Editors' Suggestion | Featured in Physics

Gell-Mann–Low Criticality in Neural Networks

Lorenzo Tiberi,[1,2,3,*] Jonas Stapmanns,[1,2,*] Tobias Kühn,[4] Thomas Luu,[3,5] David Dahmen,[1] and Moritz Helias[1,2,3]

[1]*Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, 52425 Jülich, Germany*

[2]*Institute for Theoretical Solid State Physics, RWTH Aachen University, 52074 Aachen, Germany*

[3]*Center for Advanced Simulation and Analytics, Forschungszentrum Jülich, 52425 Jülich, Germany*

[4]*Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France*

[5]*Institut für Kernphysik (IKP-3), Institute for Advanced Simulation (IAS-4) and Jülich Center for Hadron Physics, Jülich Research Centre, 52425 Jülich, Germany*

(Received 12 October 2021; revised 9 February 2022; accepted 4 March 2022; published 19 April 2022)

Criticality is deeply related to optimal computational capacity. The lack of a renormalized theory of critical brain dynamics, however, so far limits insights into this form of biological information processing to mean-field results. These methods neglect a key feature of critical systems: the interaction between degrees of freedom across all length scales, required for complex nonlinear computation. We present a renormalized theory of a prototypical neural field theory, the stochastic Wilson-Cowan equation. We compute the flow of couplings, which parametrize interactions on increasing length scales. Despite similarities with the Kardar-Parisi-Zhang model, the theory is of a Gell-Mann–Low type, the archetypal form of a renormalizable quantum field theory. Here, nonlinear couplings vanish, flowing towards the Gaussian fixed point, but logarithmically slowly, thus remaining effective on most scales. We show this critical structure of interactions to implement a desirable trade-off between linearity, optimal for information storage, and nonlinearity, required for computation.

DOI: 10.1103/PhysRevLett.128.168301

Criticality and information processing are deeply related: Statistical descriptions of the hardest-to-solve combinatorial optimization problems, for example, are right at the edge of a phase transition [1,2]. Also brain activity shows criticality [3,4], which may optimize the network's computational properties [5].

A multitude of critical processes could lie behind critical brain dynamics. Thanks to the universality paradigm of statistical physics, however, their macroscopic behavior is organized into few classes, characterized only by the structure of effective, macroscopic interactions, rather than microscopic details [6,7]. Identifying the universality classes implemented by neural networks and finding their distinctive features is key to understanding if and how the brain exploits criticality to perform computation.

Theoretical analysis of critical brain dynamics is, however, so far restricted to mean-field methods. These partly explain why memory, dynamic range, and signal separation are optimal at a critical point [8–10]. Still, a fundamental aspect of critical computation is inaccessible to these methods: the nonlinear interaction between degrees of freedom across all length scales. By approximating fluctuations as Gaussian, mean-field theory can study only the linear response of individual modes to stimuli. But a single, uncoupled mode can solve only simple computational tasks. In fact, this approximation only holds if nonlinear interactions, albeit necessary for computation, are irrelevant on macroscopic scales. Thus, mean-field criticality is a special case, in fact the simplest yet most restricted kind of criticality the brain could possibly implement. Uncovering different kinds of criticality requires more sophisticated methods.

In this Letter we analyze criticality in the stochastic Wilson-Cowan rate model, a prototypical model of brain dynamics [11–13]. We use the nonequilibrium Wilsonian renormalization group to go beyond mean-field analysis [6,7]. This technique tracks the flow of effective nonlinear couplings as one describes the system on gradually increasing length scales. This exposes the type of criticality featured by the system and its relevance for computation. The model is studied under a continuous stream of external inputs, as typical for interconnected brain areas [14,15]. While critical activity in networks driven by sparse inputs is well described by branching processes [3,16] belonging to the universality class of mean-field directed percolation [17], it is still unclear whether constantly driven brain networks are able to support criticality at all [18].

We find the Wilson-Cowan model can indeed be critical in this regime. At $d = 2$ spatial dimensions, corresponding to the planar organization of cortical networks, mean-field theory loses validity. In fact, we discover criticality of the Gell-Mann–Low kind [19] ([20], Sec. V). Nonlinear couplings here decrease logarithmically slowly, thus remaining effective on practically all length scales. This property implements a desirable balance between a linear behavior and a nonlinear one; the model optimally remembers signals presented in the past, due to its nearly linear dynamics, and at the same time can perform nonlinear classification.

