

Two types of criticality in the brain

David Dahmen,¹ Sonja Grün,^{1,2} Markus Diesmann,^{1,3,4} and Moritz Helias^{1,4}

¹*Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA BRAIN Institute I, Jülich Research Centre, Jülich, Germany*

²*Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany*

³*Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*

⁴*Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany*
(Dated: November 30, 2017)

Neural networks with tightly balanced excitatory and inhibitory coupling show high computational performance. They operate close to a critical point characterized by the joint activation of large populations of neurons. Yet, macaque motor cortex shows very different dynamics with weak fluctuations on macroscopic scales. In this light motor cortex seems to operate in a sub-optimal regime. Here we show the opposite: The large dispersion of correlations across neurons is a signature of a rich dynamical repertoire, hidden from macroscopic brain signals, but essential for high performance in such concepts as reservoir computing. Our findings suggest a refinement of the view on criticality in neural systems: network topology and heterogeneity endow the brain with two complementary substrates for critical dynamics of largely different complexity.

The brain is a dynamical system with potentially different regimes of operation. Network models and experiments suggest optimal computational performance at critical points which mark the transition between two dynamical regimes [1]. At critical points, systems exhibit universal behavior, characterized by strong concerted fluctuations between all constituents leading to effective long-range interactions despite short-range couplings.

A particular type of criticality occurs in structurally balanced neuronal networks with excitatory and inhibitory feedback on par, leading to neuronal avalanches [2]. Signatures of avalanches have been observed in indirect measures of neuronal activity in macaque motor cortex [3]. Such evidence, however, is harder to obtain from direct measures of parallel neuronal activity; for extended periods large transients of population activity and long-range temporal correlations are absent (Fig. 1). The experimental data rather show weak and fast fluctuations of the population activity, suggesting an excess of inhibitory feedback and dynamically balanced activity [4]. Average correlations are low as a consequence [5, 6]. Experimental evidence for the operation of cortical networks in such balanced states is overwhelming [7, 8, 10, 11]. Such states are also consistent [12] with the observation of wave-like activity [13].

Another type of criticality, devoid of avalanches, has been investigated in computational studies [15]: edge-of-chaos criticality. With increasing heterogeneity in network connections, the dynamics changes from regular to chaotic, as characterized by the largest Lyapunov exponent. Such states moreover possess rich dynamics: a collection of coexisting network modes leads to topologically complex responses [16]. The onset of chaotic dynamics coincides with the breakdown of linear stability of the deterministic dynamics [17]. But a direct measurement of the Lyapunov exponent is hampered by noise and the incomplete knowledge of the dynamical state; as is the distance from the breakdown of linear stability.

We therefore need a measure that indirectly assesses either of these characteristics. Correlations in the neuronal activity (Fig. 1) may be informative, because they arise from interactions between constituents, which cause the collective behavior. Pairwise covariances $c(W)$ as shown in Fig. 1, to leading order follow a unique and model invariant simple law [18–20]

$$c(W) = [1 - W]^{-1} D [1 - W^T]^{-1}, \quad (1)$$

relating them to the effective connectivity matrix W of the linearized dynamics. The covariance matrix D measures fluctuations of the single-unit activity, determined by firing rates as well as by shared and correlated external inputs. The main contribution to covariances $c(W)$ arises from interactions within the local network [21]. Therefore, the experimentally observed variability across neurons (Fig. 1D) is predominantly caused by the structural variability contained in W . A potential bias in the estimation of covariances due to a finite amount of data [8] can be accounted for [9].

The linearized network dynamics corresponding to Eq. (1) becomes unstable if one eigenvalue of W has a positive real part ($\text{Re}(\lambda) \geq 1$). Eigenvalues are, in principle, determined by all connections in the network. But their determination from experimentally observed covariances by Eq. (1) is hindered by severe subsampling; even with present massively parallel recording techniques at most hundreds of neurons can be measured simultaneously from

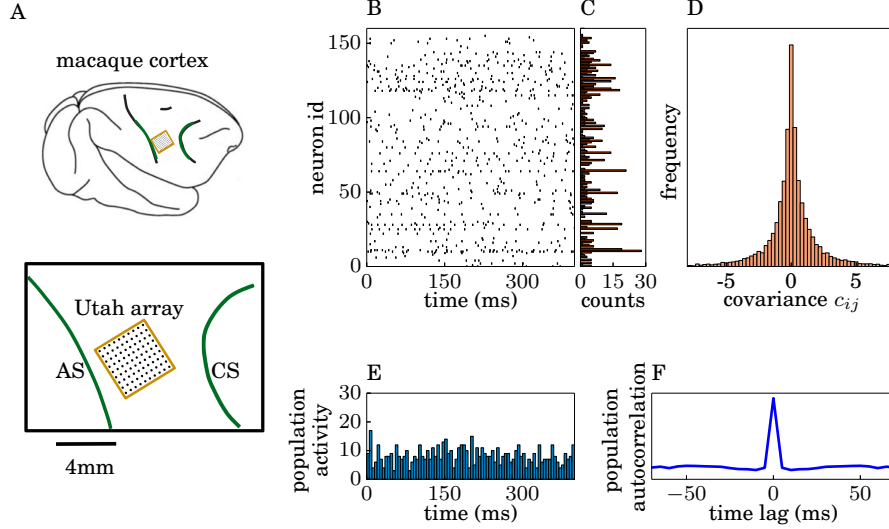


Figure 1. **Spiking activity in macaque motor cortex.** (A) 10×10 -electrode “Utah” array (Blackrock Microsystems, Salt Lake City, UT, USA; black dots) with $400 \mu\text{m}$ interelectrode distance covering an area of $4 \times 4 \text{ mm}^2$ (yellow square) of macaque motor cortex between the arcuate (AS, left green curve) and the central sulcus (CS, right green curve) of the right hemisphere. (B) Single trial raster plot of spiking activity of 155 neurons within $T = 400\text{ms}$ after trial start (TS) of a reach-to-grasp task [14]. (C) Spike counts n_i of activity within $T = 400\text{ms}$. (D) Distribution of covariances $c_{ij} = \frac{1}{T} (\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle)$ between spike counts n_i in 141 trials. (E) Time-resolved population activity binned in $t_{\text{bin}} = 5\text{ms}$. (F) Autocorrelation function of binned population activity.

the same local network [14]. We therefore need to find a characteristics of correlations that is informative about the eigenvalues, insensitive to the details of the network, and that can be estimated from a few hundred neurons at most.

The distribution of covariances is such a candidate measure. We first study how it is affected by the strength of connections (Fig. 2). With increasing connection strength, activity successively propagates over multiple synapses, leading to indirectly mediated interactions via a growing number of parallel paths (Fig. 2A) [19, 20]. The eigenvalues of the connectivity matrix approach the critical line where the dynamics loses linear stability (Fig. 2B); distributions of covariances (Eq. 1) become monomodal and broader but stay centered approximately around zero. These results expose a unique hallmark of dynamically balanced networks that operate close to linear instability: widely distributed covariances with a small mean as observed in motor cortex (Fig. 1D).

Although local cortical networks show non-random, cell-type specific, and distance-dependent connectivity, the simple model of a homogeneous random network studied in Fig. 2 is sufficient to explain gross features of the experimental data. To expose the fundamental mechanism explaining the dispersion of correlations, we change the perspective: instead of treating a single realization of the connectivity and calculating the variance of c_{ij} across pairs, we fix one pair and average over realizations of networks; a well-established technique for disordered physical systems. The disordered connectivity appears as an inverse matrix in Eq. (1), which technically complicates the analysis: no results from random matrix theory apply for this particular problem. Instead we construct a moment generating function [22] for the linearized network dynamics [9], which allows the use of spin-glass techniques [23, 24] combined with approximations for large- N field theories [25]. As a result, the disorder contained in the $\sim N^2$ entries of the connectivity formally reduces to only two fluctuating auxiliary variables that provide input to a fully symmetric all-to-all connected network. This high symmetry and the drastic reduction of dimensionality yield a mean-field theory beyond self-averaging [9]. The mean and standard deviation of variances ($i = j$) and covariances ($i \neq j$), to leading-order in the network size N , then follow as

$$c_{ij} = \left[[1 - \mu]^{-1} D_\lambda [1 - \mu^T]^{-1} \right]_{ij} \quad (2)$$

