

Statistical mechanics of phase space partitioning in large-scale spiking neuron circuits

Maximilian Puelma Touzel^{1,2,3,*} & Fred Wolf^{1,2,4}

¹*Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany*

²*Bernstein Center for Computational Neuroscience, Göttingen, Germany*

³*Laboratoire de Physique Théorique, ENS-PSL Research University, Paris, France*

⁴*Kavli Institute for Theoretical Physics, University of California Santa Barbara, Santa Barbara, USA*

Corresponding Author:

Maximilian Puelma Touzel
Laboratoire de Physique Théorique
Ecole Normale Supérieure
24 rue Lhomond
75231 Paris Cedex 05
Paris, France
Ph. +49 551 51 76 420
e-mail: `puelma@lpt.ens.fr`

Abstract

Synaptic interactions structure the phase space of the dynamics of neural circuits and constrain neural computation. Understanding how requires methods that handle those discrete interactions, yet few exist. Recently, it was discovered that even random networks exhibit dynamics that partitions the phase space into numerous attractor basins. Here we utilize this phenomenon to develop theory for the geometry of phase space partitioning in spiking neural circuits. We find basin boundaries structuring the phase space are pre-images of spike-time collision events. Formulating a statistical theory of spike-time collision events, we derive expressions for the rate of divergence of neighboring basins and for their size distribution. This theory reveals that the typical basin diameter grows with inhibitory coupling strength and shrinks with the rate of spike events. Our study provides an analytical and generalizable approach for dissecting how connectivity, coupling strength, single neuron dynamics and population activity shape the phase space geometry of spiking circuits.

Key words: *neuronal circuits, dynamics of networks, disordered systems, basins of attraction, high dimensional systems, pulse-coupled systems, sequence generation*

Computing devices, whether natural or artificial, perform their function by finely orchestrated state changes of internal dynamical variables. In nervous systems these dynamical variables are physico-chemical states of nerve cells and synapses that connect them into complex networks called neural circuits. The causal dependencies arising from the synaptic interactions between cells greatly extend the space of functions computable by the circuit, beyond that of single neurons.

Mathematical models of neural circuits have been formulated in two fundamentally distinct ways¹. Most synaptic interactions in the brain are driven by sparsely-fired nerve impulses, called spikes, each lasting only a millisecond. In spiking neural network models this fundamental granularity of neuronal interactions is explicitly represented: all interactions depend on a discrete set of spike event times. Alternatively, continuous variable models for neural circuit dynamics are formulated by assuming that a frequency of nerve impulse generation, the firing rate, represents the information-encoding variable causally relevant for neural circuit computation. Firing rate models have been commonly used to model neural circuits², theoretically study their dynamics^{3,4} and learning⁵⁻⁸, and are the basis of spectacular advances in artificial computing systems⁹. Statistical physics has played a role in this development, e.g. in clarifying the disordered phase space organization¹⁰.

From a dynamical systems perspective, attractor states and their basins of attraction play a fundamental role in theories of neural computation. While analogous in some cases^{11,12}, however, rate models are not equivalent to temporally coarse-grained versions of spiking neural networks, even if they are closely matched in structure¹³. Moreover, low firing rates (not much more than 1 Hz) in the cerebral cortex¹⁴ make it hard to imagine how continuous rate variables associated to single neurons could provide a causally accurate description on behavioral time scales (hundreds of milliseconds). Developing theory for spiking networks may well require a dedicated approach. The absence of relevant averages and even a tractable ensemble of spiking trajectories, however, has thus far limited statistical approaches. Methods to design them¹⁵ or to theoretically dissect the associated phase space organization are only starting to emerge.

Recently it has been discovered that, with dominant inhibition, even randomly wired networks partition their phase space into a complex set of basins of attraction, termed *flux tubes*^{16,17}. Here we utilize this setting to develop a statistical theory of phase space partitioning in spiking neural circuits. We first present a simulation study of flux tubes, uncovering their shape and revealing it is structured by a spike time collision event. Formulating these events, we then derive the conditions for and rate of the mutual divergence of neighboring tubes. Our main calculation is the derivation of the distribution of flux tube sizes, which we obtain from statistics of these events by leveraging the random connectivity to average over the disorder. Our analytical approach provides a transparent method to determine how coupling strength, connectivity, single neuron dynamics and population activity combine to shape the phase space geometry of spiking neural circuits.

Methods

We study a tractable instance of the inhibition-dominated regime of neural circuits. N neurons are connected by an Erdős-Rényi graph with adjacency matrix $A = (A_{mn})$. $A_{mn} = 1$ denotes a connection from neuron n to m , realized with probability, $p = K/N$. The neurons' membrane potentials, $V_n \in (-\infty, V_T]$, are governed by Leaky Integrate-and-Fire (LIF) dynamics,

$$\tau \dot{V}_n(t) = -V_n(t) + I_n(t) , \quad (1)$$

for $n \in \{1, \dots, N\}$. Here, τ is the membrane time constant and $I_n(t)$ the synaptic current received by neuron n ; when V_n reaches a threshold, $V_T = 0$, neuron n 'spikes', and V_n is reset to $V_R = -1$. At the spike time, t_s , the spiking neuron, n_s , delivers a current pulse of strength J to its $\mathcal{O}(K)$ postsynaptic neurons, $\{m | A_{mn_s} = 1\}$, (s indexes the spikes in the observation window). The total synaptic current is

$$I_n(t) = I_{\text{Ext}} + \tau J \sum_s A_{nn_s} \delta(t - t_s) , \quad (2)$$

where $I_{\text{Ext}} > 0$ is a constant external current and $J < 0$ is the recurrent coupling strength. An $\mathcal{O}(1/\sqrt{K})$ -scaling of J is chosen to maintain finite current fluctuations at large K and implies that the external drive is balanced by the recurrent input. As a consequence, firing in this network is robustly asynchronous and irregular^{18–21}. Setting $I_{\text{Ext}} = \sqrt{K}I_0$, with $I_0 > 0$, and $J = -J_0/\sqrt{K}$ with $J_0 > 0$, the corresponding stationary mean-field equation for the population-averaged firing rate, $\bar{\nu}$, is¹⁷

$$\bar{\nu} = \frac{I_0}{J_0\tau} + \mathcal{O}\left(\frac{1}{\sqrt{K}}\right). \quad (3)$$