Focusing on the key ingredients of neural networks, that are nonlinear dynamics of their constituents, noisy drive, and spatially localized nonlinear coupling, we consider a neural field following the Wilson-Cowan [11–13] equation

$$\tau \frac{dh}{dt} = -l(h) + w * f(h) + \sqrt{\tau} I, \tag{1}$$

where $h(x,t)$ is a neural activity field on the space $x \in \mathbb{R}^d$ that evolves in time t on the characteristic timescale τ. The function l describes intrinsic local dynamics, and f is a nonlinear gain function. The connectivity kernel $w(x-x')$ weighs the input from the neural state at position x' to that at position x, and $*$ is the spatial convolution. Following a common approach [21,22], the connectivity is isotropic $w(x) = w(\|x\|)$ and consists of the sum of two Gaussians $w_\pm (2\pi\sigma_\pm^2)^{-d/2} \exp(-\|x\|^2/2\sigma_\pm^2)$, with $\mathrm{sign}(w_\pm) = \pm 1$, representing excitatory and inhibitory connections. External input from remote brain areas driving the local activity is, for simplicity, modeled as Gaussian white noise with statistics $\langle I(x,t)\rangle = \mu$ and $\langle I(x,t) I(x',t')\rangle = D\delta(x-x')\delta(t-t')$. A microscopic length a characterizes the spatial resolution of the model, thus Eq. (1) is defined on a lattice with N^d sites and spacing a, eventually taking the limit $N \to \infty$. Equivalently, momenta are restricted to $|k| < \Lambda := \pi/a$. We are interested in the spatial dimension $d = 2$, on which the model is traditionally defined [12]. Thus, space points effectively represent cortical columns, connected along the direction of the cortical surface [23]. An isotropic space-dependence of the connectivity is also justified functionally by a topographical mapping of receptive fields; e.g., the retinotopic mapping onto primary visual cortex [24,25].

The computational properties of the model are later tested in a reservoir computing setting [26]: A linear readout is trained to extract a desired input-output mapping from the neural activity [Fig. 1(a)]. Collective nonlinear interactions are fundamental to achieve complex mappings. We thus need to go beyond mean-field methods to track the nonlinear interactions on gradually increasing length scales.

We make explicit all nonlinear terms in Eq. (1) by Taylor expanding $f(x) = \sum_n f_n x^n$ and likewise l. Also, the momentum dependence of the Fourier transformed coupling kernel $\hat{w}(k) = \sum_\pm w_\pm [1 - \frac{1}{2}\sigma_\pm^2 k^2 + \mathcal{O}(k^4)]$ is kept up to second order, enough to expose those terms that characterize the system on a mesoscopic length scale. We thus obtain, in the spatial domain,

$$\tau \frac{dh}{dt} = \sum_{n=1}^{\infty} [-m_n + g_n \Delta + \mathcal{O}(\Delta^2)] h^n + \sqrt{\tau} I, \tag{2}$$

where Δ is the Laplace operator and $\mathcal{O}(\Delta^2)$ denotes all terms proportional to spatial differential operators of

