Statistical Physics of Neural Systems with Nonadditive Dendritic Coupling

David Breuer,[1,2,*] Marc Timme,[1,3,4] and Raoul-Martin Memmesheimer[5]

[1]*Network Dynamics, Max Planck Institute for Dynamics and Self-Organization (MPIDS),*
37077 Göttingen, Germany
[2]*Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology (MPIMP),*
14476 Potsdam-Golm, Germany
[3]*Bernstein Center for Computational Neuroscience (BCCN), 37077 Göttingen, Germany*
[4]*Faculty of Physics, Institute for Nonlinear Dynamics, Georg August University,*
37077 Göttingen, Germany
[5]*Department of Neuroinformatics, Donders Institute, Radboud University, 6525 AJ Nijmegen, Netherlands*

How neurons process their inputs crucially determines the dynamics of biological and artificial neural networks. In such neural and neural-like systems, synaptic input is typically considered to be merely transmitted linearly or sublinearly by the dendritic compartments. Yet, single-neuron experiments report pronounced supralinear dendritic summation of sufficiently synchronous and spatially close-by inputs. Here, we provide a statistical physics approach to study the impact of such nonadditive dendritic processing on single-neuron responses and the performance of associative-memory tasks in artificial neural networks. First, we compute the effect of random input to a neuron incorporating nonlinear dendrites. This approach is independent of the details of the neuronal dynamics. Second, we use those results to study the impact of dendritic nonlinearities on the network dynamics in a paradigmatic model for associative memory, both numerically and analytically. We find that dendritic nonlinearities maintain network convergence and increase the robustness of memory performance against noise. Interestingly, an intermediate number of dendritic branches is optimal for memory functionality.

Subject Areas: Biological Physics, Complex Systems,
                                                                                                    Statistical Physics

I. INTRODUCTION: NONADDITIVE DENDRITIC INPUT PROCESSING IN NEURAL NETWORKS

Information processing in artificial and biological neural networks crucially depends on the processing of inputs in single neurons (e.g., Ref. [1]). The dendrites, branched protrusions of a biological nerve cell or the input preprocessing compartments of formal neurons, constitute the main input sites. Traditionally, dendrites are modeled as passive, cablelike conductors that integrate incoming presynaptic signals linearly or sublinearly and propagate the change in voltage to the cell body or soma, where it is subject to nonlinear transformations [2]. Accordingly, the input preprocessing in formal neurons is usually assumed to be a linear or sublinear summation.

Single-neuron experiments, however, demonstrate the occurrence of strongly supralinear dendritic amplification. Biophysically, this supralinear amplification is caused by action potentials generated in the dendrite of the neuron. Such dendritic spikes are mediated by voltage-dependent

ion channels such as sodium, calcium, and N-methyl-D-aspartate (NMDA) channels [3–7]. In particular, dendritic spikes may emerge if sufficiently synchronous inputs are received by the same branch of a dendrite. The many inputs to the dendrites can thus be processed nonadditively, depending on their spatial and temporal distribution [5,8]. These features of input processing imply crucial deviations from the classical assumptions on linear dendritic input processing as modeled, e.g., by cable equations. It has been recently shown that dendritic spikes are present and prominent all over the brain (e.g., Ref. [9]).

A number of theoretical studies already highlighted the importance of nonlinear, spiking dendrites for the input processing in single neurons: Simulations of neuron models with detailed channel density and morphology showed dendritic spike generation in agreement with neurobiological experiments [3,4,6,8,10]. Further, firing-rate models have been developed [11] that reproduce the response properties of detailed models to diverse stimuli and behave like multilayered feed-forward networks of simple-rate neurons [5,8,10]. Two- and multilayer feed-forward networks of binary, deterministic neurons have been studied using statistical physics methods [12–14]. In particular, the so-called committee machine may be seen as a neuron model incorporating a layer of dendrites with steplike activation functions, i.e., without analogous signal

---

*breuer@mpimp-golm.mpg.de

transmission [15,16]. Neurons in biological networks receive time-dependent, noisy input at high rates, which often makes a statistical description of the response properties of single neurons necessary. In Ref. [17], the authors derived such a description for linear and quadratic dendritic summation together with some numerical results for a biologically plausible, sigmoidal dendritic nonlinearity. The propagation of dendritic spikes in branched dendrites with steplike activation functions has been studied in Ref. [18], providing the somatic input as a numerical solution to a high-dimensional system of nonlinear equations.

To date, there is no efficient statistical description for neurons with biologically plausible, sigmoidal dendritic nonlinearities. In biological systems, neurons form complex, recurrent networks. Thus, a description that allows one to analytically study networks of neurons with multiple nonlinear dendrites is especially desirable.

Recent single-neuron experiments investigated the role of active dendrites in detecting specific spatiotemporal input patterns [4,19,20]. Theoretical studies showed that nonlinear dendrites improve the ability of single neurons and ensembles of single neurons to discriminate and learn different input patterns [21–24]. Besides trivially multiplying the single-neuron abilities to detect input patterns, a network of neurons can store, retrieve, and complete spatiotemporal patterns aided by its recurrent dynamics: It can function as an associative-memory device [13,25,26]. Yet, the impact of nonlinear dendrites on associative-memory networks is unknown.

So far, only a few studies have considered the impact of nonadditive dendrites on network dynamics. Selectivity and invariance of network responses to external stimuli and their intensity were analyzed in a firing-rate model [27,28]. References [28–30] proposed that NMDA-receptor-dependent dendritic nonlinearities play a crucial role in working memory, i.e., in the formation of persistent activity in unstructured networks. Nonlinear, multiplicative dendritic processing arising from spatial summation of input across the dendritic arbor was similarly shown to enable spontaneous and persistent network activity [31]. Dendritic spikes were suggested to work as coincidence detectors and provide a neuronal basis for temporal and spatial contexts in biological networks [32,33]. References [34,35] studied networks of bursting neurons, where the bursts facilitate the emergence of patterns of coordinated neuronal activity and can be explained by dendritic spikes. Further, it was shown that nonlinear dendrites can enable robust propagation of synchronous spiking in random networks with biologically plausible substructures [36] and in purely random networks [37]. Finally, dendritic spikes were related to so-called sharp-wave ripples in the hippocampus that are important for long-term memory consolidation [38].

Networks of binary neurons with linear input summation have been intensively investigated in statistical physics ("Hopfield networks" [13,25,26]), and extensions to different nonlinear and nonmonotonic transfer functions exist (cf., e.g., Refs. [39–43]). While all of these studies assumed point neurons, neural networks of arborized neurons with nonadditive coupling have not been studied in comparable setups. Hopfield networks are paradigmatic models for associative memory that may, in particular, contribute to solving two important conundrums in neuroscience: how biological neural networks achieve a high memory capacity and how they can work so reliably under the experimentally found noisy conditions. The incorporation of nonadditive dendrites into these models may therefore (1) shed light on the impact of these features on memory capacity and robustness and, at the same time, (2) allow us to understand the underlying mechanisms due to their analytical tractability.

In the first part of this article, we describe the response properties of single neurons in the presence of biologically plausible dendritic nonlinearities in a statistical framework. In the second part, we employ the results and study the effect of nonlinear dendrites on associative-memory networks. We consider networks of the Hopfield type, as this standard model for associative memory lends itself to analytical treatment and allows us to concisely work out the effects of nonlinear dendritic enhancement. We find that dendritic nonlinearities improve pattern retrieval by effectively reducing the thresholds of neurons and by increasing the robustness to noise. The improvement is strongest for intermediate numbers of dendritic branches. We quantify these effects and illustrate our analytical findings with numerical simulations.

II. RESULTS

A. Basic model for neurons with nonadditive dendrites

Consider an extended topological structure of a neuron consisting of one pointlike soma and B-independent dendritic compartments (Fig. 1). Each compartment receives its inputs from a number of presynaptic neurons and transfers its output to the soma. We assume that nonlinear dendritic integration takes place over a time window Δt. Our model can be applied to dendrites with fast or slow dendritic spikes, where Δt assumes values of 2–3 ms (fast, sodium spikes [3,4]) or 10 s of ms (slow, e.g., NMDA spikes [5,44]). The durations of the different dendritic spikes have time scales similar to their integration windows. The input arriving at a branch within Δt is denoted by u. On each branch, we capture the nonadditive input summation of the dendrites through a piecewise linear, sigmoidal transfer function

$$f(u) \coloneqq \begin{cases} u & \text{if } u < \theta \\ D & \text{otherwise.} \end{cases} \tag{1}$$

If u is smaller than a threshold θ, i.e., $u < \theta$, the inputs superpose linearly. Biologically, this linear superposition means that the inputs u has not reached the threshold θ for
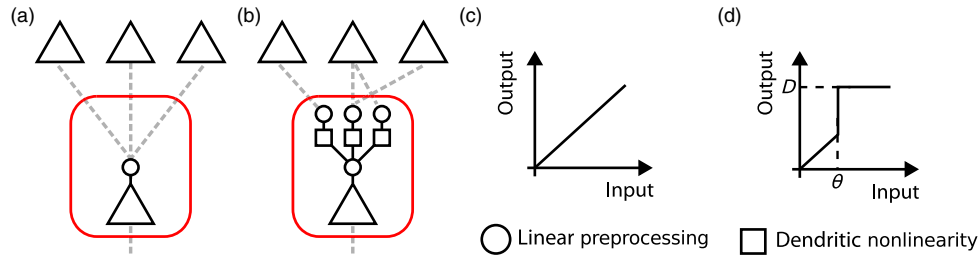
FIG. 1.   Arborized neuron with dendritic nonlinearities. Network architecture for (a) a classical point-neuron model without dendrites and (b) an arborized neuron with $B = 3$ dendrites (or dendritic branches) modeled as separate compartments. The red boxes mark the neuronal unit for each case. (c) Circles represent linear summation of inputs; triangles and squares represent somatic and dendritic processing, respectively. In a point neuron, all inputs are summed up linearly and then processed nonlinearly. In a model with nonlinear dendrites, there is an additional, preceding layer where inputs to each dendrite are summed up linearly and then subjected to a dendritic nonlinearity. (d) This nonlinearity is modeled as a piecewise linear function with threshold θ and saturation strength D, incorporating the effect of dendritic spikes. The somatic transfer function is not constrained in the model.

dendritic spike generation and that it is conventionally transferred to the soma. If the threshold is exceeded, i.e., $u \geq \theta$, the inputs superpose nonadditively and a fixed dendritic output strength D is attained. Such a fixed output strength models the effect of a dendritic spike elicited by sufficiently strong input. The summation scheme as well as the compartmentalization are in agreement with experimental findings and modeling studies [3–5,8,27,38]. Our approaches may be directly extended to neurons with multiple stages of dendritic processing (cf. Appendix C).

B. Capturing dendritic spikes by an effective somatic input

To quantify the impact of nonadditive dendritic events on the neuronal input processing, the temporal and spatial distribution of synaptic input must be taken into account. We consider a neuron with B dendritic branches $b \in \{1, ..., B\}$ and some time interval of the length of the dendritic integration window. x_b denotes the number of synapses on branch b that are active within this window. The numbers of active synapses are distributed according to $P(x_1, ..., x_B)$. Furthermore, we allow for distributed connection strengths by assigning the synapses weights w that are independently and identically distributed according to $P(w)$. Averages with respect to $P(x_1, ..., x_B)$ and $P(w)$ can be interpreted as ensemble averages or temporal averages. The ensemble average is taken over a large number of neurons at a fixed time where each neuron has numbers x_b of active inputs and weights w that are samples of $P(x_1, ..., x_B)$ and $P(w)$, respectively. Under the additional assumption of a large number of synaptic contacts on each branch, the averages may also be understood as time averages, which are taken at a fixed neuron over a suitably segmented long time interval in which the active inputs are changing. In this article, we follow the first interpretation of ensemble averages.

What is the effective input to the soma given that synaptic inputs are distributed across branches? We assume that each branch samples a volume in which

synapses of axons from S other (presynaptic) neurons can be synaptically contacted [45]. A synapse is present and active with probability p_b such that $P(x_1, ..., x_B)$ is a product of binomial distributions with means $E[x_b] = Sp_b$ and variances $\text{Var}[x_b] = Sp_b(1 - p_b)$. Alternatively, the total number of active synaptic terminals across branches might be fixed to S (e.g., due to homeostatic learning), which suggests a multinomial distribution for $P(x_1, ..., x_B)$. Then, the x_b on different branches are not independent but negatively correlated with covariances $\text{Cov}[x_b, x_c] = -Sp_b p_c$ for $b, c \in \{1, ..., B\}$ and $b \neq c$.

The input to branch b is given by the linear sum

$$u_b = \sum_{i=1}^{x_b} w_i, \qquad (2)$$

and we are interested in the distribution $P(u_1, ..., u_B)$ of input across branches. According to Wald's equation [46], the Blackwell-Girshick equation [47], and the conditional covariance formula [48], we have for $b \neq c$

$$E[u_b] = E[x_b]E[w], \qquad (3)$$

$$\text{Var}[u_b] = E[x_b]\text{Var}[w] + \text{Var}[x_b]E^2[w], \qquad (4)$$

$$\begin{aligned}
\text{Cov}[u_b, u_c] &= E(\text{Cov}[u_b, u_c | x_b, x_c]) \\
&\quad + \text{Cov}(E[u_b | x_b, x_c], E[u_c | x_b, x_c]) \\
&= \text{Cov}[x_b, x_c]E^2[w], \qquad (5)
\end{aligned}$$

where $\text{Cov}[u_b, u_c | x_b, x_c] = 0$ and $E[u_b | x_b] = x_b E[w]$ due to the independence of the synaptic weights w. $P(u_1, ..., u_B)$ is, in general, a complicated distribution (Appendix B). Since its first moments are known [Eqs. (3)–(5)], we approximate $P(u_1, ..., u_B)$ by a multivariate normal distribution, i.e., by the maximum entropy distribution for the given moments.

