Extended Anderson Criticality in Heavy-Tailed Neural Networks

Asem Wardak and Pulin Gong

School of Physics, University of Sydney, New South Wales 2006, Australia

(Dated: February 14, 2022)

We investigate the emergence of complex dynamics in networks with heavy-tailed connectivity by developing a non-Hermitian random matrix theory. We uncover the existence of an extended critical regime of spatially multifractal fluctuations between the quiescent and active phases. This multifractal critical phase combines features of localization and delocalization and differs from the edge of chaos in classical networks by the appearance of universal hallmarks of Anderson criticality over an extended region in phase space. We show that the rich nonlinear response properties of the extended critical regime can account for a variety of neural dynamics such as the diversity of timescales, providing a computational advantage for persistent classification in a reservoir setting.

Introduction.— In a diverse range of physical, biological, financial and ecological systems, complex dynamics fluctuating across multiple scales emerge from a large number of interacting, nonlinear units with heterogeneous properties. Understanding the organising principles and behavior of such complex dynamics is a longstanding topic of interest across these diverse fields [1]. In neuroscience and machine learning, neural networks with many interacting neurons likewise exhibit complex dynamics with large fluctuations which are critical for their information processing abilities on real-world inputs [2]. However, the network mechanisms and fundamental computational capabilities of complex neural dynamics remain elusive.

The classical formulation of complex dynamics in systems with many interacting elements is based on randomly connected networks of coupled, nonlinear units with homogeneous connectivity [3]. By employing mathematical approaches such as mean-field theory and random matrix theory, such networks have robustly predicted a phase of chaotic activity with global, homogeneous (i.e. delocalized) fluctuations existing adjacent to an ordered, silent regime, enabling the analysis of a wide range of systems with a characteristic scale [4]. The edge of the ordered and chaotic phases gives rise to a range of critical phenomena which is thought to be necessary for these systems to perform useful computations [5]. However, growing evidence has shown that coupling heterogeneity is widespread in complex systems such as biological [6] and artificial neural networks [7][8], underscoring the need to understand the fundamental dynamical and computational mechanisms of such heterogeneity.

Here we study the dynamics of random neural networks with heterogeneous, heavy-tailed connectivity. After describing the fixed points of the system using a Lévy mean-field approach, we develop a novel non-Hermitian random matrix theory for column-structured heavy-tailed matrices to analyze the statistical fluctuations of random neural networks around the fixed point. This theory reveals a new regime with correlated multifractal modes which are neither localized nor delocalized, but have aspects of both (Fig. [1]). Multifractality is characterized by the appearance of differing, nontrivial structures appearing simultaneously over a wide variety of scales (first proposed to explain fluctuations in turbulence by Parisi [9]), and is a hallmark of Anderson transitions (criticality) [10]. Anderson transitions were first described in the context of disordered electronic systems with localized and metallic (i.e. delocalized) phases [11], and have since been analysed in a broad sense in a wide range of systems including treelike Bethe lattices [12] and those exhibiting conventional second-order phase transitions [13].

We illustrate that the heavy-tailed heterogeneity in connectivity enables Anderson criticality to emerge in a broad parameter regime. The correlated multifractal modes characteristic of this extended critical regime are able to explain a range of realistic neural dynamics, including correlated fluctuations with low-dimensional features [13], long-range correlations [14] and a diversity of timescales [15]. Importantly, these correlated multifractal modes provide a profound computational advantage in the setting of real-time reservoir computing by allowing for a persistent form of dimensionality expansion, which is not possible in classical homogeneous systems [16].

Network model and fixed points.— We begin by extending the seminal random neural network model with interacting nonlinear units analysed by Sompolinsky et al. and others [3][4], which has the dynamics

\[ h_i(t) = -h_i(t) + g \sum_{j=1}^{N} J_{ij} \phi(h_j(t)) \]  

where \( h_i(t) \) is the input of the \( i \)-th neuron at time \( t \), \( J_{ij} \) is the strength of the connection from neuron \( j \) to neuron \( i \), and \( \phi \) is a scalar nonlinearity that determines the neural firing rate given the input. The theoretical results in this Letter apply for a wide range of \( \phi \) explained below, while the figures use \( \phi = \tanh \) for comparability with previous models [3]. The coefficient \( g \) is the gain parameter of the synaptic input. Our aim is to investigate networks with heavy-tailed heterogeneity which do not fall within the purview of the Gaussian large-size limit. Finite-size heterogeneous networks can be investigated to some degree...
using perturbative finite-order corrections to the homogeneous Gaussian limit. We instead take an approach which is exact in the large network limit by regarding each $J_{ij}$ as an independent random variable whose second moment is not finite, so that its probability density has a power-law asymptotic tail,

$$p_{J_{ij}}(x) \sim C_\alpha \frac{x^{1+\alpha}}{2N^{1+\alpha}},$$  \hspace{1cm} (2)

where $C_\alpha := \Gamma(1+\alpha) \sin(\pi\alpha/2)/\pi$ and $1 < \alpha < 2$. Such heavy-tailed connectivity has been observed in Drosophila \cite{6}, in successfully trained artificial neural networks \cite{8} and in spin-glass systems with strong disorder \cite{15}; theoretically elucidating the dynamical impact of such heterogeneity has drawn increasing attention \cite{19}. Our approach includes asymmetric Lévy matrices $J_{ij} \sim L(\alpha,0,0,1/2N)$ with independent $\alpha$-stable elements of scale parameter $1/2N$. To prevent self-interference we set $J_{ii} = 0$, which does not change the overall statistical properties of the connectivity matrix as $N \to \infty$. In the limit $\alpha \to 2$, we recover the classical rate-based random neural network \cite{3} with independent elements of variance $1/N$.

Applying the generalized central limit theorem to Eq. (1) in the large network limit yields a Lévy mean-field theory for the dynamics of the system at long times, whereby each neuron behaves as

$$h_i'(t) = -h_i(t) + \eta_i(t),$$  \hspace{1cm} (3)

where $\eta_i(t) \sim L(\alpha,0,0,g^\alpha \langle |\phi(h_j(t))|^{\alpha} \rangle_j/2)$ is a time-dependent $\alpha$-stable field generated by randomly weighted inputs from the other neurons. Setting the left hand side of Eq. (3) to zero gives the fixed points of the network, whose components are distributed as an $\alpha$-stable random variable for all values of the gain $g$. As we discover in our Jacobian analysis below, the zero fixed point is unstable so that network activity is nonzero and has a macroscopic number of nonzero fixed points at any gain $g$ for $\alpha < 2$, in contrast to homogeneous networks ($\alpha = 2$) for which the network exhibits a single stable fixed point when $g < 1$. Moreover, since the synaptic input $\eta_i(t)$ has power-law fluctuations over neurons $i$ for heavy-tailed networks with $\alpha < 2$, the synaptic input autocorrelation vector $\langle \eta_i(t)\eta_i(t+\tau) \rangle_i$ has the asymptotic tail of an $\alpha/2$-stable distribution over neurons $i$, making the averaged synaptic input autocorrelation $\langle \eta_i(t)\eta_i(t+\tau) \rangle_i, \tau$ infinite in general along with the local-field autocorrelation $\langle h_i(t)h_i(t+\tau) \rangle_i, \tau$. This non-self-averaging property of heavy-tailed network input differs greatly from classical homogeneous networks where a Gaussian dynamical mean-field theory predicts that all neurons of a sufficiently large population have the same autocorrelation function, and thus the same relaxation timescale \cite{3}. On the other hand, the heavy-tailed distribution of local-field autocorrelations over neurons in heterogeneous neural networks allows for a wide reservoir of timescales across neurons of the same population even in the large network limit.