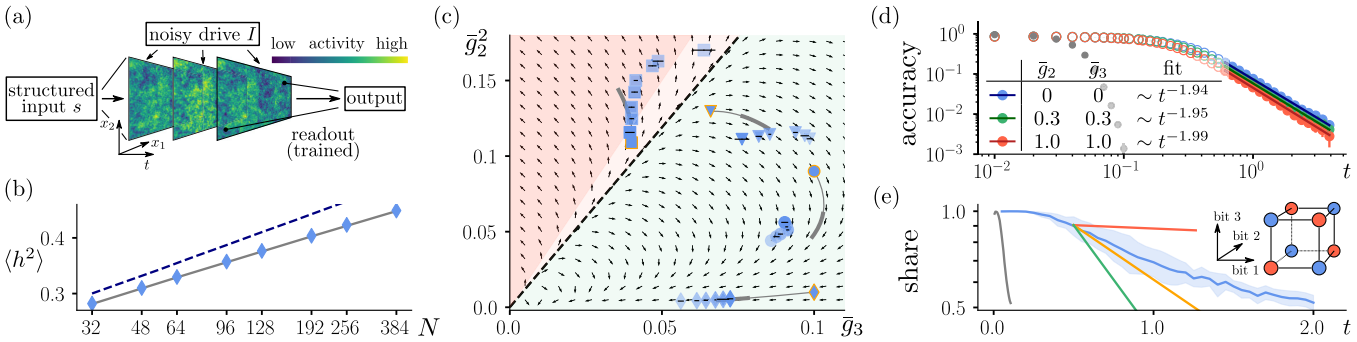

FIG. 1. (a) Model schematics. (b) Variance. Variance as a function of N for $\bar{g}_3 = 0.1$ and $\bar{g}_2^2 = 0.01$ ($\bar{g}_2^2 \ll \bar{g}_3$ regime); blue diamonds: simulation; dashed dark blue line: linear system; gray solid line: renormalized theory. (c) Flow of couplings. Direction of flow (arrows) theoretically predicted by Eqs. (5) and (6). The transition line (black, dashed) separates the converging region (green) from the diverging region (red). Region of nonphysical bistable regime in darker red. Simulated flow of couplings at $\ell = (N/2)$, with $N \in \{32, 48, 64, 96, 128, 192\}$ (dark to light shading of blue markers; squares include $N = 256$, 384) for different initial conditions (yellow edge). Predicted flow by Eqs. (5) and (6) (gray solid curves, thicker within the range of simulated ℓ). (d) Memory. Reconstruction accuracy (0: complete forgetting; 1: perfect reconstruction) over time of a Gaussian input for different strengths of the nonlinearities (legend). (e) Classification. Share of correctly assigned parity to 3-bit strings (blue curve) and standard deviation across repeated trainings (shaded area), for $\bar{g}_2 = \bar{g}_3 = 0.3$. Decay timescales of the slowest (red), middle (orange), and fastest (green) modes available at readout. Inset: spatial visualization of all 3-bit strings, colored by their parity. In (d) and (e): Gray dots or curve show performance away from criticality with a nonvanishing mass m_1, corresponding to a correlation length of a single lattice spacing.

order ≥ 4. Couplings are conveniently renamed $m_n \coloneqq l_n - f_n \sum_\pm w_\pm$, characterizing the local dynamics, and $g_n \coloneqq \frac{1}{2} f_n \sum_\pm \sigma_\pm^2 w_\pm$, quantifying the interaction across space points. Notice μ, l_0, and f_0, corresponding to a constant input causing $\langle h \rangle \neq 0$, are without loss of generality set to zero [27].

The mass term m_1 plays the role of a lower momentum cutoff [20], defining the system's spatial ($\propto m_1^{-\frac{1}{2}}$) and temporal ($\propto m_1^{-1}$) correlation lengths. Both diverge as $m_1 \to 0$, identifying a critical point of the linear model. To include the effect of nonlinearities close to this point, standard expansion methods fail: perturbative corrections diverge due to statistical fluctuations interacting on an infinite range of length scales [20].

To tackle this issue, the renormalization group (RG) [6,7,20] performs the integration of fluctuations gradually over momentum scales $\Lambda/\ell < |k| < \Lambda$; progressively increasing the flow parameter $\ell \in [1, \infty)$, we obtain a series of effective field theories, each describing only degrees of freedom on larger length scales, with $k < \Lambda/\ell$. Such theories are defined on the rescaled quantities $k_\ell \coloneqq \ell k$, $t_\ell \coloneqq \ell^{-z} t$, $\hat{h}_\ell(k_\ell, t_\ell) \coloneqq \ell^{-\zeta} \hat{h}(k, t)$, and $\hat{I}_\ell(k_\ell, t_\ell) \coloneqq \ell^{\chi} \hat{I}(k, t)$. Rescaling makes all effective field theories look formally equivalent to Eq. (2), differing only by the values of the couplings $m_n(\ell)$ and $g_n(\ell)$, which become ℓ-dependent. This dependence accounts for rescaling and for indirect interactions mediated by marginalized degrees of freedom. The couplings' flow with ℓ therefore characterizes nonlinear interactions on different length scales and, thus, the type of critical behavior featured by the system. For example, the flow running into a fixed point is a characteristic of critical systems with scale invariance.