This approximation enables us to derive an effective input to the soma that depends only on S, the number B of

branches, the probabilities p_b, and the moments $E[w]$ and $\mathrm{Var}[w]$. For only linear branches, i.e., $u_b < \theta$ on all branches, the input to the soma is simply given by the linear sum $\sum_{b=1}^{B} u_b$. For $u_b \geq \theta$, the dendritic nonlinearity sets in and branch b provides input of strength D to the soma. Since somatic preprocessing is linear, the total input to the soma is

$$F = F(u_1, \ldots, u_B) = \sum_{b=1}^{B} f(u_b). \tag{6}$$

The evaluation of the sum is numerically simple but denies analytical treatment. Yet, for Gaussian $P(u_1, \ldots, u_B)$, we may compute its mean (Appendix A) via

$$
\begin{aligned}
E[F] &= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)] \\
&\quad \times P(u_1, \ldots, u_B) \\
&= B\left(\int_{\theta}^{\infty} du D P(u) + \int_{-\infty}^{\theta} du u P(u) \right) \\
&= BP_{\mathrm{NL}} D + B(1 - P_{\mathrm{NL}}) E[u] - BC_{\mathrm{NL}},
\end{aligned} \tag{7}
$$

where we exploit that the marginal distribution $P(u)$ of the multivariate normal distribution $P(u_1, \ldots, u_B)$ is a normal distribution again and define

$$P_{\mathrm{NL}} := \frac{1}{2} \mathrm{erfc}\left(\frac{\theta - E[u]}{\sqrt{2\mathrm{Var}[u]}} \right), \tag{8}$$

$$C_{\mathrm{NL}} := \sqrt{\frac{\mathrm{Var}[u]}{2\pi}} \exp\left(-\frac{(\theta - E[u])^2}{2\mathrm{Var}[u]} \right). \tag{9}$$

In the second line of Eq. (7), we assume $p_b = p_0$, where p_0 is a constant probability independent of b. Throughout the rest of the article, we follow this choice for simplicity, although many results are independent of p_b or may be easily generalized to arbitrary p_b. Equation (7) may be interpreted as follows: The first part describes the expected number BP_{NL} of dendritic spikes of strength D, while the second part captures $B(1 - P_{\mathrm{NL}})$ linear events, and the last part BC_{NL} corrects the overestimate of the contribution of the linear branches by $E[u]$. To obtain the variance $\mathrm{Var}[F] = E[F^2] - E^2[F]$, we analogously derive the second moment (Appendix A)

$$
\begin{aligned}
E[F^2] &= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)]^2 \\
&\quad \times P(u_1, \ldots, u_B) \\
&= B \int_{-\infty}^{\infty} du f^2(u) P(u) \\
&\quad + (B^2 - B) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} du dv f(u) f(v) P(u,v). \tag{10}
\end{aligned}
$$

$P(u,v)$ is the marginal distribution of $P(u_1, \ldots, u_B)$ in two variables. For binomially distributed synapses, $P(u,v) = P(u)P(v)$ factors into two independent normal distributions, which yields

$$
\begin{aligned}
E[F^2] &= BP_{\mathrm{NL}} D^2 + B(1 - P_{\mathrm{NL}})(E^2[u] + \mathrm{Var}[u]) \\
&\quad - BC_{\mathrm{NL}}(E[u] + \theta) + (B^2 - B)B^{-2}E^2[F]. \tag{11}
\end{aligned}
$$

For multinomially distributed synapses, the double integral in Eq. (10) needs to be evaluated numerically (see Appendix A and Fig. 2).

The mean $E[F]$ and its variance $\mathrm{Var}[F]$ (as well as the expected number $E[k]$ of nonlinear branches and its variance $\mathrm{Var}[k]$; see below) may also be computed without the Gaussian approximation employing the exact expression for $P(u_1, \ldots, u_B)$ (Appendix B).

We note that our approximation can be employed to compute the input statistics to neurons with several layers of nonadditive dendritic branches by iteratively applying the formulas for $E[F]$ and $\mathrm{Var}[F]$ [Eqs. (7)–(10)] to each branching point and using the result as a new input to the next layer (cf. Appendix C for a derivation).

C. Features of the effective somatic input

We compare the input F to the soma from numerical simulations [Eq. (6) and circles in Fig. 2], from the Gaussian approximation [Eqs. (7) and (10) and solid lines in Fig. 2; Appendix A], and from the exact solution (Appendix B and dashed lines in Fig. 2) and find good agreement. We choose $p_0 = B^{-1}$ for both the binomial and multinomial cases for a direct comparability of the two. Several features of the input statistics may be noticed: The average $E[F]$ is the same for the binomial and multinomial distributions $P(x_1, \ldots, x_B)$ (Fig. 2, Gaussian approximation in a solid gray line and exact solution in a dashed black line) because their marginal distributions in one variable are the same; in particular, correlations among branches do not contribute. The variation $\mathrm{Var}[F]$ is larger in the binomial (Fig. 2, solid cyan and dashed blue lines) than in the multinomial scenario (Fig. 2, solid orange and dashed red lines) since the total number of active synaptic inputs is constant in the latter and allows less fluctuation of input across branches. In the strongly nonlinear regime, i.e., $E[u] \gg \theta$, all branches are saturated and the average $E[F]$ approaches saturation DB (Fig. 2, dotted black line). In the linear regime, i.e., $E[u] \ll \theta$, Eq. (6) becomes $F = \sum_{i=1}^{x} w_i$, where $x = \sum_{b=1}^{B} x_b$ is the total number of active synapses. Then, the average $E[F] = SE[w]$ and hence grows linearly with the number of inputs S but is independent of B (Fig. 2, dash-dotted black line). Further, the variance $\mathrm{Var}[F] = S\mathrm{Var}[w] + S(1 - B^{-1})E^2[w]$ in the binomial scenario [cf. Eq. (4) and the dash-dotted blue line in Fig. 2] because $E[x] = S$ and $\mathrm{Var}[x] = S(1 - B^{-1})$. In the multinomial scenario, $\mathrm{Var}[F] = S\mathrm{Var}[w]$ (Fig. 2,
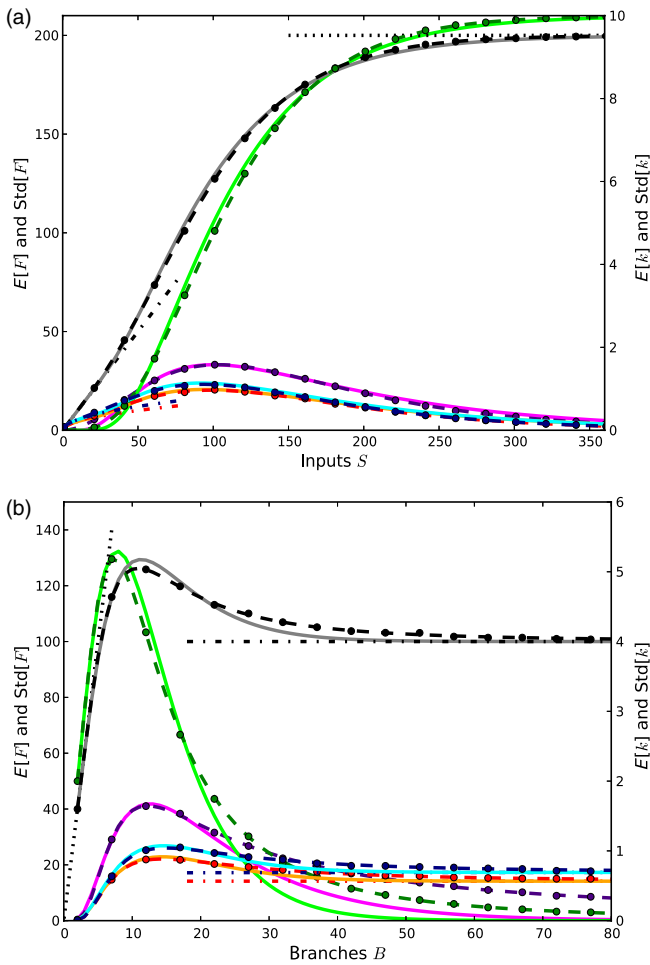
FIG. 2.   Mean and standard deviation of the somatic input F, optimal number of branches. Results for (a) fixed $B = 10$ and varying S and (b) vice versa, with $S = 100$. The remaining parameters are $\theta = 10$, $D = 20$, and $p_0 = B^{-1}$, and $P(w)$ is Gaussian with mean $E[w] = 1$ and variance $\text{Var}[w] = 2$. The effective somatic input $E[F]$, the average number of nonlinear branches $E[k]$, and their standard deviations $\text{Std}[F] = \sqrt{\text{Var}[F]}$ and $\text{Std}[k] = \sqrt{\text{Var}[k]}$ are derived numerically using 2000 realizations of binomially and multinomially distributed synapses (circles). They agree well with the Gaussian approximation (solid lines) and the exact analytical solution (dashed lines). $E[F]$ are shown in black (exact solution, dashed lines) and gray (Gaussian approximation, solid lines); the lines that represent $\text{Std}[F]$ are colored blue and cyan for binomially distributed synapses and red and orange for multinomially distributed synapses. Lines that represent averages $E[k]$ are colored green and lime, and their variances $\text{Std}[k]$ for the binomial scenario are shown in purple and pink. Dash-dotted and dotted lines represent the linear and saturated limits, respectively. See the main text for a detailed discussion.

dash-dotted red line) because the total number of active synapses is fixed so that $\text{Var}[x] = 0$.

Another important feature of the mean somatic input $E[F]$ is its maximum at an intermediate number $B = B_{F,\text{opt}}$ of branches. [In Fig. 2(b), $B_{F,\text{opt}} = 11$.] To better understand

this maximum, we compute the expectation value $E[k] = BP_{\text{NL}}$ of the number k of branches in the nonlinear regime [and its variance $\text{Var}[k] = BP_{\text{NL}}(1 - P_{\text{NL}})$ in the binomial case; Gaussian approximation, cf. Appendix A and Fig. 2, solid line and pink lines; exact solution, cf. Appendix B and Fig. 2 dashed green and purple lines]. $E[k]$ has a maximum. The presence of a maximum is plausible since the number of nonlinear branches typically starts with one for $B = 1$ because all input is concentrated on this branch, then increases when more branches are available, but goes to zero for large B. $E[k]$ assumes its maximum at approximately $B_{k,\text{opt}} \approx \frac{\text{SE}[w]}{\theta}$ because $\text{Std}[u] \ll E[u]$ for (biologically plausible) sufficiently large ratios SB^{-1} [Eqs. (3) and (4)] so that P_{NL} [Eq. (8)] approaches a step function and $E[k] = BP_{\text{NL}} \approx B\text{step}(\theta - \frac{S}{B}E[w])$. The somatic input $E[F]$ [Eq. (7)] has two contributions: The first (from nonlinearly enhanced inputs) is $DE[k]$, and the second (from linearly summed inputs) is monotonically increasing in B. Since D is comparably large, the maximum of $E[k]$ induces a maximum in $E[F]$. The latter is shifted to the right due to the monotonic increase of the linear contribution. The shift indicates that a few additional branches may further increase $E[F]$ because synapses can there provide input that would otherwise be lost on saturated, nonlinear branches. Because a further increase in the number of branches, however, leads to a substantial loss of (nonadditive) input, the maximum of $E[F]$ is close to that of $E[k]$, i.e., $B_{F,\text{opt}} \approx \frac{\text{SE}[w]}{\theta}$.

Up to now, we considered a combined processing of inhibition and excitation on the dendritic branches. Often, inhibitory synapses are found to directly target the soma [49,50]. Such input can be readily incorporated in our model by including an extra term in Eq. (7):

$$E[F] = BP_{\text{NL}}D + B(1 - P_{\text{NL}})E[u_D] - BC_{\text{NL}} + E[u_S], \quad (12)$$

where $E[u_D]$ is the average input to a dendrite and P_{NL} and C_{NL} [Eqs. (8) and (9)] are computed using $u = u_D$. $E[u_S]$ is the mean direct somatic input. Both summation scenarios may lead to different collective dynamics on the network level (see below).

Concluding, we modeled the somatic input of a neuron with nonadditive dendrites. Our findings are independent of a specific neuron model. We introduced a Gaussian approximation to describe the input irrespective of the particular distribution of active synapses across branches [Eqs. (7) and (10)]. It provides a sufficiently good description (cf. Fig. 2), simplifies calculations, and is therefore used in the remainder of this article.

D. Deterministic Hopfield networks of arborized neurons

How do nonlinear dendrites influence the dynamics of associative neural networks? Because of its analytical accessibility and its relevance in neural computation,

we consider a Hopfield network [25] of N neurons $n \in \{1, ..., N\}$ with discrete states $v_n \in \{-1, +1\}$ and asynchronous updates of one random unit at a discrete time $t \in \mathbb{N}$. We might interpret t as being measured in units of $N^{-1}\Delta t$ so that on average each neuron is updated once per Δt (approximately equaling the dendritic spike duration) and samples states that are present in other neurons for Δt (approximately equaling the dendritic integration window). The update rule for the conventional deterministic Hopfield model reads

$$v_n(t+1) = \text{sgn}\{u_n[v_1(t), ..., v_N(t)] - \Theta\}, \quad (13)$$

where n is the neuron updated at t. $\text{sgn}(x)$, with $\text{sgn}(x) = -1$ if $x < 0$ and $\text{sgn}(x) = 1$ otherwise, is the neuronal transfer function, and Θ denotes the neuronal threshold. u_n is the linear field

$$u_n = u_n(v_1, ..., v_N) = \sum_{m=1}^{N} w_{n,m} v_m. \quad (14)$$

The couplings $w_{n,m}$ between neurons n and $m \in \{1, ..., N\}$ are assumed to be symmetric $w_{n,m} = w_{m,n}$. Then, a Lyapunov function may be derived

$$E_L[v_1(t), ..., v_N(t)]$$
$$= -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} w_{n,m} v_n(t) v_m(t) + \sum_{n=1}^{N}\Theta v_n(t), \quad (15)$$

satisfying $E_L(t+1) \leq E_L(t)$ [13,25]. The equality $E_L(t+1) = E_L(t)$ only occurs if the state of the network upon update is not changed or in the rare case when u_n matches exactly the threshold Θ. Since E_L is bounded and $u_n = \Theta$ implies $v_n(t+1) = 1$, the weak Lyapunov property guarantees convergence of the system. Thus, the network converges to an asymptotically stable minimum in the energy landscape $E_L(v_1, ..., v_N)$.

To store P patterns ξ^p, $p \in \{1, ..., P\}$, the couplings in the Hopfield model are set in Hebbian manner [51]:

$$w_{n,m} = N^{-1}\sum_{p=1}^{P}\xi_n^p\xi_m^p. \quad (16)$$

Classically, the storage of random, uncorrelated patterns $\xi_n^p \in \{-1, +1\}$ is studied, where $\xi_n^p = \pm 1$ with equal probabilities. Self-coupling terms $w_{n,n}$ may lead to spurious states close to stored patterns and are usually omitted; $w_{n,n} = 0$ [13]. In this article, we adopt these conventions.

An alternative and similarly common model represents the neuronal states via $v_n \in \{0, 1\}$ (cf., e.g., Ref. [52]). For an appropriate choice of variables (cf. Ref. [13]), this representation is equivalent to the $\{-1, +1\}$ model, when introducing a dependence of the neuronal threshold Θ on the couplings w_{nm}. Since we assume constant Hebbian couplings [Eq. (16)], we can include this term into the (then

still constant) thresholds and translate a $\{0, 1\}$ model into a $\{-1, +1\}$ model. Because it is most often used in classical statistical mechanics studies of neural networks (cf. Ref. [53]), we adopt the $\{-1, +1\}$ representation.