and

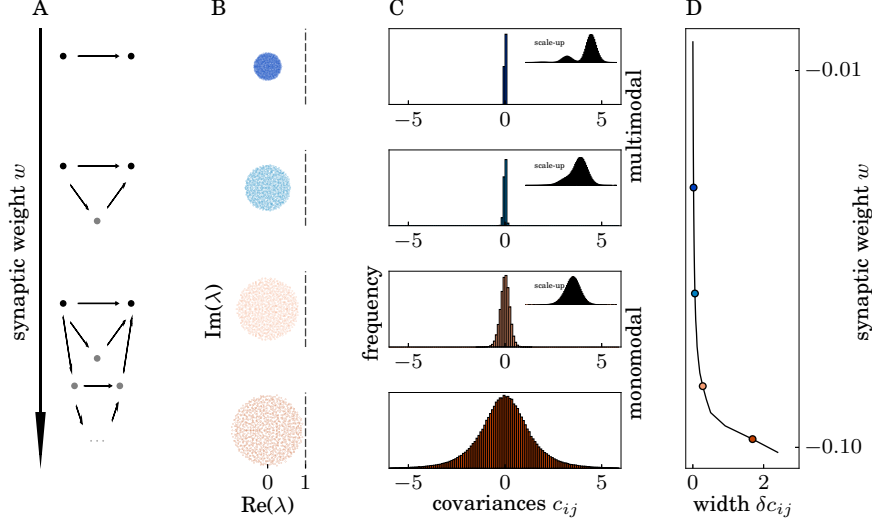


Figure 2. **Dispersion of correlations measures stability of network dynamics.** Connection strength w in a sparse random network increases from top to bottom (vertical axis). (A) Connections (arrows) between a pair of neurons (black dots); only indirect connections via intermediate neurons (gray dots) shown that predominantly contribute to correlations. (B) Bulk eigenvalues λ (colored dots) of the connectivity matrix W in the complex plane (critical line at $\text{Re}(\lambda) = 1$). (C) Distributions of covariances (Eq. 1) with low mean and large dispersion. (D) Standard deviation δc_{ij} of distribution of covariances (black curve; colored symbols correspond to distributions shown in panel C).

$$\delta c_{ij} = \sqrt{\frac{1 + \delta_{ij}}{N} \left(\left(\frac{1}{1 - \lambda_{\max}} \right)^2 - 1 \right) D_{\lambda}}, \quad (3)$$

where $\mu_{ij} \sim \mathcal{O}(1/\sqrt{N})$ is the mean and $\sigma_{ij}^2 = \lambda_{\max}/N \sim \mathcal{O}(1/N)$ the variance of connection weights in W . The latter determines the radius $\lambda_{\max}$ of the cloud of eigenvalues (Fig. 2B) and the renormalized covariance $D_{\lambda} = D/(1 - \lambda_{\max}^2)$, which accounts for the structural variability of connections (Fig. 3A,B). Eq. (2) explains the low mean $c_{ij} \sim \mathcal{O}(1/N)$ of covariances as a consequence of the dependence on $\mu < 0$: activity is decorrelated via excess inhibition [5, 6]; Eq. (3), moreover, explains the large standard deviation $\delta c_{ij} \sim \mathcal{O}(1/\sqrt{N})$ of experimentally observed distributions of covariances in motor cortex (Fig. 1D).

The theory in principle determines the largest eigenvalue from measured covariances. But there are two complications: D_{λ} is not known and cannot be measured, and it is unclear how robust the result is with regard to more realistic connectivity. Both problems are solved by considering the normalized width of the distribution of covariances $\Delta = \delta c_{ij}/c_{ii}$: This measure is predominantly determined by the network size and the most unstable eigenvalue $\lambda_{\max}$. The prediction of Δ is sufficient even for such network topologies (Fig. 3C) as excitatory-inhibitory networks (Fig. 3F) and distance-dependent connection probabilities (Fig. 3G). The latter networks also qualitatively explain the shape of the experimentally observed covariance distribution. The applicability of the theory goes beyond linear network dynamics; it predicts Δ even in spiking networks (Fig. 3H). Therefore, Δ can be used to infer the operational regime of the cortical network (Fig. 1). The distance to linear instability is determined by $\lambda_{\max}$ which, to leading order in N , is given by inversion of Eqs. (2) and (3) as

$$\lambda_{\max} = \sqrt{1 - \sqrt{\frac{1}{1 + N\Delta^2}}}. \quad (4)$$

Fig. 3D suggests that given the experimentally measured distribution of covariances (Fig. 1D) and biologically plausible N above 10^4 neurons, the network needs to operate close to instability to explain the data. The weak dependence of $\lambda_{\max}$ on N and Δ in the biologically relevant ranges further secures the robustness of the result towards variations of these experimental estimates.

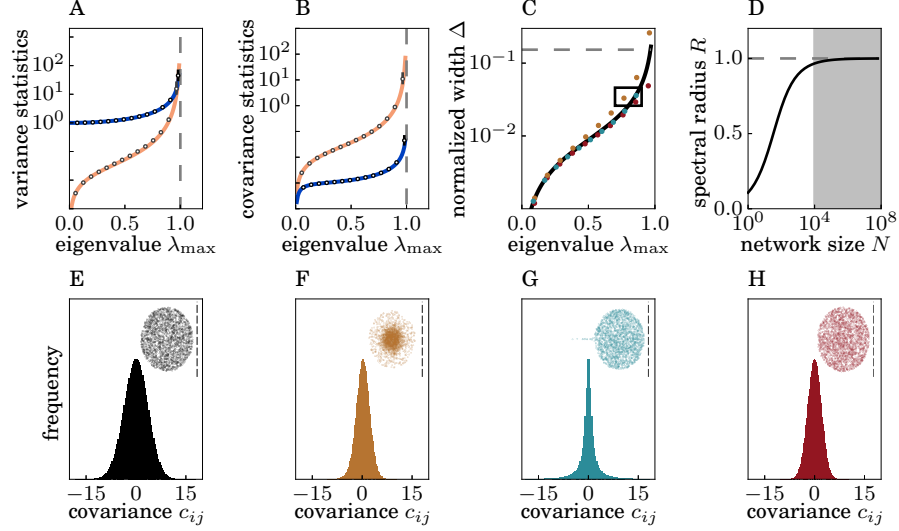


Figure 3. Large relative dispersion of covariances reveals dynamics of motor cortex close to instability. (A)-(B) Theoretical prediction for the mean (Eq. (2), blue curve) and standard deviation (Eq. (3), orange curve) of variances (panel A) and covariances (panel B) for different maximum eigenvalues $\lambda_{\max}$ compared to numerical results (Eq. (1), markers) for homogeneous inhibitory networks. Circles denote the ensemble averages for 100 network realizations, bars the average upward/downward deviation of the non-Gaussian statistics. Bars are below marker size except for rightmost data points. (C) Theoretical prediction for the normalized width Δ (black curve) of covariances compared to numerical results (Eq. (1), markers) for excitatory-inhibitory networks (yellow), inhibitory networks with distance-dependent connectivity (cyan), and numerical simulations of inhibitory networks of spiking leaky integrate-and-fire neurons (red). Distributions corresponding to data points within the black rectangle are shown in panels E-H. (D) Predicted maximum eigenvalue $\lambda_{\max}$ (Eq. 4) of the effective connectivity as a function of the number of neurons N for given moments of the covariance distribution measured in macaque motor cortex (Fig. 1D). The shaded area marks the range of biologically plausible effective network sizes corresponding to the spatial scale of the recordings. (E)-(H) Distribution of covariances (histogram) and bulk eigenvalues (dots) relative to critical line $\text{Re}(\lambda_{\max}) = 1$ (dashed black line) for a homogeneous inhibitory network model (panel E), for a network of excitatory and inhibitory neurons (panel F), for a network with distance-dependent connection probability (G), and for a network of spiking leaky integrate-and-fire neurons (panel H).

In summary, we combine methods from field theory and disordered systems to predict the dispersion of neuronal covariances in relation to the operational regime of the network. The theory goes beyond the mean-field approximation of population-averaged dynamics and, applied to massively parallel spike recordings from macaque motor cortex, unravels its dynamics close to criticality.

The dynamics in this regime is in contrast to strongly fluctuating population activity observed in avalanches [26]. These cause positive average covariances (Fig 4B) and a slowly decaying autocorrelation function of the population activity (Fig. 4C). A close structural balance between excitation and inhibition ($\mu \approx 1/N$) explains such behavior, caused by a single, nearly unstable eigenvalue that amplifies fluctuations of the population activity [27] (Fig 4A, red dot). All other modes have low amplitude and the same exponentially decaying, fast dynamics (Fig. 4C inset). The network is hence effectively one-dimensional [28].