First consider the flow due to rescaling alone. This corresponds to dimensional analysis and the mean-field approach, neglecting the contribution of fluctuations. We choose $z = 2$, $\zeta = [(d+2)/2]$, and $\chi = [(d-2)/2]$ so that g_1, τ, and the input variance D are at a fixed point (i.e., do not rescale). For $d \geq 2$ all couplings not appearing explicitly in Eq. (2) then rapidly flow to 0 with some negative power of ℓ. These couplings are termed *irrelevant* and can be neglected at large scales ℓ. The couplings $m_n(\ell) = \ell^{2-(n-1)[(d-2)/2]} m_n(1)$ diverge at $d = 2$ as $\sim \ell^2$. They are termed *relevant*, meaning they must be fine-tuned to 0 to be at a fixed point. Fine-tuning here implies balance of inhibitory and excitatory inputs [27], often observed in brain networks as a necessary condition for criticality [34]. The couplings $g_n(\ell) = \ell^{-(n-1)[(d-2)/2]} g_n(1)$ vanish for $d > 2$, $\forall\ n \geq 2$. For these spatial dimensions, mean-field theory is usually accurate: all nonlinear terms in Eq. (2) are negligible at large scales, thus fluctuations have almost no interactions and can be neglected. Dimensional analysis predicts $d = 2$ as the upper critical dimension at which the g_n instead do not scale and are thus termed *marginal*: their flow is driven by fluctuations alone and thus must be investigated with more sophisticated methods, like the RG.

The mean-field analysis above allows us to determine the form of the effective theory describing the critical system at mesoscopic scales, where irrelevant couplings are negligible

$$\tau \frac{dh}{dt} = \Delta(g_1 h + g_2 h^2 + g_3 h^3) + \sqrt{\tau} I. \tag{3}$$

At such scales, the effective field h describes neural populations exchanging activity with nearest neighbors via a diffusive process, expressed by the Laplace operator Δ [35]. Among the marginal couplings g_n, we keep only the first $n \leq n_0 = 3$. As we detail in the Supplemental Material [27], we can assume neural activity to mainly explore a limited range of the gain function [36,37], which can therefore be locally approximated with a polynomial. We choose $n_0 = 3$ to keep a minimal approach [for $n_0 = 2$, Eq. (3) would be unstable]. Equation (3) has been proposed as an alternative to the Kardar-Parisi-Zhang (KPZ) model [38,39], both describing the dynamic growth of interfaces. The original KPZ model [40] defines the KPZ universality class, where the interaction flows into a strong-coupling fixed point. Despite similarities, we show Eq. (3) to exhibit a radically different type of critical behavior.

We start from Eq. (3) to compute the fluctuation-driven part of the couplings' flow. This is conveniently done by mapping Eq. (3) to a field theory by means of the Martin-Siggia-Rose–de Dominicis-Janssen formalism [41–44]. Flow equations are then computed diagrammatically to one-loop order, within the framework of the infinitesimal momentum shell Wilsonian RG for nonequilibrium systems [6,27]. Setting τ, g_1, D, and a to unity and renaming $\bar{g}_n \coloneqq (D/g_1)^{[(n-1)/2]} (g_n/g_1)$ puts Eq. (3) in dimensionless form, so only $\bar{g}_2$ and $\bar{g}_3$ remain as parameters. Defining $s \coloneqq (2\pi)^{-1} \ln(\ell)$, the differential flow equations take the form

$$\frac{1}{g_1}\frac{dg_1}{ds} = \frac{3}{2}\bar{g}_3 - \bar{g}_2^2, \tag{4}$$

$$\frac{d\bar{g}_2^2}{ds} = -\frac{27}{2}\bar{g}_3\bar{g}_2^2 + 7\bar{g}_2^4, \tag{5}$$

$$\frac{d\bar{g}_3}{ds} = -\frac{15}{2}\bar{g}_3^2 + 14\bar{g}_3\bar{g}_2^2 - 4\bar{g}_2^4, \tag{6}$$

showing that $\bar{g}_2$ and $\bar{g}_3$ alone drive the flow. The Laplace operator in front of the nonlinear terms in Eq. (3) protects D, τ, and m_n from fluctuation corrections [27], so their flow is determined by the mean-field analysis above. Figure 1(c) shows the flow vector field in the $\bar{g}_3 - \bar{g}_2^2$ plane. A line $(\bar{g}_2^2/\bar{g}_3) = [(7 + \sqrt{145})/8]$ determines a transition between a diverging and converging behavior.