We now modify this well-known model to incorporate dendritic branches. For simplicity, we assume that each neuron has B dendritic branches. The arborization changes the network topology (Fig. 1) since neurons are now linked to branches. The coupling matrix becomes an $N \times B \times N$ "matrix" with entries $w_{n,b,m}$ that characterize the coupling of neuron m to branch b of neuron n. The input to an individual dendrite is given by the dendritic field

$$u_{n,b} = u_{n,b}(v_1, ..., v_N) = \sum_{m=1}^{N} w_{n,b,m} v_m. \quad (17)$$

The inputs are processed by the dendrites according to Eq. (1), and the somatic input is given by Eq. (6). Taken together, the update rule at time t reads

$$v_n(t+1) = \text{sgn}\{G_n[v_1(t), ..., v_N(t)] - \Theta\}, \quad (18)$$

where

$$G_n = G_n(v_1, ..., v_N)$$
$$= F[u_{n,1}(v_1, ..., v_N), ..., u_{n,B}(v_1, ..., v_N)]. \quad (19)$$

Like in the classical Hopfield model, we assume a Hebbian rule that strengthens connections $w_{n,b,m}$ between coactive neurons such that their expected value is $E[w_{n,b,m}] = B^{-1}w_{n,m}$, with $w_{n,m}$ given by Eq. (16). Because the process of adjustment of synaptic weights is subject to fluctuations [54], we further assume that the weights $w_{n,b,m}$ are distributed with variance $\text{Var}[w_{n,b,m}] = w_{n,m}^2 B^{-2}\text{Var}[w]$. The width of the distribution is proportional to the mean (with a parameter $\text{Var}[w]$), which avoids excessively large deviations for small weights.

The network is fully connected, and because input correlations across branches vanish (Appendix D), this setup can be identified with the binomial scenario introduced before, with $S = N$ and $p_b = p_0 = 1$. Equations (3) and (4) yield (Appendix D)

$$E[u_{n,b}] = B^{-1}u_n, \quad (20)$$

$$\text{Var}[u_{n,b}] = B^{-2}PN^{-1}\text{Var}[w]. \quad (21)$$

The moments are computed as ensemble averages over an ensemble of neurons with index n at a fixed network state (annealed approximation). The neural identity is preserved as it is specified by the parameters $w_{n,m}$ that determine the expectation value and the variance of the weight distribution of $w_{n,b,m}$ over which we average. An averaging over x_b is unnecessary as $p_b = p_0 = 1$, and thus $E[x_b] = N$ and

$\mathrm{Var}[x_b] = 0$. The mean somatic input at neuron n, $E[G_n]$, follows from Eqs. (7)–(9). It depends on $v_1, ..., v_N$ and n only via the mean input per branch [Eq. (20)] and thus via the linear field u_n [Eq. (14)]. We may therefore define an effective input function $\bar{F} = \bar{F}(u_n)$ as

$$\bar{F}[u_n(v_1, ..., v_N)] = E[G_n(v_1, ..., v_N)]. \qquad (22)$$

From Eqs. (7)–(9), we find

$$\bar{F}(u_n) = \mathrm{BP_{NL}}D + (1 - P_{NL})u_n - \mathrm{BC_{NL}} \qquad (23)$$

with

$$P_{NL} = \frac{1}{2}\mathrm{erfc}\left(\frac{B\theta - u_n}{\sqrt{2PN^{-1}\mathrm{Var}[w]}}\right), \qquad (24)$$

$$C_{NL} = B^{-1}\sqrt{\frac{P\mathrm{Var}[w]}{2\pi N}}\exp\left(-\frac{(B\theta - u_n)^2}{2PN^{-1}\mathrm{Var}[w]}\right). \qquad (25)$$

Analogously, Eq. (11) shows that the standard deviation of G_n, $\mathrm{Std}[G_n]$, is a function of the mean input per branch only. We may therefore define $\mathrm{Std}[F] = \mathrm{Std}[F(u_n)]$ via $\mathrm{Std}[G_n(v_1, ..., v_n)] = \mathrm{Std}\{F[u_n(v_1, ..., v_n)]\}$ and compute it from Eqs. (11) and (23).

To investigate the convergence properties of the network, we consider its state $[v_1(t), ..., v_N(t)]$ at time t and the update of a neuron n by Eq. (18). Led by Fig. 2, we neglect the fluctuations $\mathrm{Std}[F] \ll \bar{F}$ and replace G_n in Eq. (18) by its mean $\bar{F}(u_n)$. This approximation of the response of neuron n by the response of a typical neuron with identity n (as specified by the weights $w_{n,m}$) simplifies the analysis of the network dynamics. For reasonable parameter choices, the deviations are small and very rarely lead to erroneous updates of neuron states (Fig. 4). Furthermore, we may assume that the effective input function $\bar{F}(u_n)$ is strictly monotonic in u_n and thus invertible (Appendix E). This dynamics is then equivalent to the conventional Hopfield network dynamics [Eqs. (13)–(15); see Appendix E] with coupling matrix $w_{n,m}$ and an effective threshold ϑ. ϑ is determined by the intersection of the effective somatic input $\bar{F}$ and the neuronal threshold Θ (Fig. 3)

$$\vartheta = \bar{F}^{-1}(\Theta). \qquad (26)$$

In particular, for symmetric weights $w_{n,m} = w_{m,n}$, the dynamics has a Lyapunov function (Appendix E)

$$E_{NL}[v_1(t), ..., v_N(t)]$$
$$= -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}w_{n,m}v_n(t)v_m(t) + \sum_{n=1}^{N}\vartheta v_n(t). \qquad (27)$$

By construction, E_{NL} decreases in time and the system converges toward a dynamical fixed point. Thus, the supralinear dendrites effectively reduce the neuronal
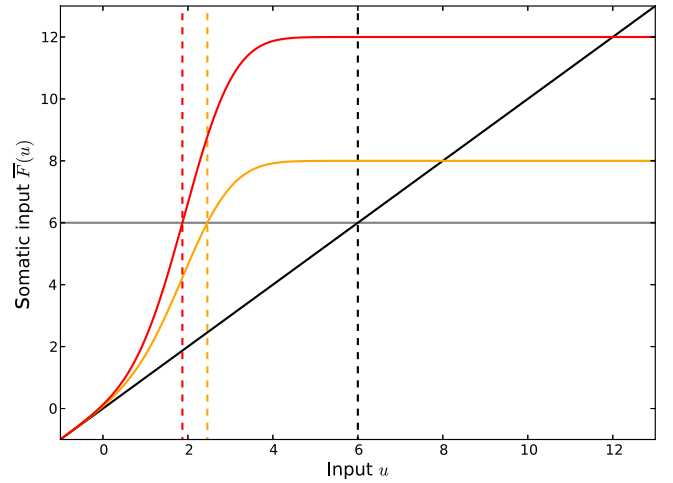

FIG. 3. Effective reduction of the neuronal threshold to $\vartheta \leq \Theta$ by nonlinear dendrites. The traditional, linear transfer function u (solid black line) crosses the threshold $\Theta = 6$ (solid gray line) at $\vartheta = 6$ (dashed black line). For $B = 2$, $PN^{-1}\mathrm{Var}[w] = 0.8$, and $\theta = 1$, dendritic nonlinearities $\bar{F}(u)$ [Eq. (23)] with strengths $D = 4$ (solid orange line) and $D = 6$ (solid red line) lead to effective thresholds $\vartheta \approx 2.5$ (dashed orange line) and $\vartheta \approx 1.9$ (dashed red line), respectively.

threshold to $\vartheta \leq \Theta$ and leave the convergence properties of the system unchanged.

To study the convergence of the extended Hopfield model, we generate a network with Hebbian synaptic weights $w_{n,m}$ according to Eq. (16) by drawing random patterns $\xi_n^p = \pm 1$ with equal probabilities. Then, $w_{n,b,m}$ are drawn from a Gaussian distribution with mean $E[w_{n,b,m}] = B^{-1}w_{n,m}$ and variance $\mathrm{Var}[w_{n,b,m}] = w_{n,m}^2 B^{-2}\mathrm{Var}[w]$ as explained above. We note that generally $\sum_{b=1}^{B}w_{n,b,m} =: w'_{n,m} \neq w'_{m,n}$, so that the neuronal connectivity in the presence of dendrites is not symmetric like in the classical Hopfield model. The energy function in Eq. (27) is derived for symmetric weights $w'_{n,m} = w_{n,m} = w_{m,n} = w'_{m,n}$, which may be seen as an approximation valid at least for slightly asymmetric couplings. Figure 4 shows that this approximate energy function correctly reflects the convergence of the network. Also, stronger deviations from the symmetric scenario (quantified by $\mathrm{Var}[w]$) leave the findings largely unchanged (Appendix F).

The results of numerical simulations displayed in Fig. 4 illustrate the convergence properties of our model. If the threshold reduction by dendrites is small, subthreshold inputs to a neuron also remain subthreshold in the presence of nonlinear dendrites and the network converges to the same state as for linear branches (Fig. 4, thick gray line). However, if the effective threshold reduction is stronger, inputs that are subthreshold may become superthreshold due to the dendritic nonlinearity and the same initial conditions tend to converge toward different attractors (Fig. 4, red lines). Since deviations of the somatic input
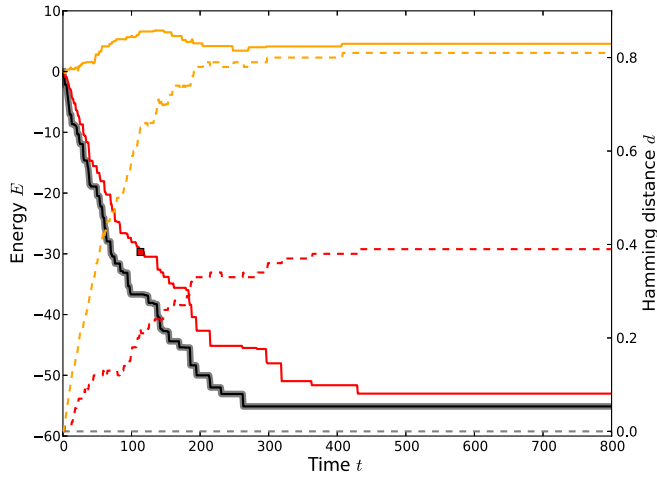
FIG. 4. Network convergence with effective threshold reduction by nonlinear dendrites. Simulation of a network of $N = 100$ binary neurons with $\Theta = 0.4$, $B = 2$, $P = 8$, and $\text{Var}[w] = 0.1$. Simulations are performed with identical initial states, topology, and identical realization of the (random) order of updates. The energy function [Eq. (15)] for linear input summation (solid black line) decreases, and the network converges toward a fixed point. Including weakly nonlinear branches (thick gray line, $\theta = 0.6$ and $D = 1$) does not alter the dynamics or the energy of the modified network [which is now given by Eq. (27)]. The unaltered dynamics is confirmed by the vanishing Hamming distance $d = \frac{1}{2N}\sum_{n=1}^{N}|v_n - v'_n|$ (dashed gray line) between the two systems. For stronger dendritic spikes (solid red line, $\theta = 0.1$ and $D = 2$), deviations of the somatic input from its ensemble average $\bar{F}$ and slightly asymmetric couplings lead to occasional increases in the energy (red square), but network convergence is preserved (checked for 1000 runs). The dynamics converges toward an attractor that is different from that of the linear network, as shown by the Hamming distance (dashed red line) between the systems. When inhibitory and excitatory inputs are processed separately by the soma and the dendrites, respectively, no energy function is known, and we exemplarily choose the energy function given by Eq. (27) with $\vartheta = \Theta$. The energy is then nonmonotonic (solid orange line, $\theta = 0.1$ and $D = 2$), but the system nonetheless reaches a stationary state (checked in simulations up to $t = 2000$) that is different from the one of the linear system; cf. the Hamming distance (dashed orange line).

from its ensemble average $\bar{F}$ may violate our approximation and the network is slightly asymmetric, the energy function can occasionally increase (Fig. 4, red square). However, for the considered parameters, these events occur rarely and do not affect the long-term convergence of the system (checked for 1000 runs, not shown).

If inhibitory synapses project directly onto the soma instead of being mingled with excitatory synapses on the dendrites, the effective somatic input $\bar{F}$ is given by Eq. (12). Since the dendritic saturation by excitatory input may be exceeded by linear inhibition, $\bar{F}$ is nonmonotonic and no Lyapunov function is apparent. The energy of the system as given by Eq. (27) with, e.g., $\vartheta = \Theta$ does not decrease monotonically in time (Fig. 4, orange lines). However,

numerical simulations suggest that the network reaches a stable fixed point nevertheless, as exemplarily shown in Fig. 4. Such network convergence despite nonmonotonic transfer functions is known from other systems [41,42,55]. These studies do not split excitation and inhibition but choose transfer functions nonmonotonic in the total, linear input u_n.

E. Capacity of stochastic Hopfield networks with nonadditive dendritic input processing

We now assess the extent to which a network of binary neurons with nonlinear dendrites is capable of storing and retrieving specific patterns. Since biological neurons are noisy, i.e., their input-output relation is not fully reliable (e.g., Ref. [56]), we generalize the above deterministic dynamics to allow for stochasticity. For the analysis of the storage capacity of the extended Hopfield network, we exploit the analogy between spin glasses and neural networks and employ statistical physics methods [13,26].

As a generalization of the deterministic update rule [Eq. (18)], we use the common Glauber dynamics [13,40,57] with asynchronous updates according to which the state of a randomly chosen neuron n is set to $v_n = \pm 1$ with probability

$$p_n(v_n) = [1 + \exp(-2\beta[G_n - \Theta]v_n)]^{-1}. \qquad (28)$$

This update rule is equivalent to flipping the state of the respective neuron with probability $p_n(v_n \to -v_n) = [1 + \exp(2\beta[G_n - \Theta]v_n)]^{-1}$. Here, $T := \beta^{-1}$ is the pseudo-temperature and a measure for the noise in the system and G_n [Eq. (19)] is the input to the neuron. We recall that P is the number of desired patterns, and the fraction $\alpha = PN^{-1}$ is called the load parameter. We obtain a temporally averaged state $\langle v_n \rangle$ of units n in the ensemble-averaged, stochastic network in mean-field theory by replacing G_n by the ensemble average $\bar{F}(u_n)$ [Eq. (23)]. Further, we replace the fluctuating argument of $\bar{F}(u_n)$ by its average, which yields $\bar{F}(\langle u_n \rangle)$, such that in the stationary state,

$$\langle v_n \rangle = (+1)p_n(+1) + (-1)p_n(-1)$$
$$= \tanh[\beta\bar{F}(\langle u_n \rangle) - \beta\Theta]$$
$$= \tanh\left[\beta\bar{F}\left(N^{-1}\sum_{m=1}^{N}w_{n,m}\langle v_m \rangle\right) - \beta\Theta\right]. \quad (29)$$

The overlap between pattern p and state $\langle v_n \rangle$ is defined by $m^p := N^{-1}\sum_{n=1}^{N}\xi_n^p\langle v_n \rangle$. Without loss of generality, we study the retrieval of pattern $p = 1$, so that $m := m^1$ estimates the quality of retrieval. m is given as an implicit solution to a set of coupled integral equations [Appendix G, Eqs. (G3), (G6), (G8), and (G9)]. In particular, we consider two limits $\alpha \approx 0$ and $T = 0$.

