and Kistler, W.M. (2002). Spiking neuron models: Single neurons, populations, plasticity. (New York: Cambridge University Press).

Goldman, M.S. (2009). Memory without feedback in a neural network. Neuron *61*, 621–634.

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. Nature *493*, 97–100.

Hennequin, G., Vogels, T.P., and Gerstner, W. (2012). Non-normal amplification in random balanced neuronal networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *86*, 011909.

Hoerzer, G.M., Legenstein, R., and Maass, W. (2014). Emergence of complex computational structures from chaotic neural networks through reward-modulated Hebbian learning. Cereb. Cortex *24*, 677–690.

Kriener, B., Tetzlaff, T., Aertsen, A., Diesmann, M., and Rotter, S. (2008). Correlations and population dynamics in cortical networks. Neural Comput. *20*, 2185–2226.

Kullmann, D.M., Moreau, A.W., Bakiri, Y., and Nicholson, E. (2012). Plasticity of inhibition. Neuron *75*, 951–962.

Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat. Neurosci. *16*, 925–933.

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat. Neurosci. *15*, 1498–1505.

Luz, Y., and Shamir, M. (2012). Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. PLoS Comput. Biol. *8*, e1002334.

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput. *14*, 2531–2560.

Mariño, J., Schummers, J., Lyon, D.C., Schwabe, L., Beck, O., Wiesing, P., Obermayer, K., and Sur, M. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. Nat. Neurosci. *8*, 194–201.

Morrison, A., Mehring, C., Geisel, T., Aertsen, A.D., and Diesmann, M. (2005). Advancing the boundaries of high-connectivity network simulation with distributed computing. Neural Comput. *17*, 1776–1801.

Murphy, B.K., and Miller, K.D. (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron *61*, 635–648.

Okun, M., and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat. Neurosci. *11*, 535–537.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat. Neurosci. *17*, 594–600.

Ozeki, H., Finn, I.M., Schaffer, E.S., Miller, K.D., and Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. Neuron *62*, 578–592.

Rajan, K., and Abbott, L.F. (2006). Eigenvalue spectra of random matrices for neural networks. Phys. Rev. Lett. *97*, 188104.

Rajan, K., Abbott, L.F., and Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *82*, 011903.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The asynchronous state in cortical circuits. Science *327*, 587–590.

Self, M.W., Kooijmans, R.N., Supèr, H., Lamme, V.A., and Roelfsema, P.R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. Proc. Natl. Acad. Sci. USA *109*, 11031–11036.

Shenoy, K.V., Kaufman, M.T., Sahani, M., and Churchland, M.M. (2011). A dynamical systems view of motor preparation: implications for neural prosthetic system design. Prog. Brain Res. *192*, 33–58.

Sompolinsky, H., Crisanti, A., and Sommers, H.J. (1988). Chaos in random neural networks. Phys. Rev. Lett. *61*, 259–262.

Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron *63*, 544–557.

Terauchi, A., and Umemori, H. (2012). Specific sets of intrinsic and extrinsic factors drive excitatory and inhibitory circuit formation. Neuroscientist *18*, 271–286.

Tsodyks, M.V., Skaggs, W.E., Sejnowski, T.J., and McNaughton, B.L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. J. Neurosci. *17*, 4382–4388.

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science *274*, 1724–1726.

Vanbiervliet, J., Vandereycken, B., Michiels, W., Vandewalle, S., and Diehl, M. (2009). The smoothed spectral abscissa for robust stability optimization. SIAM J. Optim. *20*, 156–171.

Vogels, T.P., and Abbott, L.F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nat. Neurosci. *12*, 483–491.

Vogels, T.P., Rajan, K., and Abbott, L.F. (2005). Neural network dynamics. Annu. Rev. Neurosci. *28*, 357–376.

Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science *334*, 1569–1573.

Wang, X.J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J. Neurosci. *19*, 9587–9603.

Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442–446.