Below the transition line, the couplings vanish, flowing into the Gaussian fixed point ($\bar{g}_2 = \bar{g}_3 = 0$). Differently than in a mean-field scenario, however, the flow is logarithmically slow in ℓ, characteristic of marginal couplings at the upper critical dimension. This means interactions are effectively present on all length scales. As an extreme example, from the length scale of individual neurons ($L_{\min} \sim 10\ \mu$m [45], Chap. 1) to that of the whole cortex ($L_{\max} \sim 10^{-0.5}$ m [46], Chap. 9) nonlinearities reduce only by a factor $\ln[\ell = (L_{\max}/L_{\min})] \sim 10$. This is known as Gell-Mann–Low criticality, the archetypal behavior underlying renormalizability of quantum field theories, such as quantum electrodynamics (QED) [19,20], (Sec. V). At large scales, the power law scaling exponents z, ζ, and χ maintain their mean-field values, since interactions eventually vanish. However, logarithmic corrections must be included due to the slowness of the flow. One example of this effect can be seen in the scaling of the variance $\langle h^2 \rangle$ as a function of N [Fig. 1(b)]. As we detail in the Supplemental Material [27], such logarithmic corrections may appear as weak departures of critical exponents from their mean-field value, in a way that depends on the current system's dynamical state.

A difference to a prototypical Gell-Mann–Low theory is the presence of an infinite number of marginal couplings g_n, rather than a single one, the charge in QED. This, in principle, allows for the existence of an infinite number of fixed points. Determining analytically whether these are attractive, however, is a difficult task [39]. In neglecting g_n, $\forall\ n > 3$, we are implicitly assuming the Gaussian fixed point to be stable, with a sufficiently large basin of attraction to attract the flow from any initial g_n. We therefore test the validity of the flow equations numerically, by integrating Eq. (3) with the Euler-Maruyama algorithm [47]. Simulations are restricted to $\bar{g}_2^2 < 3\bar{g}_3$ [Fig. 1(c)], due to the occurrence of an unphysical bistable regime above such boundary [27].

We simulate systems of increasing size N, which limits the extent of correlations. By measuring correlation functions of the neural field h, we extract the value of the flown couplings at different length scales $\ell = (N/2)$ and initial conditions [27], shown in Fig. 1(c). Below the transition line, we observe good qualitative agreement between the simulated and predicted flow. Quantitative departures from predictions are expected, given the approximation made to one-loop order in fluctuations and to third order in the expansion of f. This approximation suffices to qualitatively confirm the running of the flow towards the Gaussian fixed point. Higher orders could be easily included, if needed, at the relatively low cost of computing more Feynman diagrams.

Above the transition line, the flow equations in the given approximation become unreliable. Indeed, they predict the divergence of the flow, running into $(\bar{g}_2, \bar{g}_3) = (\infty, -\infty)$. The flow's divergence could signal the presence of a strong coupling fixed point somewhere above the transition line, into which the flow eventually converges. This occurs, for example, in the KPZ model for $d \geq 3$, where an analogous transition point exists [48]. Simulations, however, do not show this behavior (at least not in the entire region $\bar{g}_2^2 < 3\bar{g}_3$ that can be simulated). In fact, Fig. 1(b) (squares) suggests that the flow, even when starting above the transition line, still heads towards the Gaussian fixed point, making a u-turn similar to when starting close and below such line (triangles). This is confirmed by additional simulations considering larger initial values of the couplings, which show that the flow always runs into the $\bar{g}_3 \gg \bar{g}_2^2$, Gell-Mann–Low regime, logarithmically approaching the Gaussian fixed point; see the Supplemental Material [27], Fig. S1(b). In this regime, the flow equations in the given approximation yield reliable quantitative predictions. This is exemplified by the measured neural field variance $\langle h^2 \rangle$ as a function of N being well predicted by theory [Fig. 1(b)]. In the limit $N \to \infty$, we predict $\langle h^2 \rangle \sim [4\pi g_1(\ell)]^{-1} \ln(N/2)$, with $\ell \overset{!}{=} (N/2)$ [27]. The ℓ dependence of g_1 shows a logarithmic correction with respect to the linear case, in which $g_1(\ell) = 1\ \forall\ \ell$.