It is convenient to map the voltage dynamics to a pseudophase representation^{17,22}, $\vec{\phi}(t)$, with

$$\phi_n(t) = \frac{\tau}{T_{\text{free}}} \ln \left[\frac{I_{\text{Ext}} - V_R}{I_{\text{Ext}} - V_n(t)} \right], \quad (4)$$

where T_{free} is the oscillation period of a neuron driven only by I_{Ext} . $\phi_n(t)$ evolves linearly in time,

$$\dot{\phi}_n(t) = T_{\text{free}}^{-1}, \quad (5)$$

between spike events, *i.e.* $t \notin \{t_s\}$, and undergoes shifts given by the phase response curve, $Z(\phi)$, across input spike times where ϕ is the state at spike reception. In the large- K limit, T_{free} and $Z(\phi)$ simplify to

$$T_{\text{free}} \approx \frac{\tau}{I_{\text{Ext}}} = (\sqrt{K}J_0\bar{\nu})^{-1}, \quad (6)$$

$$Z(\phi) \approx -d\phi + \text{const.}$$

$$\text{with } d := \frac{|J|}{I_{\text{Ext}}} = (K\bar{\nu}\tau)^{-1}, \quad (7)$$

respectively (see Supplemental Methods for details). The differential phase response, $\frac{d}{d\phi}Z = -d$, is essential for the strongly dissipative nature of the collective dynamics. For $J = 0$, the dynamics (equation (5)) would preserve phase space volume. This volume, however, is strongly contracted by spikes received in the post-synaptic neurons. Consider trajectories from a small ball of initial conditions as they emit the same future spike. The ball of phases at this spike contracts by a factor $1 - d$ along each of the K dimensions of the subspace spanned by the post-synaptic neurons. The volume thus contracts by $(1 - d)^K \rightarrow e^{\lambda_{\text{inh}}}$ per spike, for $K \gg 1$, with exponential rate,

$$\lambda_{\text{inh}} \approx -Kd. \quad (8)$$

$\lambda_{\text{inh}} < 0$ is responsible for the linear stability of the dynamics given by equations (1) and (2), first shown in Refs.^{16,22}.

Phase-space partitioning

The phase space volume taken up by an ensemble of nearby trajectories at a given spike is contracted at the spike's reception. Larger phase space volumes, however, are not uniformly contracted but torn apart, with the pieces individually contracted and overall dispersed across the entire traversed phase space volume. The elementary phenomenon is illustrated in Fig. 1.

We define the *critical perturbation strength*, ϵ^* , as the flux tube's extent out from a given state $\vec{\phi}_0$ on the equilibrated trajectory, $\vec{\phi}_t$, and in a given orthogonal perturbation direction, $\vec{\xi}$,

$$\epsilon^*(\vec{\phi}_0, \vec{\xi}) := \sup \left\{ \epsilon \mid \lim_{t \rightarrow \infty} D_t(\epsilon) = 0 \right\}. \quad (9)$$

Here, $D_t(\epsilon)$ is the 1-norm distance,

$$D_t(\epsilon) := \frac{1}{N} \sum_{n=1}^N |\phi_{n,t} - \phi_{n,t}(\epsilon)|, \quad (10)$$

between $\vec{\phi}_t$ and the perturbed trajectory, $\vec{\phi}_t(\epsilon)$, evolving freely from the perturbed state, $\vec{\phi}_0(\epsilon) := \vec{\phi}_0 + \epsilon \vec{\xi}$ (reference time $t = 0$ and $||\vec{\xi}|| = 1$; see Supplemental Methods for details). ϵ^* is the largest value below which $D_t(\epsilon)$ vanishes in time. D_t initially decays exponentially, but for a supercritical perturbation, $\epsilon > \epsilon^*$, there exists a *divergence event time*, $t^* > 0$, defined and obtained as the time at which a sustained divergence in D_t begins (see Fig. 1a).

A 2D cross-section of the phase space around $\vec{\phi}_0$ (Fig. 1b) reveals that the locations of these critical perturbations form lines which intersect to form polygon-shaped basin boundaries. Before developing a theory for this phase space organization (caricatured in Fig. 1c), we first analyze two main features of the geometry of a flux tube: the punctuated exponential decay of its cross-sectional volume and the exponential separation of neighboring tubes.

Punctuated geometry of flux tubes

As expected from the typical phase space volume contraction (see equation (8)), we find along a simulated trajectory that the orthogonal phase space volume enclosed by the local flux tube exhibits exponential decay. This decay, however, is punctuated by blowup events. Figure 2a displays the spiking activity produced by the typical trajectory, $\vec{\phi}_t$. The neighborhood around $\vec{\phi}_t$ over a time window is visualized in a *folded representation* using a fixed, 2D projection of the phase space (Fig. 2b and [Supplementary Video](#); see Supplemental Information for construction details). The basin of attraction surrounding $\vec{\phi}_t$ (Fig. 2b) consists of lines which remain fixed between spike times. Across spike times, new lines appear and existing lines disappear. At irregular intervals breaking up time windows of exponential contraction, large abrupt blowup events take the boundary away from the center trajectory (Fig. 2b, c), producing jumps in the area enclosed by the boundary. It is important to note that these events do not mean that the evolving phase space volume from an ensemble of states contained in the tube would expand. Such volumes only contract and converge to the same asymptotic trajectory. The basin of attraction itself, however, does not exclusively contract with time. In fact, it should on average maintain a typical size.

The blowup events typically coincide with a divergence event time, t^* (Fig. 1a), in some perturbation direction. Two such coincidences are visible in Fig. 2c,d. We conclude that the local basin at any time extends out in phase space until the perturbed trajectory approaches the pre-image of a divergence event occurring at a future time. Flux tube shape is then determined by the statistics of such events.

Tube boundary and divergence

We analyzed a set of divergence events from simulations. We find that a collision of a pair of spikes constitutes the elementary event triggering the divergence of the perturbed trajectory. These pairs, hereon called *susceptible* spike pairs, were generated by connected pairs of neurons. Moreover, a perturbation-induced collision of a susceptible spike pair generated an abrupt spike time shift in one or both of the connected neurons' spike times. We found that the nature of the spike time shift depends on the motif by which the two neurons connect. We denote the *backward-connected* pair motif $n_{s^*} \leftarrow n_{s'}$, where s^* , the *decorrelation event index*, is the spike index of the earlier of the pair (note that $t^* \equiv t_{s^*}$), and $s' > s^*$ labels the later spike in the pair.