**Network stability and heavy-tailed random matrix theory.—** To determine the local stability of the network around the fixed points predicted by the above mean-field approach, we analyze the Jacobian matrix $-I + gJ \text{diag}_j(\phi(h_j))$ obtained from Eq. (1) where $\text{diag}_j \chi_j$ denotes the diagonal matrix with entries $\chi_j$. Shifting this Jacobian matrix yields a stability matrix obtained by scaling the columns of $gJ$ by $\phi'(h_j)$, which has the form of a column-structured non-Hermitian random matrix \cite{21}. Since the heavy-tailed matrix $J$ (and thus $J \text{diag}_j \chi_j$) has a locally treelike structure \cite{22}, we develop a new cavity approach for column-structured non-Hermitian heavy-tailed random matrices to obtain the spectral density and eigenvector localization properties of the Jacobian \cite{23}. In recent years, cavity approaches have been used in asymmetrically disordered contexts involving non-Hermitian random ensembles by mapping the problem back to a symmetric, Hermitian system of twice the dimensionality \cite{24}. Our cavity approach to the column-structured matrix $J \text{diag}_j \chi_j$ for any $\chi_j$ with $\langle |\chi_j|^{\alpha} \rangle_j < \infty$ yields the spectral density

$$\rho(z) = \frac{y^2 - 2|z|^2 \partial_x |z|^2 y_x}{\pi} \left\langle \frac{|\chi_i|^2 SS'}{|z|^2 + |\chi_i|^2 y_x S S'} \right\rangle_i$$  \hspace{1cm} (4)

and inverse participation ratio (see Ref. \cite{23} for mathematical derivations) describing the spatial localization of the left and right eigenvectors as a function of eigenvalue modulus $|z|$. $\langle \cdot \rangle_i$ denotes averaging over $i$ and any relevant random variables, while $S, S' \sim$...
$L(\alpha/2,1,0,C_\alpha/4C_{\alpha/2})$ are independent, skewed $\alpha/2$-stable random samples, and $y_\alpha$ is found by solving the equation

$$1 = \left\langle \left( \frac{|\chi_i|^2 S}{|z|^2 + y_\alpha^2 |\chi_i|^2 SS' \chi_i} \right)^{\alpha/2} \right\rangle_i.$$  

(5)

Our random matrix theory thus unifies both classical results in column-structured non-Hermitian random matrix theory [24, 25] for which $\alpha = 2$ and the random variables $S, S'$ reduce to the constant 1, and unstructured heavy-tailed non-Hermitian random matrices [22] for which $\chi = 1$.

Figure 2(a) shows the eigenvalue density $\rho(z)$ of the Jacobian and its numerical validation as a function of eigenvalue modulus $|z|$. A key characteristic of heavy-tailed neural networks is the infinite spectral radius $r_0$, given by the point at which $y_\alpha = 0$, so that the zero fixed point only occurs for zero gain. However, since the eigenvalue density of the heavy-tailed Jacobian is exponentially suppressed at large radius $|z|$, a characteristic spectral radius $r_p$ can be defined by the eigenvalue modulus $|z|$ at which the parameter $y_\alpha$ drops to a fraction $p$ of its value at $z = 0$. Meanwhile, the spectral radius $r_0$ of the Jacobian for classical networks ($\alpha = 2$) is finite and is obtained by a spectral analysis of finite-variance block-structured random matrices [4, 24]. In this sense, $y_\alpha$ behaves as an order parameter indicating the transition between microscopic and macroscopic numbers of eigenstates relative to system size $N$ at a given eigenvalue modulus $|z|$. Thus, heavy-tailed networks exhibit a quasi-ordered regime with a fixed point near zero whose magnitude is suppressed due to the microscopic number of eigenstates above the Jacobian stability line $\text{Re } z = 1$. A continuous transition parameterized by $p \ll 1$ is then defined by the value of the gain parameter $g$ at which the characteristic spectral radius $r_p = 1$, distinguishing a quiescent phase from a chaotic regime (Fig. 2(c), lower dashed line). This transition is consistent with the point at which neural activity predicted by mean-field theory (Eq. 3) deviates significantly from zero (Fig. 2(c), red line).

An extended critical phase with correlated multifractal modes.— Using our cavity approach, we find that all of the right eigenvectors of the heavy-tailed network Jacobian around the stationary state are multifractal for sigmoidal $\phi$, a hallmark of Anderson transitions (see Ref. 23 for derivation). The activity of heavy-tailed neural networks is thus dominated by multifractal chaotic fluctuations in contrast to the spatially delocalized chaos appearing in classical models. This theoretical prediction on heavy-tailed network dynamics is confirmed by simulations of temporal fluctuations of homogeneous and heavy-tailed neural activity (Fig. 2(d–e), red) which are both chaotic at $g = 1.75$. This result is remarkable from a physical standpoint as the activity itself is delocalized due to the bounded activation function $\phi = \tanh$, and does not visibly differ significantly from classical networks with Gaussian dynamics (see Ref. 23).

To investigate the behavior of the system’s multifractal modes over long timescales, we quantify the extent to which Jacobian eigenvalues $\lambda_i$ corresponding to a given eigenvector change their modulus relative to unity when new samples are chosen from the stationary distribution of neural activity. We thus consider the Jacobian average

$$\langle (|\lambda_i| - 1)^n \rangle_i = \int \rho(z) d z$$

(6)

which penalises small and rewards large eigenvalues to a degree determined by $n$, and we find a region for the gain $g$ adjacent to and above the ordered transition line which is characterised by a greater proportion of eigenvalues away from zero compared to the ordered transition (Fig. 2(c), upper dashed lines). This allows us to distinguish between an active chaotic region in which unstable fluctuations tend to be quickly suppressed in favour of new fluctuations, and a region of temporally correlated chaotic fluctuations. The continuous transition between these active and correlated chaotic regimes is parameterized by the annealing strength $n$ in Eq. 6. The theoretically predicted correlated region closes into the well-known critical point at the ordered-chaotic phase transition ($g = 1$) for classical rate-based networks ($\alpha = 2$), supporting the notion that the extended region of correlated multifractal modes is a critical regime. This extended critical phase is characterized by a significantly nonzero stationary state and a macroscopic proportion of unstable eigenstates relative to system dimensionality $N$, which is fundamentally different from the classical edge of chaos occurring around the zero fixed point when a microscopic proportion of eigenstates crosses the stability line. Consequently, this extended critical phase remains chaotic rather than existing solely at the edge of a chaotic phase as in classical networks [4].

In summary, the extended critical phase of temporally correlated, spatially multifractal fluctuations provides a demonstration of how various aspects of realistic neural dynamics may be exhibited simultaneously, such as long-range correlations [14] and low spatial dimensionality relative to system size [13]. This latter property arises from the localization of spatially multifractal fluctuations onto a small number of sites relative to system size (Fig. 1). Such behavior, along with the non-self-averaging properties characteristic of Anderson criticality [23] which we derived for local-field autocorrelation, suggests that the timescales across neurons in the heavy-tailed neural network are diverse. To validate this theoretical prediction on the extended critical region, we compute the relaxation timescales of neural autocorrelations over random networks across heavy-tailed index $\alpha$ and network gain $g$ (Fig. 2(c)). We find that the extended critical phase is characterized by diverse timescales (Fig. 2(b)): the
relaxation timescale distribution is power-law with index $-2.0$, consistent with that seen in cortical memory traces [15]. This behavior only exists in the critical regime [23].