First, we study a finite number of patterns P so that in the thermodynamic limit of large N, we have $\alpha \approx 0$ (Appendix H). The overlap m is given by the 0 s of Δm

$$\Delta m := \frac{1}{2} \tanh\{\beta[1 - P_{NL}(m)]m$$
$$+ \beta[\text{BDP}_{NL}(m) - \text{BC}_{NL}(m) - \Theta]\}$$
$$+ \frac{1}{2} \tanh\{\beta[1 - P_{NL}(-m)]m$$
$$- \beta[\text{BDP}_{NL}(-m) - \text{BC}_{NL}(-m) - \Theta]\} - m, \quad (30)$$

where the functions P_{NL} and C_{NL} are given by Eqs. (24) and (25). The solutions of the transcendental equation $\Delta m = 0$ are obtained numerically and compared to simulation results of the Hopfield network with nonlinear dendrites [Fig. 5(a)].

The dendritic nonlinearities have a strong impact on the overlap curve. They change its shape and increase the critical temperature T_c that marks the transition between functioning and nonfunctioning associative memory. They provide a discontinuous, first-order phase transition with a nonzero critical overlap $m_c := m(T_c)$. For the same parameters, the conventional Hopfield model displays a
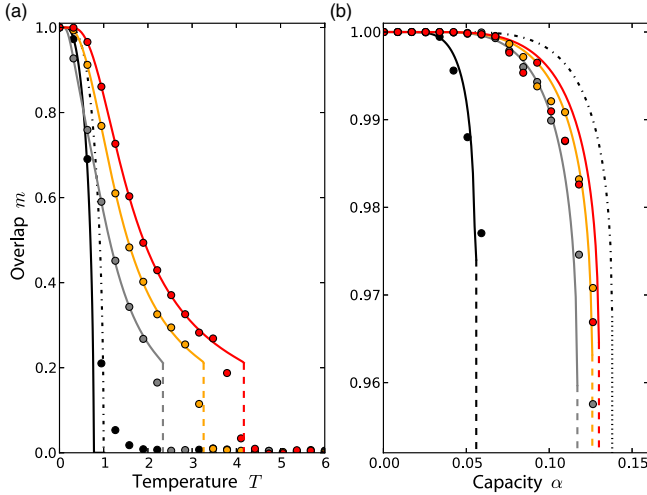
continuous, second-order phase transition. These findings may be understood by graphically solving Eq. (30): For the considered, not too large Θ ($\Theta < 0.448$), linear input processing leads to a concave $\Delta m(m)$ for high temperatures, so that the overlap continuously goes to zero when T approaches the critical temperature T_c [Fig. 6(a)]. In the presence of nonlinear dendrites, we need to take into account that $PN^{-1}\text{Var}[w] \propto \alpha \gtrsim 0$ [Eqs. (23)–(25)] is small so that the dendritic nonlinearity $\bar{F}$ sharply rises to its maximal (saturation) value at m with $B\theta - m \approx 0$ due to its dependence on $P_{NL}(m)$ [cf. Eq. (24) with $u_n = m$; $C_{NL}(m) \propto \alpha$ is small]. The second tanh is approximately constant there [because $P_{NL}(-m)$ and $C_{NL}(-m)$ are small for small $PN^{-1}\text{Var}[w]$] and may be neglected. The sharp rise in $\bar{F}$ thus induces a convex turn in the right-hand side of Eq. (30) and results in a stable and an unstable fixed point of m [Fig. 6(b), red lines]. With growing temperature T, the two fixed points vanish in a saddle-node bifurcation at nonzero $m_c \approx B\theta$ and the system undergoes a discontinuous phase transition of first order. For $\alpha = 0$, the increase in
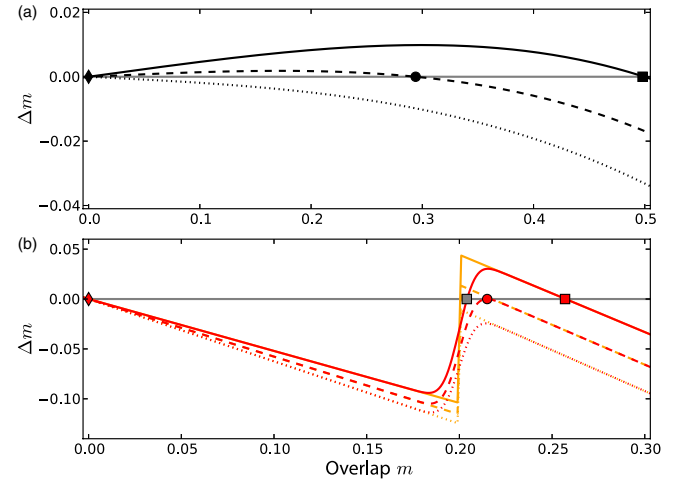

FIG. 5. Nonadditive dendritic coupling increases the overlap m of the network state with a retrieval pattern. Simulation results (circles) and analytical results (solid lines) for linear (black lines and circles) and nonlinear summations with dendritic spike strengths $D \in \{0.4, 0.6, 0.8\}$ (gray, orange, and red lines and circles). The remaining parameters are $N = 4000$, $B = 2$, $\theta = 0.1$, $\Theta = 0.4$, and $\text{Var}[w] = 0.1$. Results for linear input summation with $\Theta = 0$ are included for comparison (dash-dotted black line). The setup of the networks is the same as described in the discussion of Fig. 4. (a) The overlap m versus the temperature T for a small load $\alpha = N^{-1} \approx 0$. It decreases with increasing T and reaches zero at the critical temperature T_c above which retrieval fails. This phase transition is discontinuous for the nonlinear and continuous for the linear model of input processing. Dendritic nonlinearities increase T_c. (b) m versus α at zero temperature $T = 0$. It decreases with increasing load α and displays a discontinuous jump to zero at the critical storage capacity α_c. α_c increases with stronger dendritic nonlinearities up to a level at which the effective threshold ϑ vanishes and the linear scenario with $\Theta = 0$ is approached (dash-dotted black line).

FIG. 6. Graphical solutions to the transcendental equation for the overlap m and phase transitions. Parameters are the same as in Fig. 5, with $D = 0.4$. Equation (30) is solved graphically, with its solutions given by the zero crossings of Δm, i.e., the intersections with the solid gray line. (a) For the traditional Hopfield case with linear input summation, the stable solution for the overlap m continuously decreases (black markers, rectangle to circle to diamond) with increasing temperatures $T \in \{0.7, 0.75, 0.8\}$ (solid, dashed, and dotted black lines) and reaches $m_c = 0$ (diamond) at a critical value of $T_c \approx 0.8$. (b) For nonlinear input summation due to nonlinear dendrites and small $\alpha = N^{-1} \approx 0$, increasing temperatures $T \in \{2.0, 2.295, 2.6\}$ (solid, dashed, and dotted red lines) lead to decreasing overlaps m (red markers, from rectangle to circle to diamond) that jump discontinuously from $m_c \approx 0.22$ (circle) to zero at $T_c \approx 2.3$. The gray rectangle indicates an unstable solution. In the limit $\alpha = 0$ (orange lines) the jump in Δm is discontinuous, which slightly changes the critical temperature but leaves the system's critical behavior unchanged.

$\bar{F}$ is jumplike, so that no unstable fixed point appears, and the critical overlap is $m_c = B\theta$ [Fig. 6(b), orange lines].

The increased critical temperature T_c due to the dendrites implies an increased robustness of the network against thermal fluctuations and may be intuitively understood as follows: If the network state is close to a learned pattern, the input to the neurons is either strongly positive and thus further amplified by the dendritic nonlinearities or strongly negative and not affected by the dendrites. The overall strengthening of the stored patterns counteracts the influence of the temperature [Eq. (28)] and stabilizes the patterns against thermal fluctuations. Consequently, the nonlinear dendrites allow pattern retrieval in a temperature regime in which linear neurons fail.

Because for the conventional Hopfield model neuronal thresholds Θ decrease the critical temperature [cf. Fig. 5(a), black lines], we test if our results are a mere consequence of an effective threshold reduction $\vartheta \leq \Theta$ by the nonlinear dendrites [cf. Eq. (26)]. Repeating the above calculations and simulations with $\Theta = \vartheta = 0$, we find that results as described above hold also for vanishing neuronal thresholds (Appendix H).

Second, we consider the zero-temperature limit $T = 0$, in which thermal fluctuations cease and the binary neurons are deterministic threshold units (Appendix I). The overlap m is determined by

$$m = \frac{1}{2}\text{erf}\left(\frac{m - \vartheta}{\sqrt{2\alpha r}}\right) + \frac{1}{2}\text{erf}\left(\frac{m + \vartheta}{\sqrt{2\alpha r}}\right),$$

$$\sqrt{r} = 1 + \sqrt{\frac{1}{2\pi\alpha}}\exp\left(-\frac{(m - \vartheta)^2}{2\alpha r}\right)$$

$$+ \sqrt{\frac{1}{2\pi\alpha}}\exp\left(-\frac{(m + \vartheta)^2}{2\alpha r}\right), \qquad (31)$$

where ϑ is the effective threshold [Eq. (26)]. We solve these coupled equations numerically and compare them to the simulation data of a Hopfield network with nonlinear dendrites [Fig. 5(b)]. Equations (31) are equivalent to the order-parameter equations of the conventional Hopfield model with threshold ϑ. For $\Theta > 0$, we may therefore conclude that the dendritic branches reduce the neuronal threshold to $\vartheta \leq \Theta$ and thereby improve the critical storage capacity α_c of the network. Analogous to T_c, α_c denotes the critical load above which retrieval of patterns fails.

We note that the improved performance is a direct effect of the dendritic nonlinearity as demonstrated in Fig. 5, where the connectivity of neurons is the same for networks with linear and nonlinear dendrites, while there is a clear increase in the critical temperature T_c and load α_c in the latter case.

These findings may be understood by considering a neuronal threshold $\Theta \neq 0$ that generally introduces an asymmetry between the two states $v_n = \pm 1$ of a unit n. Since we assume the storage of random patterns $\xi_n^p = \pm 1$

with equal probabilities, the nonzero $\Theta \neq 0$ impedes the retrieval of learned patterns. For a positive threshold $\Theta > 0$, the threshold reduction by the dendritic nonlinearities attenuates the asymmetry of the network and improves the retrieval of random patterns. In the zero-temperature limit and for strong nonlinearities, our model becomes equivalent to the standard Hopfield model without threshold $\Theta = 0$ [Fig. 5(b), dash-dotted black line].

We note that the agreement of analytical and numerical results as shown in Fig. 5 is even better if the couplings $w'_{n,m} = \sum_{b=1}^{B} w_{n,b,m}$ are normalized such that the $w'_{n,m}$ exactly equal the symmetric Hebbian weights $w_{n,m}$. This improved agreement due to normalized weights is particularly pronounced for stronger dendritic spikes [Fig. 5(b), red lines] that emphasize the asymmetry. To check if the above results hold also for larger deviations from the assumption of symmetric couplings $w'_{n,m} = w'_{m,n}$, we repeat the simulations for larger $\text{Var}[w]$ [cf. Eq. (21) and Appendix F]. In the deterministic limit $T = 0$ with many patterns, we find that stronger asymmetries impede the quality of retrieval. For finite temperatures $T > 0$ and few patterns $\alpha \approx 0$, the impact of moderately asymmetric synaptic weights is negligible.

Complementing our analytical study of the limiting cases $\alpha \approx 0$ and $T = 0$, we compute the quality of retrieval in the α-T phase space numerically. We find that our associative-memory network with nonadditive dendrites enables memory functioning in a larger α-T region than the model with linear branches (cf. Appendix J).

F. Optimal number of dendritic branches for memory function

Finally, we investigate the impact of varying numbers B of dendritic branches on the performance of the extended stochastic Hopfield network. As shown above, nonadditive dendritic input processing leads to an increased storage capacity and more robust memory retrieval by amplifying strong input to the neuron. Figure 2(b) shows that when this input is fixed, the average somatic input $E[F] \cong \bar{F}$ is maximal for an intermediate number of dendrites. Since strong neuronal input may stabilize the memorized patterns we expect optimal memory performance for intermediate branch numbers.

We thus study the Hopfield network for varying B. In analogy to Fig. 5(a), we compute the overlaps m for the limiting case $\alpha \approx 0$ [Fig. 7(a)]. The critical temperature T_c [Fig. 7(b)] displays a maximum at an intermediate number of branches, here $B_{T,\text{opt}} = 30$. For larger θ, the optimal branch number is smaller (not shown). We can understand the maximum in T_c by considering the $\Delta m(m)$ curves [Fig. 7(c)]. They display maxima at $m_c \approx B\theta$ (see discussion of Fig. 6) with absolute heights determined by m_c and $\bar{F}(m_c)$ where the latter is maximal for intermediate branch numbers. In combination, they yield the maxima of $\Delta m(m \neq 0)$ that are highest for intermediate numbers $B_{T,\text{opt}}$
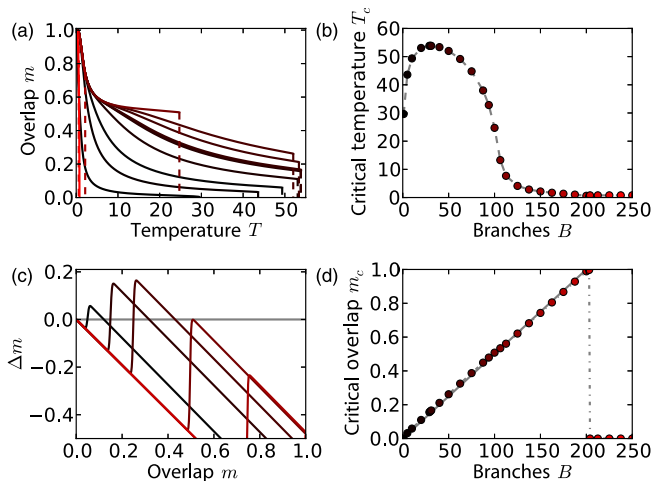
FIG. 7. Memory performance in dependence of the number B of dendritic branches. Parameters are the same as in Fig. 5, with $D = 0.6$ and $\theta = 0.005$. Branch numbers B are color-coded from black to red; cf. (b) and (d). (a),(c) Results for a few selected values of B. (a) The overlap curves $m(T)$ with $\alpha \approx 0$ are computed for varying numbers B of branches. (b) The critical temperature T_c depends nonmonotonically on B and is maximal for $B_{T,\text{opt}} = 30$ (dashed line and circles). For a fixed temperature $T = 25$, (c) shows Δm whose 0 s provide the solutions for m [Eq. (30)] and whose maxima display a nonmonotonical dependency on B. (d) The critical overlap m_c (dotted line and circles) grows approximately linearly with B as indicated by the linear function $m_c = \theta B$ (solid line) and jumps to zero at $m_c \approx 1$.

of branches. Upon increasing temperature, the corresponding curves are thus the last to fall entirely below zero at their $T_c(B)$ so that T_c is highest for such branch numbers.

Another notable feature is the growing critical overlap m_c at the critical temperature T_c with increasing numbers of branches [Fig. 7(d)]. As shown in the discussion of Fig. 6, the critical overlap for small α is given by $m_c \approx B\theta$. The argument is correct for strong nonlinearities D and moderate θB but breaks down for $\theta B \approx m_c > 1$ since the overlap is naturally bounded by 1 and only the solution $m_c = 0$ remains, if a larger overlap would be needed to reach the upturn point of Δm [cf. Fig. 7(c)]. In particular, for large B, the behavior of the linear scenario with $m_c = 0$ is reobtained.