For $\epsilon \lesssim \epsilon^*$, the presynaptic spike time, $t_{s'=s^*+1}$, is advanced with increasing ϵ relative to the postsynaptic spike time, t_{s^*} , until the two spikes collide (see Fig. 3). At collision ($\epsilon = \epsilon^*$), the pulsed inhibition and the rate of approach to voltage threshold cause an abrupt delay of t_{s^*} by Δt_{jump} . Using equation (3), we obtain

$$\Delta t_{\text{jump}} = \tau \ln [1 + d] \approx \tau d = (K\bar{v})^{-1} \quad (11)$$

for $d \ll 1$. Further details and the other two motifs (forward-connected and symmetrical) are discussed in the Supplemental Notes.

For each spike in the network sequence, the rate of its susceptible spike partners is

$$\lambda_{\text{sus}} = p/\overline{\Delta t} = K\bar{\nu}, \quad (12)$$

where $\overline{\Delta t} = (N\bar{\nu})^{-1}$ is the average distance between successive spikes. Since $\Delta t_{\text{jump}} \approx \lambda_{\text{sus}}^{-1}$, the spike time of neuron n_{s^*} is shifted forward typically as far as its next nearest susceptible partner spike. Thus, one collision event will typically induce another in at least one of the $\mathcal{O}(K)$ neurons to which the involved pair of neurons are presynaptic. A cascade of collision events then follows with near certainty (see Supplemental Notes for details).

The shift in t_{s^*} by Δt_{jump} is carried forward to all future spike times of n_{s^*} , so that n_{s^*} becomes a source of collision events. The total collision rate is then λ_{sus} multiplied by the number of source neurons, which approximately increments with each collision in the cascade. Averaging over realizations of the cascade (reference time $t^* = 0$), the average number of collisions, \bar{m} , grows as $\frac{d}{dt}\bar{m} = \lambda_{\text{sus}}\bar{m}$. Finally, since each collision produces a jump in distance of equal size, we obtain the pseudoLyapunov exponent, $\lambda_p = \lambda_{\text{sus}}$ from its implicit definition, $\bar{D}_t = \bar{D}_0 \exp[\lambda_p t]$ (see Supplemental Notes), as the exponential rate at which flux tubes diverge.

Statistical theory of flux tube diameter

The geometry of a flux tube is captured by the *flux tube indicator function*, $\mathbb{1}_{\text{FT}}(\epsilon) = \Theta(\epsilon^*(\vec{\phi}_0, \vec{\xi}) - \epsilon)$, evaluated across network states, $\vec{\phi}_0$, of its contained attracting trajectory and perturbation directions, $\vec{\xi}$. Using the Heaviside function, $\Theta(x)$, $\mathbb{1}_{\text{FT}}(\epsilon) = 1$ for perturbations remaining in the tube ($\epsilon < \epsilon^*$), and 0 otherwise. The average of $\mathbb{1}_{\text{FT}}(\epsilon)$ over $\vec{\phi}_0$ and $\vec{\xi}$,

$$\hat{S}(\epsilon) = [\mathbb{1}_{\text{FT}}(\epsilon)]_{\rho(\vec{\phi}_0, \vec{\xi})}, \quad (13)$$

is the *survival function*: the probability that an ϵ -sized perturbation does not lead to a divergence event later in the perturbed trajectory, i.e. $\epsilon < \epsilon^*$, and is formally defined as $\hat{S}(\epsilon) := 1 - \int_0^\epsilon \rho(\epsilon^*) d\epsilon^*$, with $\rho(\epsilon^*)$ the transformed density over ϵ^* . $\hat{S}(0) = 1$ and decays to 0 as $\epsilon \rightarrow \infty$. The scale of this decay defines the typical flux tube size. Calculating $\hat{S}(\epsilon)$ requires two steps: firstly, establishing a tractable representation of $\epsilon^*(\vec{\phi}_0, \vec{\xi})$ and secondly, performing the average in equation (13). Both of these in general pose intricate problems. However, as we will see next, both substantially simplify when generic properties of the asynchronous, irregular state are taken into account.

Perturbed spike intervals are obtained using the spike time deviations, $\delta t_s(\epsilon) := t_s(\epsilon) - t_s(0)$, $s = 1, 2, \dots$,

$$\Delta t_s(\epsilon) = t_s(\epsilon) - t_{s-1}(\epsilon) = \Delta t_s(0) + \delta t_s(\epsilon) - \delta t_{s-1}(\epsilon), \quad s \geq 2. \quad (14)$$

In a linear approximation we find,

$$\delta t_s(\epsilon) \approx -C a_s \epsilon, \quad (15)$$

where $C = \frac{T_{\text{free}}}{\sqrt{N}}$ converts network phase deviation to spike time deviation and a_s is a dimensionless susceptibility that depends on the adjacency matrix, $A = (A_{mn})$, derivatives of the phase response curve evaluated at the network states at past spike times, $\{\vec{\phi}_{s'} = \vec{\phi}_{t_{s'}}\}$ for $s' < s$, and the perturbation direction $\vec{\xi}$ (see Supplemental Notes for its derivation). Substituting equation (15) into equation (14) gives

$$\Delta t_s(\epsilon) \approx \Delta t_s(0) - C \Delta a_s \epsilon, \quad (16)$$

with $\Delta t_s = \Delta t_s(0)$. Note that $\Delta t_s(\epsilon)$ can have a zero, i.e. a spike time collision only when $\Delta a_s = a_s - a_{s-1} > 0$.

To obtain the scaling behavior of the flux tube geometry it is sufficient to examine the statistics of flux tube borders using the corresponding divergence events generated by collisions of backward-connected susceptible spike pairs in the perturbed trajectory (Fig. 3). In these cases, the perturbation strength $\epsilon \rightarrow \epsilon^*$ as the network spike interval $\Delta t_{s^*}(\epsilon) \rightarrow 0$ for $A_{n_{s^*} n_{s^*+1}} = 1$. In fact, the latter condition serves in these cases as an implicit definition of ϵ^* and s^* .