**Persistent reservoir computing.**— To explore the computational implications of the extended critical regime of correlated multifractal chaos, we consider a reservoir computing task described in [16] exploiting the chaotic dimensionality expansion of neural representations $\Delta H_{12} = H^{(s)}_{12} - H^{(n)}_{12}$ between signal ($H^{(s)}_{12}$) and noise ($H^{(n)}_{12}$) distances (see Ref. [23] for setup details). Heavy-tailed networks make use of the temporal correlations of multifractal Jacobian eigenvectors above the stability line to enact a persistent form of real-time computation in a reservoir computing context. Because the extended critical regime is chaotic in heavy-tailed networks with a significantly nonzero fixed point, this regime is able to perform dimensionality expansion on its input in contrast to the classical edge of chaos which resides around the zero fixed point. At the same time, the correlated multifractal Jacobian eigenvectors work to hold off the onset of mixing (Fig. 3, yellow) to perform persistent chaotic dimensionality expansion on its input, allowing the computed result to remain in the system and thus the classification performance to stay above the baseline (0.02) for a longer period of time (Fig. 3, bottom). The extended critical regime of correlated multifractal chaos is thus able to produce efficient neural representations balancing the dimensional compression of stimuli (Fig. 3 top, red and blue), which is useful for generalization [24], and the separation of stimuli, in order to enact a form of persistent real-time computation.

**Discussion.**— Our theory rigorously demonstrates that heterogeneous, heavy-tailed connectivity can endow neural circuits with Anderson criticality over an extended parameter region, thus eliminating the fine-tuning needed in homogeneously connected neural networks. The Anderson criticality is characterized by correlated, low-dimensional fluctuations [13, 14] and a diverse reservoir of timescales [15] as observed in biological neural systems; these observations otherwise remain unexplained in a unified manner in conventional theories. Moreover, the extended Anderson criticality provides a unique mechanism for combining robust real-time computation with long-term memory of the computed output. Both homogeneous and heavy-tailed networks utilize chaos to enhance the separation of inputs for linear classification, but after a transient period of high separability, mixing dominates and erases the computed output from the classical homogeneous system. The recent observations on the ubiquity of heavy-tailed coupling in pretrained deep neural networks [8] suggest that our theory would be powerful for revealing the shared dynamical principles for persistent computation in both biological and artificial neural networks.
By making a novel link between Anderson criticality and the highly fluctuating complex dynamics of neural networks, our results suggest that complex systems operating over multiple scales should display a degree of multifractality at some level of fluctuations of system activity, even when the activity itself is bounded due to physical constraints. Multifractal phenomena have indeed been seen in a wide variety of natural systems such as turbulence. In statistical and condensed matter physics, the Anderson transition appears with multifractality around the boundary of two phases characterised by localized and delocalized eigenvectors. Our heavy-tailed neural network may thus be viewed as an inversion of these canonical models: a structurally extended Anderson regime $0 < \alpha < 2$ is bounded by delocalized Jacobian modes appearing in classical homogeneous networks with $\alpha = 2$, and localized eigenstates appearing in the limit $\alpha \to 0$ corresponding to sparse matrices and directed random graphs with a small average degree. Hence our random matrix theory could be applied to understand how complex dynamics emerge in physical systems.

---

* E-mail: pulin.gong@sydney.edu.au

[1] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. 80, 1275 (2008).
[2] H. E. Stanley and P. Meakin, Multifractal phenomena in physics and chemistry, Nature 335, 405 (1988).
[3] S. Ghoshghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Turbulent cascades in foreign exchange markets, Nature 381, 767 (1996).
[4] D. R. Chialvo, Emergent complex neural dynamics, Nature Physics 6, 744 (2010).
[5] B. Poole, S. Lahiri, M. Raghuv, J. Sohl-Dickstein, and S. Ganguli, Exponential expressivity in deep neural networks through transient chaos, in Advances in Neural Information Processing Systems, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016).
[6] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. 61, 259 (1988).
[7] J. Kadmon and H. Sompolinsky, Transition to chaos in random neuronal networks, Phys. Rev. X 5, 041030 (2015).
[8] J. Aljadeff, M. Stern, and T. Sharpee, Transition to chaos in random networks with cell-type-specific connectivity, Phys. Rev. Lett. 114, 088101 (2015).
[9] M. Stern, H. Sompolinsky, and L. F. Abbott, Dynamics of random neural networks with bistable units, Phys. Rev. E 90, 062710 (2014).
[10] K. Rajan, L. F. Abbott, and H. Sompolinsky, Stimulus-dependent suppression of chaos in recurrent neural networks, Phys. Rev. E 82, 011903 (2010).
[11] G. Wainrib and J. Touboul, Topological and dynamical complexity of random neural networks, Phys. Rev. Lett. 110, 118101 (2013).
[12] N. Bertschinger and T. Natschläger, Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks, Neural Computation 16, 1413 (2004).
[13] C.-T. Shih, Y.-J. Lin, C.-T. Wang, T.-Y. Wang, C.-C. Chen, T.-S. Su, C.-C. Lo, and A.-S. Chiang, Diverse community structures in the neuronal-level local connectome of the drosophila brain, Neuroinformatics 18, 267 (2020).
[14] L. K. Scheffer, C. S. Xu, M. Januszewski, Z. Lu, S.-y. Takekura, K. J. Hayworth, G. B. Huang, K. Shinomiya, J. Maitlin-Shepard, S. Berg, et al., A connectome and analysis of the adult Drosophila central brain, eLife 9, e57443 (2020).
[15] G. Buzsáki and K. Mizuseki, The dynamic brain: how skewed distributions affect network operations, Nature Reviews Neuroscience 15, 264 (2014).
[16] N. Perez-Nieves, V. C. H. Leung, P. L. Dragotti, and D. F. M. Goodman, Neural heterogeneity promotes robust learning, Nature Communications 12, 5791 (2021).
[17] C. H. Martin, T. S. Peng, and M. W. Mahoney, Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data, Nature Communications 12, 4122 (2021).
[18] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, On the multifractal nature of fully developed turbulence and chaotic systems, Journal of Physics A: Mathematical and General 17, 3521 (1984).
[19] F. Evers and A. D. Mirlin, Anderson transitions, Rev. Mod. Phys. 80, 1355 (2008).
[20] P. W. Anderson, Absence of diffusion in certain random lattices, Phys. Rev. 109, 1492 (1958).
[21] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson, A selfconsistent theory of localization, Journal of Physics C: Solid State Physics 6, 1734 (1973).
[22] G. Parisi, S. Pascacio, F. Pietracaprina, V. Ros, and A. Scardicchio, Anderson transition on the bethe lattice: an approach with real energies, Journal of Physics A: Mathematical and Theoretical 53, 014003 (2019).
[23] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron 99, 609 (2018).
[24] C. Huang, D. A. Ruff,
R. Pyle, R. Rosenbaum, M. R. Cohen, and B. Doiron, Circuit models of low-dimensional shared variability in cortical networks, Neuron 101, 337 (2019).

[14] B. J. He, Scale-free brain activity: past, present, and future, Trends in Cognitive Sciences 18, 480 (2014).

[15] A. Bernacchia, H. Seo, D. Lee, and X.-J. Wang, A reservoir of time constants for memory traces in cortical neurons, Nature Neuroscience 14, 366 (2011).

[16] C. Keup, T. Kühn, D. Dahmen, and M. Helias, Transient chaotic dimensionality expansion by recurrent networks, Phys. Rev. X 11, 021064 (2021).

[17] M. Helias and D. Dahmen, Statistical Field Theory for Neural Networks (Springer International Publishing, Cham, 2020); A. van Meegen, T. Kühn, and M. Helias, Large-deviation approach to random recurrent neuronal networks: Parameter inference and fluctuation-induced transitions, Phys. Rev. Lett. 127, 158302 (2021).

[18] E. Tarquini, G. Biroli, and M. Tarzia, Level statistics and localization transitions of levy matrices, Phys. Rev. Lett. 116, 010601 (2016).