Thus, the performance of the memory network depends nontrivially on the number of dendritic branches B. Additionally, depending on the purpose of the memory network, its robustness against noise (specified by T_c) may be balanced against the quality of retrieval (for which m_c gives a worst-case measure) [cf. Fig. 7(a)].

III. DISCUSSION: NONLINEAR DENDRITES IMPROVE PATTERN RETRIEVAL

Nonadditive processing of synaptic input is an important feature of biological neurons and may have severe consequences for neural processing in single neurons and

networks. In this work, we first studied the influence of dendritic spikes in a neuron with a variable number of dendritic branches and sufficiently synchronous spiking inputs of variable strength, independently of a specific neuron model. We derived an approximation for the somatic input in the presence of nonlinear dendrites [Eqs. (7)–(10) and Appendix C]. This approximation allows the analytical investigation of dendritic summation phenomena in networks of arbitrary connectivity. Second, we extended the well-known Hopfield model to include neurons with branches that process inputs nonadditively. Employing the results from the first part, we constructed networks that are, at each neuron, ensemble averaged over the nonlinear dendrites and their inputs, such that the overall connectivity and thus the neural identities in the networks are preserved. These networks could be analyzed analytically with statistical physics methods. We used them to approximate the full dynamics of networks with nonlinear dendrites. We find that, for a deterministic Hopfield network, the dendritic nonlinearities reduce the neuronal thresholds [Eq. (26)], and the network still converges to a dynamical fixed point [Eq. (27)]. Separate processing of inhibition and excitation [Eq. (12)] can break the monotonic decrease of common energy functions. A mean-field analysis for a stochastic Hopfield network revealed an improved memory-storage capacity and a greater robustness against thermal fluctuations due to nonadditive dendritic input processing [Eqs. (30) and (31) and Appendix J]. An intermediate number of dendritic branches was shown to optimally support the memory functionality of the associative network.

Our findings help to advance the understanding of the role of nonlinear dendrites in three respects: (i) Earlier works studied the ability of arborized neurons to discriminate patterns [22,23]. They focused on a combinatorial approach of counting the numbers of different input-output functions of single neurons with multiple dendrites. In contrast, our study assumes a dynamical perspective. We derived an expression for the approximate somatic input that may be readily used to investigate the dynamics of networks of neurons with nonlinear dendrites for arbitrary connectivity and synaptic weights. (ii) We applied our results and studied the capability of networks with non-additive dendrites to serve as memory devices. Our work shows that nonlinear dendrites can increase the capacity and the robustness of memory retrieval against thermal fluctuations in recurrent, dynamic associative-memory networks. (iii) Finally, our theoretical results suggest that there might be an intermediate number of dendritic branches that is optimal for network functionality [22,58]. Such a gain in performance may have severe implications for biological neural circuits featuring nonadditive dendrites, e.g., in the hippocampus. Since nonadditive dendritic integration may take place in sliding windowlike segments of the dendritic tree, a precise number of independent dendritic

compartments is unlikely to be found [5]. Biological studies suggest around 50 to 100 independent sites capable of generating dendritic spikes per neuron, depending on the neuron type and function [19,59]. Our model shows optimal memory performance for such numbers of branches, depending on the dendritic parameters (see discussion of Fig. 7).

For biological neural networks, we suggest that nonlinear dendrites may serve to stabilize memory recall against noisy network background activity. In such networks, memories might be stored in so-called Hebbian cell assemblies with higher internal connectivities or connection strengths that display elevated firing rates when presented with a specific input pattern [60,61]. Similar to our Hopfield model, matching patterns of activity provide a larger input to the other cells of the assembly, which is amplified by the dendritic nonlinearities. In our model, we restricted ourselves to Hebbian learning of coupling strengths, and noise and patterns were equally nonlinearly enhanced. Biological neural networks can change their wiring as well as the dendritic nonlinearities in an activity-dependent manner [5,20,62]. Further, some kinds of dendritic spikes amplify only temporally highly coordinated inputs [3]. Both features may contribute to a selective nonlinear enhancement of pattern activity and may increase the stabilizing effects of nonadditive dendrites in memory networks.

On a theoretical level of statistical physics, our work may be continued in several directions: Preliminary calculations suggest that for certain parameters, new phenomena arise, such as improved memory retrieval for moderate noise levels. (This improved performance in presence of noise is reminiscent of stochastic resonance [63,64].) Further, previous studies on nonmonotonic transfer functions found beneficial effects for memory performance in artificial and biological neural systems [25,41,42,65]. The novel kind of nonmonotonicity that is due to dendritic reception of excitatory input and somatic reception of inhibitory input [cf. Eq. (12)] should be further explored and linked to these findings. Finally, many studies suggest that neural plasticity exploits dendritic spikes and dendritic compartmentalization [5,20,62]. Non-Hebbian learning rules that are tailored to utilize dendritic spikes and branches were shown to increase the memory capabilities of single neurons or ensembles of such neurons [15,21–23,66]. Therefore, such dendrite-based learning is expected to also boost the performance of associative-memory networks and is a particularly important target of future studies. Our statistical treatment [Eqs. (7)–(10)] and the numerical approach can be directly applied to address these issues. Gaining insight into these matters will help to better understand and utilize the full power of dendritic computation.

ACKNOWLEDGMENTS

APPENDIX A: APPROXIMATE MEAN AND VARIANCE OF THE EFFECTIVE SOMATIC INPUT AND THE NUMBER OF NONLINEAR BRANCHES

We compute the first moments of the somatic input F [Eq. (6)] using a Gaussian approximation of the dendritic input distribution $P(u_1, ..., u_B)$ and assuming statistically identical branches, i.e., $p_b = p_0$, $b \in \{1, ..., B\}$. Means $E[u_b]$, variances $\text{Var}[u_b]$, and covariances $\text{Cov}[u_b, u_c]$ of $P(u_1, ..., u_B)$ are given by Eqs. (3)–(5). We start with

$$E[F] = \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)] P(u_1, ..., u_B) \tag{A1}$$

and pick $u := u_1$ without loss of generality to obtain

$$E[F] = B \int_{-\infty}^{\infty} du f(u) \int_{-\infty}^{\infty} du_2 \cdots \int_{-\infty}^{\infty} du_B P(u, u_2, ..., u_B) = B \int_{-\infty}^{\infty} du f(u) P(u), \tag{A2}$$

where $P(u)$ is the marginal distribution of $P(u, u_2, ..., u_B)$ and thus Gaussian with mean $E[u]$ and variance $\text{Var}[u]$. Using the definition of the dendritic transfer function f [Eq. (1)], we split

$$E[F] = B \left(\int_{\theta}^{\infty} du D P(u) + \int_{-\infty}^{\theta} du u P(u) \right) = B P_{\text{NL}} D + B(1 - P_{\text{NL}}) E[u] - B C_{\text{NL}}, \tag{A3}$$

where we use partial integration in the second line and the definitions

$$P_{\text{NL}} := \frac{1}{2} \text{erfc} \left(\frac{\theta - E[u]}{\sqrt{2 \text{Var}[u]}} \right), \tag{A4}$$

$$C_{\mathrm{NL}} := \sqrt{\frac{\mathrm{Var}[u]}{2\pi}} \exp\left(-\frac{(\theta - E[u])^2}{2\mathrm{Var}[u]}\right). \tag{A5}$$

The second moment is computed similarly:

$$
\begin{aligned}
E[F^2] &= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)]^2 P(u_1, ..., u_B) \\
&= B\int_{-\infty}^{\infty} du f^2(u)P(u) + (B^2 - B)\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv f(u)f(v)P(u, v) \\
&= B\left(\int_{\theta}^{\infty} du D^2 P(u) + \int_{-\infty}^{\theta} du u^2 P(u)\right) + (B^2 - B)\left(\int_{\theta}^{\infty} du \int_{\theta}^{\infty} dv D^2 P(u, v)\right. \\
&\quad \left. + 2\int_{\theta}^{\infty} du \int_{-\infty}^{\theta} dv DuP(u, v) + \int_{-\infty}^{\theta} du \int_{-\infty}^{\theta} dv uvP(u, v)\right) \\
&= BP_{\mathrm{NL}}D^2 + B(1 - P_{\mathrm{NL}})(E^2[u] + \mathrm{Var}[u]) - BC_{\mathrm{NL}}(E[u] + \theta) + (B^2 - B)I_F. \tag{A6}
\end{aligned}
$$

Here, $v := u_2$ and the marginal distribution $P(u, v)$ is again Gaussian with means $(E[u], E[v])$, variances $(\mathrm{Var}[u], \mathrm{Var}[v])$, and covariance $\mathrm{Cov}[u, v]$. For binomially distributed numbers of active synapses per branch, $P(u, v) = P(u)P(v)$, and therefore

$$
\begin{aligned}
I_F &= (P_{\mathrm{NL}}D)^2 + 2(P_{\mathrm{NL}}D)\{(1 - P_{\mathrm{NL}})E[u] - C_{\mathrm{NL}}\} + \{(1 - P_{\mathrm{NL}})E[u] - C_{\mathrm{NL}}\}^2 \\
&= \{P_{\mathrm{NL}}D + (1 - P_{\mathrm{NL}})E[u] - C_{\mathrm{NL}}\}^2 \\
&= (B^{-1}E[F])^2. \tag{A7}
\end{aligned}
$$

For a multinomial distribution of active synapses across branches, the double integral I_F needs to be computed numerically (see Fig. 2). These results are discussed in the main text and in the caption of Fig. 2. The calculations may be easily extended to cover nonuniform branch probabilities p_b, dendritic thresholds θ_b, and strengths D_b.

Similarly to $E[F]$, we compute the expected number $E[k]$ of branches k in the nonlinear regime. The dendritic transfer function $f(u)$ in Eq. (A1) is replaced by a step function $\mathrm{step}(u - \theta)$ to count the number of branches above threshold θ. Here, we define $\mathrm{step}(x) = 0$ if $x < 0$ and $\mathrm{step}(x) = 1$ otherwise. Then,

$$
\begin{aligned}
E[k] &= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [\mathrm{step}(u_1 - \theta) + \cdots \\
&\quad + \mathrm{step}(u_B - \theta)]P(u_1, ..., u_B) \\
&= B\int_{\theta}^{\infty} du P(u) \\
&= BP_{\mathrm{NL}}. \tag{A8}
\end{aligned}
$$

Since the step function satisfies $\mathrm{step}^2(x) = \mathrm{step}(x)$, we derive the second moment of the distribution of the number k of nonlinear branches via

$$
\begin{aligned}
E[k^2] &= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [\mathrm{step}(u_1 - \theta) + \cdots \\
&\quad + \mathrm{step}(u_B - \theta)]^2 P(u_1, ..., u_B) \\
&= B\int_{\theta}^{\infty} du P(u) + (B^2 - B)\int_{\theta}^{\infty} du \int_{\theta}^{\infty} dv P(u, v) \\
&= BP_{\mathrm{NL}} + (B^2 - B)I_k. \tag{A9}
\end{aligned}
$$

The binomial distribution of synapses among branches provides $P(u, v) = P(u)P(v)$, and thus

$$I_k = P_{\mathrm{NL}}^2 = (B^{-1}E[k])^2, \tag{A10}$$

while I_k needs to be computed numerically in the multinomial case.

APPENDIX B: EXACT MEAN AND VARIANCE OF THE EFFECTIVE SOMATIC INPUT AND THE NUMBER OF NONLINEAR BRANCHES

We now derive the mean somatic input $E[F]$ and its variance $\mathrm{Var}[F]$ for the exact distribution $P(u_1, ..., u_B)$ of input where $u_b = \sum_{i=1}^{x_b} w_i$ [cf. Eq. (2)] with Gaussian $P(w)$. We decompose $P(u_1, ..., u_B)$ into

$$
\begin{aligned}
P(u_1, ..., u_B) &= \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(u_1, ..., u_B | x_1, ..., x_B) \\
&\quad \times P(x_1, ..., x_B), \tag{B1}
\end{aligned}
$$

where

$$P(u_1, ..., u_B | x_1, ..., x_B) = P(u_1 | x_1) \cdots P(u_B | x_B) \tag{B2}$$

since u_b depends only on x_b. The $P(u_b | x_b)$ are Gaussian distributed with means $x_b E[w]$ and variances $x_b \mathrm{Var}[w]$, and $P(u_1, ..., u_B)$ is a weighted superposition of Gaussian distributions. For non-Gaussian $P(w)$ and large numbers of small inputs, we may employ a central limit theorem to establish the Gaussianity of the $P(u_b | x_b)$. $E[F]$ is computed as

$$E[F] = \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)] P(u_1, ..., u_B)$$

$$= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B \sum_{b=1}^{B} f(u_b) \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) P(u_1, ..., u_B | x_1, ..., x_B)$$

$$= \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B \sum_{b=1}^{B} f(u_b) \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) P(u_1 | x_1) \cdots P(u_B | x_B)$$

$$= \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) \sum_{b=1}^{B} \int_{-\infty}^{\infty} du_b f(u_b) P(u_b | x_b) \prod_{c \neq b} \int_{-\infty}^{\infty} du_c P(u_c | x_c)$$

$$= \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) \sum_{b=1}^{B} \int_{-\infty}^{\infty} du_b f(u_b) P(u_b | x_b) \cdot 1$$

$$= \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) \sum_{b=1}^{B} \{ P_{\mathrm{NL},u_b|x_b} D + (1 - P_{\mathrm{NL},u_b|x_b}) E[u_b | x_b] - C_{\mathrm{NL},u_b|x_b} \}$$

$$= \sum_{x=1}^{S} P(x) B \{ P_{\mathrm{NL},u|x} D + (1 - P_{\mathrm{NL},u|x}) E[u|x] - C_{\mathrm{NL},u|x} \}, \tag{B3}$$

where we assume identical statistics for all branches only in the last line and $P(x)$ is the marginal distribution of $P(x_1, ..., x_B)$. Similarly to Appendix A, we define

$$P_{\mathrm{NL},u|x} := \frac{1}{2} \mathrm{erfc} \left(\frac{\theta - E[u|x]}{\sqrt{2\mathrm{Var}[u|x]}} \right), \tag{B4}$$

$$C_{\mathrm{NL},u|x} := \sqrt{\frac{\mathrm{Var}[u|x]}{2\pi}} \exp \left(-\frac{(\theta - E[u|x])^2}{2\mathrm{Var}[u|x]} \right), \tag{B5}$$

with

$$E[u|x] = x E[w], \tag{B6}$$

$$\mathrm{Var}[u|x] = x \mathrm{Var}[w]. \tag{B7}$$

The second moment of F is given by

$$E[F^2] = \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [f(u_1) + \cdots + f(u_B)]^2 P(u_1, ..., u_B)$$

$$= \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) \sum_{b=1}^{B} \int_{-\infty}^{\infty} du_b f^2(u_b) P(u_b | x_b)$$

$$+ \sum_{x_1=1}^{S} \cdots \sum_{x_B=1}^{S} P(x_1, ..., x_B) \sum_{c=1}^{B} \sum_{b \neq c}^{B} \int_{-\infty}^{\infty} du_b \int_{-\infty}^{\infty} du_c f(u_b) f(u_c) P(u_b | x_b) P(u_c | x_c)$$

$$= \sum_{x=1}^{S} P(x) B [P_{\mathrm{NL},u|x} D^2 + (1 - P_{\mathrm{NL},u|x})(E^2[u|x] + \mathrm{Var}[u|x]) - C_{\mathrm{NL},u|x}(E[u|x] + \theta)]$$

$$+ \sum_{x=1}^{S} \sum_{y=1}^{S} P(x, y)(B^2 - B) \{ P_{\mathrm{NL},u|x} D + (1 - P_{\mathrm{NL},u|x}) E[u|x] - C_{\mathrm{NL},u|x} \}$$

$$\times \{ P_{\mathrm{NL},u|y} D + (1 - P_{\mathrm{NL},u|y}) E[u|y] - C_{\mathrm{NL},u|y} \}, \tag{B8}$$

where the assumption of identical branch statistics is used in the last step and $P(x, y)$ is the marginal distribution of $P(x_1, ..., x_B)$.

