

Tailored ensembles of neural networks optimize sensitivity to stimulus statistics

Johannes Zierenberg^{1,2}, Jens Wilting,¹ Viola Priesemann,^{1,2,*}† and Anna Levina^{3,4,*}‡

¹Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077 Göttingen, Germany

²Bernstein Center for Computational Neuroscience, Am Fassberg 17, 37077 Göttingen, Germany

³University of Tübingen, Max Planck Ring 8, 72076 Tübingen, Germany

⁴Max Planck Institute for Biological Cybernetics, Max Planck Ring 8, 72076 Tübingen, Germany



(Received 12 April 2019; revised manuscript received 21 October 2019; accepted 17 December 2019; published 3 February 2020)

The capability of a living organism to process stimuli with nontrivial intensity distributions cannot be explained by the proficiency of a single neural network. Moreover, it is not sufficient to maximize the dynamic range of the neural response; it is also necessary to tune the response to the intervals of stimulus intensities that should be reliably discriminated. We derive a class of neural networks where these intervals can be tuned to the desired interval. This allows us to tailor ensembles of networks optimized for arbitrary stimulus intensity distributions. We discuss potential applications in machine learning.

DOI: [10.1103/PhysRevResearch.2.013115](https://doi.org/10.1103/PhysRevResearch.2.013115)

I. INTRODUCTION

Living organisms are constantly exposed to sensory stimuli with nontrivial intensity distributions that can cover multiple orders of magnitude [1–3]. Properly encoding these distributions is essential for an organism to survive. Stimulus intensity can be faithfully encoded in neural firing rates, with stronger intensities eliciting higher firing rate responses. Examples of such reliable encoding can be found for hearing [4–7], odor discrimination [8,9], or motion detection [10]. However, typical neural responses show a steep increase in firing rate, resulting in a narrow range of discriminable intensities. This raises the question of how organisms can process broad intensity distributions.

In principle, the capability to process a broad distribution of stimulus intensities can be quantified by the dynamic range. The dynamic range of a neural network is defined as the log-ratio between the strongest and weakest stimulus intensities that are reliably encoded by the neural firing rate. For example, the dynamic range in cat primary auditory nerve fibers has been experimentally measured to extend from 40 to 50 dB [11]. This, however, clearly does not cover the full range of hearing (for humans approximately 0–120 dB sound pressure level), resulting in the so-called dynamic-range problem [3,12].

How can the organism increase its dynamic range? One line of solutions argues that the dynamic range of a single recurrent neural network, driven by external stimuli, is maximal at the critical point of the nondriven network [13] or

close to the bifurcation in an inhibitory network [14]. This theoretical concept is supported by experiments on cultured cortical slices [15] and somatosensory cortex in rats [16]. Thereby criticality would foster flexible information processing [17,18], while leaving freedom to adjust the dynamic range through network topology [19,20] or inhibition modulation [16]. In another line of argument, it is proposed that the dynamic range could be increased by combining networks with different responses to stimulus intensities, as suggested for the olfactory system [9,21], or by directly adjusting the neurons' response to a temporal intensity distribution, as observed in the auditory midbrain [7].

However, a large dynamic range is not sufficient to efficiently process stimulus intensities from nontrivial distributions, e.g., from bimodal distributions. An efficient encoding would represent relevant intensity intervals by more sensitive neural responses manifested in a larger slope of the network response function. Indeed, it was found in auditory midbrain that neurons adjust their response to the statistics of the sound level distribution [7]. Finding the neural network implementation of such an optimized response is crucial to explaining how an organism can deal with arbitrary intensity distributions. At the same time this mechanism can be used by an organism for tuning its sensory processing system to focus on specific elements required for a given computational task.

In this paper we show that, in general, processing capability cannot be optimized by a single homogeneous recurrent network, but instead requires an ensemble of specialized neural networks that is tailored to stimulus statistics. We derive a class of neural network implementations that allow tuning the interval of stimulus intensities that are reliably discriminated by the network (the discriminable interval). Combining these networks in an ensemble can generate optimal sensitivity to arbitrary intensity distributions.

II. RESULTS

No matter what type of system one analyzes, the dynamic range Δ and the discriminable interval are response measures

*These authors contributed equally to this work.