According to equation (16), $\hat{S}(\epsilon)$ in principle depends on the adjacency matrix, $A = (A_{mn})$, of the network realization. Removing this dependence by averaging over the ensemble of graphs, $P_A((A_{mn}))$, simplifies the calculation of the survival function,

$$S(\epsilon) = \left[\hat{S}(\epsilon) \right]_{P_A((A_{mn}))} . \quad (17)$$

Evaluating the right-hand side of equation (17) using the perturbed spike intervals, linearized in ϵ , requires knowledge of the joint probability density of all variables present in equation (16),

$$\rho_T = \rho(\{\Delta a_s\}, \{\Delta t_s\}, M, \vec{\phi}_0 | \vec{\xi}, (A_{mn})) \rho(\vec{\xi}) P_A((A_{mn})), \quad (18)$$

where we have chosen the perturbation direction, $\vec{\xi}$, to be statistically independent of the state, $\vec{\phi}_0$, being perturbed at $t = 0$. Here, the unperturbed spike pattern is represented by two random variables: M , the number of spikes in the time interval $[0, T]$ after the perturbation, and $\{\Delta t_s\}$, the set of all $M - 1$ inter-spike intervals in this window. It is well understood that in the large-system limit in a sparse graph, $1 \ll K \ll N$, the currents driving individual neurons in the network converge to independent, stationary Gaussian random functions²³. For low average firing rates, this implies that the pattern of network spikes $(M, \{\Delta t_s\})$ resembles a Poisson process²⁴. Furthermore, the susceptibility becomes state-independent in this limit. Neglecting the weak dependence between the distribution of network spike patterns and $A = (A_{mn})$, the full density, ρ_T (equation (18)), approximately factorizes,

$$\rho_T \sim P_{A_{mn}}(A_{mn}) P_T(M) \prod_{s=2}^M \rho_t(\Delta t) 2\Theta(\Delta a_s) \rho_a(\Delta a_s), \quad (19)$$

with distribution of a single adjacency matrix element, $P_{A_{mn}}(A_{mn} = 1) = p$, $P_{A_{mn}}(A_{mn} = 0) = 1 - p$, count distribution of spikes in the observation window, $P_T(M)$, and distribution of single inter-spike interval $\rho_t(\Delta t)$. The latter is exponential with rate $\overline{\Delta t}$. With these approximations (see Supplementary Notes for details), all dependencies on the distribution of perturbation direction are mediated by the susceptibilities, $\{\Delta a_s\}$. For any isotropic $\rho(\vec{\xi})$ having finite-variance, $\rho_a(\Delta a_s)$ has zero mean and standard deviation proportional to $\exp[\frac{\lambda_{\text{inh}}}{N} s]$, with the average contraction rate per neuron, $\frac{\lambda_{\text{inh}}}{N} = -\frac{Kd}{N} = -pd$, due to the inhibition. The factor $2\Theta(\Delta a_s)$ places support only positive values of Δa_s as required.

As ρ_T factorizes, so does $S(\epsilon)$,

$$S(\epsilon) = \lim_{T \rightarrow \infty} \left[\prod_{s=1}^M S_s(\epsilon) \right]_{P(M)} = \prod_{s=1}^{\infty} S_s(\epsilon), \quad (20)$$

where $S_s(\epsilon)$ is the probability that a perturbation of strength ϵ does not lead to a collision event involving the s^{th} spike. With the above simplifications,

$$S_s(\epsilon) = \left[\Theta(\Delta t - C\Delta a_s \epsilon)^{A_{mn}} \right]_{\rho_t(\Delta t) \rho_a(\Delta a_s) P_{A_{mn}}(A_{mn})}. \quad (21)$$

Evaluating equation (21) (see Supplementary Notes for the derivation), we find

$$S_s(\epsilon) = 1 + p(\text{Erfcx}[x_s] - 1), \quad (22)$$

where $x_s = \frac{C}{\Delta t} e^{\frac{\lambda_{\text{inh}}}{N} s} \epsilon \leq \epsilon/\sqrt{p}$, and $\text{Erfcx}[x] = e^{x^2} \left(1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \right)$ is the scaled complementary error function. $\text{Erfcx}[x_s] - 1 \approx -x_s$ for $\epsilon/\sqrt{p} \ll 1$, so that finally

$$S(\epsilon) \approx \prod_{s=1}^{\infty} \left(1 - C\lambda_{\text{sus}} e^{\frac{\lambda_{\text{inh}}}{N} s} \epsilon \right),$$

where we have identified $\lambda_{\text{sus}} = \frac{p}{\Delta t}$. Employing the logarithm and $C\lambda_{\text{sus}} \epsilon \propto \sqrt{p}\epsilon \ll 1$,

$$S(\epsilon) \approx \exp \left[\frac{\lambda_{\text{sus}}}{\frac{\lambda_{\text{inh}}}{N}} C \epsilon \right] = \exp \left[-\frac{\epsilon}{\epsilon^*} \right] \quad (23)$$

with

$$\begin{aligned} \bar{\epsilon}^* &= \frac{1}{C} \frac{\left| \frac{\lambda_{\text{inh}}}{N} \right|}{\lambda_{\text{sus}}} = \frac{\sqrt{N}}{T_{\text{free}}} \frac{pd}{p/\Delta t} = \sqrt{N\Delta t} \frac{d}{T_{\text{free}}} \\ &= \sqrt{N\Delta t} \frac{|J|/I_{\text{Ext}}}{\tau/I_{\text{Ext}}} = \sqrt{N\Delta t} \frac{|J|}{\tau} = \frac{\sqrt{N}}{\tau} \frac{|-J_0/\sqrt{K}|}{N\bar{\nu}} \end{aligned} \quad (24)$$

$$\bar{\epsilon}^* = \frac{J_0}{\sqrt{KN\bar{\nu}\tau}}, \quad (25)$$

where we have used equations (6) and (7) in the second line and note the cancellation of p and I_{Ext} . Equation (23) shows for $1 \ll K \ll N$ that the basin diameter, ϵ^* , is exponentially distributed and so completely determined by its characteristic scale, $\bar{\epsilon}^*$ (equation (25)), that is smaller for larger network size, higher average in-degree, higher population activity, and larger membrane time constant, τ . $\bar{\epsilon}^*$ grows, however, with the synaptic coupling strength, J_0 . In Fig. 4b, we see quantitative agreement in simulations between the definition of $\hat{S}(\epsilon)$ (equation (13) using the definition of ϵ^* , equation (9)) and its approximate microstate parametrization (equations (20), (21)). These also confirm the exponential form of our reduced expression (equations (23), (25)) and a scaling dependence on J_0 (Fig. 4c). The latter holds until J is no longer of size $\mathcal{O}(1/\sqrt{K})$. The other scalings were reported in Ref. ¹⁷. A derivation of only the characteristic scaling of $\bar{\epsilon}^*$, but not depending on the Poisson spiking assumption, is given in the Supplemental Notes.