[19] L. Kuśmierz, S. Ogawa, and T. Toyoizumi, Edge of chaos and avalanches in neural networks with heavy-tailed synaptic weight distribution, Phys. Rev. Lett. 125, 028101 (2020). A. Wardak and P. Gong, Fractional diffusion theory of balanced heterogeneous neural networks, Phys. Rev. Research 3, 013083 (2021).

[20] D. Schertzer, M. Larchevêque, J. Duan, V. V. Yanovsky, and S. Lovejoy, Fractional fokker–planck equation for nonlinear stochastic differential equations driven by non-gaussian levy stable noises, Journal of Mathematical Physics 42, 290 (2001).

[21] For clarity, we refer to $gJ \text{diag}(\phi'(h_j))$ as the Jacobian matrix in the remainder of the Letter as its eigenvectors remain unchanged and its eigenvalues are shifted one unit to the right.

[22] C. Bordenave, P. Caputo, and D. Chafaï, Spectrum of non-hermitian heavy tailed random matrices, Communications in Mathematical Physics 307, 513 (2011).

[23] See Supplemental Material.

[24] F. L. Metz, I. Neri, and T. Rogers, Spectral theory of sparse non-hermitian random matrices, Journal of Physics A: Mathematical and Theoretical 52, 434003 (2019).

[25] K. Rajan and L. F. Abbott, Eigenvalue spectra of random matrices for neural networks, Phys. Rev. Lett. 97, 188104 (2006). Y. Wei, Eigenvalue spectra of asymmetric random matrices for multicomponent neural networks, Phys. Rev. E 85, 066116 (2012).

[26] A. Solórzano, L. F. Santos, and E. J. Torres-Herrera, Multifractality and self-averaging at the many-body localization transition, Phys. Rev. Research 3, L032030 (2021).

[27] C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, and K. D. Harris, High-dimensional geometry of population responses in visual cortex, Nature 571, 361 (2019).

[28] F. L. Metz and I. Neri, Localization and universality of eigenvectors in directed random graphs, Phys. Rev. Lett. 126, 040604 (2021).
In this Supplemental Material, we expand on the mathematical results underpinning this Letter. We begin with a review of non-Hermitian random matrix theory and the bipartisation method. We then present new theoretical results by developing a cavity approach on column-structured non-Hermitian heavy-tailed random matrices, allowing us to analytically treat the Jacobian matrices in the main text. Finally, we present some dynamical Lévy mean-field derivations which apply to heavy-tailed random neural networks, along with additional background information on some of the results of the Letter.

In the following, the notation \( \langle f(x_i) \rangle_i = \frac{1}{N} \sum_{i=1}^{N} f(x_i) \) applies in the limit of large matrix size.

**CONTENTS**

I. Elements of non-Hermitian random matrix theory 1
   A. Hermitian matrices 2
   B. Bipartised matrix: quaternionic resolvent 2

II. The cavity method for column-structured non-Hermitian heavy-tailed random matrices 3
   A. Cavity method for scalar and vector treelike matrices 3
   B. Spectral theory of column-structured non-Hermitian heavy-tailed random matrices 4
   C. Eigenvector localisation for various column structures 6

III. Additional results and details 7
   A. Lévy dynamical mean-field theory 7
   B. The chaotic fluctuations of classical homogeneous random neural networks are delocalised 8
   C. Details of the reservoir computing task 8

References 9

**I. ELEMENTS OF NON-HERMITIAN RANDOM MATRIX THEORY**

In this section we review the random matrix theory required for understanding the cavity method [1, 2]. The eigenvalues \( \lambda_k \) of an \( N \times N \) matrix \( A \) are described by the spectral density \( \rho_A(z) := \langle \delta(z - \lambda_k) \rangle_k \), which can be computed using the resolvent \( G_A(z) := (A - z I)^{-1} \) via the relation \( \pi \rho_A(z) = -\partial_z \langle G_A(z)_{ii} \rangle_i \), where \( \partial_z := (\partial_x - i \partial_y)/2 \) for \( z = x + iy \). Writing the resolvent in its canonical form

\[
G_A(z) = \sum_{k=1}^{N} \frac{\langle v_k | u_k \rangle}{\lambda_k - z}
\]  

where \( A |v_k \rangle = \lambda_k |v_k \rangle \) and \( \langle u_k | A = \lambda_k \langle u_k | \) are the right and left eigenvectors of \( A \), one may evaluate the resolvent in the neighbourhood of an eigenvalue to deduce the correlations between the corresponding left and right eigenvectors.

* pulin.gong@sydney.edu.au
A. Hermitian matrices

This canonical form simplifies for a Hermitian matrix $H$ to

$$G_H(\lambda + i\epsilon)_{ii} = \sum_{k=1}^{N} \frac{|\langle i|v_k \rangle|^2}{\lambda_k - \lambda - i\epsilon} = \sum_{k=1}^{N} \frac{|\langle i|v_k \rangle|^2(\lambda_k - \lambda + i\epsilon)}{(\lambda_k - \lambda)^2 + \epsilon^2} \xrightarrow{\epsilon \to 0} G_H(\lambda)_{ii} + i\pi N \rho(\lambda)|\langle i|v(\lambda) \rangle|^2,$$

yielding a simpler form of the spectral density $\pi\rho_H(\lambda) = \lim_{\epsilon \to 0}(\text{Im } G_H(\lambda + i\epsilon)_{ii})$, along with the inverse participation ratio $\text{IPR}_q : = \sum_i |\langle i|v \rangle|^{2q}$ used to characterise the localisation of the eigenvectors,

$$\text{IPR}_{q,H}(\lambda) = N^{1-q} \lim_{\epsilon \to 0} \frac{\langle \text{Im } G_H(\lambda + i\epsilon)_{ii} \rangle^q}{\langle \text{Im } G_H(\lambda + i\epsilon)_{ii} \rangle^q}.$$

Denoting the fractal dimension $D_q$ by $\text{IPR}_q \sim N^{D_q(1-q)} [3]$, delocalised and localised eigenvectors are characterised by $D_q$ equalling unity and zero respectively, while multifractal eigenvectors are characterised by $D_q$ being a nontrivial function of $q$. To determine the exact value of the IPR when the eigenvectors are localised, another technique may be used by taking the lowest-order expansion of the $q$-th power of the sum when $q > 1$,

$$|G_H(\lambda + i\epsilon)_{ii}|^q = \sum_k \left| \frac{|\langle i|v_k \rangle|^{2q}}{|\lambda_k - \lambda - i\epsilon|^q} \right| + \cdots = \sum_k \frac{|\langle i|v_k \rangle|^{2q}}{(\lambda_k - \lambda)^2 + \epsilon^2} + \cdots = \sum_k |\langle i|v_k \rangle|^{2q} c_{q,\epsilon} \delta(\lambda_k - \lambda)$$

where $c_{q,\epsilon} := \int_{-\infty}^{\infty} dx/(x^2 + \epsilon^2)^{q/2} = \sqrt{\pi} \epsilon^{-1-q} \Gamma(\frac{q+1}{2})/\Gamma(\frac{q}{2})$, yielding for $q > 1$ (cf. Eq. (5) of [4])

$$\text{IPR}_{q,H}(\lambda) = \lim_{\epsilon \to 0} \frac{1}{\rho(\lambda)c_{q,\epsilon}^{1/q}} \langle |G_H(\lambda + i\epsilon)_{ii}|^q \rangle_i.$$