†viola.priesemann@ds.mpg.de

‡anna.levina@uni-tuebingen.de

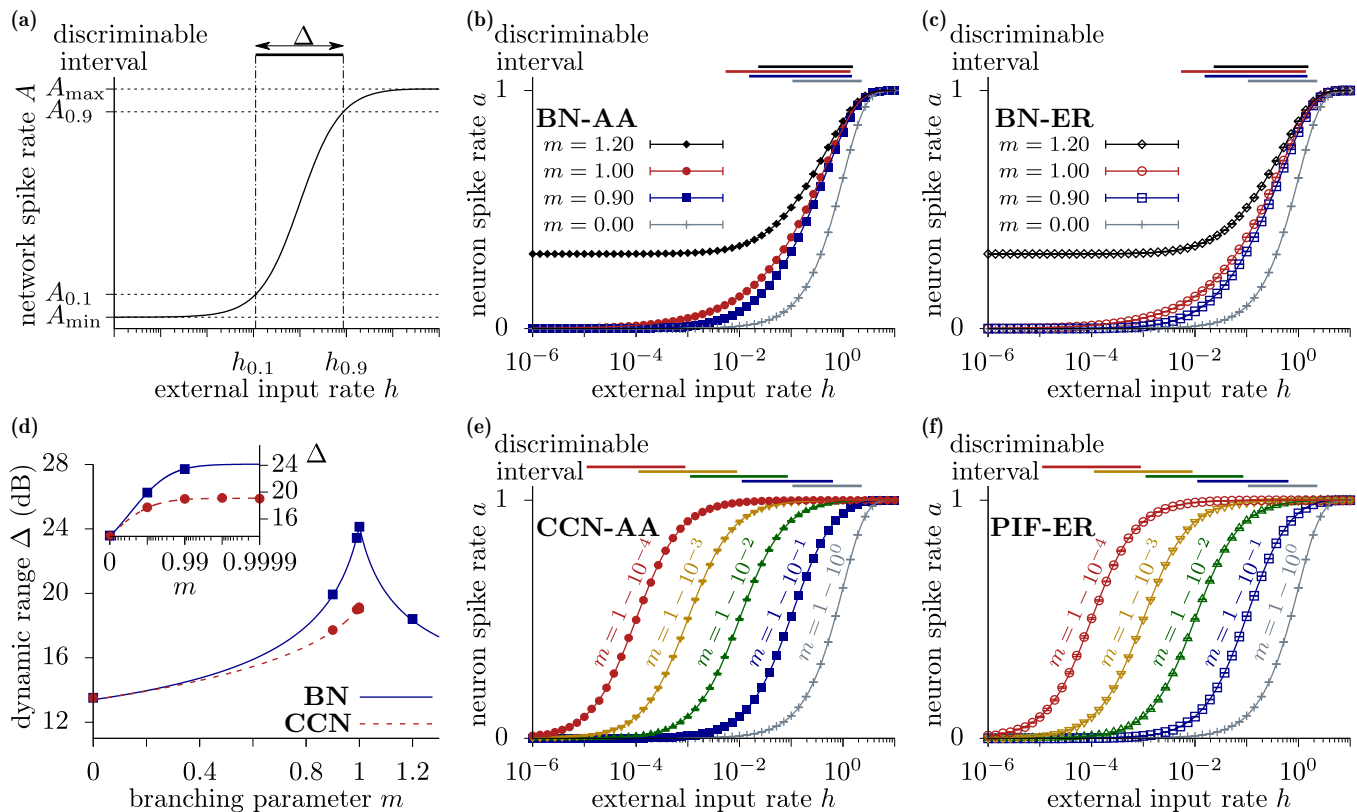


FIG. 1. Dynamic range Δ and discriminable interval $[h_{0.1}, h_{0.9}]$ for several models (lines are analytic solutions and data points with error bars are results from numerical simulations). (a) Typical network spike rate as a function of the external input rate with the definition of the dynamic range Δ as the width of the discriminable interval. (b) Branching network on a fully connected network ($N = 10^4$). (c) Branching network on a sparse ER network ($N = 10^4$, $p = 10^{-3}$, and $\bar{K} = 10$). (d) Dynamic range as a function of the branching parameter for the branching network (BN) and the coalescence-compensating network (CCN) on a fully connected network (see Appendix A). The inset shows the dynamic range as a function of distance to the critical point ($m = 1$) in logarithmic scale. (e) Coalescence-compensating network on a fully connected network ($N = 10^2$). (f) Probabilistic integrate-and-fire (PIF) network on a sparse ER network ($N = 10^3$, $p = 10^{-2}$, and $\bar{K} = 10$). For log-log plots of (b) and (e), see Appendix B.

defined in terms of the range of stimulus intensities that can be reliably discriminated in the systems' output [Fig. 1(a)]. For a neural network, we consider as input Poisson activity with rate h (per neuron). The response is measured by a network spike rate $A(h)$. For a zero input rate, the network typically produces its minimal baseline rate A_{\min} , whereas for very high input rates the network response saturates. Hence, the network rate covers the interval $[A_{\min}, A_{\max}]$. The discriminable interval is defined as the 10th to 90th percentiles $[h(A_{0.1}), h(A_{0.9})] = [h_{0.1}, h_{0.9}]$. Then the dynamic range is defined as $\Delta = 10 \log_{10}[h(A_{0.9})/h(A_{0.1})]$ in decibels. These response measures characterize the range of input rates that can be discriminated by the system, in contrast to response measures that characterize how well a system can detect changes in the input rate [22].

To start, we show for branching networks, the prime example for a network with maximal dynamic range at criticality [13,23,24], that the discriminable interval cannot be tuned. The network consists of N binary units that are updated at discrete time steps Δt . Each neuron i can be either silent ($s^i = 0$) or excited to spike ($s^i = 1$). It can be excited by (i) an external Poisson input with rate h , such that the transition probability is $\lambda(h) = 1 - \exp(-h\Delta t)$, or (ii) a presynaptic

neuron j (which was excited in the previous time step) with transition probability w^{ij} . If no external or internal input reaches an excited neuron, it returns to a silent state in the next time step (no refractory period). In this model, the control parameter is the branching parameter m , which is sometimes said to characterize the average number of excitations created by a single excited neuron. For analytical tractability, we focus on a fully connected [all-to-all (AA)] network such that $w^{ij} = w = m/N$. In addition, we consider a sparse random Erdős-Rényi (ER) network, where connections are randomly drawn with probability p resulting in a distribution of outgoing/incoming connections $K_{\text{out}}/K_{\text{in}}$ with average degree \bar{K} . For the ER branching network, $w^{ij} = w^j = m/K_{\text{out}}^j$. The network activity at each time step, $A_t = \sum_i s_t^i$, is thus determined by both internal (m) and external (h) activation.

For the fully connected branching network, we can analytically derive the neuron spike rate as a function of external input rate [Fig. 1(b)]. In the following, we sketch the main steps and refer to Ref. [25] for a detailed derivation. Given a network activity A_t at time t , the probability for any neuron to be excited in the next time step is given by $P[s_{t+1}^i = 1 | A_t, w, h] = 1 - (1 - w)^{A_t} [1 - \lambda(h)] = p(A_t)$. The network activity A_{t+1} is then binomially distributed with

expectation value $\langle A_{t+1}|A_t \rangle = Np(A_t)$. Demanding stationary activity $A = \langle A_t \rangle = \langle \langle A_{t+1}|A_t \rangle \rangle$ and neglecting fluctuations in a mean-field approximation $\langle (1 - m/N)^{A_t} \rangle \approx (1 - m/N)^{\langle A_t \rangle}$ implies

$$A = N - N(1 - w)^A [1 - \lambda(h)]. \quad (1)$$

Using the Lambert- W function defined as $W(z)e^{W(z)} = z$ [26], Eq. (1) is solved by $A(h) = N - W((1 - w)^N \ln(1 - w)N[1 - \lambda(h)]) / \ln(1 - w)$. For $w = m/N$ and $N \rightarrow \infty$, we can identify $(1 - m/N)^N \rightarrow e^{-m}$ and expand $\ln(1 - m/N)N = -m + O(N^{-1})$ to obtain the solution for the expected neuron spike rate

$$a_{\text{BN}}(m, h) = \frac{A(h)}{N} = 1 + \frac{W(-me^{-m}[1 - \lambda(h)])}{m}, \quad (2)$$

which turns out to be system-size independent for sufficiently large N . This solution describes our numerical results for fully connected networks accurately [Fig. 1(b)] and describes our numerical results for sparse ER networks approximately [Fig. 1(c)]. As expected for the universality class of directed percolation, the spike rate scales with the external input as $a \sim h$ ($m < 1$) and $a \sim h^{0.5}$ ($m = 1$) (see Appendix B).

From the neuron rate, we readily obtain mean-field solutions for the dynamic range and discriminable interval. Inverting Eq. (2) yields

$$h_{\text{BN}}(m, a) = -\frac{1}{\Delta t} \ln[(1 - a)e^{ma}], \quad (3)$$

which allows us to compute $h(a_{0.1})$ and $h(a_{0.9})$. In agreement with previous results, we find that the dynamic range is maximal at criticality [Fig. 1(d)]. However, it does not diverge for $\epsilon = 1 - m \rightarrow 0$ [19], which can be seen clearly from the figure inset.