The geometry of phase space partitioning

Figure 5 presents the phase space organization of these spiking circuits as we have revealed it, replacing the caricature of Fig. 1c. For a perturbation made to a stable trajectory, the geometry of the determining collision event is shown in Figure 5a, in a folded representation. The pre-images of this event determine the flux tube boundary back to the perturbation. Our results also provide a global, i.e. non-folded geometry of the partitioning (Fig. 5b(left)). Susceptible spike collisions are edges of the N -dimensional unit hypercube of phases where the corresponding voltages of two connected neurons both approach threshold. The Poincare section obtained by projecting the dynamics orthogonal to the trajectory (since no motion exists orthogonal to this subspace) then reveals the intrinsic partition. Here, the polygon basin boundaries arise as the pre-images of the projections of susceptible edges lying nearby the trajectory at future spike times (Fig. 5b(right)).

Discussion

We have developed a theory of phase space partitioning in spiking neural circuits, exemplified using the phenomenon of flux tubes. Importantly, the approach yields the dependence on various control parameters. We find the flux tube diameter contracts with the rate of volume contraction per neuron, $\lambda_{\text{inh}}/N = (N\bar{\nu}\tau)^{-1}$, due to the inhibition received across the post-synaptic subspace of each spike. This contraction is punctuated, however, by collision events between susceptible spikes, i.e. those from pairs of connected neurons, occurring at rate $\lambda_{\text{sus}} = (K\bar{\nu})^{-1}$ and across which the basin volume expands out to a pre-image of the next collision event. For some neighboring tube, this collision event sets off a cascade of such events with exponential rate, λ_{sus} that is responsible for their mutual divergence. Using these collision events to identify the spiking trajectories lying on flux tube boundaries, we were able calculate the size distribution of these basins. The average size is controlled by the ratio of these two exponential rates. Leaving out a factor converting shifts in spike time to shifts in state,

$$\bar{\epsilon}^* \propto \frac{|J|}{N\bar{\nu}} \equiv \frac{\text{stabilizing inhibitory coupling strength}}{\text{destabilizing rate of spikes}}.$$

The final scaling, $\bar{\epsilon}^* = J_0 / (\sqrt{KN}\bar{\nu}\tau)$, thus combines the contraction from the single neuron dynamics responsible for the dissipative dynamics, with the overall rate of spikes, which appears since each spike can be involved in a destabilizing collision event. Both contracting and expanding rates scale with the probability of connection, p , so we intuitively expect p to appear in $\bar{\epsilon}^*$ only implicitly through J and, reassuringly, p indeed cancels out.

Our framework motivates a variety of extensions. Our calculations can be performed for different disordered connectivity ensembles (e.g. correlated entries from annealed dilution processes²⁵ and structured second-order statistics²⁶), different activity regimes (e.g. non-Markovian spike interval processes²⁷), and different single neuron models (e.g. any threshold neuron with known phase response curve). We have applied the theory to an instability caused by abrupt changes in spike time due to an inhibitory input near voltage threshold, a scenario that can also be analyzed in neuron models with smooth thresholds (e.g. the rapid theta-neuron²⁸ that has the LIF neuron as a limit). The theory may also apply to other, as yet unknown instabilities involving spike collision events. Finally, while the linear stability of the dynamics precludes finite, asymptotic (Kolmogorov-Sinai) entropy production, the partition refinement picture we provide in Fig. 5b suggests a transient production of information about the perturbation on timescales of the order of the divergence event time, t^* . Making this connection to ergodic theory more precise is an interesting direction for future research.

Applying our approach in a relatively idealized context allowed for a tractable assessment of phase space organization. Despite its simplicity, however, the LIF neuron accurately captures many properties of cortical neurons, such as their dynamic response²⁹. We have also neglected heterogeneity in many properties. For instance, in contrast to the locally stable regime studied here, mixed networks of excitatory and inhibitory neurons can instead be conventionally chaotic³⁰. This chaos can nevertheless be suppressed in the ubiquitous presence of fluctuating external drive^{31,32} or with spatially-structured connectivity³³, suggesting a generality to locally stable dynamics and phase space partitioning in neural computation. Our approach, in particular the way we have quantified the ensemble of perturbed spiking trajectories, can inform formulations of local stability in these more elaborate contexts. Of particular interest are extensions where a macroscopic fraction of tubes remain large enough to realize encoding schemes tolerant of intrinsic and stimulus noise. For example, extensions to random dynamical systems^{34,35} could provide theoretical control over spiking dynamic variants of rate network-based learning schemes to generate stable, input-specific trajectories⁷.

Recent advances in experimental neuroscience have allowed for probes of the finite-size stability properties of cortical circuit dynamics call for *in vivo*. For example, simultaneous intra- and extra-cellular recordings in the whisker motion-sensing system of the rat reveal that the addition of a single spike makes a measurable impact on the underlying spiking dynamics of the local cortical area³⁶. Indeed, rats can be trained to detect perturbations to single spikes emitted in this area³⁷. Representative toy theories, such as the one we provide, can guide this work by highlighting the features of spiking neural circuits that contribute to phase space partitioning. The combined effort promises to elucidate the dynamical substrate for neural computation at the level at which the neuronal interactions actually operate.

Acknowledgements

M.P.T. would like to acknowledge discussions with Michael Monteforte, Sven Jahnke and Rainer Engelken. This work was supported by BMBF (01GQ07113, 01GQ08111, 01GQ0922, 01GQ1005B), GIF (906-17.1/2006), DFG (SFB 889), VW-Stiftung (ZN2632), and the Max Planck Society.

Author Contributions

M.P.T. conceived the project, developed the concepts, and wrote the manuscript. F.W. supervised the project, discussed the results and edited the manuscript.

Additional Information

The authors declare no competing financial interests.

1. Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, **334**, 1569–73, 2011.
2. Wilson, H. R. and Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, **12**, 1–24, 1972.
3. Sompolinsky, H., Crisanti, A., and Sommers, H. J. Chaos in random neural networks. *Physical Review Letters*, **61**, 259–262, 1988.
4. Kadmon, J. and Sompolinsky, H. Transition to chaos in random neuronal networks. *Physical Review X*, **5**, 1–28, 2015.
5. Hopfield, J. J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554–2558, 1982.
6. Sussillo, D. and Abbott, L. F. Generating Coherent Patterns of Activity from Chaotic Neural Networks. *Neuron*, **63**, 544–557, 2009.
7. Laje, R. and Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nature Neuroscience*, **16**, 925–933, 2013.
8. Brunel, N. Is cortical connectivity optimized for storing information? *Nature Neuroscience*, **19**, 749–755, 2016.
9. LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature*, **521**, 436–444, 2015.
10. Gardner, E. Optimal basins of attraction in randomly sparse neural network models. *Journal of Physics A: Mathematical and General*, **22**, 1969–1974, 1988.
11. Harish, O. and Hansel, D. Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Computational Biology*, **11**, e1004266, 2015.
12. Mastrogiuseppe, F. and Ostojic, S. Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *PLOS Computational Biology*, **13**, 1–40, 2017.
13. Engelken, R., Farkhooi, F., Hansel, D., van Vreeswijk, C., and Wolf, F. A reanalysis of “Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons”. *F1000Research*, **5**, 2043, 2016.
14. Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and Van Vreeswijk, C. On the distribution of firing rates in networks of cortical neurons. *J Neurosci*, **31**, 16217–26, 2011.
15. Memmesheimer, R. M., Rubin, R., Ölveczky, B., and Sompolinsky, H. Learning Precisely Timed Spikes. *Neuron*, **82**, 925–938, 2014.
16. Jahnke, S., Memmesheimer, R.-M., and Timme, M. Stable Irregular Dynamics in Complex Neural Networks. *Physical Review Letters*, **100**, 2–5, 2008.
17. Monteforte, M. and Wolf, F. Dynamic flux tubes form reservoirs of stability in neuronal circuits. , **2**, 041007, 2012.
18. van Vreeswijk, C. and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science (New York, N.Y.)*, **274**, 1724–6, 1996.
19. Brunel, N. and Hakim, V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, **11**, 1621–71, 1999.
20. Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. The asynchronous state in cortical circuits. *Science*, **327**, 587–590, 2010.

21. Barral, J. and D Reyes, A. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, **19** , 1690–1696, 2016.
22. Jin, D. Fast Convergence of Spike Sequences to Periodic Patterns in Recurrent Networks. *Physical Review Letters*, **89** , 1–4, 2002.
23. Tuckwell, H. *Introduction to Theoretical Neurobiology vols. 1 and 2*. Cambridge University Press, 1988.
24. Lindner, B. Superposition of many independent spike trains is generally not a Poisson process. *Physical Review E*, **73** , 1–4, 2006.
25. Boutent, M., Engels, A., Komodat, A., and Serneelst, R. Quenched versus annealed dilution in neural networks. *J. Phys. A: Math. Gen.*, **23** , 4643–4657, 1990.
26. Zhao, L., Beverlin, B., Netoff, T., and Nykamp, D. Q. Synchronization from second order network connectivity statistics. *Frontiers in computational neuroscience*, **5** , 28, 2011.
27. Schwalger, T., Droste, F., and Lindner, B. Statistical structure of neural spiking under non-poissonian or other non-white stimulation. *Journal of Computational Neuroscience*, **39** , 29–51, 2015.
28. Monteforte, M. *Chaotic Dynamics in Networks of Spiking Neurons in the Balanced State*. PhD thesis, 2011.
29. Brette, R. What Is the Most Realistic Single-Compartment Model of Spike Initiation? *PLOS Computational Biology*, **11** , e1004114, 2015.
30. Monteforte, M. and Wolf, F. Dynamical Entropy Production in Spiking Neuron Networks in the Balanced State. *Phys. Rev. Lett.*, **105(26)** , 1–4, 2010.
31. Molgedey, L., Schuchhardt, J., and Schuster, H. G. Suppressing chaos in neural networks by noise. *Physical Review Letters*, **69** , 3717–3719, 1992.
32. Goedeke, S., Schuecker, J., and Helias, M. Noise dynamically suppresses chaos in neural networks. , pages 1–5, 2016.
33. Rosenbaum, R. and Doiron, B. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X*, **4** , 1–9, 2014.
34. Lajoie, G., Thivierge, J., and Shea-Brown, E. Structured chaos shapes spike-response noise entropy in balanced neural networks. *Frontiers in computational neuroscience*, **8** , 2014.
35. Arnold, L. *Random Dynamical Systems*. Springer, 1998.
36. London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, **466** , 123–127, 2010.
37. Houweling, A. R. and Brecht, M. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, **451** , 65–8, 2008.