B. Bipartised matrix: quaternionic resolvent

However, for non-Hermitian matrices whose bulk spectra have a nonempty interior, the resolvent cannot be evaluated in any neighbourhood of these eigenvalues due to the presence of nearby poles. Moreover, one cannot separate the left and right eigenvector elements in the canonical form of the resolvent. These issues can be alleviated by using the resolvent of an auxiliary matrix of twice the dimensionality via a bipartisation [1, 5] or Hermitisation [2, 6, 7] method. To see this, observe that the $2N \times 2N$ Hermitian matrix

$$H(z, \eta) = \begin{pmatrix} -\eta I_N & A - zI_N \\ A^\dagger - \bar{z}I_N & -\eta I_N \end{pmatrix}$$

has inverse

$$H^{-1}(z, \eta) = G_{H(z,0)}(\eta) = \begin{pmatrix} \eta X & X(A - zI_N) \\ (A^\dagger - \bar{z}I_N)X & \eta Y \end{pmatrix}$$

where $X = ((A - z)(A^\dagger - z) - \eta^2)^{-1}$ and $Y = ((A^\dagger - \bar{z})(A - z) - \eta^2)^{-1}$. The lower left block of $H^{-1}(z, \eta)$ is then equal to $G_{A}(z) + O(|\eta|^2)$ for small $|\eta|$, so that the spectral density of $A$ satisfies $\pi\rho_A(z) = -\lim_{|\eta| \to 0} \partial_{\eta} \langle H^{-1}(z, \eta)(i+N),i \rangle_i$. Meanwhile, for all $c_1, c_2 \in \mathbb{C}$,

$$H(\lambda_k(A), \eta) \begin{pmatrix} c_1 |u_k(A) \rangle \\ c_2 |v_k(A) \rangle \end{pmatrix} = -\eta \begin{pmatrix} c_1 |u_k(A) \rangle \\ c_2 |v_k(A) \rangle \end{pmatrix}.$$

Thus the left and right eigenvectors of $A$ are given by the diagonal elements of the top and bottom diagonal blocks of $H^{-1}(z, \eta)$ respectively, while the eigenvalues of $A$ are given by the diagonal of either off-diagonal block of $H^{-1}(z, \eta)$. Permuting the nodes of $H$ yields an $N \times N$ matrix $A$ with quaternionic elements

$$A_{ij} = \begin{pmatrix} 0 & A_{ij} \\ \bar{A}_{ji} & 0 \end{pmatrix}, \quad U(z, \eta) = \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix}.$$
The eigenvalues and eigenvectors of $A$ can then be recovered with a quaternionic notion of the resolvent, $G_A(U) := (A - U \otimes I_N)^{-1}$, which is related to the usual resolvent via $G_A(U(z, \eta)) = G_{A - U(z,0) \otimes I_N}(\eta)$. To see this, denote the diagonal elements of the quaternionic resolvent by

$$
G_A(U)_{ii} = \begin{pmatrix} a_i(z, \eta) & b_i(z, \eta) \\ b_i^*(z, \eta) & c_i(z, \eta) \end{pmatrix}.
$$

(11)

If $\eta \in i\mathbb{R}^+$, then $a_i, c_i \in i\mathbb{R}^+$ (i.e. $a_i = i|a_i|$), and $b_i^* = \overline{b_i}$ (Lemma 2.2 of [5]). Then $\pi \rho_A(z) = -\lim_{t \to 0} (\partial_z b_i(z, it))_i$ and the inverse participation ratios of the left and right eigenvectors of $A$ are

$$
\text{IPR}^{(l)}_{q,A}(z) = N^{1-q} \lim_{t \to 0} \frac{\langle \text{Im} a_i(z, it) \rangle_i}{\langle \text{Im} a_i(z, it) \rangle_i^2}, \quad \text{IPR}^{(r)}_{q,A}(z) = N^{1-q} \lim_{t \to 0} \frac{\langle \text{Im} c_i(z, it) \rangle_i}{\langle \text{Im} c_i(z, it) \rangle_i^2}.
$$

(12)

II. THE CAVITY METHOD FOR COLUMN-STRUCTURED NON-HERMITIAN HEAVY-TAILED RANDOM MATRICES

In this section we introduce a new cavity method for obtaining the resolvent of random matrices obtained by scaling the columns of an asymmetric heavy-tailed random matrix by a set of values, deterministic or random, such as those appearing in Jacobian matrices of heavy-tailed random neural networks. This will be used to compute the spectral density and eigenvector localisation of the Jacobian matrices of heavy-tailed neural networks considered in the main text.

To compute the diagonal of the resolvent for matrices with scalar or vector entries, one may employ the blockwise inversion formula

$$
\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} s_d^{-1} & -s_d^{-1}bd^{-1} \\ -d^{-1}cs_d^{-1} & s_a^{-1} \end{pmatrix},
$$

(13)

where $s_d = a - bd^{-1}c$ is the Schur complement of $d$ and $s_a = d - ca^{-1}b$ is the Schur complement of $a$. The diagonal elements of the quaternionic resolvent of $A$ thus satisfy the recursion relation

$$
G_A(U)_{ii} = (A_{ii} - U - \sum_{j,k \neq i} A_{ij} G^{(i)}_{jk} A_{kj})^{-1},
$$

(14)

where $G^{(i)}_{jk} = G_{A^{(i)}}(U)_{jk}$ is the resolvent of the adjacency matrix $A^{(i)}$ of the graph with node $i$ removed.

A. Cavity method for scalar and vector treelike matrices

If $A$ is the adjacency matrix of a weighted tree, then so too is $A$ and the sum is restricted to $j, k \in \partial_i$, where $\partial_i$ are the neighbouring nodes of $i$. Moreover, since $A^{(i)}$ is the adjacency matrix of a forest of isolated trees with roots $j \in \partial_i$, the relation $j \neq k$ implies that $j, k$ belong to distinct trees in $A^{(i)}$ and $G^{(i)}_{jk} = 0$ (seen by permuting the blockwise inversion formula with $b, c = 0$). Thus

$$
G_A(U)_{ii} = \left(A_{ii} - U - \sum_{j \in \partial_i} A_{ij} G^{(i)}_{jj} A_{ji}\right)^{-1}.
$$

(15)

However, $G_A(z)_{jj} \neq G_{A^{(i)}}(z)_{jj}$ in general. Removing another node from $A^{(i)}$ brings the recursion to a second step:

$$
G^{(i)}_{jj} = \left(A_{jj} - U - \sum_{k \in \partial_j \setminus i} A_{jk} G^{(i)}_{kk} A_{kj}\right)^{-1},
$$

(16)

where $j \in \partial_i$. To close the recursion observe that $G^{(i,j)}_{kk} = G^{(j)}_{kk}$ in Fig. 1, and the subtrees rooted at $j$ and $k$ are
FIG. 1. The cavity method.

statistically identical over \( \partial_i \) and \( \partial_j \) respectively. Thus \((G_{ij}^{(i)})_{j \in \partial_i} \) and \((G_{jk}^{(j)})_{k \in \partial_j \setminus i}\) have the same distribution for all \( j' \in \partial_i \). In particular,

\[
\langle f(G_{ij}^{(i)}) \rangle_{j \in \partial_i} = \langle f(G_{kk}^{(j)}) \rangle_{k \in \partial_j \setminus i}
\]

for any function \( f \).