Importantly, the discriminable interval of branching networks is constrained, i.e., the location barely changes for small $\epsilon = 1 - m$ [Fig. 1(b)]. We can derive the end points of the discriminable interval $h(a_x)$, $a_x \in \{a_{0.1}, a_{0.9}\}$, by expanding $\ln(1 - a) = -\sum_{n=1}^{\infty} a^n/n$ and rewriting Eq. (3) as $h_{\text{BN}}(m, a_x) \approx (a_x \epsilon + \sum_{n=2}^{\infty} a_x^n/n) / \Delta t$ (see also Appendix A). For all ϵ sufficiently smaller than $a_{0.1}$, which holds in the vicinity of the critical point, the end points are barely distinguishable (Fig. 2). This result is valid for sufficiently large system sizes and limits the discriminable interval (and thereby the dynamic range) even in the infinite system limit. As a consequence, cortical networks with identified branching parameters varying around $m \approx 0.98$ [27,28] would have strongly overlapping discriminable intervals.

In the following, we propose a framework that allows us to tune the discriminable interval. We notice that the constrained discriminable interval is a result of the nonlinearity in the excitation probability $p(A_t)$ [cf. Eq. (1)]. This nonlinearity, in turn, is a result of coalescence (the simultaneous excitation of the same neuron from multiple sources) in the branching network [25]. Coalescence violates the assumption that each excited neuron on average excites m new neurons in the next time step, i.e., the autoregressive representation $\langle A_{t+1}|A_t \rangle = mA_t$ is not strictly true for the branching network. This autoregressive representation, however, is the basis of the branching process [29], which has been shown to approximate spike propagation in neuronal tissue well [13,23,27,30–33] and was the historic

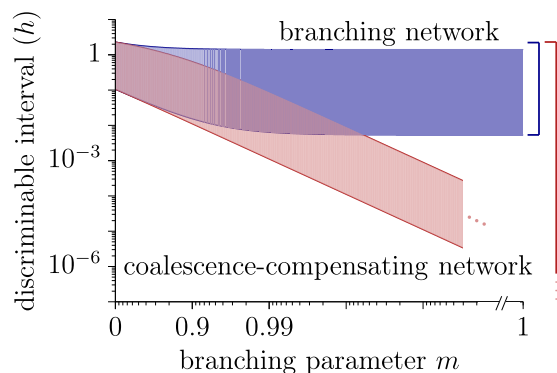


FIG. 2. Influence of compensating coalescence on the discriminable interval. For the branching network, the discriminable interval barely changes, especially for branching parameters close to critical-like dynamics ($m = 1$). For the coalescence-compensating network, the discriminable interval becomes a function of the branching parameter.

motivation to model neural spiking activity with a branching network. For the fully connected branching network, we instead find $\langle A_{t+1}|A_t \rangle = Np(A_t) = m_{\text{eff}}(w, A_t)A_t + N\lambda(h)$ with an effective branching parameter

$$m_{\text{eff}}(w, A_t) = \left(\frac{N}{A_t}\right) [1 - (1 - w)^{A_t}] [1 - \lambda(h)]. \quad (4)$$

For sparse ER networks, a similar result can be derived [13,25], which coincides to leading order with Eq. (4). In short, coalescence reduces the effective branching parameter $m_{\text{eff}}(w, A_t) < m = wN$.

Next we show that the bounds of the discriminable interval $[h_{0.1}, h_{0.9}]$ can be fine-tuned in models without coalescence. We assume that the network may at most compensate coalescence with internal sources, because compensating coalescence with external sources would require knowledge of the external input rate h (a logical contradiction as one of the network's tasks is to infer this rate). Then we can set $\lambda(h) = 0$ and compensate coalescence by inverting Eq. (4) for a target branching parameter $m_{\text{eff}}(w_{\text{CC}}, A_t) = m$, obtaining adaptive (coalescence-compensating) synaptic weights

$$w_{\text{CC}}(m, A_t) = \left[1 - \left(1 - \frac{mA_t}{N}\right)^{1/A_t}\right]. \quad (5)$$

For finite networks in the close vicinity of the critical point, we need to introduce a finite-network cutoff at $w_{\text{CC}}(m, N) = \ln(N)/N$, which avoids an additional absorbing boundary at $A = N$ (see Appendix C). Compensating internal coalescence with adaptive weights [Eq. (5)] implies $m_{\text{eff}}(w_{\text{CC}}(A_t), A_t) = m[1 - \lambda(h)]$, such that for small input h we approximate the target branching parameter.

The resulting coalescence-compensating network retains a maximal dynamic range at criticality [Fig. 1(d)] but allows us to fine-tune the discriminable interval via the branching parameter m [Fig. 1(e)]. To show this we compute the neuron spike rate for an autoregressive representation with $m_{\text{eff}} = m[1 - \lambda(h)]$. For stationary activity we obtain $A = \langle \langle A_{t+1}|A_t \rangle \rangle = m[1 - \lambda(h)]A + N\lambda(h)$, which is linear in A and

h [cf. Eq. (1)]. For the neuron spike rate we obtain

$$a_{CC}(m, h) = \frac{\lambda(h)}{1 - m[1 - \lambda(h)]}. \quad (6)$$

This solution describes our numerical results for fully connected networks accurately [Fig. 1(e)] and further shows that the coalescence-compensating model with external input is only nontrivial for $m < 1$, because for $m \rightarrow 1$ the average spike rate $a_{CC}(1, h) \rightarrow 1$ for all values of h . Similarly, for all $m > 1$ we would observe a fully excited network even for $h \rightarrow 0$. The saturation is caused by the divergence of correlation length and time for $m \rightarrow 1$, as expected for a driven branching process [29], and no longer allows us to define a dynamic range for $m \geq 1$ [Fig. 1(c)]. When we invert Eq. (6), we obtain $h_{CC}(m, a) = -\log[1 - (1 - m)a/(1 - ma)]/\Delta t$ such that the bounds of the discriminable interval $h_{CC}(m, a_x) \approx (1 - m)a_x/(1 - ma_x)\Delta t \simeq a_x(1 - m)$ can be tuned as a function of m (Fig. 2).

Our result that the discriminable interval can be tuned when compensating coalescence on fully connected networks can be generalized to any network structure by compensating coalescence locally. For this we define a local linear autoregressive representation for neuron i in terms of the average activity $a^i = \langle P[s_{t+1}^i = 1] \rangle = ma_{in}^i + h$, where a_{in}^i is the average number of excited neurons that project to neuron i . A straightforward solution for such a local linear autoregressive representation is a probabilistic integrate-and-fire mechanism with a linear transfer function [34], defined by $P[s_{t+1}^i = 1] = \sum w^{ij} s_t^j$, where $w^{ij} = m/K_{in}^i$ for all K_{in}^i neurons projecting to i and $w^{ij} = 0$ otherwise. The neuron spike rate for this dynamics on sparse ER networks is accurately described by our analytical solution for coalescence-compensating dynamics [Fig. 1(f)]. We conclude that the discriminable interval can be controlled in a biologically plausible manner, namely, by linear probabilistic integrate-and-fire neurons, where only the local presynaptic strength of each neuron needs to be adjusted.