Neuron, Volume *82*

Supplemental Information

Optimal Control of Transient Dynamics in Balanced Networks Supports Generation of Complex Movements

Guillaume Hennequin, Tim P. Vogels, and Wulfram Gerstner

Optimal control of transient dynamics in balanced networks supports generation of complex movements

Supplemental Information

Guillaume Hennequin, Tim P. Vogels and Wulfram Gerstner

Contents

1 Supplemental Figures

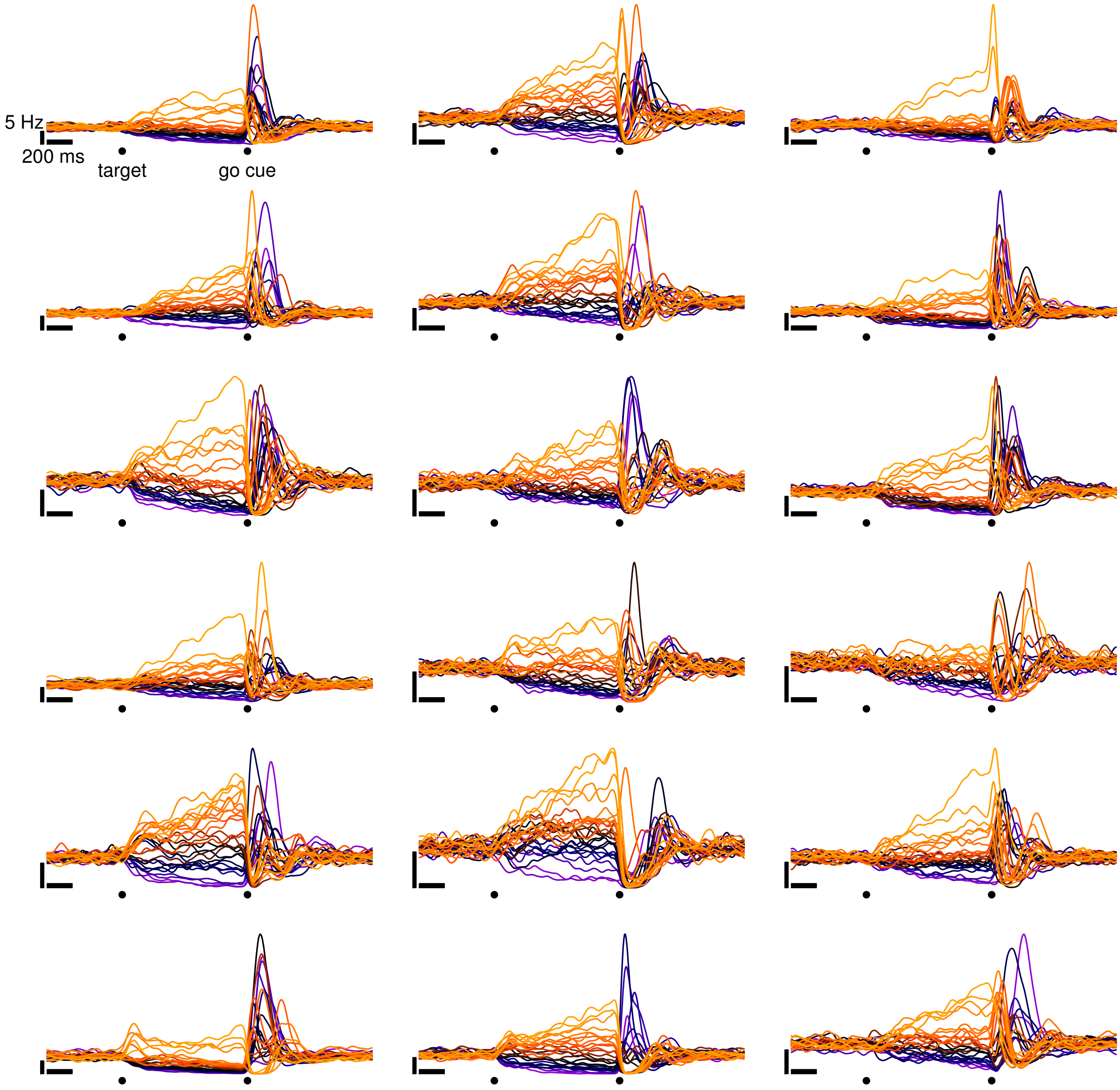

Figure S1, related to Figure 7 of the main text – Diversity of single-cell transient responses in the spiking SOC. Shown here are trial-averaged responses (1'000 trials) of the first cell for each of the first 18 subgroups (one panel per subgroup). The preparatory period drives the network in one of 27 random combinations (color coded) of its top 2 preferred initial conditions. The color code is meaningless and not consistent across panels, but only intended to ease visualization. Rate traces were smoothed using a Gaussian kernel of width 20 ms.

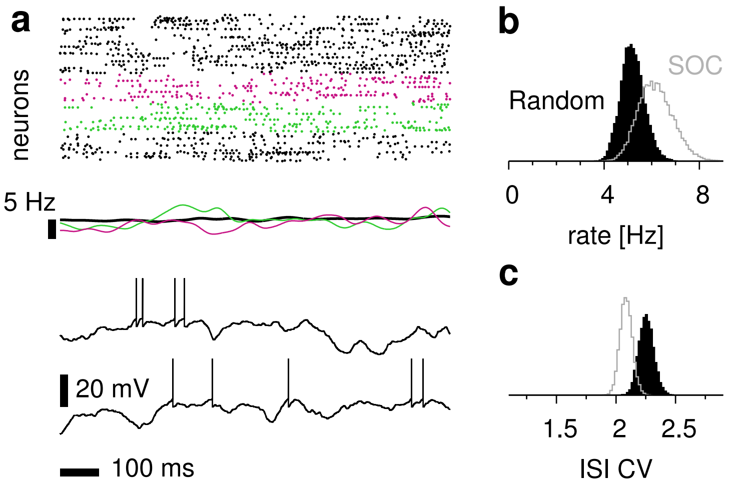

Figure S2, related to Figure 8 of the main text – Control random spiking network. **(a)** Top: raster plot of spiking activity in the control network. Middle: average momentary network activity (black), and activities of the two colored subgroups. Note that the concept of subgroups is absent from the connectivity; subgroups are completely arbitrary here, and only meant to provide a comparison with Figure 8 of the main manuscript. Bottom: voltage traces for two randomly chosen cells. **(b)** Distribution of average firing rates in the random network (black) and in the SOC (gray). Averate rates were computed from a 1'000 second-long simulation. **(c)** Same as in (b), for the CVs of the ISI distributions.

2 Supplementary Movie caption

Supplementary Movie The eigenvalues of the original random balanced network are shown as fixed gray dots in the complex plane. Roughly 50% of them lie to the right of the critical line defined by $\mathsf{Re}(\lambda) = 1$ (purple line), thus corresponding to unstable eigenmodes. The moving dots show the evolution of the eigenvalues as the inhibitory connectivity is being refined to produce a SOC. Each frame corresponds to one step of the gradient descent (cf. Experimental Procedures). Purple dots indicate unstable eigenmodes.

3 Supplemental Experimental Procedures

We present here additional information regarding both the theoretical underpinnings of the stabilization algorithm we have used to build Stability-Optimized Circuits (SOCs), and all details of its implementation. We also give a complete description of our spiking network model.