The diagonal elements \( G_{\mathbf{A}}(U)_{ii} \) of the quaternionic resolvent are thus obtained by solving the recursive distributional equation

\[
\begin{pmatrix}
    a_{ij}^{(i)} \\
    b_{ij}^{(i)} \\
    c_{ij}^{(i)}
\end{pmatrix} = -\begin{pmatrix}
    \eta & z - A_{jj} \\
    \bar{z} - A_{jj} & \eta
\end{pmatrix} + \sum_{k \in \partial_j \setminus i} \left( \frac{|A_{jk}|^2 c_{kj}^{(j)}}{A_{kj} A_{jk}} A_{kj} \right) \left( \frac{|A_{kj}|^2 a_{kj}^{(j)}}{A_{kj} A_{jk}} A_{kj} \right)^{-1}.
\]

(18)

B. Spectral theory of column-structured non-Hermitian heavy-tailed random matrices

Consider the case \( \eta = it \in i \mathbb{R}_+ \) and let \( A_{ij} = J_{ij} \chi_j \) be the column-structured non-Hermitian heavy-tailed random matrix defined in the main text. Assume that \( \langle |\chi_i|^{\alpha} \rangle_i \) is finite and, if \( \chi \) is a random variable, that \( \chi_j \) is independent of \( J_{ij} \). Since \( J \) converges to a random operator on a tree for large \( N \) [5], and

\[
p_{xx}(x) \xrightarrow{x \to \infty^+} \frac{c_\alpha}{2n|x|^{1+\alpha/2}} = \frac{c_\alpha}{2c_\alpha/2} \frac{c_\alpha/2}{n|x|^{1+\alpha/2}}
\]

(19)

for \( x > 0 \), the generalised central limit theorem gives the recursive distributional equation

\[
\begin{pmatrix}
    a_{ij}^{(i)} \\
    b_{ij}^{(i)} \\
    c_{ij}^{(i)}
\end{pmatrix} = \frac{1}{D} \left( i|\chi_j|^2 \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a \langle |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c \right)
\]

(20)

in the limit \( t \to 0 \), where \( S_a^{(j)}, S_c^{(j)} \sim L(\alpha/2, 1, 0, C_\alpha/4C_{\alpha/2}) \) are independent (as in the operator limit \( A_{ij} = 0 \) whenever \( A_{ij} \) is not [5]), and

\[
D = |z|^2 + |\chi_j|^2 \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a^{(j)} \langle |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c^{(j)}.
\]

(21)

The equivalence of the distributions \((G_{ij}^{(i)})_{j \in \partial_i}\) and \((G_{jk}^{(j)})_{k \in \partial_j \setminus i}\) for any node \( j' \in \partial_i \) gives the following coupled equations for \( \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} \) and \( \langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} \),

\[
\langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} = \langle |a_{ij}^{(i)}|^{\alpha/2} \rangle_{j \in \partial_i} = \left( \frac{|\chi_j|^2 \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a^{(j)} \langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c^{(j)}}{|z|^2 + |\chi_j|^2 \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a^{(j)} \langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c^{(j)}} \right)^{\alpha/2},
\]

(22)

\[
\langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} = \langle |\chi_j|^\alpha |c_{ij}^{(i)}|^{\alpha/2} \rangle_{j \in \partial_i} = \left( \frac{|\chi_j|^2 \langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a^{(j)} \langle |\chi_j|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c^{(j)}}{|z|^2 + |\chi_j|^2 \langle |a_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_a^{(j)} \langle |\chi_k|^\alpha |c_{ij}^{(j)}|^{\alpha/2} \rangle_{k \in \partial_j \setminus i} S_c^{(j)}} \right)^{\alpha/2},
\]

(23)
The invariance of these two equations under the symmetry $\langle |a(j)_k|^\alpha/2 \rangle_{k \in \partial_j \backslash i} \leftrightarrow \langle |x_k|^\alpha |c(j)_k|^\alpha/2 \rangle_{k \in \partial_j \backslash i} =: y_*$, where

$$y_*^{\alpha/2} = \left( \left( \frac{|y_*|^2 |\chi_j|^2 S}{|z|^2 + |\chi_j|^2 S^2} \right)^{\alpha/2} \right)_{j \in \partial_i}.$$  \hspace{1cm} (24)

Hence the diagonal elements of the quaternionic resolvent are

$$\left( a_i b_i \right) = \lim_{t \to 0} \frac{1}{|z|^2 + (t + y_* |\chi_i|^2 S)(t + y_* S^2)} \left( it + iy_* |\chi_i|^2 S - z \bar{z} \right),$$ \hspace{1cm} (25)

This gives

$$\rho(z) = \frac{1}{\pi} (y_*^2 - 2|z|^2 y_* \partial_\partial y_*) \left( \frac{|\chi_i|^2 SS'}{|z|^2 + |\chi_i|^2 y_* S^2 S'} \right).$$  \hspace{1cm} (26)

for the eigenvalues of $A$ and

$$\text{IPR}^{(l)}_q(z) = c_q N^{1-q} \left( \frac{|y_*|^2 |\chi_i|^2 S}{|z|^2 + |\chi_i|^2 S^2} \right)^q,$$ \hspace{1cm} (27)

$$\text{IPR}^{(r)}_q(z) = c_q N^{1-q} \left( \frac{|y_*|^2 |\chi_i|^2 S}{|z|^2 + |\chi_i|^2 y_* S^2 S'} \right)^q,$$

for the left and right IPRs, where $c_q$ is a constant factor determined by the local eigenvalue statistics. To compute $\partial_\partial y_*$, use the implicit function theorem. Defining

$$G(y, r) = \left| \frac{|\chi_i|^2 S}{r + y^2 |\chi_i|^2 S^2} \right|^{\alpha/2}$$  \hspace{1cm} (28)

yields $\partial_\partial y_* = -\langle \partial_y G(y_*,|z|^2) \rangle/\langle \partial_y G(y_*,|z|^2) \rangle$, where

$$\partial_y G(y, r) = \frac{-\langle \alpha/2 |\chi_i|^{\alpha/2} S \rangle}{(r + y^2 |\chi_i|^2 S^2)^{\alpha/2+1}}, \hspace{1cm} \partial_y G(y, r) = 2y |\chi_i|^2 S S' \partial_r G(y, r).$$  \hspace{1cm} (29)

The density of eigenvalues at $z = 0$ is

$$\rho(0) = y_*^{-2} |\chi_i|^{-2} \langle (S^{-1})^2 \rangle,$$  \hspace{1cm} (30)

which may be infinite due to $\chi$. To obtain $y_*^{-2}|z| = 0$, observe that $y_*^{\alpha}|z| = \langle S^{-\alpha/2} \rangle$ along with the fact that if $S_D \sim L(\beta, 1, 0, D)$ with $\beta < 1$, then

$$\langle S_D^{-\beta} \rangle = \frac{\cos(\beta \pi/2)}{D \Gamma(1 + \beta)}.$$  \hspace{1cm} (31)

Noting the constraint $\langle |\chi_i|^\alpha \rangle_i < \infty$, it can be shown using similar techniques to [5] that the spectral radius $r_0$ is in general infinite, and that the eigenvalue density is exponentially suppressed at large modulus in stark contrast to the power-law distribution of singular values.

An analogous derivation for the classical finite-variance case $\alpha = 2$ yields the above results with the transformation $S \to 1$. In this case, the eigenvalue density for column-structured matrices of finite-variance elements is recovered along with the spectral radius $\langle |\chi_i|^2 \rangle_i^{1/2}$ as known previously [8–10], while a new result on the localisation of column-structured matrices of finite-variance elements arises from the analogue to Eq. (27).

1 In the case of block-structured connectivity $g_{i,i} \mapsto g_{i,c_j} J_{j,i}$, which describes multipopulation neural networks with varying cell types, this fixed-point equation for $y_*$ is replaced by a set of coupled equations $y_*, c$ over the cell types.
C. Eigenvector localisation for various column structures

Let us now characterise the localisation properties of the IPR for various distributions of $\chi$. In the case of non-Hermitian heavy-tailed random matrices [5] ($\chi = 1$), the left and right eigenvectors are delocalised and can be directly evaluated using Eq. (27). However, since many common sigmoidal transfer functions are bounded or flatten out for large inputs, and local-field stationary distributions have power-law tails, $\chi = \phi'/(h)$ is more likely to have significant mass near zero.