In general, each eigenvalue corresponds to a specific linear combination of neuron activities, a mode (Fig. 4A, D). The deterministic network structure (structural balance, populations, hubs, etc) determines only a small subset of eigenvalues, whereas heterogeneity of connections across neuron pairs shapes the bulk. If sufficiently large, this variability causes a multitude of modes close to instability (Fig. 4D). The broad distribution of covariances in the cortical data (Fig. 1D, Fig. 4E) implies such a critical state. The almost vanishing mean covariances and the weakly fluctuating and quickly decaying population activity (Fig. 1F, Fig. 4F), however, show that the network is dynamically balanced [5, 6] rather than structurally balanced (Fig. 4D, red dot). Each of the critical modes exhibits a slowly decaying autocorrelation function and evolves in a specific direction within the high-dimensional space of all neurons that is different from the population activity (Fig. 4F). The experimental identification of these hidden modes presents a major challenge for the future due to the strong subsampling of simultaneous single-neuron activities.

We expect implications of the operation in the dynamically balanced, critical regime ($\lambda_{\max} \lesssim 1$) for learning and information processing. Neurons belonging to a critical mode show pairwise covariances that strongly exceed

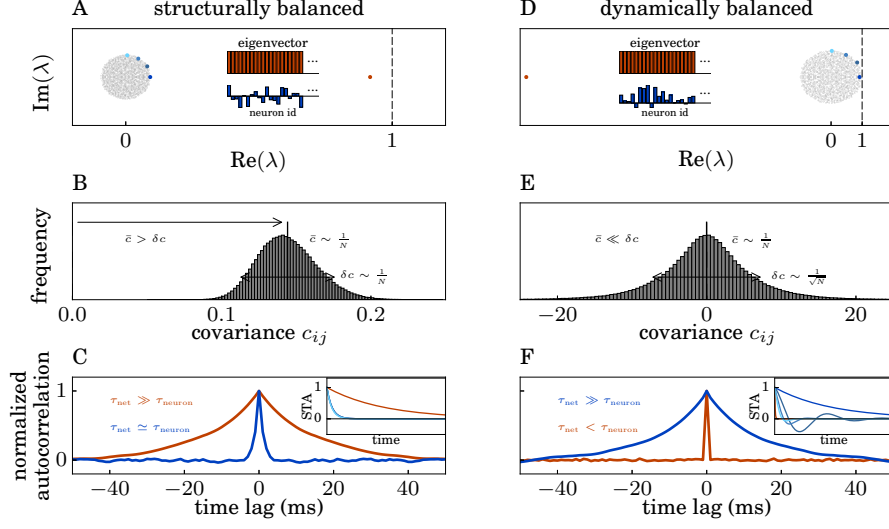


Figure 4. **Two types of criticality.** Left: Structurally balanced network with virtually unstable linearized dynamics on the population level. (A) Spectrum of eigenvalues with positive outlier (red dot) and corresponding eigenvector (red bars, population activity) generated by deterministic connectivity structure. (B) Distribution of covariances with positive mean and small standard deviation. (C) Autocorrelation functions of the population activity (red curve) and the activity projected onto the eigendirection corresponding to the largest bulk eigenvalue (blue curve, cf. dark blue dot in panel A). Inset: Time course of network modes corresponding to colored eigenvalues in panel A. Right: Dynamically balanced network with virtually unstable linearized dynamics hidden in specific linear combinations of neuron activities. (D) Spectrum of eigenvalues with negative outlier (red dot) and nearly critical bulk eigenvalue (blue dot) and corresponding eigenvector (blue bars) generated by heterogeneity in connections across neuron pairs. (E) Distribution of covariances with almost vanishing mean and large standard deviation. (F) Autocorrelation functions of the population activity (red curve, cf. red dot in panel D) and the activity projected onto the eigendirection corresponding to the largest real bulk eigenvalue (blue curve). Inset: Time course of network modes corresponding to colored eigenvalues in panel D.

the average (Fig. 4E). Such covariances interact with spike-timing dependent synaptic plasticity [29], leading to an increased interaction between neuronal and synaptic dynamics. Weak external inputs to the network are, moreover, sufficient to shift a macroscopic number of eigenvalues across the edge of stability [16] and are therefore potent to modulate the integrative properties of the network. Critical modes have a multitude of characteristic shapes and life times (Fig. 4F inset, [30]); they here arise despite the stereotypical and fast dynamics of individual neurons as a result of the disordered network structure. This diversity is in stark contrast to the small number of different modes at the point of close structural balance (here: two modes, Fig. 4C inset). The enriched repertoire enables the parallel integration and maintenance of information over prolonged time scales. Such networks provide a wealth of transformations on the input and therefore may serve as an exhaustive reservoir for computation [31, 32].

Finally, the two types of criticality are not mutually exclusive as they are governed by different mechanisms. They can coexist in different brain regions or even in the same local network; networks may hence dynamically move into either regime, adapting brain function to momentary demands.

ACKNOWLEDGMENTS

We are grateful to Alexa Riehle and Thomas Brochier for providing the experimental data. We thank Vahid Rostami for help with the analysis of the multi-channel data. The research was carried out within the scope of the International Associated Laboratory “Vision for Action - LIA V4A” of INT (CNRS, AMU), Marseilles and INM-6, Jülich. This work was partially supported by HGF young investigator’s group VH-NG-1028, Helmholtz portfolio theme SMHB, and EU Grants 604102 and 720270 (Human Brain Project, HBP).

-
- [1] W. L. Shew, D. Plenz, *Neuroscientist* **19**, 88 (2013).
 - [2] J. M. Beggs, D. Plenz, *J. Neurosci.* **23**, 11167 (2003).
 - [3] T. Petermann, *et al.*, *Proc. Nat. Acad. Sci. USA* **106**, 15921 (2009).
 - [4] C. van Vreeswijk, H. Sompolinsky, *Science* **274**, 1724 (1996).
 - [5] A. Renart, *et al.*, *Science* **327**, 587 (2010).
 - [6] T. Tetzlaff, M. Helias, G. Einevoll, M. Diesmann **8**, e1002596 (2012).
 - [7] M. Okun, I. Lampl, *Nat. Neurosci.* **11**, 535 (2008).
 - [8] A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, *Science* **327**, 584 (2010).
 - [9] *See supplementary material.*
 - [10] K. Reinhold, A. D. Lien, M. Scanziani, *Nat. Neurosci.* **18**, 1789 (2015).
 - [11] N. Dehghani, *et al.*, *Scientific reports* **6** (2016).
 - [12] R. Pyle, R. Rosenbaum, *Phys. Rev. Lett.* **118**, 018103 (2017).
 - [13] K. Takahashi, *et al.*, *Nature Communications* **6** (2015).
 - [14] A. Riehle, S. Wirtsohn, S. Grün, T. Brochier, *Frontiers in Neural Circuits* **7**, 48 (2013). Doi: 10.3389/fncir.2013.00048.
 - [15] N. Bertschinger, T. Natschläger, *Neural Comput.* **16**, 1413 (2004).
 - [16] G. Wainrib, J. Touboul, *Phys. Rev. Lett.* **110**, 118101 (2013).
 - [17] H. Sompolinsky, A. Crisanti, H. J. Sommers, *Phys. Rev. Lett.* **61**, 259 (1988).
 - [18] B. Lindner, B. Doiron, A. Longtin, *Phys. Rev. E* **72**, 061919 (2005).
 - [19] V. Pernice, B. Staude, S. Cardanobile, S. Rotter **7**, e1002059 (2011).
 - [20] J. Trousdale, Y. Hu, E. Shea-Brown, K. Josic **8**, e1002408 (2012).
 - [21] M. Helias, T. Tetzlaff, M. Diesmann **10**, e1003428 (2014).
 - [22] C. Chow, M. Buice, *The Journal of Mathematical Neuroscience* **5** (2015).
 - [23] C. De Dominicis, *Phys. Rev. B* **18**, 4913 (1978).
 - [24] H. Sompolinsky, A. Zippelius, *Phys. Rev. B* **25**, 6860 (1982).
 - [25] M. Moshe, J. Zinn-Justin, *Physics Reports* **385**, 69 (2003).
 - [26] G. Tkacik, *et al.*, *Proc. Nat. Acad. Sci. USA* **112**, 11508 (2015).
 - [27] D. B. Larremore, W. L. Shew, J. G. Restrepo, *Phys. Rev. Lett.* **106**, 058101 (2011).
 - [28] G. Tkacik, *et al.*, *PLoS Comp Biol* **10**, e1003408 (2014).
 - [29] H. Markram, J. Lübke, M. Frotscher, B. Sakmann, *Science* **275**, 213 (1997).
 - [30] G. Hennequin, T. Vogels, W. Gerstner **82**, 1394 (2014).
 - [31] W. Maass, T. Natschläger, H. Markram, *Neural Comput.* **14**, 2531 (2002).
 - [32] H. Jaeger, H. Haas, *Science* **304**, 78 (2004).
 - [33] J. Aljadeff, M. Stern, T. Sharpee, *Phys. Rev. Lett.* **114**, 088101 (2015).
 - [34] J. Kadmon, H. Sompolinsky, *Phys. Rev. X* **5**, 041030 (2015).
 - [35] D. Grytskyy, T. Tetzlaff, M. Diesmann, M. Helias **7**, 131 (2013).
 - [36] I. Ginzburg, H. Sompolinsky, *Phys. Rev. E* **50**, 3171 (1994).
 - [37] J. A. Hertz, Y. Roudi, P. Sollich, *Journal of Physics A: Mathematical and Theoretical* **50**, 033001 (2017).
 - [38] K. H. Fischer, J. A. Hertz, *Spin Glasses* (Cambridge University Press, 1991).
 - [39] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, no. 13 in Springer Series in Synergetics (Springer-Verlag, Berlin, 1985), second edn.
 - [40] D. Sherrington, S. Kirkpatrick, *Phys. Rev. Lett.* **35**, 1792 (1975).
 - [41] C. De Dominicis, L. Peliti, *Phys. Rev. B* **18**, 353 (1978).
 - [42] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing An Introduction* (Clarendon Press, Oxford, 2001).
 - [43] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, no. 13 in Springer Series in Synergetics (Springer-Verlag, Berlin, 1983).
 - [44] C. van Vreeswijk, H. Sompolinsky, *Neural Comput.* **10**, 1321 (1998).
 - [45] D. J. Amit, *Field Theory, the Renormalization Group, and Critical Phenomena; 3rd ed.* (World Scientific, Singapore, 2005).
 - [46] Y. Roudi, J. Hertz, *Physical Review Letters* **106**, 048702 (2011).
 - [47] R. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (John Wiley & Sons, 2013).
 - [48] E. L. Lehmann, G. Casella, *Theory of point estimation* (Springer Science & Business Media, 2006).