The advantage of the coalescence-compensating model can be further illustrated by an extended power-law scaling of the avalanche-size distributions at criticality (Fig. 3). In the branching network the power law is cut off at $s \approx N$, for which typical avalanches (Fig. 3, inset, blue) with a bell shape $A(t, d) = d\mathcal{F}(t/d)$ [35] and avalanche duration d imply that the avalanche peak value scales as $A^{\text{peak}} \sim d \sim \sqrt{N}$ (see Appendix D). In the coalescence-compensating network, this maximum is now extended to the nonabsorbing boundary $A^{\text{peak}} \sim d \sim N$ (Fig. 3, inset, red) such that the power-law characteristics extend until $s \approx N^2$. This means that for the convergence-compensating network the avalanche-size distribution covers twice as many orders of magnitude as for the branching network. Similarly, the power-law scaling of the avalanche-duration distribution is extended from \sqrt{N} to N (see Appendix D).

III. CONCLUSION

Our results show that an ensemble of coalescence-compensating networks can be tailored to optimize sensitivity to arbitrary stimulus intensity distributions. For example, a bimodal distribution of stimulus intensities can be well processed by at least two networks with disjoint

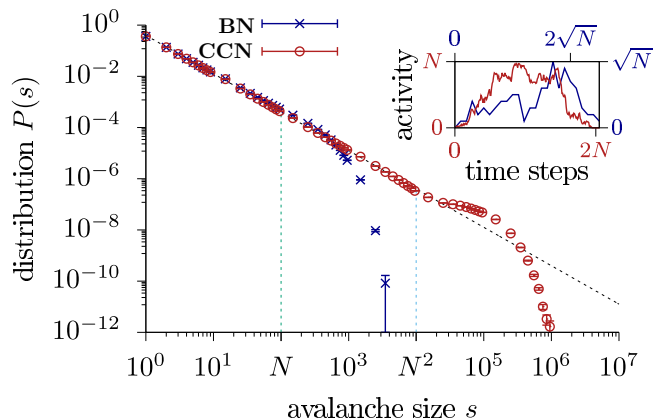


FIG. 3. Avalanche-size distribution of a branching network and a coalescence-compensating network with $N = 100$ (cf. Fig. 7 for $N = 10^4$) in the separation of timescale regime, i.e., each avalanche is triggered manually ($h \rightarrow 0$ limit). The inset shows example avalanches at the typical upper bounds of the power-law distribution (vertical dashed line in main plot).

discriminable intervals (Fig. 4), consistent with the heterogeneous single-neuron responses observed in auditory mid-brain when presented with bimodal sound level intensities [7]. Processing a bimodal distribution of stimulus intensities may also be relevant for higher cortical areas that receive input from areas with up and down states [36–39] or that react to complex behavior such as bimodal escape sequence duration of drosophila [40]. If, however, stimulus statistics cannot be anticipated, surprises are best dealt with by assuming a uniform broad distribution of stimulus intensities. In this case, a good strategy is to maximize the dynamic range by an ensemble of networks with sufficiently overlapping discriminable intervals [Figs. 1(e) and 4]. For our models, logarithmically spaced choices of $\epsilon = 1 - m$ yield a homogeneous overlap of discriminable intervals [Fig. 1(e)].

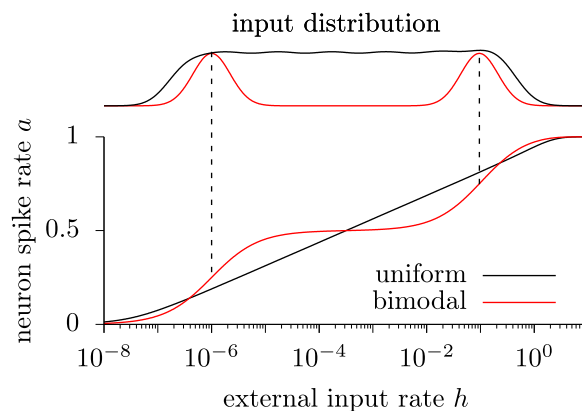


FIG. 4. Responses for tailored ensembles of coalescence-compensating networks to particular stimulus statistics. A broad uniform distribution of intensities can be represented by networks with overlapping discriminable intervals. A bimodal distribution of intensities can be represented by networks with discriminable intervals around the peaks of the distribution.

The optimized sensitivity of the ensemble to arbitrary stimulus intensity distributions can be exploited in machine-learning applications. A potential application could be a tailored ensemble of coalescence-compensating recurrent networks for reservoir computing [41–47]. Stacking these reservoirs will perform optimal separation of stimulus intensities within the union of discriminable intervals of participating networks.

Our results suggest an alternative strategy to solve the discrepancy between the dynamic range of a single network compared to the large range of sensory stimulus intensities. First, networks probably normalize the input statistics, as observed in the early stages of the visual pathway [48]. Second, a single coalescence-compensating network could implement an adaptable discriminable interval (see, e.g., Ref. [7]) by tuning its distance to criticality $(1 - m)$ [28]. Supporting this possibility, rapid adjustment of neuronal sensitivity to stimulus intensity was demonstrated in a recent study of the *Drosophila* hearing system [49]. We hypothesize that the brain combines all strategies for maximal robustness.

In summary, we have shown for coalescence-compensating networks that the discriminable intervals can be tuned by tuning m (see Fig. 2). That offers a flexible manner to extend the dynamic range, either by tuning networks to task or by combining multiple networks, in general. The necessary compensation of coalescence can be achieved fully locally with probabilistic integrate-and-fire neurons even in sparse networks. We believe that this is relevant for both biology and artificial intelligence.

ACKNOWLEDGMENTS

The authors received support from the Max Planck Society. J.Z. and V.P. received financial support from the German Ministry of Education and Research via the Bernstein Center for Computational Neuroscience Göttingen under Grant No. 01GQ1005B. J.W. was financially supported by Gertrud-Reemtsma-Stiftung. A.L. received funding from a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation, endowed by the Federal Ministry of Education and Research.

APPENDIX A: ADAPTIVE-WEIGHT NETWORKS HAVE DYNAMIC RANGE AND DISCRIMINABLE INTERVALS WITH PROPERTIES FROM BOTH THE BRANCHING PROCESS AND BRANCHING NETWORK

If compensating for coalescence extends the range of avalanche statistics (cf. Fig. 3), one could conclude that the resulting coalescence-compensating model is “more critical” than the branching network. Intuitively, one may expect that this leads to more optimal information processing and other benefits that come with operating with critical dynamics. For example, one could expect that the dynamic range, which is maximal for critical-like dynamics [13], is even larger in coalescence-compensating networks with critical dynamics. In the following, we will show that on the level of a single network, the dynamic range is not improved when we compensate for coalescence.