We have so far demonstrated that the system showcases criticality of the Gell-Mann–Low kind. The question remains on why, among types of criticality, this one would be especially useful for computation. Information processing may benefit from a balance between a linear behavior, which naturally supports information storage and transmission, and a nonlinear one, necessary for computation. The Gell-Mann–Low criticality implements such a balance by sitting in-between a mean-field and strong coupling fixed point scenario. We exemplify this by training the system Eq. (3) to solve example tasks by what is known as reservoir computing [Fig. 1(a)] [26]: A structured input s is applied at some time t_{in} as a perturbation $h(x, t_{\text{in}}) + s(x)$. At a later time t_{out}, a linear readout $\sum_x W(x) h(x, t_{\text{out}})$ is taken and the parameters $W(x)$ are trained with gradient descent.

We first focus on memory; for concreteness, consider the Fischer memory curve. In the linear case, it decays with time as t^{-2} and is expected to be optimal [49]. Intuitively speaking, the signal distortion caused by nonlinearities makes it harder to retrieve information on past inputs (some network structures may, of course, be fine-tuned to have an optimal interplay with nonlinearities which increases memory, for example by active noise suppression [50], but this does not occur in our model). In the Gell-Mann–Low regime, nonlinearities flow towards zero, rather than to a strong coupling fixed point. We thus expect them to not worsen the power law decay found in the optimal linear case; rather causing only small logarithmic corrections. Inspired by these theoretical grounds, we train N^2 linear readouts to reconstruct a Gaussian-shaped input at some time t after injection, recording the reconstruction accuracy [27] [Fig. 1(d)]. As expected, increasing the strength

of nonlinearities, the accuracy's power law exponent does not deviate appreciably from the mean field, linear case: performance is only worsened by a constant shift in double logarithmic scale.

In contrast to a mean-field scenario, however, nonlinear interactions vanish only logarithmically slowly. They thus remain effective on practically all length scales, allowing the system's degrees of freedom to interact and to collectively perform computation. We exemplify this by training two linear readouts to correctly classify the parity of 3-bit strings [27]. The task is not linearly separable: no plane can correctly separate the strings in the two categories [see Fig. 1(e) inset]. Nonlinear dynamics are therefore necessary to expand the input's dimensionality, enabling linear separation. The input string's nth bit is encoded by the sign of a perturbation of the Fourier mode of momentum $k_n = (\pi - n4\pi/N)\hat{e}_1$, right below the high momentum cutoff π. The linear readout is taken on a low-pass filtered h, restricted to momenta $< |k_3|$. Thus nonlinear dynamics are further necessary to transfer information from the input to the readout modes. Figure 1(e) indeed shows successful training. Performance decays on a timescale between those of the slowest and the fastest modes, suggesting their collective interaction to be employed to perform the task. The modes' responses shown in the Supplemental Material [27] further imply this—see Fig. S3.

In conclusion, we provide a renormalized theory of neural network dynamics to uncover the structure of nonlinear interactions across scales, so far inaccessible by mean-field methods. This framework opens the door to the wealth of universality classes [6,7], beyond mean field, that may lie behind critical brain activity. We argue that beyond mean-field behavior is fundamental for computation, and provide tools to quantitatively address nonlinear signal transformations at a dynamical critical point. Applied to the stochastic Wilson-Cowan model, we find a new form of criticality, which is robust to the biologically plausible external drive: At the relevant dimension $d = 2$, critical behavior is of the Gell-Mann–Low kind; in contrast to mean-field behavior, nonlinear couplings, fundamental for computation, here remain relevant on all length scales. Their slowly vanishing flow, however, does not alter the mean-field critical exponents to leading order. This may explain why previous works might have overlooked beyond mean-field behavior.