Supplementary Materials

I. MEAN-FIELD THEORY FOR META-STATISTICS BEYOND SELF-AVERAGING

Mean-field theory most-commonly employs the thermodynamic limit ($N \rightarrow \infty$), reducing the collective dynamics of the N interacting units to N pairwise independent units, each subject to a self-consistently determined auxiliary field [4, 17, 33, 34]. Covariances of individual neurons are self-averaging in this limit, so they are identical for all units and independent of the realization of the randomness in the network connectivity. In particular, cross-covariances vanish. In the absence of external stimuli and in the weakly correlated regime, covariances can be understood in linear response theory [35]. For such a linearized network model, interactions between neurons can be included as a finite-size correction within this self-averaging framework [5, 6, 20, 21, 36]. While this procedure yields covariances averaged over many pairs of units, the experimentally observed neuron to neuron variability [8] is lost. Here, we develop a theory beyond self-averaging covariances that captures their statistics. It exploits the large size of biologically realistic local networks, but includes finite-size fluctuations of auxiliary fields which derive from the quenched disorder of the couplings to explain the variability of covariances. To perform this qualitative step, we need to combine methods from different fields: We use a functional formulation of Gaussian processes that originates from statistical field theory [22, 23, 37] and combine it with methods typically used for disordered systems in the large N limit [25], such as spin glasses [24, 38]. We outline these steps in detail below.

A. Moment generating functional for the network dynamics

In the following, we consider time-lag integrated covariances

$$c_{ij} = \int_{-\infty}^{\infty} c_{ij}(\tau) d\tau = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle x_i(t+\tau) x_j(t) \rangle_x dt d\tau = \langle X_i(0) X_j(0) \rangle_x, \quad (5)$$

calculated as averages $\langle \rangle_x$ across different trials of the linearized network dynamics $x(t)$ or its Fourier transform $X(\omega)$ evaluated at $\omega = 0$ (Wiener-Khinchin theorem, [39, sec. 1.4.2]). The linearized dynamics can be modeled as a set of coupled Ornstein-Uhlenbeck processes

$$\tau \frac{dx(t)}{dt} = -x(t) + W \cdot x(t) + \xi(t) \quad (6)$$

with the moment-generating functional

$$Z[j] = \int \mathcal{D}x \int \mathcal{D}\tilde{x} \exp \left(S_0[x, \tilde{x}] + j^T x \right) \\ \text{with } S_0[x, \tilde{x}] = \tilde{x}^T (\partial_t + 1 - W) x + \frac{D}{2} \tilde{x}^T \tilde{x}. \quad (7)$$

Here $\xi(t)$ is a Gaussian white noise with variance $\langle \xi_i(t) \xi_j(t') \rangle = D \delta_{ij} \delta(t-t')$, $\tilde{x}(t)$ is a purely imaginary response field, $\int \mathcal{D}\tilde{x}$, $\int \mathcal{D}x$ are suitably defined path integral measures, and $\tilde{x}^T \tilde{x} = \sum_i \int \tilde{x}_i(t) \tilde{x}_i(t) dt$ is a scalar product [22, 23, 37]. The generating functional can easily be interpreted in Fourier domain due to the linearity of Eq. (6) and the invariance of scalar products under unitary transforms

$$Z[J] = \int \mathcal{D}X \int \mathcal{D}\tilde{X} \exp \left(S_0[X, \tilde{X}] + J^T X \right) \\ \text{with } S_0[X, \tilde{X}] = \tilde{X}^T (i\omega + 1 - W) X + \frac{D}{2} \tilde{X}^T \tilde{X}, \quad (8)$$

with Fourier transformed variables denoted by capital letters. The scalar product in frequency domain reads $\tilde{X}^T X = \sum_i \int \tilde{X}_i(-\omega) X_i(\omega) d\omega$. The generating functional factorizes into generating functions for each frequency ω . As shown in Eq. (5), time-lag integrated covariances only require the knowledge of $X(0)$. In the following, we will therefore only discuss zero frequency. After integration over all non-zero frequencies one obtains the generating function for

zero frequency

$$\begin{aligned} Z(J) &= \det(1 - W) \int \mathcal{D}X \int \mathcal{D}\tilde{X} \exp \left(S_0(X, \tilde{X}) + J^T X \right) \\ &= \exp \left(\frac{1}{2} J^T (1 - W)^{-1} D (1 - W^T)^{-1} J \right) \end{aligned} \quad (9)$$

$$\text{with } S_0(X, \tilde{X}) = \tilde{X}^T (1 - W) X + \frac{D}{2} \tilde{X}^T \tilde{X}, \quad (10)$$

with the single-frequency ($\omega = 0$) scalar product defined as $\tilde{X}^T X = \sum_i \tilde{X}_i X_i$, and integration measures $\int \mathcal{D}X = \prod_j \int_{-\infty}^{\infty} dX_j$ and $\int \mathcal{D}\tilde{X} = \prod_j \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\tilde{X}_j$. The determinant in Eq. (9) follows from the normalization condition $Z[J=0] = 1$. The time-lag integral of the covariance functions follows as

$$c(W) = [1 - W]^{-1} D [1 - W^T]^{-1}. \quad (11)$$

B. Self-averaging meta-statistics

Equation (11) relates covariances between individual pairs of neurons to the connectivity matrix W . These, however, change between realizations of the random connectivity. In contrast, the meta-statistics, by which we denote the moments of the distribution of covariances, can be assumed constant across realizations (self-averaging, [38]). We therefore seek for an expression relating the moments of the distribution of covariances to the statistics of the connectivity W .