To analytically calculate the dynamic range of the branching network, we start with the mean-field approximation (1) and solve for the external input

$$h_{\text{BN}}(m, A) = -\frac{1}{\Delta t} \ln \left[\frac{1 - \frac{A}{N}}{\left(1 - \frac{m}{N}\right)^A} \right]. \quad (\text{A1})$$

For sufficiently large N , we approximate $(1 - m/N)^A \rightarrow e^{-ma}$ with neuron rate $a = A/N$ and find the system-size independent result

$$h_{\text{BN}}(m, a) = -\frac{1}{\Delta t} \ln[(1 - a)e^{ma}]. \quad (\text{A2})$$

Recalling our result for the system-size independent neuron rate $a_{\text{BN}}(m, h)$, Eq. (2), we continue with the bounds $a_{\text{min}} = a(m, h \rightarrow 0) = 1 + W(-me^{-m})/m$ and $a_{\text{max}} = a(m, h \rightarrow \infty) = 1 + W(0)/m = 1$ such that $a_x = 1 + (1 - x)W(-me^{-m})/m$. Inserting this into Eq. (A2), we obtain

$$h_{\text{BN}}(m, a_x) = -\frac{1}{\Delta t} \ln[(1 - x)e^{-xW(-me^{-m})}] \quad (\text{A3})$$

$$= \frac{xW(-me^{-m}) - \ln(1 - x)}{\Delta t}. \quad (\text{A4})$$

With this we can calculate the dynamic range of the branching network

$$\Delta_{\text{BN}}(m) = 10 \log_{10} \left[\frac{0.9W(-me^{-m}) - \ln(0.1)}{0.1W(-me^{-m}) - \ln(0.9)} \right]. \quad (\text{A5})$$

For the coalescence-compensating network, we first need to calculate the neuron rate as a response to the external input rate. Assuming stationary activity $A \approx \langle A_{t+1} | A_t \rangle$, we can use Eqs. (4) and (5) to write down the mean-field approximation

$$A_{\text{CC}} = m[1 - \lambda(h)]A_{\text{CC}} + N\lambda(h). \quad (\text{A6})$$

Solving this for the neuron rate, we obtain

$$a_{\text{CC}}(m, h) = \frac{\lambda(h)}{1 - m[1 - \lambda(h)]}. \quad (\text{A7})$$

This rate is only finite and non-negative for $m < 1$. In this range, $a_{\text{min}} = 0$ and $a_{\text{max}} = 1$ always, such that $a_x = x$. Calculating the inverse of Eq. (A7),

$$h_{\text{CC}}(m, a) = -\frac{1}{\Delta t} \ln \left[1 - \frac{(1 - m)a}{1 - ma} \right], \quad (\text{A8})$$

we find the dynamic range for coalescence-compensating networks ($m < 1$)

$$\Delta_{\text{CC}}(m) = 10 \log_{10} \left[\frac{\ln \left(1 - \frac{(1-m)0.9}{1-m0.9} \right)}{\ln \left(1 - \frac{(1-m)0.1}{1-m0.1} \right)} \right]. \quad (\text{A9})$$

For coalescence-compensating networks the interval of discriminable input rates is a function of the branching parameter (Fig. 2).

The results for the coalescence-compensating network are consistent with a modified branching process. Consider a branching process with upper-bound population activity $A_t \leq N$ and external input per time step $Nh\Delta t$. The stationary activity is $A_{\text{BP}}(m, h) = Nh\Delta t / (1 - m)$ [29]. The discriminating

activity is defined as $A_x = xN$ such that the inverse of the activity yields

$$h_{\text{BP}}(A_x) = (1 - m)x/\Delta t \quad (\text{A10})$$

and the dynamic range is independent of the branching parameter

$$\Delta_{\text{BP}}(m) = 10 \log_{10}(0.9/0.1) \approx 9.5. \quad (\text{A11})$$

Importantly, the discriminable interval $[h_{\text{BP}}(A_{0.1}), h_{\text{BP}}(A_{0.9})]$ is highly dependent on m .

APPENDIX B: POWER-LAW BEHAVIOR OF RESPONSE CURVES FOR DIFFERENT MODELS

Figure 5 shows the response curves from Fig. 1 in log-log scale. As expected [13], subcritical response curves scale as $a \sim h$ and only at the critical point ($m = 1$) we observe $a \sim h^{0.5}$ up to finite-size effects. The different scalings can be explained by a Taylor expansion of Eq. (3), i.e.,

$$h_{\text{BN}}(m, a) = -\frac{1}{\Delta t} \ln[(1 - a)e^{ma}] \quad (\text{B1})$$

$$\approx -\frac{1}{\Delta t} (ma - a - a^2 + \dots) \quad (\text{B2})$$

$$= \frac{1}{\Delta t} [a(1 - m) + a^2 + \dots]. \quad (\text{B3})$$

Only for $m = 1$ the leading linear term vanishes such that $h \sim a^2$ or $a \sim h^{0.5}$; otherwise $h \sim a$ or $a \sim h$.

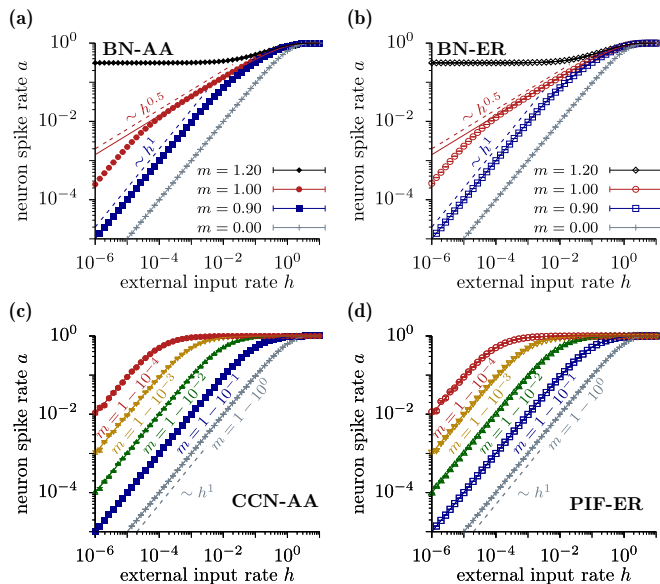


FIG. 5. Power-law behavior of response curves for different models. (a) Branching network (with coalescence) on an all-to-all connected network ($N = 10^4$). (b) Branching network (with coalescence) on a sparse random network ($N = 10^4$ and $p = 10^{-3}$). Results are very similar to [13]. (c) Coalescence-compensating networks on an all-to-all connected network ($N = 100$). (d) Probabilistic integrate-and-fire networks with linear firing probability on a sparse random network ($N = 10^3$ and $p = 10^{-2}$).

APPENDIX C: ADAPTIVE WEIGHTS COMPENSATE INTERNAL COALESCENCE

Internal and external coalescence reduce the effective branching parameter for static transition probabilities (connection weights) $w = m/N$ [25]. Thereby, macroscopic branching parameters \hat{m} , estimated from the network rate, differ from the model branching parameter m . For a detailed discussion, we refer the reader to Ref. [25]. We will now exploit the analytical insight on coalescence to construct microscopic dynamics that compensates for internal coalescence. The basic idea is simple: Adjust the microscopic dynamics w such that the effective branching parameter matches the desired macroscopic branching parameter.