3.1 Preferred initial states

Given a network with connectivity matrix $\mathbf{W}$ and network dynamics of the form

$$\tau \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}\Delta\mathbf{r}(\mathbf{x}, t) + \mathbf{S}(t) + \boldsymbol{\xi}(t) \tag{1}$$

(cf. Equation (1) in the main text), we ask: what initial condition $\mathbf{a} \equiv \Delta\mathbf{r}(t = 0)$ will give rise to the strongest transient response? How "strong" a transient response is is quantified by the energy $\mathcal{E}(\mathbf{a})$, defined in Equation (4) of the main text, which we repeat here for convenience:

$$\mathcal{E}(\mathbf{a}) = \frac{2}{\tau} \int_0^\infty \|\Delta\mathbf{r}(t)\|^2 \, \mathrm{d}t \tag{2}$$

We restrict the analysis to the linear regime in which $\Delta r_i \simeq x_i$, and we assume that the network dynamics run freely without noise ($\boldsymbol{\xi}(t) = 0$) from $t = 0$ on. In Equation (2), $2/\tau$ is a normalizing factor such that $\mathcal{E}(\mathbf{a}) = 1$ for an unconnected network ($\mathbf{W} = 0$), irrespective of the (unit-norm) initial condition $\mathbf{a}$ (in which case $\|\Delta\mathbf{r}(t)\|^2 = \exp(-2t/\tau)$). Since the SOC is linearly stable, $\mathcal{E}$ is finite, in the sense that any initial condition is bound to decay (exponentially) after sufficiently long periods of time.

The "best" input direction is then defined as the initial condition $\mathbf{a}_1$ that maximizes $\mathcal{E}(\mathbf{a})$. In the linear regime, this maximization can be performed analytically. Indeed, Equation (2) can be rewritten as

$$\begin{aligned} \mathcal{E}(\mathbf{a}) &= \mathbf{a}^\top \left[\frac{2}{\tau} \int_0^\infty \mathrm{d}t \; e^{(\mathbf{W}^\top - \mathbf{I})\, t/\tau} \; e^{(\mathbf{W} - \mathbf{I})\, t/\tau} \right] \mathbf{a} & (3) \\ &\equiv \mathbf{a}^\top \mathbf{Q}\, \mathbf{a} & (4) \end{aligned}$$

where $(\cdot)^\top$ denotes the matrix transpose. The last equality defines $\mathbf{Q}$ as the matrix integral inside square brackets. $\mathbf{Q}$ is a symmetric, positive-definite matrix, so its principal eigenvector is precisely the initial condition $\mathbf{a}_1$ that maximizes the evoked energy, which is then given by the corresponding principal eigenvalue of $\mathbf{Q}$. In fact, the full eigenbasis of $\mathbf{Q}$, ranked in decreasing order of the associated eigenvalues, defines a collection $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N)$ of N orthogonal input states that each maximize the evoked energy within the subspace orthogonal to all previous best input directions (cf. Figure 3a in the main manuscript). Again, the eigenvalues are the corresponding evoked energies. We use this energy formalism again below to explain the optimal inhibitory stabilization algorithm. Note that in the linear regime, $\mathbf{a}_k$ and $-\mathbf{a}_k$ evoke the exact same energy. In the non-linear network, this needs not be the case, and we resolve this sign ambiguity by picking the condition that evoked most energy. Note also that $\mathbf{Q}$ is the solution to the Lyapunov equation

$$(\mathbf{W} - \mathbf{I})^\top \mathbf{Q} + \mathbf{Q}\,(\mathbf{W} - \mathbf{I}) = -2 \cdot \mathbf{I} \tag{5}$$

which is easily solved numerically (Bartels and Stewart, 1972), e.g. using the Matlab function `lyap`.

3.2 Construction of the SOC architecture

Random connectivity matrices of size $N = 2M$, with M positive (excitatory) columns and M negative (inhibitory) columns, were generated as in Hennequin et al. (2012) with connectivity density $p = 0.1$. Non-zero excitatory (resp. inhibitory) weights were set to $w_0/\sqrt{N}$ (resp. $-\gamma w_0/\sqrt{N}$), where $w_0 = R/\sqrt{p(1-p)(1+\gamma^2)/2}$ and R is the desired spectral radius (before stability optimization, Rajan and Abbott (2006)).