Analogously, we may compute the exact average number $E[k]$ of nonlinear branches. Similarly to Appendix A, we use a step function to count the number k of branches in the nonlinear regime

$$E[k] = \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [\text{step}(u_1 - \theta) + \cdots$$
$$+ \text{step}(u_B - \theta)]P(u_1, ..., u_B)$$
$$= \sum_{x=1}^{S} P(x)BP_{\text{NL},u|x}. \tag{B9}$$

The second moment yields

$$E[k^2] = \int_{-\infty}^{\infty} du_1 \cdots \int_{-\infty}^{\infty} du_B [\text{step}(u_1 - \theta) + \cdots$$
$$+ \text{step}(u_B - \theta)]^2 P(u_1, ..., u_B)$$
$$= \sum_{x=1}^{S} P(x)BP_{\text{NL},u|x}$$
$$+ \sum_{x=1}^{S} \sum_{y=1}^{S} P(x, y)(B^2 - B)P_{\text{NL},u|x}P_{\text{NL},u|y}. \tag{B10}$$

The exact expressions for $E[F]$ and $\text{Std}[F] = \sqrt{E[F^2] - E^2[F]}$ as well as $E[k]$ and $\text{Std}[k] = \sqrt{E[k^2] - E^2[k]}$ are compared to simulation results and the Gaussian approximation of F (see Appendix A) in Fig. 2.

APPENDIX C: EFFECTIVE SOMATIC INPUT APPROXIMATION FOR NEURONS WITH MULTIPLE LAYERS OF DENDRITIC BRANCHES

The approximation for the somatic input F [Eqs. (7)–(10)] may be readily employed to compute the somatic input for more complex dendritic arbors and multiple steps of non-additive dendritic processing (Fig. 8). For this approximation, whenever necessary, we assume that the distribution of inputs may be approximated by the maximum entropy distribution for given mean and variance, i.e., by a normal distribution [cf. discussion of Eqs. (3)–(5)]. For simplicity of presentation, we assume that the number of sub-branches B_l at a level $l \in \{1, ..., L\}$ is identical for all branches $b_l \in \{1, ..., B_l\}$ at the level, that inputs arrive only at the terminal branches and sufficiently synchronously, and that the distribution of synapses is identical across the terminal branches. To compute the neuronal input for such a neuron, we may start from the soma and recursively work toward the terminal branches: The first moments of the effective somatic input are given by
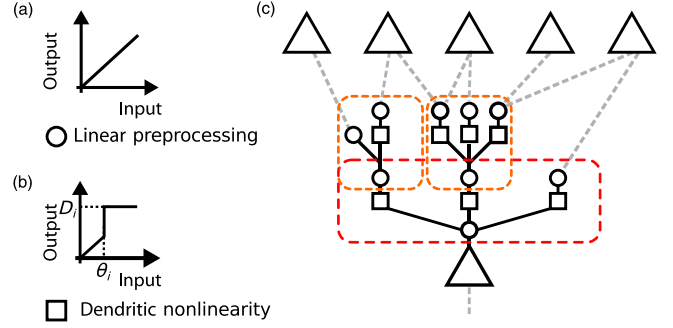

FIG. 8. Neuron model with multiple layers of dendritic branches and nonadditive processing. Similarly to Fig. 1, different steps of (a) linear and (b) nonadditive dendritic input processing are represented by circles and squares, respectively. A neuron with two levels of nonadditive dendritic branches is displayed in (c). As is exemplarily shown, the numbers of sub-branches may vary (at each level and between branches), and there may be synaptic input to linear and nonlinear dendrites at all levels.

$$E[F] = B_1 D P_{\text{NL}} + B_1(1 - P_{\text{NL},1})E[u_1] - B_1 C_{\text{NL},1}, \tag{C1}$$

$$P_{\text{NL},1} = P_{\text{NL}}(E[u_1], \text{Var}[u_1]), \tag{C2}$$

$$C_{\text{NL},1} = C_{\text{NL}}(E[u_1], \text{Var}[u_1]), \tag{C3}$$

as defined in Eqs. (7)–(9), and $\text{Var}[F]$ can be computed similarly; cf. Eq. (10). The appearing mean input per branch $E[u_1]$ and its variance $\text{Var}[u_1]$ are now given by an analogous approximation that captures the nonadditive processing of the preceding layer. Indeed, the mean input per branch $E[u_{l-1}]$ for any layer $l - 1$, $L \geq l > 1$, is given by

$$E[u_{l-1}] = B_l D P_{\text{NL}} + B_l(1 - P_{\text{NL},l})E[u_l] - B_l C_{\text{NL},l}, \tag{C4}$$

$$P_{\text{NL},l} = P_{\text{NL}}(E[u_l], \text{Var}[u_l]), \tag{C5}$$

$$C_{\text{NL},l} = C_{\text{NL}}(E[u_l], \text{Var}[u_l]), \tag{C6}$$

and $\text{Var}[u_{l-1}]$ analogously; cf. Eq. (10). In this nomenclature, $E[u_0] = E[F]$ and $\text{Var}[u_0] = \text{Var}[F]$ capture the somatic input F, and $l = L$ indexes the layer of terminal branches that receives the synaptic input so that $E[u_L] = E[u]$ and $\text{Var}[u_L] = \text{Var}[u]$ [cf. Eqs. (3)–(5)]. We note that one can also introduce factors implementing branch-coupling strengths at this point.

Figure 9 shows good agreement of our analytical predictions with simulation results, both for the mean somatic input $E[F]$ and its standard deviation $\text{Std}[F]$. Dendritic thresholds and dendritic spike strengths in layer $l = 1$ are increased to compensate for large input strengths in the absence of branch-coupling factors. We note that, as for the single-layered neuron, the
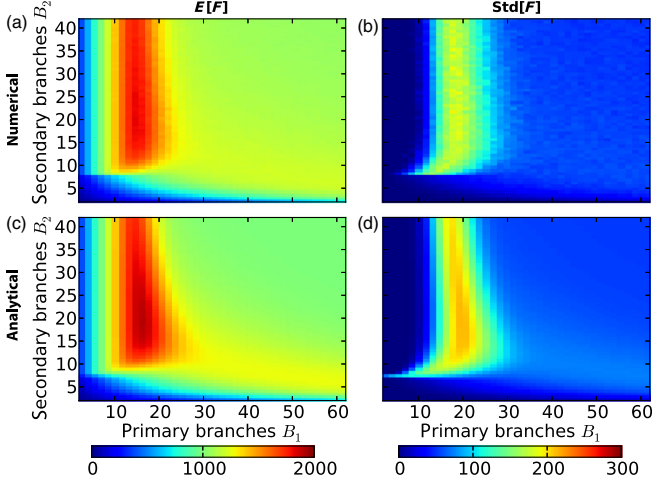
FIG. 9. Somatic input for a neuron model with two layers of nonadditive dendritic branches. Parameters are $\theta_1 = 100$, $D_1 = 200$, $\theta_2 = 5$, $D_2 = 10$, $E[w] = 1$, $\text{Var}[w] = 2$, and $S = 1000$; the synaptic input arrives at the terminal branches. Numerical results for (a) the mean somatic input and (b) its standard deviation for different numbers of branches on the first (B_1) and the second (B_2) levels are obtained from 400 realizations of multinomially distributed synapses. The corresponding analytical results from Eqs. (C3) and (C6) are highlighted in (c) and (d). Analytical and numerical results agree well. The input is largest for intermediate numbers of branches $B_{1,\text{opt}}$ and $B_{2,\text{opt}}$.

input is largest for intermediate numbers of branches $B_{1,\text{opt}}$ and $B_{2,\text{opt}}$ (cf. discussion of Fig. 2).

The derivation can be directly generalized to cover nonidentical branches, i.e., branches with different probabilities for the formation of synapses, different numbers of sub-branches (cf. Appendixes A and B), linear branches, and additional external inputs on intermediate-level branches. To incorporate linear branches, we may set the dendritic threshold to infinity; to incorporate external inputs to intermediate-level dendrites, we can add an additional, linear input branch to the considered dendrite, where the external inputs arrive. We may thus conclude that our approach covers arbitrary treelike dendritic structures.

APPENDIX D: MEAN AND VARIANCE OF THE INPUT PER BRANCH IN A NETWORK WITH DENDRITIC BRANCHES

We now derive the expected input to branch b of neuron n in the extended Hopfield model when the states of the neurons are fixed and the average is taken over the weight distribution. By construction, we have

$$
\begin{aligned}
E[u_{n,b}] &= E\left[\sum_{m=1}^{N} w_{n,b,m} v_m\right] \\
&= \sum_{m=1}^{N} E[w_{n,b,m}] v_m \\
&= \sum_{m=1}^{N} B^{-1} w_{n,m} v_m \\
&= B^{-1} u_n,
\end{aligned} \tag{D1}
$$

so that the mean input to the branch depends only on the field u_n of the classical Hopfield model. The choice $E[w_{n,b,m}] = B^{-1} w_{n,m}$ in line three with $w_{n,m}$ given by Eq. (16) is justified by assuming Hebbian learning. The variance of the input per branch is given by

$$
\begin{aligned}
\text{Var}[u_{n,b}] &= \text{Var}\left[\sum_{m=1}^{N} w_{n,b,m} v_m\right] \\
&= \sum_{m=1}^{N} v_m^2 \text{Var}[w_{n,b,m}] + \sum_{m \neq k}^{N} v_m v_k \text{Cov}[w_{n,b,m}, w_{n,b,k}] \\
&= \sum_{m=1}^{N} \text{Var}[w_{n,b,m}] + 0 \\
&= \sum_{m=1}^{N} w_{n,m}^2 B^{-2} \text{Var}[w] \\
&= \text{Var}[w] B^{-2} N E[w_{n,m}^2] \\
&= \text{Var}[w] B^{-2} N (\text{Var}[w_{n,m}] + E^2[w_{n,m}]),
\end{aligned} \tag{D2}
$$

where we use the independence of the $w_{n,b,m}$ in the third line. In the fourth line, we employ $\text{Var}[w_{n,b,m}] = w_{n,m}^2 B^{-2} \text{Var}[w]$ with a parameter $\text{Var}[w]$; cf. the discussion preceding Eqs. (20) and (21). For the Hopfield network with P random patterns to be stored [Eq. (16)], $E[w_{n,m}] = 0$ and $\text{Var}[w_{n,m}] = P N^{-2} = N^{-1} \alpha$ (with the load $\alpha = P N^{-1}$) since $w_{n,m}$ is a sum of P contributions $\pm N^{-1}$ with equal probabilities. Finally, in our extended Hopfield model, the correlation of input between different branches vanishes:

$$
\begin{aligned}
\text{Cov}[u_{n,b}, u_{n,c}] &= \text{Cov}\left[\sum_{m=1}^{N} u_{n,b,m} v_m, \sum_{k=1}^{N} u_{n,c,k} v_k\right] \\
&= \sum_{m=1}^{N} \sum_{k=1}^{N} v_m v_k \text{Cov}[u_{n,b,m}, u_{n,c,k}] \\
&= 0,
\end{aligned} \tag{D3}
$$

where $b \neq c$. In the third line, we use that for $m = k$ the $u_{n,b,m}$ are independently chosen from (the same) Gaussian distributions with means $B^{-1} w_{n,m}$ and variances $B^{-2} w_{n,m}^2 \text{Var}[w]$ and for $m \neq k$ they are independently

chosen from their respective (in general, different) distributions. Because the input across branches is uncorrelated, we may employ the results for the somatic input we have derived for binomially distributed active synapses [Eqs. (7) and (11)].

APPENDIX E: CONVERGENCE OF A HOPFIELD NETWORK WITH DENDRITIC NONLINEARITIES

Here, we show that the dynamics of the deterministic Hopfield network with nonlinear dendrites is equivalent to that of the classical Hopfield model with reduced threshold. In particular, network convergence is guaranteed. Assuming that $\bar{F}$ [Eq. (23)] is strictly monotonically increasing and therefore also invertible, we find that

$$v_n(t+1) = \mathrm{sgn}\{\bar{F}[u_n(t)] - \Theta\} = \mathrm{sgn}[u_n(t) - \vartheta], \quad (E1)$$

with the effective threshold

$$\vartheta = \bar{F}^{-1}(\Theta); \quad (E2)$$

see also Fig. 3. This update rule is equivalent to the conventional one with threshold ϑ. Hence, the energy function E_{NL} of the system is obtained from E_L [Eq. (15)] by substituting Θ by ϑ:

$$E_{\mathrm{NL}}[v_1(t), \ldots, v_N(t)]$$
$$= -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} w_{n,m} v_n(t) v_m(t) + \sum_{m=1}^{N} \vartheta v_m(t). \quad (E3)$$

For symmetric couplings $w_{n,m} = w_{m,n}$, E_{NL} is monotonically decreasing and the system converges to a steady state. More explicitly, by assuming that neuron n is updated, we have

$$E_{\mathrm{NL}}(t+1) - E_{\mathrm{NL}}(t)$$
$$= -[v_n(t+1) - v_n(t)][u_n(t) - \vartheta] \leq 0. \quad (E4)$$

Equality holds only for $v_n(t+1) = v_n(t)$ or $u_n(t) = \vartheta$, where the latter implies $v_n(t+1) = 1$. Therefore, the energy E_{NL} either decreases in time or remains constant only if the state of the network does not change or the updated neuron n is set to $v_n = 1$. Since the energy is bounded from below due to

$$|E_{\mathrm{NL}}| \leq \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} |w_{n,m}| + N\vartheta, \quad (E5)$$

the network converges to a stable state that is given by a minimum in the energy landscape $E_{\mathrm{NL}}(v_1, \ldots, v_N)$. Thus, the effective nonlinearity reduces the neuronal
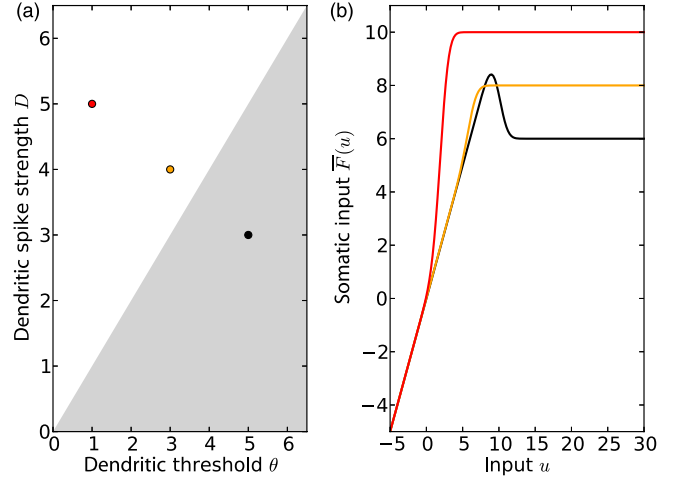

FIG. 10. Monotonicity of the effective somatic input $\bar{F}$. (a) The monotonicity of $\bar{F}$ [Eq. (23)] for $B = 2$, $PN^{-1}\mathrm{Var}[w] = 0.8$ (cf. the caption of Fig. 3), with varying dendritic thresholds θ and dendritic spike strengths D. It displays two regions, one where $\bar{F}$ is strictly monotonic (white) and one where it is nonmonotonic (gray). They are separated by $D = \theta$. (b) The monotonic or nonmonotonic shape of $\bar{F}$ for different dendritic parameters $(\theta, D) \in \{(5,3), (3,4), (1,5)\}$ (black, orange, and red lines).

threshold to $\vartheta \leq \Theta$ as compared to linear input summation [Eq. (15)] but maintains network convergence.