To compensate for coalescence, we adjust the weights w such that $m_{\text{eff}}(w, A_t) = m$ for all A_t and thereby tune the model parameter equal to the macroscopic branching parameter $m = \hat{m}$. Inserting Eq. (4), we obtain activity-dependent weights

$$\tilde{w}_{\text{CC}}(A_t) = 1 - \left(1 - \frac{mA_t}{N[1 - \lambda(h)]}\right)^{1/A_t}, \quad (\text{C1})$$

which would compensate for internal and external coalescence. We now assume that the network has a mechanism to communicate the current activity A_t to each neuron but that it cannot have information about the external input rate h . As a result, we neglect the factor $[1 - \lambda(h)]$ and obtain the adaptive (coalescence-compensating) weights

$$w_{\text{CC}}(m, A_t) = 1 - \left(1 - \frac{mA_t}{N}\right)^{1/A_t}, \quad (\text{C2})$$

which compensate only for internal coalescence. Inserted as weight in Eq. (4), the remaining external coalescence leads to

$$m_{\text{eff}}(w_{\text{CC}}(m, A_t), A_t) = m[1 - \lambda(h)], \quad (\text{C3})$$

which determines the $N \rightarrow \infty$ limit of linear-regression estimates [25].

The adaptive (activity-dependent) weight (C2) bridges the gap to the macroscopic description (Fig. 6). The conditional expectation value $\langle A_{t+1} | A_t \rangle$ for coalescence-compensating networks with adaptive weights now coincides with that of a branching process (dashed line), which is different from the conditional expectation value of the branching network subject to coalescence (solid line) (see Ref. [25]).

We note that Eq. (C2) is well defined only for $m < 1$, because for $m = 1$ the adaptive weights diverge at $A_t = N$. This is explained by the fact that for critical-like dynamics ($m = 1$) the compensation of coalescence induces an avalanche-size distribution with a perfect power-law behavior. A perfect power law, however, includes avalanches of all sizes. For a finite network with coalescence compensation, this mathematically results in a second absorbing state at full network activity.

We need to avoid the absorbing state at full network activity for $m \rightarrow 1$, because the microscopic dynamics would not allow us to ever leave this state again. For this we will derive a cutoff weight $w_{\text{CC}}(m, N) = w_{\text{CC}}^\dagger$ such that the probability to transition away from $A_t = N$ is nonzero.

To derive the cutoff weight, we need to calculate the probability that all neurons are activated if in the previous

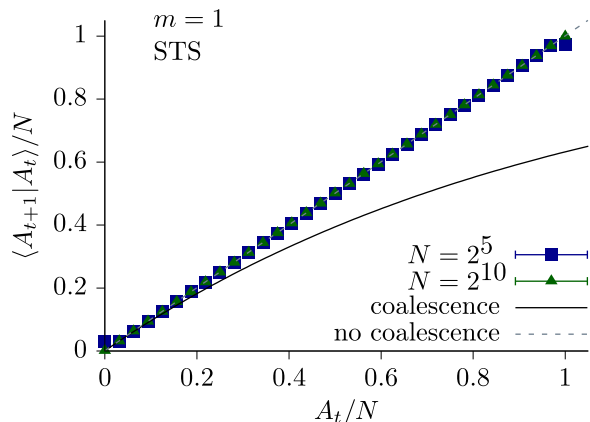


FIG. 6. Universal scaling function for effective spreading of network activity as discussed in Ref. [25] [$m = 1$ in the separation-of-timescale (STS) regime]. For the branching network (solid line), the scaling function is nonlinear, causing a bias in linear estimates of the branching parameter through the network rate. However, for coalescence-compensating networks (data points), the conditional expectation value is again linear and is described by the scaling function of a branching process (dashed line).

time step already all neurons were active. Recall that $P[s_t^i = 1 | A_t, w_{CC}, h] = 1 - (1 - w_{CC})^{A_t} [1 - \lambda(h)]$ for the probability to activate a single neuron given that A_t neurons are active. Now we restrict our discussion to the divergent case $m A_t = N$ in Eq. (C2) [defining, for $m = 1$, $w_{CC}^\dagger = w_{CC}(1, N)$] and further neglect the external input rate ($h \rightarrow 0$), because the divergence only occurs for $m \rightarrow 1$, where the external input rate has to be small for persistent activity. Then the probability to activate all neurons in the network, given that all neurons were activated already in the previous time step, is

$$p_{\text{all}} = [1 - (1 - w_{CC}^\dagger)^N]^N. \quad (\text{C4})$$

The probability to transition away from $A_t = N$ is given by the probability to not activate at least one neuron, i.e., by $1 - p_{\text{all}}$.

We now ask how to choose w_{CC}^\dagger such that once the network is fully active, there is a nonvanishing probability to transition away from $A_t = N$, i.e., that $p_{\text{all}} < 1$. For this we solve Eq. (C4) for w_{CC}^\dagger and obtain

$$w_{CC}^\dagger = 1 - (1 - p_{\text{all}}^{1/N})^{1/N} = 1 - (1 - e^{\alpha/N})^{1/N}, \quad (\text{C5})$$

where we have introduced $\alpha = \ln(p_{\text{all}}) < 0$. Next we aim for an asymptotic expansion around $N \rightarrow \infty$. For large N we can expand the exponential function $e^{\alpha/N} = 1 + \alpha/N + (\alpha/N)^2/2 + \dots$ such that

$$\begin{aligned} (1 - e^{\alpha/N})^{1/N} &= (1 - 1 - \alpha/N - (\alpha/N)^2/2 + \dots)^{1/N} \\ &= (-\alpha/N)^{1/N} [1 - (\alpha/N)/2 + \dots]^{1/N}. \end{aligned} \quad (\text{C6})$$

Restricting ourselves to the leading order, we get

$$w_{CC}^\dagger \simeq 1 - (-\alpha/N)^{1/N} = 1 - e^{\ln(-\alpha/N)/N}. \quad (\text{C7})$$

Ensuring that $\lim_{N \rightarrow \infty} \frac{1}{N} \ln(-\alpha/N) = 0$ according to l'Hôpital's rule and expanding $e^x \simeq 1 + x$, we obtain

$$w_{CC}^\dagger \simeq \frac{1}{N} \ln \left(\frac{N}{-\ln(p_{\text{all}})} \right). \quad (\text{C8})$$

Equation (C8) allows us to set the probability to transition away from the state of full network activity. We choose

$$w_{CC}^\dagger = \frac{\ln(N)}{N} \quad \text{such that } 1 - p_{\text{all}} = 1 - e^{-1} \approx 0.632, \quad (\text{C9})$$

such that the state of full network activity is guaranteed not to be absorbing. Our explicit choice for w_{CC}^\dagger has several motivations: First, it is the simplest choice; second, for $m = 1$, Eq. (C2) yields $w_{CC}(m = 1, N - 1) = 1 - (1 - \frac{N-1}{N})^{1/(N-1)} = 1 - \exp[\frac{1}{N-1} \ln(1/N)] \approx \ln(N)/N$ such that $w_{CC}^\dagger \approx w_{CC}(m = 1, N - 1)$; and last, w_{CC}^\dagger highlights the interpretation in terms of a random Erdős-Rényi network, namely, that in the limit $N \rightarrow \infty$ weights with probability $\ln(N)/N$ set us at the transition between disconnected graphs ($p_{\text{all}} = 0$) and fully connected graphs ($p_{\text{all}} = 1$).