We denote with $\overline{}$ the empirical average, with $\langle \rangle_x$ the expectation over realizations of the processes x , and with $\langle \rangle$ the ensemble average over the disordered connectivity W . Exchanging the order of differentiation and averaging allows expressing second moments of covariances as derivatives of a single disorder-averaged generating function $\langle Z(J) \rangle$: We first assume the empirical average to be self-averaging

$$\overline{c_{ii}^2} = \langle \overline{c_{ii}^2} \rangle. \quad (12)$$

We then use its definition as $\overline{c_{ii}^2} = \frac{1}{N} \sum_{i=1}^N c_{ii}^2$, and that of the second cumulant $c_{ij} = \langle X_i X_j \rangle_x$ of the zero frequency components $X = X(\omega = 0) = \int_{-\infty}^{\infty} x(t) dt$ of Ornstein-Uhlenbeck processes x to obtain

$$\overline{c_{ii}^2} = \frac{1}{N} \sum_{i=1}^N \langle \langle X_i X_i \rangle_x^2 \rangle. \quad (13)$$

As the action Eq. (9) for a single realization of W is quadratic, Wick's theorem applies, $\langle X_i X_i X_j X_j \rangle_x = 2 \langle X_i X_j \rangle_x^2 + \langle X_i X_i \rangle_x \langle X_j X_j \rangle_x = 2c_{ij}^2 + c_{ii}c_{jj}$. For the special case $i = j$, we may identify the squared second moment (13) with a fourth moment. The latter can be expressed with the help of the generating function

$$\begin{aligned} \overline{c_{ii}^2} &= \frac{1}{N} \sum_{i=1}^N \langle \langle X_i X_i \rangle_x^2 \rangle \stackrel{\text{Wick's th.}}{=} \frac{1}{N} \sum_{i=1}^N \frac{1}{3} \langle \langle X_i X_i X_i X_i \rangle_x \rangle \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{3} \left\langle \frac{d^4}{dJ_i^4} Z(J) \right\rangle \Big|_{J=0} = \frac{1}{3} \frac{d^4}{dJ_i^4} \langle Z(J) \rangle \Big|_{J=0}. \end{aligned} \quad (14)$$

In the last step, we exchanged the order of derivatives and the expectation value over network realizations and used the symmetry of the disorder-averaged network over units. Analogously follows for $i \neq j$ and $N - 1 \approx N$,

$$\overline{c_{ij}^2} = \frac{1}{2} \frac{d^4}{dJ_i^2 dJ_j^2} \langle Z(J) \rangle \Big|_{J=0} - \frac{1}{2} \langle \overline{c_{ii}} \rangle^2. \quad (15)$$

C. Disorder-averaged generating function

Ignoring insignificant variations in the normalization $\det(1 - W)$ of $Z[J]$, the disorder average only affects the coupling term in Eq. (9)

$$\begin{aligned} \langle \exp(\tilde{X}^T W X) \rangle &= \left\langle \exp \left(\sum_{i,j} W_{ij} \tilde{X}_i X_j \right) \right\rangle \\ &= \prod_{i,j} \langle \exp(W_{ij} \tilde{X}_i X_j) \rangle \\ &= \prod_{i,j} \exp \left(\sum_{k=1}^{\infty} \frac{\kappa_k}{k!} (\tilde{X}_i X_j)^k \right). \end{aligned}$$

For clarity, we here assume independent and identically distributed weights W_{ij} . In the resulting cumulant expansion [38, 40–42] κ_k is the k -th cumulant for a single connection W_{ij} [43]. For fixed connection probability p , the number of inputs to a neuron scales with the network size N . To keep the input and its fluctuations within a certain dynamic range when increasing the network size, we require synaptic weights to scale with $1/\sqrt{N}$ [4, 44], such that the cumulant expansion is an expansion in $1/\sqrt{N}$. A truncation at the second cumulant ($\propto N^{-1}$) maps W to a Gaussian connectivity $\mathcal{N}(\mu, \sigma^2/N)$ so that

$$\begin{aligned} \langle Z(J) \rangle &\sim \int \mathcal{D}X \int \mathcal{D}\tilde{X} \exp \left(S_0(X, \tilde{X}) + \frac{\sigma^2}{2N} V(X, \tilde{X}) + J^T X \right), \\ S_0(X, \tilde{X}) &= \tilde{X}^T (1 - \mu) X + \frac{D}{2} \tilde{X}^T \tilde{X}, \\ V(X, \tilde{X}) &= \tilde{X}^T \tilde{X} X^T X, \end{aligned} \tag{16}$$

with a homogeneous mean connection weight $\mu_{ij} = \mu = \mathcal{O}(1/\sqrt{N})$. The second cumulant (σ^2/N) is the first non-trivial contribution to the second moment of covariances. While higher cumulants of the connectivity have an impact on higher moments of the distribution of covariances, their effect on the first two moments is suppressed by the large network size.

D. Auxiliary-field formalism

The interaction term V prevents an exact calculation of the disorder-averaged generating function. A converging perturbation series can be obtained in the auxiliary-field formulation [25], where a field $Q_1 = \frac{\sigma^2}{N} X^T X$ is introduced for the sum of a large number of statistically equivalent activity variables. Using the Hubbard-Stratonovich transformation

$$\begin{aligned} e^{\frac{\sigma^2}{2N} \tilde{X}^T \tilde{X} X^T X} &= \int_{-\infty}^{\infty} dQ_1 \delta \left(Q_1 - \frac{\sigma^2}{N} X^T X \right) e^{\frac{1}{2} Q_1 \tilde{X}^T \tilde{X}} \\ &= \underbrace{\frac{1}{2\pi i} \frac{N}{\sigma^2} \int_{-\infty}^{\infty} dQ_1 \int_{-i\infty}^{i\infty} dQ_2}_{\int \mathcal{D}Q} e^{-\frac{N}{\sigma^2} Q_1 Q_2 + \frac{1}{2} Q_1 \tilde{X}^T \tilde{X} + Q_2 X^T X}, \end{aligned}$$

one obtains a free theory, which is a quadratic action in the activity (X) and response variables ($\tilde{X}$), on the background of fluctuating fields Q

$$\begin{aligned} \langle Z(J) \rangle &\sim \int \mathcal{D}Q \exp \left(-\frac{N}{\sigma^2} Q_1 Q_2 + \ln(Z_Q(J)) \right), \\ Z_Q(J) &= \int \mathcal{D}X \int \mathcal{D}\tilde{X} \exp \left(S_0(X, \tilde{X}) + \frac{1}{2} Q_1 \tilde{X}^T \tilde{X} + Q_2 X^T X + J^T X \right). \end{aligned} \tag{17}$$

The high dimensional integrals of the free theory $Z_Q(J)$ can be solved analytically yielding a two-dimensional interacting theory in the auxiliary fields Q_1 and Q_2 . The auxiliary field formalism translates the high-dimensional ensemble average over W to a low-dimensional average over Q ; it maps the local disorder in the connections to fluctuations

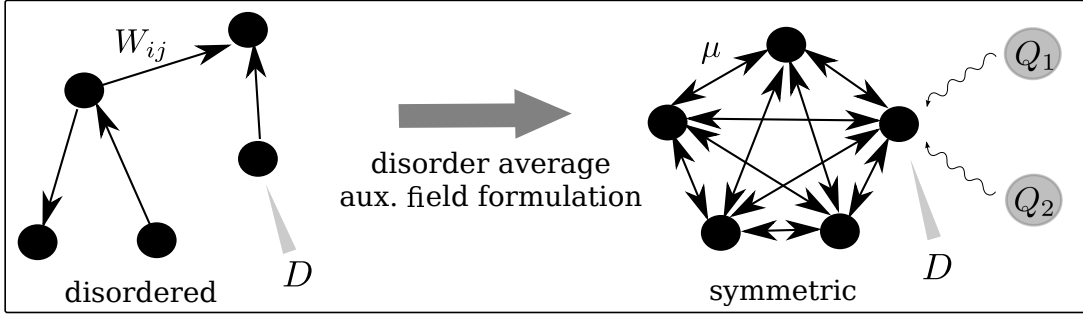


Figure S1. **Reduction of a disordered to a regular network.** Disorder average maps network with frozen variability in connections (left) to highly symmetric network on the background of fluctuating auxiliary fields Q (right). Their fluctuations contribute globally to the covariance that drives the network fluctuations (illustrated by wavy arrows).

of global fields Q interacting with a highly symmetric all-to-all connected network, illustrated in Fig. S1. Only in the special case of vanishing mean connection strength $\mu = 0$, the system factorizes into N unconnected units, each interacting with the same set of fields Q . The all-to-all network not only captures the autocovariance of a single neuron, but also the cross-covariance with any other neuron.