To generate a SOC, we generated such a random connectivity matrix with $R = 10$, producing unstable, chaotic network behavior. After the creation of the initial $\mathbf{W}$, all excitatory connections remained fixed. To achieve robust linear stability of the dynamics (i.e. quench chaos), we refined the inhibitory synapses to minimize the "smoothed spectral abscissa" $\tilde{\alpha}_\epsilon(\mathbf{W})$, a relaxation of the spectral abscissa (the largest real part in the eigenvalues of $\mathbf{W}$) that – among other advantages – leads to tractable optimization (Vanbiervliet et al., 2009). We describe it in more details below, but in short, the inhibitory weights followed a gradient descent on $\tilde{\alpha}_\epsilon(\mathbf{W})$, subject to three constraints. First, we constrained the inhibitory weights to remain inhibitory, i.e. negative. Second, we enforced a constant ratio between the average magnitude of the inhibitory weights, and its excitatory counterpart ($\gamma = 3$). Third, the density of inhibitory connections was restricted to less than 40%. This constrained gradient descent usually converged within a few hundred iterations (cf. also our online Supplemental Movie).

In more technical terms, the smoothed spectral abscissa can be introduced by way of considering, as above, the energy evoked through linear network dynamics by some initial condition $\mathbf{a}$:

$$\mathcal{E}(\mathbf{W}, \mathbf{a}) = \mathbf{a}^\top \mathbf{Q}(1)\mathbf{a} \tag{6}$$

where $\mathbf{Q}(s)$ is defined more generally as

$$\mathbf{Q}(s) = \frac{2}{\tau}\int_0^\infty e^{(\mathbf{W}^\top - s\mathbf{I})t/\tau}\, e^{(\mathbf{W}-s\mathbf{I})t/\tau}\, \mathrm{d}t \quad . \tag{7}$$

For the network to be stable, the energy in Equation (6) must remain finite for any initial condition $\mathbf{a}$. Since $\mathbf{Q}(1)$ is positive definite, its maximum eigenvalue is smaller than the sum of its eigenvalues. Thus, for any $\mathbf{a}$, $\mathcal{E}(\mathbf{W}, \mathbf{a})$ is upper-bounded by the trace of $\mathbf{Q}(1)$. Consequently, if $\operatorname{trace}[\mathbf{Q}(1)] < \epsilon^{-1}$ for some given $\epsilon > 0$, then the energy evoked by any $\mathbf{a}$ is less than ϵ^{-1}, so the network dynamics of Equation (1) are guaranteed to be stable[1]. In a network that is not (yet) linearly stable, we ask: how far must the system be "shifted", $\mathbf{W} \mapsto \mathbf{W} - s\mathbf{I}$ (cf. Equation (7)), for $\operatorname{trace}[\mathbf{Q}(s)]$ to become smaller than ϵ^{-1}? The ϵ-smoothed spectral abscissa answers this question. Mathematically, $\tilde{\alpha}_\epsilon(\mathbf{W})$ is the unique root of $s \mapsto \operatorname{trace}[\mathbf{Q}(s)] - \epsilon^{-1}$, which is a monotonically decreasing function of s. If the shift s is smaller than the spectral abscissa $\alpha(\mathbf{W})$, some of the eigenvalues of $\mathbf{W} - s\mathbf{I}$ will have positive real parts, causing $\operatorname{trace}[\mathbf{Q}(s)]$ to diverge. The smoothed spectral abscissa $\tilde{\alpha}_\epsilon(\mathbf{W})$ is therefore necessarily greater than $\alpha(\mathbf{W})$, which means that stability can be achieved by minimizing $\tilde{\alpha}_\epsilon(\mathbf{W})$ instead of $\alpha(\mathbf{W})$. This is an advantageous replacement as, unlike α, $\tilde{\alpha}_\epsilon$ is a smooth function of the synaptic weights (Vanbiervliet et al., 2009).