For ϑ to be uniquely defined, the dendritic nonlinearities have to be strong enough $BD > \Theta$, so that $\bar{F}$ intersects the constant function Θ (Fig. 3). Analytical calculations show that for $D > \theta$, the transfer function $\bar{F}$ is strictly monotonic (cf. Fig. 10). Since experiments demonstrate supralinear dendritic amplification, e.g., with thresholds of $\theta \approx 3.8$ mV and spike amplitudes of $D \approx 10$ mV [3], this parameter regime is biologically plausible.

APPENDIX F: ASYMMETRIC COUPLINGS AND CONVERGENCE OF A HOPFIELD NETWORK WITH NONLINEAR DENDRITES

The couplings of the classical Hopfield network are symmetric $w_{n,m} = w_{m,n}$, so that convergence is guaranteed by a Lyapunov function [Eq. (15)]. In the extended model, we argue that the coupling weights $w_{n,b,m}$ to the dendritic branches obey $E[w_{n,b,m}] = B^{-1}w_{n,m}$ or, equivalently, $BE[w_{n,b,m}] = w_{n,m}$ due to Hebbian learning [Eq. (20)]. To account for fluctuations in the learned weights, we assume a variance $\mathrm{Var}[w_{n,b,m}] = w_{n,m}^2 B^{-2}\mathrm{Var}[w]$ [Eq. (21)]. In a particular network realization with a finite number of branches and fluctuations, the dendritic weights do therefore not sum up to the expected Hebbian weight precisely:

$$\sum_{b=1}^{B} w_{n,b,m} =: w'_{n,m} \neq w_{n,m}. \quad (F1)$$
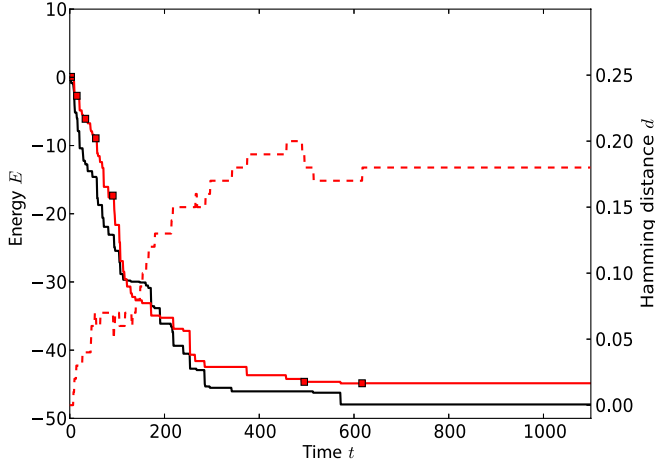
FIG. 11.   Network convergence despite more asymmetric couplings. Parameters are the same as in Fig. 4, with $\text{Var}[w] = 0.5$. Simulations are performed with identical initial states, topology, and order of updates. As the linear Hopfield model is not affected by $\text{Var}[w]$, its energy (solid black line) decreases monotonically and reaches a fixed point. For the extended Hopfield model with nonadditive dendrites and more asymmetric couplings, the energy decreases (solid red line) with rare events of increasing energy (red squares) due to deviations from the mean-field approach or asymmetric couplings [cf. the discussion of Eq. (27)]. The convergence of the system is preserved (checked for 1000 runs). The Hamming distance $d = \frac{1}{2N}\sum_{n=1}^{N}|v_n - v'_n|$ (dashed red line) between the systems shows that they settle into different attractors.

Generally, we have $w'_{n,m} \neq w'_{m,n}$, and the magnitude of the deviation from symmetric couplings is determined by $\text{Var}[w]$. To check if our analytical calculations are applicable despite larger asymmetries, we redo the simulations from the main part of the paper for larger $\text{Var}[w]$.

Our simulations indicate that although convergence of the extended deterministic Hopfield network is not guaranteed by a Lyapunov function for asymmetric couplings, it reaches a fixed point, as shown exemplarily by Fig. 11 and also confirmed for 1000 runs (not shown) for larger asymmetries $\text{Var}[w]$.

To study the impact of asymmetric couplings on the memory performance of the extended stochastic Hopfield network, we repeat the simulations shown in Fig. 5 for larger $\text{Var}[w]$. For a small load $\alpha \approx 0$ and nonzero temperatures $T > 0$, the analytical calculations for symmetric couplings agree well with the simulation results [Fig. 12(a)]. Yet, in the zero-temperature limit $T = 0$, the asymmetries decrease the storage capacity of the network compared to the symmetric case [Fig. 12(b)].

APPENDIX G: MEAN-FIELD CALCULATIONS FOR A STOCHASTIC HOPFIELD MODEL WITH NONLINEAR DENDRITIC BRANCHES

We now derive the mean-field equations for the overlap $m := m^1 = N^{-1}\sum_{n=1}^{N}\xi_n^1\langle v_n\rangle$ of the network state
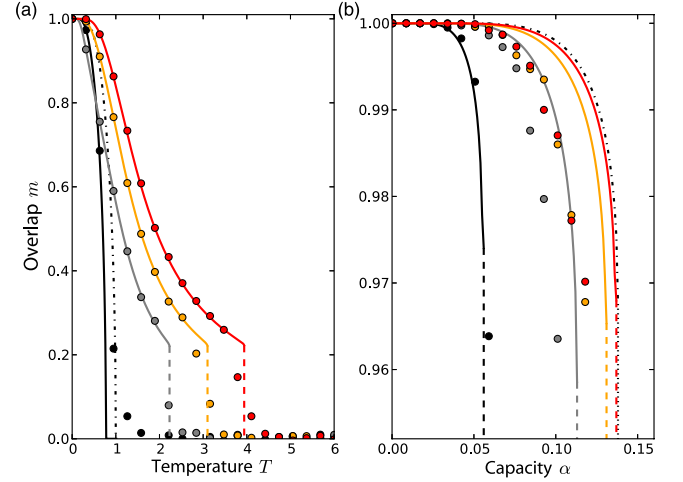

FIG. 12.   Memory performance of the stochastic Hopfield model with nonlinear dendritic branches and more asymmetric couplings. Parameters are the same as in Fig. 5, with $\text{Var}[w] = 0.5$. The figure compares analytical results (solid lines) for linear summation (black lines and symbols) and nonlinear summation with dendritic spike strengths $D \in \{0.4, 0.6, 0.8\}$ (gray, orange, and red lines and symbols) and symmetric couplings to simulation results (circles) for asymmetric couplings. Results for linear input summation with $\Theta = 0$ are included for comparison (dash-dotted black line). (a) The overlap m versus the temperature T for a small load $\alpha = N^{-1} \approx 0$. The more asymmetric couplings change m only slightly, and the simulation results agree well with the analytics of the symmetric couplings. (b) m versus α for $T = 0$. The stronger asymmetries impair the memory function, and the overlap m displays a drop at lower loads α compared to the symmetric case.

$(v_1, \ldots, v_N)$ with pattern $p = 1$, i.e., $(\xi_1^1, \ldots, \xi_N^1)$. The following calculations go along those provided in Refs. [13,67]. From

$$\langle u_n\rangle = \sum_{k=1}^{N} w_{n,k}\langle v_k\rangle = \sum_{k=1}^{N} N^{-1}\sum_{p=1}^{P}\xi_n^p\xi_k^p\langle v_k\rangle$$

$$= \sum_{p=1}^{P}\xi_n^p m^p =: u_n \qquad (G1)$$

and Eq. (29), we find

$$m^q = N^{-1}\sum_{n=1}^{N}\xi_n^q \tanh\{\beta[\bar{F}(u_n) - \Theta]\}$$

$$= N^{-1}\sum_{n=1}^{N}\xi_n^q\xi_n^1 \tanh\{\beta\xi_n^1[\bar{F}(u_n) - \Theta]\}, \qquad (G2)$$

where we employ the point symmetry of $\tanh(x) = -\tanh(-x)$ in the second line. We now assume that the number P of patterns is large, of order $\mathcal{O}(N)$. We define the mean square overlap $r := \alpha^{-1}\sum_{q\neq1}^{P}(m^q)^2$ and assume that the m^q are independent, zero-centered random variables

with variance $\alpha r P^{-1}$. Then, the sum $\xi_n^1 \sum_{p \neq 1}^{P} \xi_n^p m^p$ can be seen as a Gaussian noise term of variance αr and the sum $N^{-1} \sum_{n=1}^{N}$ may be treated as an average over this noise. Since $\xi_n^1 = \pm 1$ with equal probabilities, the overlap with the first pattern is [Eq. (G2)]

$$m = \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh\{\beta[\bar{F}(m + \sqrt{\alpha r}z) - \Theta]\}$$
$$+ \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh\{-\beta[\bar{F}(-m - \sqrt{\alpha r}z) - \Theta]\},$$

(G3)

with the effective somatic input $\bar{F}$ given by Eqs. (23)–(25). Next, the correlations quantified by r must be determined self-consistently. We define $\bar{u}_n^q := \sum_{p \neq q}^{P} \xi_n^p m^p$, and because $\xi_n^q m^q = u_n - \bar{u}_n^q$ is small, of order $\mathcal{O}(N^{-1/2})$, we expand Eq. (G2) into a Taylor series to first order:

$$m^q \approx N^{-1} \sum_{n=1}^{N} \xi_n^q \xi_n^1 \tanh\{\beta \xi_n^1 [\bar{F}(\bar{u}_n^q) - \Theta]\}$$
$$+ \beta N^{-1} \sum_{n=1}^{N} (1 - \tanh^2\{\beta \xi_n^1 [\bar{F}(\bar{u}_n^q) - \Theta]\}) \bar{F}'(\bar{u}_n^q) m^q,$$

(G4)

where $f'(x_0) = \frac{df(x)}{dx}\big|_{x=x_0}$ denotes the first derivative of a function $f(x)$ evaluated at $x = x_0$. Similar to the derivation of Eq. (G3), we approximate $\xi_n^1 \sum_{p \neq 1, q}^{P} \xi_n^p m^p$ by a zero-centered Gaussian distribution with variance αr and in the second term of Eq. (G4) treat the sum $N^{-1} \sum_{n=1}^{N}$ as an average so that

$$m^q = N^{-1} \sum_{n=1}^{N} \xi_n^q \xi_n^1 \tanh\{\beta \xi_n^1 [\bar{F}(\bar{u}_n^q) - \Theta]\} + \beta C m^q, \quad (G5)$$

with

$$C := \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} (1 - \tanh^2\{\beta[\bar{F}(m + \sqrt{\alpha r}z) - \Theta]\})$$
$$\times \bar{F}'(m + \sqrt{\alpha r}z) + \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2}$$
$$\times (1 - \tanh^2\{-\beta[\bar{F}(-m - \sqrt{\alpha r}z) - \Theta]\})$$
$$\times \bar{F}'(-m - \sqrt{\alpha r}z),$$

(G6)

where we take into account again that $\xi_n^1 = \pm 1$ with equal probabilities. Solving Eq. (G5) for m^q, squaring it, and averaging over all patterns q yields

$$r = (1 - \beta C)^{-2} N P^{-1} \sum_{q \neq 1}^{P} N^{-2} \sum_{n,k=1}^{N} \xi_n^q \xi_n^1 \xi_k^q \xi_k^1$$
$$\cdot \tanh\{\beta \xi_n^1 [\bar{F}(\bar{u}_n^q) - \Theta]\} \tanh\{\beta \xi_k^1 [\bar{F}(\bar{u}_k^q) - \Theta]\}$$
$$= (1 - \beta C)^{-2} N^{-1} \sum_{n=1}^{N} \tanh^2\{\beta \xi_n^1 [\bar{F}(\bar{u}_n^2) - \Theta]\}. \quad (G7)$$

Here, we use that the arguments of the tanh are independent of q, and in the average $P^{-1} \sum_{q \neq 1}^{P}$, only terms $n = k$ survive. We set $\bar{u}_n^q = \bar{u}_n^2$ without loss of generality. Employing a Gaussian approximation of the sum $\xi_n^1 \sum_{p \neq 1,2}^{P} \xi_n^p m^p$ like in Eq. (G5) yields

$$r = (1 - \beta C)^{-2} s, \quad (G8)$$

with

$$s := \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh^2\{\beta[\bar{F}(m + \sqrt{\alpha r}z) - \Theta]\}$$
$$+ \frac{1}{2} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh^2\{-\beta[\bar{F}(-m - \sqrt{\alpha r}z) - \Theta]\}.$$

(G9)

The above Gaussian approximations hold for α of order $\mathcal{O}(1)$ and smaller. Equations (G3), (G6), (G8), and (G9) constitute a set of nonlinear, coupled integral equations for the order parameters m, r, and s and can be solved numerically or in limiting cases.
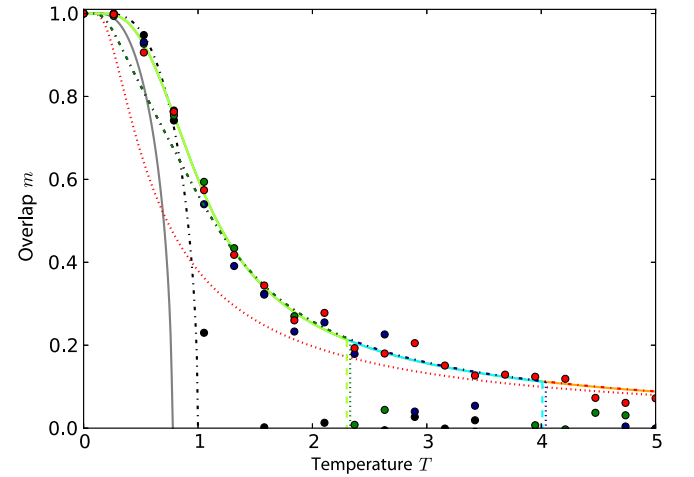

FIG. 13. Critical behavior of the network for a few patterns $\alpha \approx 0$ and varying neuronal thresholds Θ and dendritic thresholds θ. Parameters are the same as in Fig. 5, with $D = 0.4$ and different dendritic thresholds $\theta \in \{\infty, 0.1, 0.05, 0\}$. Simulation results (circles) and analytical results (dash-dotted lines) for vanishing neuronal thresholds $\Theta = 0$ are compared to the analytical results for $\Theta = 0.4$ (solid lines; cf. Fig. 5). $\theta = \infty$ (dash-dotted black and solid gray lines) corresponds to the linear Hopfield model. Decreasing the dendritic threshold θ increases the critical temperature ($\theta = 0.1$ in green and lime lines; $\theta = 0.05$ in blue and cyan lines). For $\theta = 0$ (red and orange lines), there is no phase transition anymore, as $m(T)$ asymptotically follows Eq. (H2) (dotted red line).