E. Saddle-point approximation

Performing a change of variables $Q = Q^*(J) + \frac{\delta Q}{\sqrt{N}}$ in Eq. (17), where $Q^*(J)$ are the saddle-points of the exponent $-\frac{N}{\sigma^2} Q_1 Q_2 + \ln(Z_Q(J)) =: -N \cdot Y(Q_1, Q_2, J)$ determined by the saddle-point equations

$$0 = \left. \frac{\partial}{\partial Q_\alpha} Y(Q_1, Q_2, J) \right|_{Q=Q^*(J)} \quad \alpha \in \{1, 2\}, \quad (18)$$

and expanding $Y(Q_1, Q_2, J)$ around $Q^*(J)$, yields an expansion in powers of $1/N$ [25, 45]. We focus only on the leading-order contribution $Y(Q_1^*(J), Q_2^*(J), J)$ which describes tree-level diagrams in the Q -theory and drop all source dependence in higher Taylor coefficients

$$\langle Z(J) \rangle \sim \exp(-N \cdot Y(Q_1^*(J), Q_2^*(J), J)), \quad (19)$$

with the corresponding action

$$S_{Q_\alpha^*(J)}(X, \tilde{X}) = \tilde{X}^T (1 - \mu) X + \frac{D + Q_1^*(J)}{2} \tilde{X}^T \tilde{X} + Q_2^*(J) X^T X. \quad (20)$$

Note that this Gaussian theory in $X, \tilde{X}$ still contains in the latter two terms contributions from the quartic interaction term V of the original theory. Variability in $Q^*(J)$, through their source dependence, therefore gives rise to cumulants of the activity variables beyond second order. The dependence of $Q^*(J)$ on external sources J was neglected in prior work [24] since it scales to leading order as $1/N$. This approximation, however, yields a Gaussian theory (see Eq. (20) for $J = 0$), which does not generate fourth cumulants and therefore no distribution of covariances. Taking into account the J dependence of auxiliary fields in combination with the relation between fourth cumulants and distributions of covariances Eq. (14) and Eq. (15) thus extends mean-field theory beyond self-averaging covariances.

The action (20) shows that $Q_1^*(J)$ acts as a global contribution to the Gaussian noise whereas $Q_2^*(J)$ directly contributes to the inverse of the covariance matrix. By integrating out the response variables $\tilde{X}$, one can alternatively interpret both $Q_\alpha^*(J)$ as contributions to the covariance matrix of the noise.

Saddle points $Q^*(J)$ are obtained self-consistently from Eq. (18) which reduces to the set of equations

$$\begin{aligned} Q_1^*(J) &= \frac{\sigma^2}{N} \langle X^T X \rangle_{Q_\alpha^*(J)}(J) \\ Q_2^*(J) &= \frac{\sigma^2}{2N} \langle \tilde{X}^T \tilde{X} \rangle_{Q_\alpha^*(J)}(J) \end{aligned}$$

with $\langle \circ \rangle_{Q_\alpha^*(J)}(J) := \frac{\int \mathcal{D}X \int \mathcal{D}\tilde{X} \circ \exp(S_{Q_\alpha^*(J)}(X, \tilde{X}) + J^T X)}{\int \mathcal{D}X \int \mathcal{D}\tilde{X} \exp(S_{Q_\alpha^*(J)}(X, \tilde{X}) + J^T X)} \quad (21)$

The above set of equations cannot be solved analytically for $Q^*(J)$. However, moments of activity result from derivatives of $\langle Z(J) \rangle$ evaluated at $J = 0$. The derivatives act not only on the source term $J^T X$, but also on the J -dependence of the saddle-point. Therefore, moments are determined by saddle-points and their derivatives evaluated at zero source. Setting $J = 0$ in Eq. (21) yields the self-consistent result

$$\langle \tilde{X}_i \tilde{X}_j \rangle_{Q_\alpha^*} = 0, \quad (22)$$

$$\langle X_i X_j \rangle_{Q_\alpha^*} = \left[(1 - \mu)^{-1} (D + Q_1^*) (1 - \mu^T)^{-1} \right]_{ij}, \quad (23)$$

$$\langle \tilde{X}_i X_j \rangle_{Q_\alpha^*} = \left[(1 - \mu)^{-1} \right]_{ij}, \quad (24)$$

with

$$\begin{aligned} Q_1^* &:= Q_1^*(0) = \frac{R^2}{1 - R^2} D, \\ Q_2^* &:= Q_2^*(0) = 0, \end{aligned} \quad (25)$$

$\langle \circ \rangle_{Q_\alpha^*} := \langle \circ \rangle_{Q_\alpha^*(0)}$, $R = \sqrt{1 + \gamma} \sigma$ and $\gamma = \mathcal{O}(1/N)$ resulting from subleading terms in Eq. (23) containing μ . Eq. (23) shows that Q_1^* leads to a renormalization of the noise D as $D_r = D + Q_1^* = D/(1 - R^2)$.

Analogously, second derivatives of saddle-points with respect to sources can be calculated from Eq. (21) with the final result

$$\begin{aligned} \left. \frac{d^2 Q_1^*(J)}{dJ_k dJ_l} \right|_{J=0} &= \frac{\sigma^2}{\beta N} \sum_i \langle X_i X_k \rangle_{Q_\alpha^*} \langle X_i X_l \rangle_{Q_\alpha^*} + \frac{\sigma^2}{\beta N} \sum_{i,a} \langle X_i X_a \rangle_{Q_\alpha^*}^2 \left. \frac{d^2 Q_2^*(J)}{dJ_k dJ_l} \right|_{J=0}, \\ \left. \frac{d^2 Q_2^*(J)}{dJ_k dJ_l} \right|_{J=0} &= \frac{\sigma^2}{\beta N} \sum_i \langle \tilde{X}_i X_k \rangle_{Q_\alpha^*} \langle \tilde{X}_i X_l \rangle_{Q_\alpha^*}, \quad \text{with } \beta = 1 - \frac{\sigma^2}{N} \sum_{i,a} \langle X_i \tilde{X}_a \rangle_{Q_\alpha^*}^2, \end{aligned}$$

where we used that $\left. \frac{dQ_\alpha^*(J)}{dJ_k} \right|_{J=0} = 0$ since three-point correlators vanish.

Using partial derivative calculus $\frac{d}{dJ_i} \langle Z(J) \rangle = \left(\frac{\partial Q_1^*}{\partial J_i} \frac{\partial}{\partial Q_1^*} + \frac{\partial Q_2^*}{\partial J_i} \frac{\partial}{\partial Q_2^*} + \frac{\partial}{\partial J_i} \right) \langle Z(Q_1^*, Q_2^*, J) \rangle$ and the saddle-point condition $\frac{\partial}{\partial Q_\alpha^*} \langle Z(Q_1^*, Q_2^*, J) \rangle = 0$ which follows from Eq. (18) and Eq. (19), we find that second moments of activity, the average covariances, are not influenced by the source dependence of saddle points

$$\langle \langle X_i X_j \rangle_x \rangle = \left. \frac{d^2}{dJ_i dJ_j} \langle Z(J) \rangle \right|_{J=0} = \langle X_i X_j \rangle_{Q_\alpha^*} = \left[(1 - \mu)^{-1} D_r (1 - \mu^T)^{-1} \right]_{ij}. \quad (26)$$

The non-zero mean connection strength $\mu = \mathcal{O}(1/\sqrt{N})$ yields cross-covariances between neurons due to the finite size of the network. In addition, the fourth moments of activity are crucial for the non-vanishing variance of covariances: Formally, two of the four derivatives act on $Q^*(J)$ while the other two act on the source term $J^T X$ to yield

$$\begin{aligned} \langle \langle X_i X_j X_i X_j \rangle_x \rangle &= \left. \frac{d^4}{dJ_i dJ_j dJ_i dJ_j} \langle Z(J) \rangle \right|_{J=0} \\ &= 4 \left. \frac{d^2 Q_1^*(J)}{dJ_i dJ_j} \right|_{J=0} \sum_a \langle X_i \tilde{X}_a \rangle_{Q_\alpha^*} \langle X_j \tilde{X}_a \rangle_{Q_\alpha^*} + 2 \left. \frac{d^2 Q_1^*(J)}{dJ_i dJ_i} \right|_{J=0} \sum_a \langle X_j \tilde{X}_a \rangle_{Q_\alpha^*} \langle X_j \tilde{X}_a \rangle_{Q_\alpha^*} \\ &\quad + 4 \left. \frac{d^2 Q_2^*(J)}{dJ_i dJ_j} \right|_{J=0} \sum_a \langle X_i X_a \rangle_{Q_\alpha^*} \langle X_j X_a \rangle_{Q_\alpha^*} + 2 \left. \frac{d^2 Q_2^*(J)}{dJ_i dJ_i} \right|_{J=0} \sum_a \langle X_j X_a \rangle_{Q_\alpha^*} \langle X_j X_a \rangle_{Q_\alpha^*} \\ &\quad + 2 \langle X_i X_j \rangle_{Q_\alpha^*} \langle X_i X_j \rangle_{Q_\alpha^*} + \langle X_i X_i \rangle_{Q_\alpha^*} \langle X_j X_j \rangle_{Q_\alpha^*}. \end{aligned} \quad (27)$$

The latter two terms in Eq. (27) correspond to the trivial Wick decomposition from the Gaussian part of the theory at $Q_\alpha^* = Q_\alpha^*(J = 0)$, whereas the second derivatives of saddle points $\left. \frac{d^2 Q_\alpha^*(J)}{dJ_i dJ_j} \right|_{J=0} = \mathcal{O}(1/N)$ evaluated at zero source determine the non-vanishing fourth cumulants. Note that no quartic derivatives of saddle points appear in Eq. (27) due to the saddle-point condition $\frac{\partial}{\partial Q_\alpha^*} \langle Z(Q_1^*, Q_2^*, J) \rangle = 0$.