The tractability of the approach stems from the computability of $\tilde{\alpha}_\epsilon(\mathbf{W})$ and its derivatives w.r.t $\mathbf{W}$. For any $s > \alpha(\mathbf{W})$, the matrix $\mathbf{Q}(s)$ defined in Equation (7) is known to be the solution to the following Lyapunov equation

$$(\mathbf{W} - s\mathbf{I})^\top \mathbf{Q}(s) + \mathbf{Q}(s)(\mathbf{W} - s\mathbf{I}) = -2 \cdot \mathbf{I} \tag{8}$$

Solving this equation numerically can be done efficiently (Bartels and Stewart, 1972). Knowing that $\operatorname{trace}[\mathbf{Q}(s)] - \epsilon^{-1}$ is a decreasing function of s, one can apply standard root-finding methods

[1] globally stable in the linear case, locally stable around background state otherwise

to identify $\tilde{\alpha}_\epsilon(\mathbf{W})$. Finally, Vanbiervliet et al. (2009) also provide its derivatives w.r.t the synaptic weights, needed to perform our gradient-based minimization of $\tilde{\alpha}_\epsilon$:

$$\frac{\partial\,\tilde{\alpha}_\epsilon(\mathbf{W})}{\partial \mathbf{W}} = \frac{\mathbf{Q}(\tilde{\alpha}_\epsilon)\mathbf{P}(\tilde{\alpha}_\epsilon)}{\text{trace}\left[\mathbf{Q}(\tilde{\alpha}_\epsilon)\mathbf{P}(\tilde{\alpha}_\epsilon)\right]} \tag{9}$$

where

$$\mathbf{P}(s) = 2\int_0^\infty e^{t(\mathbf{W}-s\mathbf{I})}\; e^{t(\mathbf{W}-s\mathbf{I})^\top}\, dt \tag{10}$$

solves a Lyapunov equation analogous to Equation (8):

$$\left(\mathbf{W}-s\mathbf{I}\right)\mathbf{P}(s) + \mathbf{P}(s)\left(\mathbf{W}-s\mathbf{I}\right)^\top = -2\cdot\mathbf{I} \tag{11}$$

The iterative constrained gradient descent on $\tilde{\alpha}_\epsilon(\mathbf{W})$ entails the following steps:

1. Compute the current value of the smoothed spectral abscissa $\tilde{\alpha}_\epsilon(\mathbf{W})$. This implies multiple iterations of a numerical root-finding method (e.g. bisection) on $s \mapsto \text{trace}[\mathbf{Q}(s)] - \epsilon^{-1}$. The main cost here is to solve Equation (8) for $\mathbf{Q}(s)$ numerically, but we show below that the present step can in fact be bypassed.

2. Solve Equation (8) and Equation (11) with $s = \tilde{\alpha}_\epsilon$ found in step 1. This gives matrices $\mathbf{Q}(\tilde{\alpha}_\epsilon)$ and $\mathbf{P}(\tilde{\alpha}_\epsilon)$, which must be multiplied to form the desired gradient (Equation (9)).

3. Move the inhibitory weights by a small amount in the direction of the negative gradient. That is, for every *existing* inhibitory synapse W_{ij} (only 40% of all possible inhibitory connections exist at any given time, see step 6), set $W_{ij} \leftarrow W_{ij} - \eta\,[\partial\tilde{\alpha}_\epsilon/\partial\mathbf{W}]_{ij}$. Here $\eta = 10$ is a learning rate.