APPENDIX D: ADAPTIVE WEIGHTS EXTEND THE RANGE OF CRITICAL BRANCHING STATISTICS

Above we argued that compensating for internal coalescence extends the range over which a finite network reproduces the statistics of a true branching process. In the following, we show that this is indeed the case even for critical dynamics ($m = 1$).

We here complement Fig. 3 ($N = 100$) by large networks with $N = 10^4$ neurons and critical dynamics ($m = 1$) in the separation-of-timescale regime (Fig. 7). Compared to a branching network of the same size, the avalanche-size distribution and avalanche-duration distribution in the coalescence-compensating model show a drastically extended power law. In particular, it appears that the power-law regime is extended from avalanche size $s \sim N$ to $s \sim N^2$ and from avalanche duration $d \sim \sqrt{N}$ to $d \sim N$. The results are preserved in the driven regime with small external input rate $h \ll 1$ (not shown).

We can understand the extended range of critical branching statistics by considering the universal shape of avalanches. It was shown that the average time development $A(t, d)$ of neural avalanches with the same duration d collapses on a universal shape $\mathcal{F}(x)$, if properly rescaled as $A(t, d) = d\mathcal{F}(t/d)$ for our model [35]. It can be anticipated that the power-law characteristics in finite-size branching networks extend up to $s_{\text{max}} = O(N)$. Considering any reasonable parabolic avalanche shape $\mathcal{F}(x)$ between rectangular and triangular, the area always scales as $s \sim d A^{\text{peak}}$ where the peak itself scales as $A^{\text{peak}} \sim d$ (according to the universal avalanche shape where $\mathcal{F}^{\text{peak}} = \text{const}$) such that $s \sim d^2$. With a maximal avalanche size $s_{\text{max}} = O(N)$, we thus expect a maximal avalanche duration $d_{\text{max}} = O(\sqrt{N})$. This expectation nicely agrees with our numerical observation (Figs. 7 and 3). In particular, this means that, due to coalescence, typical avalanches of duration d_{max} have a maximum number of simultaneously activated neurons that scales as $A_{\text{max}}^{\text{peak}} \sim d_{\text{max}} \sim \sqrt{N}$.

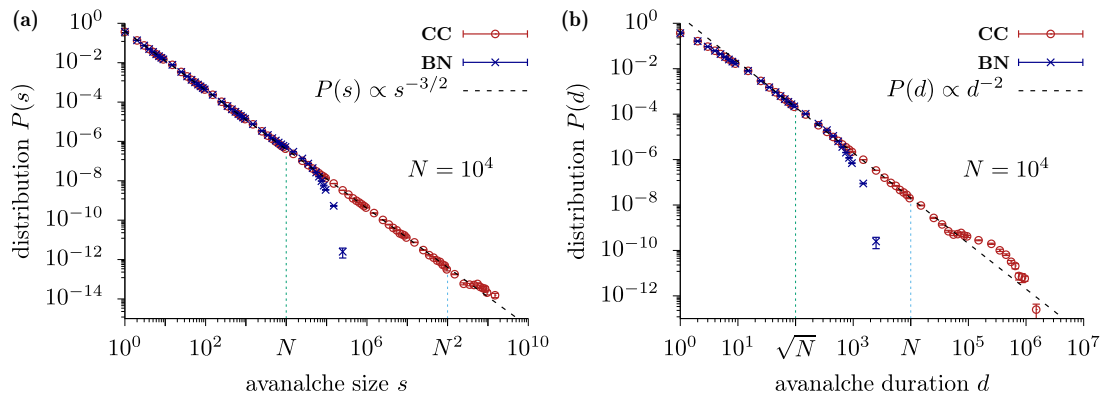


FIG. 7. (a) Avalanche-size and (b) avalanche-duration distributions for a branching network and a coalescence-compensating network with critical dynamics ($m = 1$) in the separation-of-timescale regime. Compensating coalescence extends the power-law behavior of the avalanche-size distribution from $s \sim N$ to $s \sim N^2$ and the power-law behavior of the avalanche-duration distribution from $d \sim \sqrt{N}$ to $d \sim N$.

Compensating coalescence now shifts the potential maximum number of simultaneously activated neurons from $A_{\max}^{\text{peak}} \sim \sqrt{N}$ to $A_{\max}^{\text{peak}} \sim N$. Because of the universal avalanche

shape, the maximum duration consequently scales as $d_{\max} \sim A_{\max}^{\text{peak}} \sim N$ and the maximal size then scales as $s_{\max} \sim d_{\max} A_{\max}^{\text{peak}} \sim N^2$.