It would be interesting to explore whether this, or other types of criticality emerge in more complex models. For example, in models that feature multiple cortical layers, collective behavior within a column may influence criticality along the cortical surface. The structure of hierarchical networks is also a parameter, other than the synaptic strength here considered, that can tune systems into criticality and interestingly extend it to broad regions in parameter space [51]. In general, the here presented methodology is applicable to any stochastic dynamical system; specifically to neural network models, some of which are even formulated as field theories from the outset [35,52]. Of particular interest are models that reproduce neural avalanches in simulations [35], or predict their exponents from scaling considerations [52], which have so far been restricted to the mean-field regime, for $d > 4$. State-dependent departures from mean-field exponents in a Gell-Mann–Low theory are reminiscent of the recently observed dependence of neural avalanche exponents on the neural activity's coefficient of variation [53]. It would thus be interesting to see if Gell-Mann–Low criticality also emerges in these models.

RG techniques are also being applied to artificial neural networks [54,55]. In the context of reservoir computing, we argue that Gell-Mann–Low criticality supports a computationally optimal balance between linear and nonlinear dynamics. Our analysis may also be extended to recurrent networks with trainable hidden weights that are amenable to field-theoretical formulations [56–58], further investigating the relation between criticality and computation. In general, the here exposed link between neural field theory, the KPZ model central to nonequilibrium statistical physics, and quantum field theory presents a stepping stone to transfer expertise from these fields, where RG methods are widely used.

We are grateful for various helpful discussions with Carsten Honerkamp. This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreement No. 945539 (Human Brain Project SGA3); the Helmholtz Association: Young Investigator's Grant VH-NG-1028; the Jülich-Aachen Research Alliance Center for Simulation and Data Science (JARA-CSD) School for Simulation and Data Science (SSD); the German Federal Ministry for Education and Research (BMBF Grant No. 01IS19077A to Jülich); L. T. received funding as Vernetzungsdoktorand: "Collective behavior in stochastic neuronal dynamics".

[*] These authors contributed equally to this work.

[1] P. Cheeseman, B. Kanefsky, and W. M. Taylor, in *Proceedings of the 12th International Joint Conference on Artificial Intelligence—Volume 1* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 1991), IJCAI'91, pp. 331–337, ISBN 1558601600.

[2] L. Saitta, A. Giordana, and A. Cornuéjols, *Phase Transitions in Machine Learning* (Cambridge University Press, Cambridge, England, 2011).

[3] J. M. Beggs and D. Plenz, J. Neurosci. **23**, 11167 (2003).

[4] D. R. Chialvo, Nat. Phys. **6**, 744 (2010).

[5] C. G. Langton, Physica (Amsterdam) **42D**, 12 (1990).

[6] P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).

[7] U. C. Taeuber, *Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior* (Cambridge University Press, Cambridge, England, 2014), ISBN 9780521842235.