4. Enforce the negativity constraint (inhibitory weights remain inhibitory), by clipping all positive inhibitory weights to zero

5. Enforce the constraint that inhibition must be $\gamma = 3$ times stronger than excitation on average. This is done by writing the connectivity matrix $\mathbf{W}$ block-wise as

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}^{E\to E} & \mathbf{W}^{I\to E} \\ \mathbf{W}^{E\to I} & \mathbf{W}^{I\to I} \end{pmatrix}, \tag{12}$$

   and multiplicatively rescaling both inhibitory blocks by $-\gamma\overline{\mathbf{W}}^{E\to E}/\overline{\mathbf{W}}^{I\to E}$ and $-\gamma\overline{\mathbf{W}}^{E\to I}/\overline{\mathbf{W}}^{I\to I}$ respectively, where $\overline{\mathbf{W}}^{X\to Y}$ denotes the average over all matrix elements. This step is not necessary for stability optimization, but is essential to make sure that the high correlation of excitatory and inhibitory input currents that emerges from optimization is not overwhelmed by the baseline correlation contributed by shared population fluctuations (see Discussion in the main text, and Renart et al. (2010); Hennequin et al. (2012)).

6. Enforce the maximum density of inhibitory connections, by removing any existing W_{ij} that step 4 may have set to zero, and replacing it by a new connection W_{ik} (which does not exist yet) where the presynaptic inhibitory neuron k is chosen randomly. Set the strength of these new connections to zero initially. Again, this constraint is not strictly required, but adds to the biological plausibility of the resulting connectivity. Note that this step ensures that the number of "existing" inhibitory synapses per neuron is left unchanged. The desired number of allowed synapses must therefore be decided before starting this iterative procedure; here we "tag" 40% of all inhibitory synapses as "existing", picking the 10% of non-zero synapses that are already present in the initial chaotic network and completing the remaining 30% with randomly chosen synapses, whose weights are initially set to zero.

Steps 1 through 6 are then repeated until convergence of the spectral abscissa (Supplemental Movie).

Choice of ϵ In this procedure, ϵ modulates the distance between the spectral abscissa α and its upper bound $\tilde{\alpha}_\epsilon$: if ϵ is decreased, $\tilde{\alpha}_\epsilon$ becomes a tighter upper bound to the spectral abscissa. In pilot studies, we realized that stability could be reached much faster if ϵ was set to decrease progressively during the course of the gradient descent. Empirically, it seemed a good idea to keep the ratio $\tilde{\alpha}_\epsilon/\alpha$ constant, and to adjust ϵ in every iteration to meet this need. Mathematically, this means that the cost function $(\mathbf{W} \mapsto \tilde{\alpha}_\epsilon(\mathbf{W}))$ keeps moving, but it becomes a progressively tighter upper bound on α, and crucially, we no longer need to compute $\tilde{\alpha}_\epsilon$ in step 1! We thus capitalized on this observation and set $\tilde{\alpha}_\epsilon(\mathbf{W}) = C\alpha(\mathbf{W})$ in every iteration, with $C = 1.5$, empirically found to be a good choice. Note that this automatically constrains ϵ to a value of $1/\text{trace}[\mathbf{Q}(C\alpha)]$, where $\mathbf{Q}(\cdot)$ is defined in Equation (7). Steps 2 to 6 are then performed as prescribed above. Note that computationally, the cost is still of order N^3, but the large multiplicative constant implied by the iterative root-finding method of Step 1 is eliminated. Note also that Step 2 requires solving Equation (8) and Equation (11), for which we need to compute a Schur decomposition of $\mathbf{W}$ only once (Bartels and Stewart, 1972). As a byproduct, the Schur decomposition also returns the spectral abscissa at no further cost, so α needs not be computed separately. The above simplified procedure is very effective, except for one small detail. As $\tilde{\alpha}_\epsilon$ gets closer to α (which is bound to happen as learning progresses, since $\tilde{\alpha}_\epsilon = C\alpha$ and $\tilde{\alpha}_\epsilon$ decreases), it also becomes increasingly non-smooth as a function of $\mathbf{W}$. For its gradient to remain well-behaved, we therefore kept $\tilde{\alpha}_\epsilon$ some safe margin away from α, by setting $\tilde{\alpha}_\epsilon$ to $C\alpha$ or $\alpha + B$ in every iteration, whichever was the greatest. We used $B = 0.2$.