- [1] S. Hecht, The visual discrimination of intensity and the Weber-Fechner law, *J. Gen. Physiol.* **7**, 235 (1924).
- [2] G. Borg, H. Diamant, L. Ström, and Y. Zotterman, The relation between neural and perceptual intensity: A comparative study on the neural and psychophysical response to taste stimuli, *J. Physiol.* **192**, 13 (1967).
- [3] N. F. Viemeister, Intensity coding and the dynamic range problem, *Hear. Res.* **34**, 267 (1988).
- [4] E. Rouiller, Y. de Ribaupierre, A. Morel, and F. de Ribaupierre, Intensity functions of single unit responses to tone in the medial geniculate body of cat, *Hear. Res.* **11**, 235 (1983).
- [5] N. Viemeister, Auditory intensity discrimination at high frequencies in the presence of noise, *Science* **221**, 1206 (1983).
- [6] C. Schreiner, J. Mendelson, and M. Sutter, Functional topography of cat primary auditory cortex: Representation of tone intensity, *Exp. Brain Res.* **92**, 105 (1992).
- [7] I. Dean, N. S. Harper, and D. McAlpine, Neural population coding of sound level adapts to stimulus statistics, *Nat. Neurosci.* **8**, 1684 (2005).
- [8] M. Wachowiak and L. B. Cohen, Representation of odorants by receptor neuron input to the mouse olfactory bulb, *Neuron* **32**, 723 (2001).
- [9] T. A. Cleland and C. Linstér, Computation in the olfactory system, *Chem. Senses* **30**, 801 (2005).
- [10] K. Britten, M. Shadlen, W. Newsome, and J. Movshon, The analysis of visual motion: A comparison of neuronal and psychophysical performance, *J. Neurosci.* **12**, 4745 (1992).
- [11] E. Evans and A. Palmer, Relationship between the dynamic range of cochlear nerve fibres and their spontaneous activity, *Exp. Brain Res.* **40**, 115 (1980).
- [12] E. F. Evans, in *Neuronal Mechanisms of Hearing*, edited by J. Syka and L. Aitkin (Springer US, Boston, 1981), pp. 69–85.
- [13] O. Kinouchi and M. Copelli, Optimal dynamical range of excitable networks at criticality, *Nat. Phys.* **2**, 348 (2006).
- [14] C. L. Buckley and T. Nowotny, Multiscale Model of an Inhibitory Network Shows Optimal Properties near Bifurcation, *Phys. Rev. Lett.* **106**, 238109 (2011).
- [15] W. L. Shew, H. Yang, T. Petermann, R. Roy, and D. Plenz, Neuronal avalanches imply maximum dynamic range in cortical networks at criticality, *J. Neurosci.* **29**, 15595 (2009).
- [16] S. H. Gautam, T. T. Hoang, K. McClanahan, S. K. Grady, and W. L. Shew, Maximizing sensory dynamic range by tuning the cortical state to criticality, *PLoS Comput. Biol.* **11**, e1004576 (2015).
- [17] J. M. Beggs, The criticality hypothesis: How local cortical networks might optimize information processing, *Philos. Trans. R. Soc. A* **366**, 329 (2008).
- [18] M. A. Muñoz, Colloquium: Criticality and dynamical scaling in living systems, *Rev. Mod. Phys.* **90**, 031001 (2018).
- [19] A.-C. Wu, X.-J. Xu, and Y.-H. Wang, Excitable Greenberg-Hastings cellular automaton model on scale-free networks, *Phys. Rev. E* **75**, 032901 (2007).
- [20] D. B. Larremore, W. L. Shew, and J. G. Restrepo, Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology, *Phys. Rev. Lett.* **106**, 058101 (2011).
- [21] T. A. Cleland and C. Linstér, Concentration tuning mediated by spare receptor capacity in olfactory sensory neurons: A Theoretical Study, *Neural Comput.* **11**, 1673 (1999).
- [22] J. Chevallier, Stimulus sensitivity of a spiking neural network model, *J. Stat. Phys.* **170**, 800 (2018).
- [23] C. Haldeman and J. M. Beggs, Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States, *Phys. Rev. Lett.* **94**, 058101 (2005).
- [24] J. Zierenberg, J. Wilting, and V. Priesemann, Homeostatic Plasticity and External Input Shape Neural Network Dynamics, *Phys. Rev. X* **8**, 031018 (2018).
- [25] J. Zierenberg, J. Wilting, V. Priesemann, and A. Levina, Description of spreading dynamics by microscopic network

- models and macroscopic branching processes can differ due to coalescence, *Phys. Rev. E* **101**, 022301 (2020).
- [26] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, On the LambertW function, *Adv. Comput. Math.* **5**, 329 (1996).
- [27] J. Wilting and V. Priesemann, Inferring collective dynamical states from widely unobserved systems, *Nat. Commun.* **9**, 2325 (2018).
- [28] J. Wilting, J. Dehning, J. Pinheiro Neto, L. Rudelt, M. Wibral, J. Zierenberg, and V. Priesemann, Operating in a reverberating regime enables rapid tuning of network states to task requirements, *Front. Syst. Neurosci.* **12**, 55 (2018).
- [29] T. E. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963).
- [30] J. M. Beggs and D. Plenz, Neuronal avalanches in neocortical circuits, *J. Neurosci.* **23**, 11167 (2003).
- [31] V. Priesemann, M. Wibral, M. Valderrama, R. Pröpper, M. Le Van Quyen, T. Geisel, J. Triesch, D. Nikolić, and M. H. J. Munk, Spike avalanches *in vivo* suggest a driven, slightly subcritical brain state, *Front. Syst. Neurosci.* **8**, 108 (2014).
- [32] A. Levina, J. M. Herrmann, and M. Denker, Critical branching processes in neural networks, *PAMM* **7**, 1030701 (2008).
- [33] A. Levina, U. Ernst, and J. M. Herrmann, Criticality of avalanche dynamics in adaptive recurrent networks, *Neurocomputing* **70**, 1877 (2007).
- [34] D. B. Larremore, W. L. Shew, E. Ott, F. Sorrentino, and J. G. Restrepo, Inhibition Causes Ceaseless Dynamics in Networks of Excitable Nodes, *Phys. Rev. Lett.* **112**, 138103 (2014).
- [35] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal Critical Dynamics in High Resolution Neuronal Avalanche Data, *Phys. Rev. Lett.* **108**, 208102 (2012).
- [36] C. Wilson, Up and down states, *Scholarpedia* **3**, 1410 (2008).
- [37] E. A. Stern, A. E. Kincaid, and C. J. Wilson, Spontaneous subthreshold membrane potential fluctuations and action potential variability of rat corticostriatal and striatal neurons *in vivo*, *J. Neurophysiol.* **77**, 1697 (1997).
- [38] D. Holcman and M. Tsodyks, The emergence of up and down states in cortical networks, *PLoS Comput. Biol.* **2**, e23 (2006).
- [39] D. Millman, S. Mihalas, A. Kirkwood, and E. Niebur, Self-organized criticality occurs in non-conservative neuronal networks during 'up' states, *Nat. Phys.* **6**, 801 (2010).
- [40] C. R. von Reyn, P. Breads, M. Y. Peek, G. Z. Zheng, W. R. Williamson, A. L. Yee, A. Leonardo, and G. M. Card, A spike-timing mechanism for action selection, *Nat. Neurosci.* **17**, 962 (2014).
- [41] D. V. Buonomano and M. M. Merzenich, Temporal information transformed into a spatial code by a neural network with realistic properties, *Science* **267**, 1028 (1995).
- [42] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
- [43] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [44] U. D. Schiller and J. J. Steil, Analyzing the weight dynamics of recurrent learning algorithms, *Neurocomputing* **63**, 5 (2005).
- [45] H. Jaeger, W. Maass, and J. Principe, Special issue on echo state networks and liquid state machines, *Neural Netw.* **20**, 287 (2007).
- [46] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, Information processing in echo state networks at the edge of chaos, *Theory Biosci.* **131**, 205 (2012).
- [47] B. Cramer, D. Stöckel, M. Kreft, J. Schemmel, K. Meier, and V. Priesemann, Control of criticality and computation in spiking neuromorphic networks with plasticity, [arXiv:1909.08418](https://arxiv.org/abs/1909.08418).
- [48] D. J. Heeger, Normalization of cell responses in cat striate cortex, *Vis. Neurosci.* **9**, 181 (1992).
- [49] J. Clemens, N. Ozeri-Engelhard, and M. Murthy, Fast intensity adaptation enhances the encoding of sound in *Drosophila*, *Nat. Commun.* **9**, 134 (2018).