First consider the test case $^2 P(\chi_i = 0) > 0$ and consider the sum $\sum_{i=1}^N |c_i(z,i0)|^q$ using Eq. (25),

$$\sum_{i=1}^N |c_i(z,i0)|^q = \sum_{i=1}^N |\text{Im} c_i(z,i0)|^q = \sum_{i=1}^N \left|\frac{y_i S'}{|z|^2 + t + y_i S'}^q\right|.$$

In this sum, $P(\chi_i = 0)N$ terms are of the form $(y_i S'/|z|^2)^q$, which has the asymptotic tail of an $\alpha/2q$-stable random variable if $\alpha/2q < 1$, and has finite mean otherwise. Hence the sum scales as $N^{2q/\alpha}$ if $q > \alpha/2$, and as $N$ otherwise. Thus the right IPR scales as $N^{2q/\alpha}/(N^{2/\alpha})^q = N^0$ when $q > \alpha/2$ and $N/(N^{2/\alpha})^q = N^{1-2q/\alpha}$ otherwise. As a result the multifractal dimension $D_q = 0$ when $q > \alpha/2$, and $D_q = (1-2q/\alpha)/(1-q)$ otherwise, i.e. the right eigenvectors are multifractal. Meanwhile, the left eigenvector remains delocalised and can be directly evaluated using Eq. (27).

Now consider the general case where a certain asymptotic scaling around zero. Suppose the quantity $P(|\chi|^2 S < \epsilon)$ is known for small $\epsilon$. Of the sum $\sum_{i=1}^N |c_i(z,it)|^q$, $P(|y_i|S' < t/N)N$ terms are in the interval

$$\left(\frac{|t + y_i S'|}{|z|^2 + t + y_i S'}^q\right).$$

Of these terms, $P(|y_i|S' < t/N)P(ty_i S' > |z|^2) = P(|\chi|^2 S < t/y_i)N(|z|^2/ty_i)^{-\alpha/2}$ terms have the scaling $1/t^\alpha$. Noting that the stability of the average of the diagonal elements of the scalar-valued resolvent upon the addition of a small imaginary part is classically used as a criterion for localisation for the Anderson transition [11], this average for the right eigenvector elements scales over large $N$ and small $t$ as $P(|\chi|^2 S < t)t^{\alpha/2-1}N^0$. If this average remains finite,\(^3\)

$$P(|\chi|^2 S < 1/N) = O(N^{\alpha/2-1}) \quad \text{(deloc.)}$$

then the state is considered delocalised (for example, if the density of $|\chi|^2$ at zero is finite or if $P(|\chi|^2 S < t) \sim t^a$ with $a > 1 - \alpha/2 > 0$). However, if this average is unstable for small $t$,

$$P(|\chi|^2 S < 1/N) \neq O(N^{\alpha/2-1}) \quad \text{(not deloc.)}$$

(e.g. $P(|\chi|^2 S < t) \sim t^a$ with $0 < a < 1 - \alpha/2$ or $|\chi|^2$ has sufficient mass near zero) then the state is localised to some degree as $\text{Im} G_{\chi_U}(\epsilon \Omega \phi h)_{2(1)}_{2(2)}$ diverges when $t \to 0$. In this case the second IPR definition yields the right IPR scaling to be $t^\alpha N^0$ for $q > 1$. Since any configuration of $\chi$ yields a greater or equal localisation than the test case of a delta function at zero, the right eigenvector is multifractal in this case and the multifractal dimension satisfies

$$D_q \geq \frac{1 - 2q/\alpha}{1 - q}$$

when $q < \alpha/2$ ,

$$D_q \geq 0$$

when $\alpha/2 < q < 1$ ,

$$D_q = 0$$

when $q > 1$ .

For what transfer functions $\phi$ does the multifractality condition $P(\phi'(h)^2 S < 1/N) \neq O(N^{\alpha/2-1})$ hold? If $\phi$ is sigmoidal (bounded, one inflection point, non-negative derivative), we can observe that asymptotically for large $N$

$$P(\phi'(h)^2 S < 1/N) > P(\phi'(h)^2 < 1/N)P(S < 1)$$

$$\sim P(\phi'(h) < 1/\sqrt{N})$$

$$= P(h > (\phi')^{-1}(1/\sqrt{N}))$$

$$\sim ((\phi')^{-1}(1/\sqrt{N}))^{-\alpha}$$

\(^2\) This case arises, for example, when $\phi = \Theta(x - \theta)$ is a binary activation function or has a hard threshold.

\(^3\) Note that $P(|\chi|^2 S < 1/N)$ is monotonically decreasing as $N \to \infty$, and that the exponent of $N^{\alpha/2-1}$ is negative.
The derivatives of bounded transfer functions must have a tail heavier than $x^{-1}$. Suppose that $\phi'(x) = O(x^{-p})$ as $x \to \infty$ for some $p > 1$. Then $\phi'(N^{1/2p}) = O(1/\sqrt{N})$ and so $(\phi')^{-1}(1/\sqrt{N}) = O(N^{1/2p})$ as $\phi'$ is asymptotically monotone decreasing. Therefore asymptotically

$$\mathbb{P}(\phi'(h)^2 S < 1/N) \sim ((\phi')^{-1}(1/\sqrt{N}))^{-\alpha}$$

$$\gtrsim N^{-\alpha/2p} \neq O(N^{\alpha/2 - 1})$$

if $-\alpha/2p > \alpha/2 - 1$, i.e.

$$\alpha(1 + 1/p) < 2 .$$

This has a number of important implications. Firstly, any sigmoidal transfer function is guaranteed to yield multifractal Jacobian eigenvectors in the stationary state for $\alpha \leq 1$. We have seen above that binary transfer functions or those with a hard threshold already yield multifractality for all $\alpha < 2$. To determine the multifractality of Jacobian eigenvectors for $1 < \alpha < 2$ for the remaining sigmoidal transfer functions $\phi$, it suffices to check the asymptotic behaviour of $\phi'(x)$ as $x \to \infty$. For example, the figures in this Letter have used $\phi = \tanh$, for which $\phi'(x) = \text{sech}(x)^2$. Since $\text{sech}(x) = O(e^{-x})$, it follows that $\text{sech}(x)^2 = O(x^{-p})$ for all $p > 1$, and so the stationary Jacobian eigenvectors are multifractal for all $\alpha < 2$ under the transfer function $\phi = \tanh$.

III. ADDITIONAL RESULTS AND DETAILS

A. Lévy dynamical mean-field theory

In the stationary state, a randomly selected local field

$$h_i \sim L(\alpha, 0, 0, g^\alpha (|\phi|^\alpha)/2)$$

is distributed as an $\alpha$-stable random variable whose scale parameter is computed self-consistently via

$$\langle |\phi|^\alpha \rangle = \int |\phi(h)|^\alpha p_{L(\alpha, 0, 0, g^\alpha (|\phi|^\alpha)/2)}(h) dh .$$

This gives the distribution of $(\phi'(h_i))_j$ for use in the Jacobian cavity equations. Taking the Fourier transform of the dynamic mean-field equation gives

$$i\omega h_i(\omega) = -h_i(\omega) + \sum_j J_{ij} F[\phi \circ h_j](\omega)$$

where $F$ is the Fourier operator and $h_i(\omega) := F[h_i](\omega)$. This complex equation may be broken into its real and imaginary parts to yield

$$\text{Re} \, h_i(\omega) - \omega \text{Im} \, h_i(\omega) = \sum_j J_{ij} \text{Re} \, F[\phi \circ h_j](\omega) ,$$

$$\text{Im} \, h_i(\omega) + \omega \text{Re} \, h_i(\omega) = \sum_j J_{ij} \text{Im} \, F[\phi \circ h_j](\omega) .$$
Using the independence of the real and imaginary parts of the Fourier transform of \( h_i(t) \) [12], along with the generalised central limit theorem, yields

\[
D_\alpha (\text{Re} h_i(\omega))_i + |\omega|^\alpha D_\alpha (\text{Im} h_i(\omega))_i = \langle |\text{Re} F(\phi \circ h_j(\omega))|^{\alpha} \rangle_j / 2 ,
\]