3.3 Generation of SOC matrices for spiking networks

SOC matrices for spiking networks were generated in a similar manner as described above for rate-based networks, except for a few simple variations to account for the effective gains of the excitatory and inhibitory synaptic pathways between subgroups. To calculate these gains, we isolated two subgroups of spiking neurons, n (sender) and m (receiver), and assumed (self-consistently) that the rest of the network fired with Poisson statistics at 5 Hz. We then numerically estimated the effect of small firing rate deviations in group n, Δ_n, on the firing rate response Δ_m of the receiver (thus linearizing the joint dynamics of subgroups around the background state). Due do the E/I balance, this is a roughly linear function of the connection probability $\widehat{W}_{mn}$ from group n to group m:

$$\frac{\mathrm{d}\Delta_m}{\mathrm{d}\Delta_n} \simeq \begin{cases} \beta_{\text{exc.}}\,\widehat{W}_{mn} & \text{if } n \text{ is excitatory} \\ \beta_{\text{inh.}}\,\widehat{W}_{mn} & \text{if } n \text{ is inhibitory} \end{cases} \tag{13}$$

Using a simple least-squares fit, we extracted the coefficients of proportionality $\beta_{\text{exc.}} \simeq 5.2$ and $\beta_{\text{inh.}} \simeq -21.9$. These constants were then used to rescale the excitatory and inhibitory parts of the normalized matrix of connection probabilities, $\widehat{\mathbf{W}}$, yielding an effective connectivity matrix $\mathbf{W}$ such that linear dynamics of the form

$$\tau \frac{\mathrm{d}\Delta\mathbf{r}}{\mathrm{d}t} = -\Delta\mathbf{r} + \mathbf{W}\Delta\mathbf{r} + \text{noise} \tag{14}$$

approximated the joint dynamics of our spiking subgroups (which is shown in Figure 8b of the main manuscript). This suggests that the threshold process of spiking in combination with the finite size of the randomly connected subgroups leads to apparent stochasticity that can, to a high degree of accuracy, be treated as noise around an effective population rate of each subgroup (see also Brunel (2000)).

We used the smoothed spectral abscissa method described above to modify the connections probabilities $\widehat{W}_{mn}$ $(200 < n \leq 400)$ that involved inhibitory presynaptic subgroups to achieve stability in an initially unstable, random sparse matrix. In each gradient step, the inhibitory part of each row of $\widehat{\mathbf{W}}$ was re-normalized to sum to 1, to preserve the probabilistic interpretation of $\widehat{\mathbf{W}}$. Second, we used a straightforward application of the chain rule to propagate the gradient of the smoothed spectral abscissa of $\mathbf{W}$ to $\widehat{\mathbf{W}}$ given the value of $\beta_{\text{inh.}}$.

We used the effective weight matrix $\mathbf{W}$ to compute the preferred initial states of the spiking SOC, akin to the rate-based case with $\Delta r_i \propto x_i$ (i.e. assuming linear dynamics among subgroups given by Equation (14)). During the preparatory period, each LIF neuron received a ramping input current that depended only on the subgroup it belonged to.

4 References

Bartels, R. H. and Stewart, G. W. (1972). Solution of the matrix equation AX+XB=C. *Communications of the ACM*, 15:820–826.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, 8:183âĂŞ–208.

Hennequin, G., Vogels, T. P., and Gerstner, W. (2012). Non-normal amplification in random balanced neuronal networks. *Phys. Rev. E*, 86:011909.

Rajan, K. and Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett.*, 97:188104.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. (2010). The asynchronous state in cortical circuits. *Science*, 327:587.

Vanbiervliet, J., Vandereycken, B., Michiels, W., Vandewalle, S., and Diehl, M. (2009). The smoothed spectral abscissa for robust stability optimization. *SIAM Journal on Optimization*, 20:156–171.