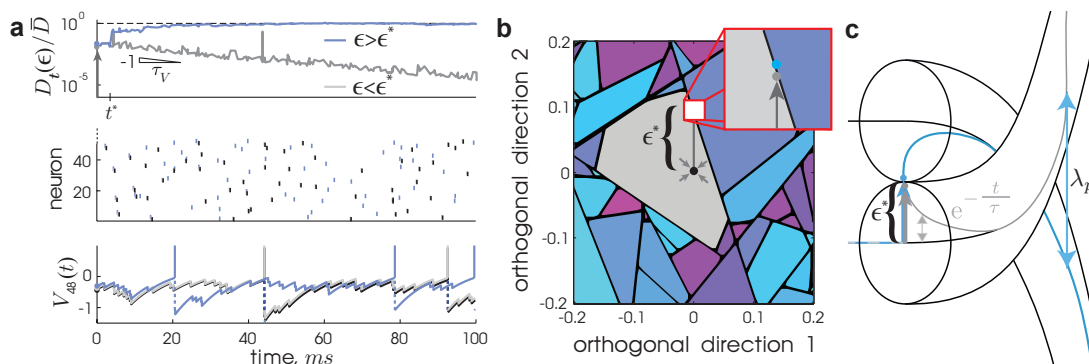


Figure 1 Finite-size perturbation instability and phase space partitioning in spiking networks. The three panels display the same slightly subcritical and supercritical perturbation of strength $\epsilon^* \pm \delta$, $\delta \gtrsim 0$, respectively, applied once at $t = 0$ and in a random direction away from an equilibrated trajectory. **(a)** Temporal responses of the system. *Top*: The corresponding distance time series, $D_t(\epsilon)$, between the perturbed and unperturbed trajectories (gray: sub-critical, blue: super-critical; arrows in all three panels indicate the respective the perturbation). The divergence of $D_t(\epsilon^* + \delta)$ begins at $t^* \approx 3$ ms, and saturates at the average distance between randomly chosen trajectories, \bar{D} (dashed line)¹⁷, while $D_t(\epsilon^* - \delta)$ only decays exponentially. *Middle*: The spike times as vertical ticks of the first 50 randomly labeled neurons from the network. The unperturbed sequence is shown in black. *Bottom*: The subthreshold voltage time course of an example neuron. The spike sequence and membrane potentials of the sub and supercritical trajectories decorrelate after t^* . **(b)** A 2D cross-section $(\delta\phi_1, \delta\phi_2)$ of the pseudophase representation of the phase space, orthogonal to and centered on the unperturbed trajectory from (a) at $t = 0$ (see also Ref. 15). The black dot at the origin indicates the latter, whose attractor basin is colored gray. The other colors distinguish basins in the local neighborhood. The two perturbed trajectories from (a) were initiated from $(\delta\phi_1, \delta\phi_2) = (0, \epsilon^* \pm \delta)$, respectively (shown as gray and blue dots, respectively, in the inset, in (a,Top and Bottom), and in (c)). **(c)** Schematic phase space caricature of two neighboring flux tubes with subcritical perturbations decaying on the order of the membrane time constant, τ , and typical basin diameter, ϵ^* . The pseudoLyapunov exponent, λ_p , is the rate at which neighboring tubes separate from each other (parameters: $N = 200$, $K = 50$, $\bar{\nu} = 10$ Hz, $\tau = 10$ ms, $J_0 = 1$).

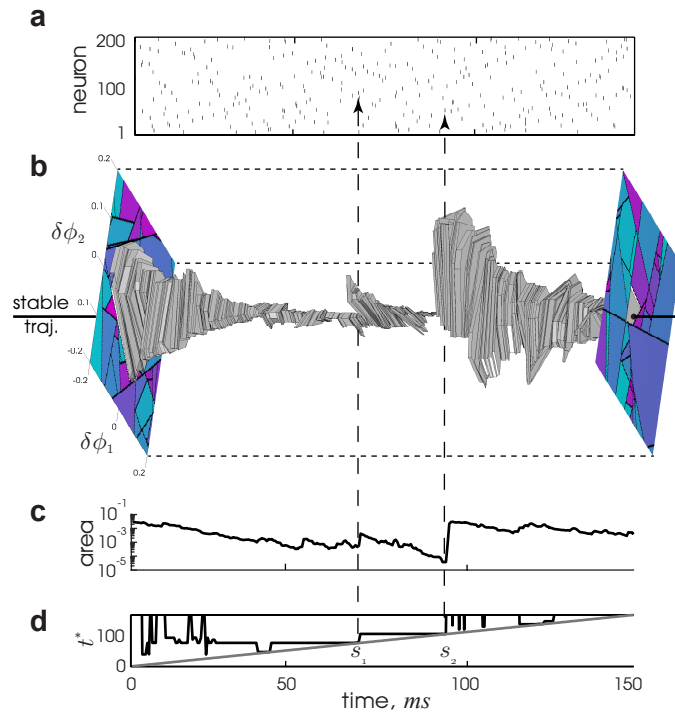


Figure 2 The basin boundary contracts towards and can blowup away from the stable trajectory within it. **(a)** Spike times from all neurons of the simulated trajectory, $\vec{\phi}_t$, in a 150 ms window. **(b)** 2+1D folded phase space volume, $(\delta\phi_1, \delta\phi_2, t)$, centered around $\vec{\phi}_t$ located at $(0, 0, t)$ (black line) and extended in two fixed, random directions, $\delta\phi_1$ and $\delta\phi_2$. The center tube is filled gray in this volume, and the two cross-sections, $(\delta\phi_1, \delta\phi_2, 0)$ and $(\delta\phi_1, \delta\phi_2, 150)$, are shown. **(c)** Cross-sectional area of the center tube from (b) versus time. The area decays exponentially but can undergo abrupt expansions at blow-up times, e.g. at spikes s_1 and s_2 (note the logarithmic scale on the ordinate). **(d)** The absolute time of the next divergence event, t^* (see Fig. 1a, top), versus time, for perturbations along $\delta\phi_1$. Note the step increase coincident with the blowup events seen in (b,c) (vertical, dashed lines). (Same parameters as Fig. 1.)

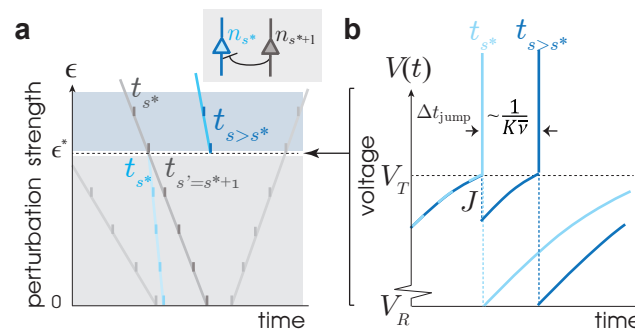


Figure 3 The collision of a susceptible spike pair causes an abrupt change in spike time. **(a)** A schematic illustration of the collision event for the backward-connected pair motif (shown in inset). For this motif, the interval vanishes as $\epsilon \rightarrow \epsilon^*$ from below. Perturbation strength, ϵ , is plotted versus time, where the timings of spikes for every perturbation strength are indicated as ticks on lines. The spike times shift continuously for $\epsilon < \epsilon^*$. As the next input spike time, $t_{s^*+1}(\epsilon^* - \delta)$, is advanced over $t_{s^*}(\epsilon^* + \delta)$ a discontinuous jump of size Δt_{jump} occurs in the spike time of the post-synaptic neuron, n_{s^*} (light to dark blue) from $t_{s^*}(\epsilon^* - \delta)$ to $t_{s' > s^*}(\epsilon^* + \delta)$, $\delta \gtrsim 0$. **(b)** Schematic illustration of the voltage of the n_{s^*} neuron versus time for $\epsilon^* \pm \delta$. The inhibitory kick of size $J = -J_0/\sqrt{K}$ (not shown to scale) delays the spike time by an amount $\Delta t_{\text{jump}} \sim (K\bar{v})^{-1}$.