[8] N. Bertschinger and T. Natschläger, Neural Comput. **16**, 1413 (2004).
[9] N. Bertschinger, T. Natschläger, and R. A. Legenstein, in *Advances in neural information processing systems* (MIT Press, Cambridge, 2005), pp. 145–152.
[10] O. Kinouchi and M. Copelli, Nat. Phys. **2**, 348 (2006).
[11] H. R. Wilson and J. D. Cowan, Biophys. J. **12**, 1 (1972).
[12] H. R. Wilson and J. D. Cowan, Kybernetik **13**, 55 (1973).
[13] A. Destexhe and T. J. Sejnowski, Biol. Cybern. **101**, 1 (2009).
[14] A. Schüz and V. Braitenberg, in *Cortical Areas: Unity and Diversity*, edited by A. Schüz and R. Miller (Taylor and Francis, London, 2002), Chap. 16, pp. 377–385.
[15] A. Destexhe, M. Rudolph, and D. Pare, Nat. Rev. Neurosci. **4**, 739 (2003).
[16] C. Haldeman and J. M. Beggs, Phys. Rev. Lett. **94**, 058101 (2005).
[17] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions* (Springer, Netherlands, 2008), ISBN 978-1-4020-8764-6, https://www.springer.com/gp/book/9781402087646.
[18] L. J. Fosque, R. V. Williams-García, J. M. Beggs, and G. Ortiz, Phys. Rev. Lett. **126**, 098101 (2021).
[19] M. Gell-Mann and F. E. Low, Phys. Rev. **95**, 1300 (1954).
[20] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
[21] B. Ermentrout, Rep. Prog. Phys. **61**, 353 (1998).
[22] P. C. Bressloff, Phys. Rev. Lett. **89**, 088101 (2002).
[23] S. Coombes, Biol. Cybern. **93**, 91 (2005).
[24] P. C. Bressloff, J. Phys. A **45**, 033001 (2012).
[25] E. L. Schwartz, Biol. Cybern. **25**, 181 (1977).
[26] M. Lukoševičius and H. Jaeger, Comput. Sci. Rev. **3**, 127 (2009).
[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.128.168301 for further details on theoretical calculations, numerical simulations, and computational tasks, which additionally includes Refs. [28–33].
[28] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, 1996).
[29] A. Coolen, in *Neuro-Informatics and Neural Modelling*, edited by F. Moss and S. Gielen, Vol. 4 of Handbook of Biological Physics (North-Holland, Amsterdam, 2001), pp. 619–684.
[30] J. Kiefer and J. Wolfowitz, Ann. Math. Stat. **23**, 462 (1952).
[31] G. E. Hinton, Artif. Intell. **40**, 185 (1989).
[32] E. Medina, T. Hwa, M. Kardar, and Y.-C. Zhang, Phys. Rev. A **39**, 3053 (1989).
[33] M. Helias, J. Phys. A **53**, 445004 (2020).
[34] W. L. Shew, H. Yang, T. Petermann, R. Roy, and D. Plenz, J. Neurosci. **29**, 15595 (2009).
[35] S. di Santo, P. Villegas, R. Burioni, and M. A. Munoz, Proc. Natl. Acad. Sci. U.S.A. **115**, E1356 (2018).
[36] S. Ostojic and N. Brunel, PLoS Comput. Biol. **7**, e1001056 (2011).
[37] A. Roxin, N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk, J. Neurosci. **31**, 16217 (2011).
[38] S. Pavlik, Zh. Eksp. Teor. Fiz. **106**, 553 (1994).
[39] N. V. Antonov and A. N. Vasiliev, Zh. Éksp. Teor. Fiz. **81**, 485 (1995).
[40] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[41] P. Martin, E. Siggia, and H. Rose, Phys. Rev. A **8**, 423 (1973).
[42] C. De Dominicis, J. Phys. Colloques **37**, C1 (1976).
[43] J. A. Hertz, Y. Roudi, and P. Sollich, J. Phys. A **50**, 033001 (2017).
[44] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks* (Springer International Publishing, New York, 2020), Vol. 970.
[45] *Dendrites*, edited by G. Stuart, N. Spruston, and M. Häusser (Oxford University Press, New York, 2007), 10.1093/acprof:oso/9780198566564.001.0001.
[46] W. Gerstner and W. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge University Press, Cambridge, England, 2002), ISBN (paperback) 0521890799.
[47] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1992).
[48] H. C. Fogedby and W. Ren, Phys. Rev. E **80**, 041116 (2009).
[49] S. Ganguli, D. Huh, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **105**, 18970 (2008).
[50] T. Toyoizumi, Neural Comput. **24**, 2678 (2012).
[51] P. Moretti and M. A. Muñoz, Nat. Commun. **4**, 2521 (2013).
[52] M. A. Buice and J. D. Cowan, Phys. Rev. E **75**, 051919 (2007).
[53] A. J. Fontenele, N. A. P. de Vasconcelos, T. Feliciano, L. A. A. Aguiar, C. Soares-Cunha, B. Coimbra, L. Dalla Porta, S. Ribeiro, A. J. Rodrigues, N. Sousa *et al.*, Phys. Rev. Lett. **122**, 208101 (2019).
[54] P. Mehta and D. J. Schwab, arXiv:1410.3831.
[55] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks* (Cambridge University Press, Cambridge, England, 2022).
[56] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Phys. Rev. Lett. **61**, 259 (1988).
[57] J. Schuecker, S. Goedeke, and M. Helias, Phys. Rev. X **8**, 041029 (2018).
[58] K. Krishnamurthy, T. Can, and D. J. Schwab, Phys. Rev. X **12**, 011011 (2022).