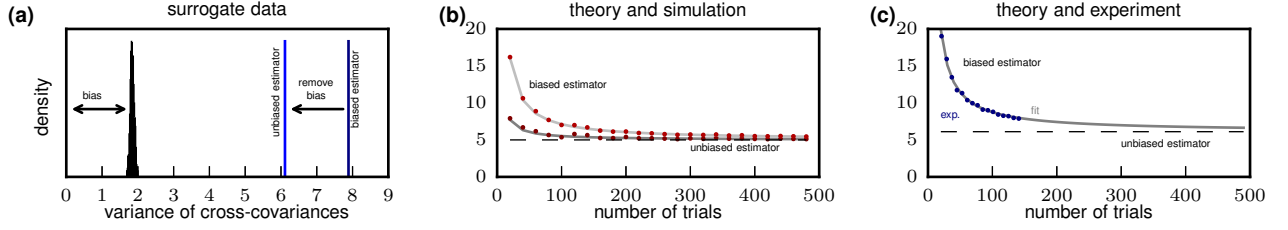


Figure S2. **Assessment and correction of the bias in estimator of variance of cross-covariances.** (a) Variance of cross-covariances (black distribution) obtained from trial-shuffled data compared to the biased estimator (Eq. (33), dark blue) and unbiased estimator (light blue) of the true variance as obtained from experimental data. (b) Convergence of biased estimator for small (dark gray) and large (light gray) mean covariances towards true variance given by unbiased estimator (dashed line). Curves indicate the theoretical prediction Eq. (33), dots numerical simulations for artificial data. (c) True value 6.11 of variance of cross-covariances (dashed line) derived from fitting (gray curve) Eq. (33) to the biased estimator (blue dots) obtained from subsampled experimental data with different numbers of trials.

F. Mean and variance of the covariance distribution

We obtain the mean integral covariances (see Eq. (26))

$$\overline{c_{ij}} = \left[(1 - \mu)^{-1} \frac{D}{1 - R^2} (1 - \mu^T)^{-1} \right]_{ij} = D_r \gamma_{ij} \quad (28)$$

and the variance of integral covariances (see Eq. (14), Eq. (15) and Eq. (27))

$$\overline{\delta c_{ij}^2} = R^2 \left[\frac{1}{(1 - R^2)^2} + \frac{1}{1 - R^2} \right] D_r^2 \chi_{ij}. \quad (29)$$

The mean connectivity μ enters the coefficients $\gamma_{ij} = \delta_{ij} + \gamma$ (with δ_{ij} the Kronecker symbol), $\chi_{ij} = \frac{1}{N} (1 + \delta_{ij} + \mathcal{O}(1/N))$ and $\gamma = \mathcal{O}(1/N)$ only in their sub-leading corrections, and acts as a negative feedback in inhibitory or inhibition-dominated networks [6]. While this feedback suppresses mean cross-covariances (28) which consequently scale as $\overline{c_{ij}} \sim \frac{1}{N}$, it only yields a subleading contribution to the dispersion (29). The spread of individual cross-covariances is determined by fluctuations in connection weights and shows a scaling as $\sqrt{\overline{\delta c_{ij}^2}} \sim \frac{1}{\sqrt{N}}$. These fluctuations, which formally originate from the variability of the auxiliary fields Q , are therefore much larger than the mean; they cause broad distributions of cross-covariances of both signs even in a homogeneous network. The expressions therefore explain the first two moments of the experimentally observed distribution of cross-covariances: mean cross-covariances scale as $\mathcal{O}(1/N)$, whereas the standard deviation only scales as $\mathcal{O}(1/\sqrt{N})$. The width of the distribution is thus much larger than the mean for large networks and the distribution is centered approximately around zero.

We note that this approach is inherently different from mean-field theories for single realizations of network connectivities which keep the site dependence to infer relations between covariances and connections on the level of individual neurons [46]. In contrast, we derive a relation between the statistics of the structure and the statistics of the dynamics using a quenched average. The crucial feature of the presented theory is that heterogeneity across neurons can still be extracted from the ensemble description. We here show that the heterogeneity is closely linked to fluctuations of the auxiliary fields, which are accessible by studying their source dependence.

II. BIAS AND ERROR OF ESTIMATING THE DISPERSION IN EXPERIMENTAL DATA

Due to limited numbers of trials, the estimator for the dispersion of the covariance may be biased towards larger variances. We seek to find a correction for this bias, as illustrated in Fig. S2a. To this end we derive the dependence of experimentally estimated moments of the activities on the number of neurons and trials. The derivation here follows the standard approach of correcting the estimation, sometimes called Bessel's correction [47].

We assume a probability distribution $p(n^1, \dots, n^{N_T})$, $n^k \in \mathbb{N}_0^N$, of activities n_i^k of neuron i in trial $k \in \{1, \dots, N_T\}$. We assume that the spike counts are sufficiently large so that their statistics is to leading order described by its first two cumulants, which is in line with our findings in Section I. We further assume that there are no correlations

between different trials, so that the probability distribution factorizes over trials and the moment generating function reads

$$\phi(l^1, \dots, l^{N_T}) = \prod_{k=1}^{N_T} \exp \left(m^T l^k + \frac{1}{2} l^{k,T} c l^k \right), \quad (30)$$

where m is the vector of mean activities with mean $\overline{m}$ and variance $\overline{\delta m^2}$ across neurons, c is the covariance matrix with mean autocovariance $\overline{a}$, variance of autocovariances $\overline{\delta a^2}$, mean cross-covariance $\overline{c}$ and variance of cross-covariances $\overline{\delta c^2}$ across neurons. The meta-statistics are assumed to be identical across trials. Furthermore, we assume the meta-statistics to be the same for all realizations of distributions of mean activities and covariances across neurons. In the following, $\langle \rangle$ denotes the average over these realizations obtained from the averaged moment generating function

$$\begin{aligned} \langle \phi(l^1, \dots, l^{N_T}) \rangle &= \exp \left(\sum_{k=1}^{N_T} \left(\overline{m} \sum_i l_i^k + \frac{1}{2} \overline{a} \sum_i l_i^k l_i^k + \frac{1}{2} \overline{c} \sum_{i \neq j} l_i^k l_j^k \right) \right. \\ &\quad \left. + \sum_{k,l=1}^{N_T} \left(\frac{1}{2} \overline{\delta m^2} \sum_i l_i^k l_i^l + \frac{1}{8} \overline{\delta a^2} \sum_i l_i^k l_i^k l_i^l + \frac{1}{8} \overline{\delta c^2} \sum_{i \neq j} l_i^k l_j^k l_i^l \right) \right) \end{aligned} \quad (31)$$

and $\hat{\cdot}$ denotes the empirical estimates of mean activities and covariances from N recorded neurons in N_T trials of the experiment. Note that the average across realizations formally introduces correlations between different trials (mixed k, l terms in (31)) which allow us to calculate corrections due to the finite number of trials. Using Eq. (31), it is straight-forward to show and well known that an empirical covariance defined as $\hat{c}_{ij}^b = \frac{1}{N_T} \sum_{k=1}^{N_T} (n_i^k - \hat{m}_i)(n_j^k - \hat{m}_j)$ yields a biased estimator:

$$\begin{aligned} \langle \hat{c}_{ij}^b \rangle &= \frac{1}{N_T} \sum_{k=1}^{N_T} \langle (n_i^k - \hat{m}_i)(n_j^k - \hat{m}_j) \rangle = \frac{1}{N_T} \sum_{k=1}^{N_T} \langle n_i^k n_j^k \rangle - \langle \hat{m}_i \hat{m}_j \rangle = \frac{1}{N_T} \sum_{k=1}^{N_T} \langle n_i^k n_j^k \rangle - \frac{1}{N_T^2} \sum_{k,l=1}^{N_T} \langle n_i^k n_i^l \rangle \\ &= \frac{N_T - 1}{N_T} (\delta_{ij} \overline{a} + (1 - \delta_{ij}) \overline{c}) \end{aligned} \quad (32)$$

with $\hat{m}_i = \frac{1}{N_T} \sum_{k=1}^{N_T} n_i^k$ and $\langle n_i^k n_j^l \rangle = \delta_{ij}(\delta_{kl} \overline{a} + \overline{\delta m^2}) + (1 - \delta_{ij}) \delta_{kl} \overline{c} + \overline{m^2}$. The unbiased estimator is therefore given by $\hat{c}_{ij} = \frac{N_T}{N_T - 1} \hat{c}_{ij}^b = \frac{1}{N_T - 1} \sum_{k=1}^{N_T} (n_i^k - \hat{m}_i)(n_j^k - \hat{m}_j)$ [47]. Using this definition, along the same lines a lengthy, but straightforward analogous calculation shows that the variance of cross-covariances ($i \neq j$) defined as $\hat{\delta c^2} = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N (\hat{c}_{ij} - \hat{c})^2$ with the mean cross-covariance across neurons $\hat{c} = \frac{1}{N(N-1)} \sum_{i \neq j=1}^N \hat{c}_{ij}$ yields the biased estimator

$$\langle \hat{\delta c^2} \rangle = \left(1 - \frac{2}{N(N-1)} \right) \left(\overline{\delta c^2} + \frac{\overline{a^2} - \overline{c^2}}{N_T - 1} \right) \quad (33)$$

of the true variance of cross-covariances $\overline{\delta c^2}$. For a finite number of trials, there is a significant bias of $\langle \hat{\delta c^2} \rangle$ caused primarily by the average variance $\overline{a}$ of spike-counts across trials (see Fig. S2a for an empirical estimate of the bias with trial-shuffled data).

The aim is hence to correct for the bias of the estimator. This can be formally done using Eq. (33), which we validate for synthetic data in Fig. S2b. A fit of Eq. (33) to the variance of cross-covariances obtained from subsampled numbers of trials of the experimental data shows that the true variance of cross-covariances can be approximately obtained by subtracting the mean of the distribution of the surrogate data in Fig. S2a.

In conclusion, the finite number of trials yields a significant bias to the variance of cross-covariances, but not to the mean autocovariances, if properly defined (see Eq. (32)). Fitting data to Eq. (33) and extrapolating for $N_T \rightarrow \infty$ yields the unbiased estimate which we use in the main text. However, even neglecting this correction, the order of magnitude of the ratio Δ between the two estimates, which determines the spectral radius (Eq. (4) in the main text) and the operational regime, is not changed by the bias.

In addition to the bias in the estimation of the variance, the estimator comes with a statistical uncertainty due to the limited number of recorded neurons. We observe $n = N(N-1)$ cross-covariances of which we want to estimate the true variance. The relative standard error of the variance is therefore given by $\sqrt{2/(n-1)} \approx 0.009 < 1\%$ [48]. Error propagation to the estimation of R^2 shows that the relative error of R^2 is even smaller for $R^2 > 1/3$, and hence in the critical regime close to $R^2 \simeq 1$ this error is negligible.

III. NUMERICAL SIMULATIONS

A. Figure 2

To illustrate the general mechanism relating network heterogeneity, eigenvalues, and distributions of covariances, we consider a simple network model with a sparsely and randomly connected inhibitory population of size $N = 1000$ and covariance of single-unit fluctuations $D = 1$. Connections $W_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{B}(p, w)$ are drawn independently from the same Bernoulli distribution with connection probability $p = 0.1$ and weight $w = -1/\sqrt{N}$ (first row), $w = -2/\sqrt{N}$ (second row), $w = -2.5/\sqrt{N}$ (third row) and $w = -3/\sqrt{N}$ (fourth row).

B. Figure 3

All networks have sparse random connections of uniform strength drawn from Bernoulli distributions. Furthermore, neurons have a fixed number of incoming connections (in-degree) and do neither connect to themselves (no autapses) nor are they connected multiple times to other neurons (no multapses). The maximum eigenvalue on the abscissa (panels A-C) is varied by the choice of the connection strength.

Homogeneous inhibitory network: We consider inhibitory populations of size $N = 1000$ (panels A,B) or $N = 10000$ (panels C,E), $D = 1$, connection probability $p = 0.1$ and uniform non-zero weights varied in the range $w = -0.1, \dots, -0.001$ (panels A,B) or $w = -0.03, \dots, -0.003$ (panel C). Panel E shows data for $w = -0.0285$.

Excitatory-inhibitory network: We consider a network of $N_E = 8000$ excitatory and $N_I = 2000$ inhibitory neurons with $D = 1$, connection probability $p = 0.1$, uniform excitatory connection strengths varied in the range $w_E = 0.001, \dots, 0.01$ and uniform inhibitory connection strengths varied in the range $w_I = -0.06, \dots, -0.006$. Panel F shows data for $w_E = 0.009$ and $w_I = -0.05$.

Inhibitory network with distance-dependent connectivity: We consider a network of $N = 10000$ inhibitory neurons ($D = 1$) randomly positioned on a $1 \text{ mm} \times 1 \text{ mm}$ sheet. Each neuron receives $K = 100$ incoming connections of uniform strength varied in the range $w = -0.1, \dots, -0.01$. Connections are drawn from a connection profile $p(x) \sim \exp(-x^2/(2\Delta^2))$ where x is the Euclidean distance between the presynaptic and postsynaptic neuron, and $\Delta = 50 \mu\text{m}$ the space constant.

Leaky integrate-and-fire network: We consider a network of $N = 10000$ inhibitory leaky integrate-and-fire model neurons with delta-shaped postsynaptic currents. The membrane potential of each neuron follows the differential equation

$$\tau_m \frac{dV_i}{dt} = -V_i + \tau_m \sum_j J_{ij} s_j(t-h),$$

where $s_j(t) = \sum_k \delta(t - t_k^j)$ is the spike train of the j -th neuron and t_k^j denotes the k -th spike of neuron j , which occurs whenever V_j exceeds the threshold θ . The membrane potential is reset $V_j(t_k^j+) \leftarrow V_r$ to the reset potential V_r after each such event and held at this level for the absolute refractory time τ_r . The time lag h equals the resolution of the time-driven simulation. We choose a connection probability $p = 0.1$ and uniform weights varied in the range $J = -1.1, \dots, -0.1$. Neuron and simulation parameters are shown in Table T1.

C. Figure 4

Linear instability is determined by an eigenvalue close to the critical line $\text{Re}(\lambda) = 1$. For networks with close structural balance (Fig. 4 left), this eigenvalue corresponds to the population activity and arises from a marginal net excitation of order $1/N$, where N is the network size. The simplest network that shows this feature is a network of excitatory neurons. We consider a sparse, random network of $N = 1000$ neurons with independent and identically distributed connections $W_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{B}(p, w)$. The connection probability $p = 0.1$ and weight $w = 9.17/N$ are chosen such that $\mu = pw = \mathcal{O}(1/N)$ and $\lambda_{\max} = N\mu = 0.917 \lesssim 1$.

For the dynamically balanced network close to the critical point, we consider a sparse, random network of $N = 1000$ inhibitory neurons with independent and identically distributed connections $W_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{B}(p, w)$. The connection probability $p = 0.1$ and weight $w = -3.1/\sqrt{N}$ are chosen such that the theoretical prediction for the largest eigenvalue

Neuron model		
Name	Value	Description
τ_m	20 ms	membrane time constant
τ_r	2 ms	absolute refractory period
V_r	0 mV	reset potential
θ	15 mV	fixed firing threshold
E_{leak}	0 mV	leak potential
Simulation parameters		
Name	Value	Description
h	0.1 ms	simulation time step
T	1000s	simulation time after initial transients
T_{trial}	1s	time window to compute spike counts

Table T1. Specification of neuron and simulation parameters for network of LIF model neurons shown in Fig. 3 of the main text.

is $\lambda_{\max} = \sqrt{Np(1-p)w^2} = 0.93 \lesssim 1$. Due to the finite size of the network, the largest real eigenvalue, which is used in Fig. 4F, slightly differs from this result and is approximately the same as for the excitatory network ($\lambda \approx 0.917$).