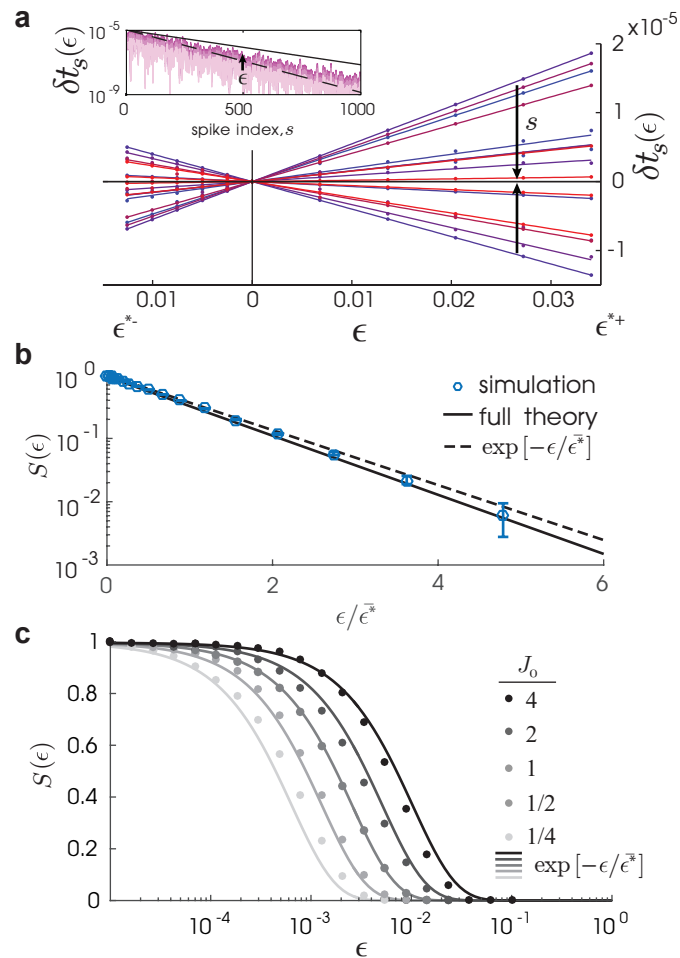


Figure 4 The flux tube indicator function, once expressed with microstate variables and averaged, gives the survival probability to remain in the containing flux tube. **(a)** Spike-time deviations, $\delta t_s(\epsilon)$ (dots), as a function of perturbation strength up to the positive and negative critical strength, ϵ^{*-} and ϵ^{*+} , respectively, for $s = 1, \dots, 15$ (colors) with their linear approximation (lines) given by equation (15). Inset: $\delta t_s(\epsilon)$ as a function of s (shown for $\epsilon = 0.2\epsilon^{*\pm}, 0.4\epsilon^{*\pm}, 0.6\epsilon^{*\pm}, 0.8\epsilon^{*\pm}$) decays exponentially at a rate near the maximum and mean Lyapunov exponent, λ_{\max} (black line) and λ_{mean} (black-dashed line) respectively¹⁷. **(b)** The survival probability function $S(\epsilon)$ from simulations (dots, equation (13); bars are standard error), theory (line, equations (20),(21)), and the simplified theory at large K , $\exp[-\epsilon/\bar{\epsilon}^*]$ (dotted line, equation (23)), where $\bar{\epsilon}^* = \left(\sqrt{KN}\bar{\nu}\tau/J_0\right)^{-1}$. **(c)** $S(\epsilon)$ from simulations (dots) and $\exp[-\epsilon/\bar{\epsilon}^*]$ (lines) for $J_0 = 2^n$, $n = -2, -1, 0, 1, 2$. (Same parameters as Fig. 1 except $N = 10^4$, $K = 10^3$.)

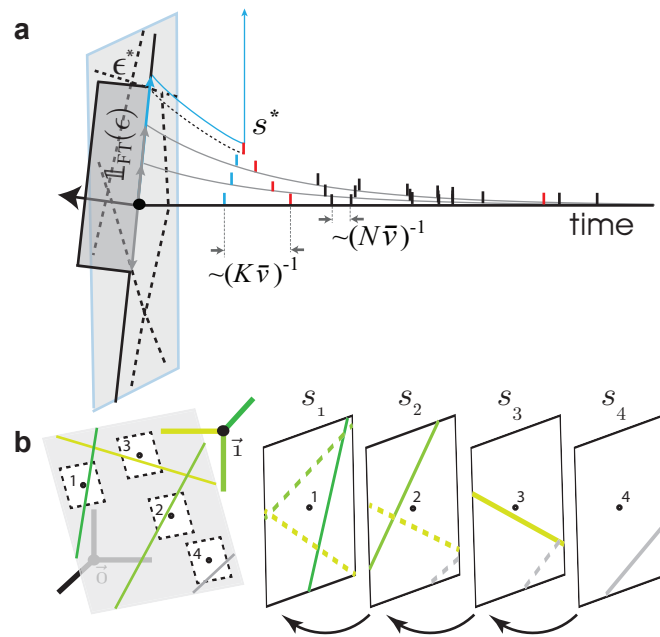


Figure 5 Flux tube boundaries are the pre-images of future susceptible spike collisions. **(a)** A folded phase space representation of a susceptible spike collision. Spikes (ticks) occur at a rate $N\bar{v}$ in the unperturbed trajectory (black line). For an example spike (blue tick), its susceptible spike partners (red ticks) occur at lower rate $K\bar{v}$. Small perturbations (gray arrows) lead to trajectories (gray) exhibiting spike time deviations that decay over time (tick alignment). A larger perturbation (example just beyond the critical perturbation strength: blue arrow) can push a spike and one of its susceptible partner spikes in the subsequent trajectory (blue and red, respectively) to collide, generating a divergence event at spike s^* . The indicator function, $\mathbb{1}_{FT}(\epsilon)$, has support (dark gray) only over the local tube. **(b)** Constructing the local flux-tube partition in the non-folded phase space. *Left:* Susceptible spikes are represented by susceptible edges (thick green lines) of the unit hypercube having $\vec{1} = (1, \dots, 1)$ (black dot) as an endpoint. An intrinsic random partition (thin green lines) is generated by projecting these edges onto the hyper-plane orthogonal to $\vec{1}$ (light gray). A given trajectory (labeled sequence of small dots) and its local neighborhood (within black dashed lines) is shown. *Right:* The flux tube partition for this trajectory at a given spike (here s_1) is obtained from back-iterating the intrinsic partition from all future spikes (here only partitions from s_2, s_3 , and s_4 are back-iterated; dashed lines). The partition at sufficiently distant future spikes (here the gray edge at s_4) will no longer refine the partition in the local neighborhood at s_1 , since the expansive backwards dynamics maps the projected edges outside the neighborhood. A concrete example obtained from simulations is presented in the Supplemental Notes.