(51)

\[
D_\alpha (\text{Im} h_i(\omega))_i + |\omega|^\alpha D_\alpha (\text{Re} h_i(\omega))_i = \langle |\text{Im} F(\phi \circ h_j(\omega))|^{\alpha} \rangle_j / 2 ,
\]

(52)

where \( D_\alpha \) is the scale parameter of the \( \alpha \)-stable random variable with the same asymptotic tail as the argument. When \( \alpha = 2 \), the operator \( D_\alpha \) is reduced to become half the variance of its argument. Adding these two equations together and defining \( \Delta(\tau) = 2F_\omega^{-1}[D_\alpha (\text{Re} h_i(\omega))_i + D_\alpha (\text{Im} h_i(\omega))_i](\tau) \) (which reduces to the regular autocorrelation \( \Delta(\tau) = F_\omega^{-1}[|(h_i(\omega)|^2)_i](\tau) = \langle h_i(t)h_i(t+\tau) \rangle_i, \tau \) when \( \alpha = 2 \) gives

\[
\Delta(\tau) + (-\Delta_\tau)^{\alpha/2}\Delta(\tau) = F_\omega^{-1}[|\text{Re} F(\phi \circ h_j(\omega))|^{\alpha} + |\text{Im} F(\phi \circ h_j(\omega))|^{\alpha}](\tau) ,
\]

(53)

where \( (-\Delta_\tau)^{\alpha/2} \) is the fractional Laplacian [13]. This reduces to the classical Newtonian equation [14] when \( \alpha = 2 \),

\[
\Delta(\tau) - \frac{\partial^2}{\partial \tau^2}\Delta(\tau) = F_\omega^{-1}[|\text{Re} F(\phi \circ h_j(\omega))|^{2}](\tau) = \langle \phi(h_j(t))\phi(h_j(t+\tau)) \rangle_j, \tau .
\]

(54)

**B. The chaotic fluctuations of classical homogeneous random neural networks are delocalised**

The localization of dominant Jacobian eigenmodes determines the spatial profile of network fluctuations and is described by the inverse participation ratio \( \text{IPR}_q(v) = \sum_j |v_j|^q \) for a normalized vector \( v \). Defining the multifractal dimension \( D_q \) by the asymptotic relation \( \text{IPR}_q(v) \sim N^{(1-q)D_q} \) for large system size \( N \), localized and delocalized states correspond to \( D_q \) equalizing zero and one respectively, while multifractal states are characterized by \( D_q \) being a nontrivial function of \( q \) (Fig. 1 of the main text), which are a hallmark of Anderson transitions [3]. Standard results in random matrix theory have established that the eigenvectors of random matrices with independent sub-exponential entries are delocalized [15], which includes the Jacobian matrices of classical homogeneous random neural networks (\( \alpha = 2 \)). The chaotic fluctuations of classical random neural networks are thus delocalized, with the IPR (Fig. 2(d) of the main text, blue) staying close to the value \( N^{1-q} \) attained for a constant, maximally delocalized vector \((1/\sqrt{N}, \ldots, 1/\sqrt{N})\), and the \( D_q \) estimate staying close to that of a random \( N \)-dimensional spherical vector, which is delocalized with overwhelming probability [16] (Fig. 2(e) of the main text, blue and black); these two \( D_q \) estimates remain slightly curved due to finite-size effects.

**C. Details of the reservoir computing task**

The extended critical regime of correlated multifractal chaos has important computational implications for networks with heavy-tailed connectivity. To illustrate this, we consider a pattern classification task [17] with \( P \) fixed patterns of randomly drawn binary vectors of length \( L \). From these, \( P \) pattern classes are formed by generating noisy realizations of each pattern with the addition of independent Gaussian vectors of variance \( \sigma^2 = 0.1^2 \). Each realization forms the initial state of the first \( L \) neurons of the network with the connectivity and initial state of the remaining \( N - L \) neurons remaining quenched for the task. At each timestep, a linear readout is trained for each pattern class by
linear regression to provide a unit output for the corresponding pattern and zero for the others, with the strongest readout signal determining classification. This task relies on the signal distance $H_{12}^{(s)}$, which is the Hamming distance between patterns of different classes, being greater than the noise distance $H_{12}^{(n)}$ between patterns of the same class. The difference between the signal and noise distances, $\Delta H_{12}$, is known as the linear separability of pattern classes, which is initially low [17]. Classical random neural networks are known to undergo transient chaotic dimensionality expansion (Fig. 3(b) of the main text, yellow) to increase class separability $\Delta H_{12}$ which aids computation in a reservoir computing context. However, the onset of chaotic mixing reduces separability $\Delta H_{12}$ and thus classification performance (Fig. 3 of the main text, bottom).

[1] C. Bordenave and D. Chafaï, Around the circular law, Probability Surveys 9, 1 (2012).
[2] F. L. Metz, I. Neri, and T. Rogers, Spectral theory of sparse non-hermitian random matrices, Journal of Physics A: Mathematical and Theoretical 52, 434003 (2019).
[3] F. Evers and A. D. Mirlin, Anderson transitions, Rev. Mod. Phys. 80, 1355 (2008).
[4] F. L. Metz, I. Neri, and D. Bollé, Localization transition in symmetric random matrices, Phys. Rev. E 82, 031135 (2010).
[5] C. Bordenave, P. Caputo, and D. Chafaï, Spectrum of non-hermitian heavy tailed random matrices, Communications in Mathematical Physics 307, 513 (2011).
[6] T. Rogers and I. P. Castillo, Cavity approach to the spectral density of non-hermitian sparse matrices, Phys. Rev. E 79, 012101 (2009).
[7] I. Neri and F. L. Metz, Spectra of sparse non-hermitian random matrices: An analytical solution, Phys. Rev. Lett. 109, 030602 (2012).
[8] Y. Wei, Eigenvalue spectra of asymmetric random matrices for multicomponent neural networks, Phys. Rev. E 85, 066116 (2012).
[9] Y. Ahmadian, F. Fumarola, and K. D. Miller, Properties of networks with partially structured and partially random connectivity, Phys. Rev. E 91, 012820 (2015).
[10] J. Aljadeff, M. Stern, and T. Sharpee, Transition to chaos in random networks with cell-type-specific connectivity, Phys. Rev. Lett. 114, 088101 (2015).
[11] E. Tarquini, G. Biroli, and M. Tarzia, Level statistics and localization transitions of lèvy matrices, Phys. Rev. Lett. 116, 010601 (2016).
[12] M. Stern, H. Sompolinsky, and L. F. Abbott, Dynamics of random neural networks with bistable units, Phys. Rev. E 90, 062710 (2014).
[13] D. Schertzer, M. Larchevêque, J. Duan, V. V. Yanovsky, and S. Lovejoy, Fractional fokker–planck equation for nonlinear stochastic differential equations driven by non-gaussian lèvy stable noises, Journal of Mathematical Physics 42, 200 (2001).
[14] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Chaos in random neural networks, Phys. Rev. Lett. 61, 259 (1988).
[15] M. Rudelson and R. Vershynin, Delocalization of eigenvectors of random matrices with independent entries, Duke Mathematical Journal 164, 2507 (2015).
[16] S. O’Rourke, V. Vu, and K. Wang, Eigenvectors of random matrices: A survey, Journal of Combinatorial Theory, Series A 144, 361 (2016).
[17] C. Keup, T. Kühn, D. Dahmen, and M. Helias, Transient chaotic dimensionality expansion by recurrent networks, Phys. Rev. X 11, 021064 (2021).