APPENDIX H: MEMORY CAPACITY OF A STOCHASTIC HOPFIELD NETWORK WITH NONLINEAR DENDRITIC BRANCHES IN THE THERMODYNAMIC LIMIT FOR A FINITE NUMBER OF PATTERNS

We now consider the quality of pattern retrieval estimated by the overlap m in the thermodynamic limit of large N with finitely many patterns P, i.e., $\alpha \approx 0$. Because m^q, $q \neq 1$, is of order $\mathcal{O}(N^{-1/2})$ and P is finite, we may write $u_n = \sum_{p=1}^{P} \xi_n^p m^p \approx \xi_n^1 m$. Starting from Eq. (G2) and using the definition of the effective somatic input $\bar{F}$ [Eqs. (23)–(25)],

$$
\begin{aligned}
m &= N^{-1} \sum_{n=1}^{N} \tanh\{\beta[1 - P_{\mathrm{NL}}(u_n)]\xi_n^1 u_n + \beta \xi_n^1 [\mathrm{BDP}_{\mathrm{NL}}(u_n) - \mathrm{BC}_{\mathrm{NL}}(u_n) - \Theta]\} \\
&\approx N^{-1} \sum_{n=1}^{N} \tanh\{\beta[1 - P_{\mathrm{NL}}(\xi_n^1 m)]m + \beta \xi_n^1 [\mathrm{BDP}_{\mathrm{NL}}(\xi_n^1 m) - \mathrm{BC}_{\mathrm{NL}}(\xi_n^1 m) - \Theta]\} \\
&= \frac{1}{2}\tanh\{\beta[1 - P_{\mathrm{NL}}(m)]m + \beta[BDP_{\mathrm{NL}}(m) - \mathrm{BC}_{\mathrm{NL}}(m) - \Theta]\} \\
&\quad + \frac{1}{2}\tanh\{\beta[1 - P_{\mathrm{NL}}(-m)]m - \beta[\mathrm{BDP}_{\mathrm{NL}}(-m) - \mathrm{BC}_{\mathrm{NL}}(-m) - \Theta]\},
\end{aligned}
\tag{H1}
$$

where we use that $\xi_n^1 = \pm 1$ with equal probabilities in the third line. This transcendental equation for m is solved numerically (Fig. 5).

It is shown in the main text that dendritic nonlinearities elevate the critical temperature T_c above which retrieval fails (Fig. 5). The increased critical temperature T_c may result partially from the effectively reduced neuronal threshold $\vartheta \leq \Theta$ [Eq. (26)]. To exclude this effect, we study the impact of the nonlinearity on the critical temperature T_c for vanishing neuronal threshold $\Theta = 0$. We find that when changing Θ from $\Theta = 0.4$ to $\Theta = 0$, the critical temperature T_c is altered only slightly (Fig. 13; see the differences between the dashed and dotted vertical lines) and the behavior of the system remains the same.
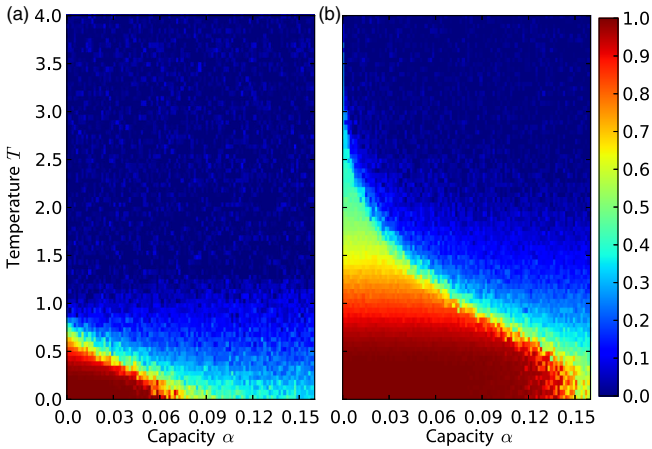
In contrast, the threshold θ of the dendritic nonlinearity has a strong influence on the critical behavior of the network. For $\theta = \infty$, we reobtain linear input summation (Fig. 13, dash-dotted black and solid gray lines). For finite $0 < \theta < \infty$, the critical temperature T_c is increased (Fig. 13, green and lime, blue and cyan lines). For $\theta = 0$, there is no phase transition at all (Fig. 13, red and orange lines) and memory retrieval at arbitrarily high temperatures T (although with smaller and smaller overlaps m) is possible. The absence of a phase transition may be understood by assuming $\theta = 0$ and $T = \beta^{-1} \to \infty$ in Eq. (H1). Then,

$$
m \to \frac{1}{2}\tanh[T^{-1}(BD - \Theta)] + \frac{1}{2}\tanh(T^{-1}\Theta) \tag{H2}
$$

because $P_{\mathrm{NL}}(m) \approx 1$ and $C_{\mathrm{NL}}(m) \approx P_{\mathrm{NL}}(-m) \approx C_{\mathrm{NL}}(-m) \approx 0$ for $\theta = 0$ and $m \to 0$ for $T \to \infty$. Since $BD > \Theta$ (Appendix E), we have $m > 0$ for $T > 0$ and no phase transition occurs (Fig. 13, dotted red line).

APPENDIX I: MEMORY CAPACITY OF A STOCHASTIC HOPFIELD NETWORK WITH NONLINEAR DENDRITIC BRANCHES IN THE ZERO-TEMPERATURE LIMIT

We now compute the overlap m in the zero-temperature limit $T = 0$. The starting points are Eqs. (G3), (G6), (G8), and (G9). For $\beta = T^{-1} \to \infty$, we may simplify

$$
\begin{aligned}
\lim_{\beta \to \infty} \tanh\{&\pm\beta[\bar{F}(\pm m \pm \sqrt{\alpha r}z) - \Theta]\} \\
&= \mathrm{sgn}[\pm\bar{F}(\pm m \pm \sqrt{\alpha r}z) \mp \Theta] \\
&= \mathrm{sgn}[m + \sqrt{\alpha r}z \mp \vartheta], \tag{I1}
\end{aligned}
$$

by definition of the effective threshold ϑ [see Eq. (26)], $\bar{F}(\vartheta) = \Theta$. Using the dominated convergence theorem, we may compute [Eq. (G3)]

FIG. 14. Quality of retrieval in dependence of capacity α and temperature T. Numerical results for the overlap m, averaged over 20 realizations. Parameters are the same as in Fig. 5 with $D = 0.8$. The network model with nonadditive dendrites displays (b) a larger region of stable memory retrieval than (a) the original Hopfield network.

$$\lim_{\beta\to\infty} m = \frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\mathrm{sgn}[m+\sqrt{\alpha r}z-\vartheta]+\frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\mathrm{sgn}[m+\sqrt{\alpha r}z+\vartheta]$$

$$=\frac{1}{2}\mathrm{erf}\left(\frac{m-\vartheta}{\sqrt{2\alpha r}}\right)+\frac{1}{2}\mathrm{erf}\left(\frac{m+\vartheta}{\sqrt{2\alpha r}}\right) \tag{I2}$$

and [Eq. (G9)]

$$\lim_{\beta\to\infty} s = \frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\mathrm{sgn}^2[m+\sqrt{\alpha r}z-\vartheta]+\frac{1}{2}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\mathrm{sgn}^2[m+\sqrt{\alpha r}z+\vartheta]=1. \tag{I3}$$

To derive r, we further compute βC [Eq. (G6)]. We use that $\frac{\beta}{2}(1-\tanh^2[\beta x])$ approaches the Dirac delta function $\delta(x)$ for $\beta\to\infty$:

$$\lim_{\beta\to\infty}\beta C = \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\delta[\bar{F}(m+\sqrt{\alpha r}z)-\Theta]\bar{F}'(m+\sqrt{\alpha r}z)$$

$$+\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\delta[-\bar{F}(-m-\sqrt{\alpha r}z)+\Theta]\bar{F}'(-m-\sqrt{\alpha r}z)$$

$$=\frac{1}{|\bar{F}'(\vartheta)|\sqrt{\alpha r}}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\delta\left[z+\frac{m-\vartheta}{\sqrt{\alpha r}}\right]\bar{F}'(m+\sqrt{\alpha r}z)+\frac{1}{|\bar{F}'(\vartheta)|\sqrt{\alpha r}}$$

$$\times\int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\delta\left[z+\frac{m+\vartheta}{\sqrt{\alpha r}}\right]\bar{F}'(-m-\sqrt{\alpha r}z)$$

$$=\sqrt{\frac{1}{2\pi\alpha r}}\left[\exp\left(-\frac{(m-\vartheta)^2}{2\alpha r}\right)+\exp\left(-\frac{(m+\vartheta)^2}{2\alpha r}\right)\right]. \tag{I4}$$

In the second line, we use $\delta[f(x)]=|f'(x_0)|^{-1}\delta(x-x_0)$ with the (single) root x_0, $f(x_0)=0$, and again $\bar{F}(\vartheta)=\Theta$. There and in the third line, we use that $\bar{F}$ is monotonically increasing, i.e., $\bar{F}'\geq 0$. From Eqs. (G8) and (I4), we find

$$\sqrt{r}=1+\sqrt{\frac{1}{2\pi\alpha}}\exp\left(-\frac{(m-\vartheta)^2}{2\alpha r}\right)$$

$$+\sqrt{\frac{1}{2\pi\alpha}}\exp\left(-\frac{(m+\vartheta)^2}{2\alpha r}\right). \tag{I5}$$

Equations (I2) and (I5) provide coupled, implicit equations for the order parameters m and r that can be solved numerically. Solutions for varying effective thresholds ϑ are shown in Fig. 5.

APPENDIX J: PHASE DIAGRAMS OF ASSOCIATIVE-MEMORY NETWORKS OF NEURONS WITH ADDITIVE AND NONADDITIVE DENDRITIC PROCESSING

For a further comparison of the memory performance of networks with arborized neurons with the classical Hopfield model, we compute the quality of retrieval m in the α-T plane.

We find both analytically and numerically that in the limits $\alpha\approx 0$ and $T=0$, the critical temperature and capacity of the network with nonadditive dendrites are higher than for the linear model; cf. Fig. 5. Complementing these findings, numerical simulations show that the α-T region of successful memory retrieval is larger for nonadditive dendrites [Fig. 14(b)] than for linear dendrites [Fig. 14(a)].

[1] C. Koch and I. Segev, *The Role of Single Neurons in Information Processing*, Nat. Neurosci. **3**, 1171 (2000).

[2] G. Stuart, N. Spruston, and M. Häusser, *Dendrites* (Oxford University Press, New York, 2007).

[3] G. Ariav, A. Polsky, and J. Schiller, *Submillisecond Precision of the Input-Output Transformation Function Mediated by Fast Sodium Dendritic Spikes in Basal Dendrites of CA1 Pyramidal Neurons*, J. Neurosci. **23**, 7750 (2003).

[4] S. Gasparini, M. Migliore, and J. C. Magee, *On the Initiation and Propagation of Dendritic Spikes in CA1 Pyramidal Neurons*, J. Neurosci. **24**, 11046 (2004).

[5] A. Polsky, B. W. Mel, and J. Schiller, *Computational Subunits in Thin Dendrites of Pyramidal Cells*, Nat. Neurosci. **7**, 621 (2004).

[6] T. Nevian, M. E. Larkum, A. Polsky, and J. Schiller, *Properties of Basal Dendrites of Layer 5 Pyramidal Neurons: A Direct Patch-Clamp Recording Study*, Nat. Neurosci. **10**, 206 (2007).

[7] M. E. Larkum and T. Nevian, *Synaptic Clustering by Dendritic Signalling Mechanisms*, Curr. Opin. Neurobiol. **18**, 321 (2008).

[8] P. Poirazi, T. Brannon, and B. W. Mel, *Pyramidal Neuron as Two-Layer Neural Network*, Neuron **37**, 989 (2003).

[9] M. London and M. Häusser, *Dendritic Computation*, Annu. Rev. Neurosci. **28**, 503 (2005).

[10] P. Poirazi, T. Brannon, and B. W. Mel, *Arithmetic of Subthreshold Synaptic Summation in a Model CA1 Pyramidal Cell*, Neuron **37**, 977 (2003).

[11] B. W. Mel, *Advances in Neural Information Processing Systems* (Morgan Kaufmann, San Mateo, CA, 1992), Vol. 4, p. 35–42.

[12] E. Barkai, D. Hansel, and I. Kanter, *Statistical Mechanics of a Multilayered Neural Network*, Phys. Rev. Lett. **65**, 2312 (1990).

[13] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Westview, Boulder, CO, 1991).

[14] M. Biehl, M. Ahr, and E. Schlösser, *Statistical Physics of Learning: Phase Transitions in Multilayered Neural Networks*, Advances in Solid State Physics Vol. 40 (Springer, Berlin, 2000).

[15] A. Engel, H.-M. Köhler, F. Tschepke, H. Vollmayr, and A. Zippelius, *Storage Capacity and Learning Algorithms for Two-Layer Neural Networks*, Phys. Rev. A **45**, 7590 (1992).

[16] R. Urbanczik, *Storage Capacity of the Fully-Connected Committee Machine*, J. Phys. A **30**, L387 (1997).

[17] B. Ujfalussy, T. Kiss, and P. Érdi, *Parallel Computational Subunits in Dentate Granule Cells Generate Multiple Place Fields*, PLoS Comput. Biol. **5**, e1000500 (2009).

[18] L. L. Gollo, O. Kinouchi, and M. Copelli, *Statistical Physics Approach to Dendritic Computation: The Excitable-Wave Mean-Field Approximation*, Phys. Rev. E **85**, 011911 (2012).

[19] J. Schiller and Y. Schiller, *NMDA Receptor-Mediated Dendritic Spikes and Coincident Signal Amplification*, Curr. Opin. Neurobiol. **11**, 343 (2001).

[20] A. Losonczy, J. K. Makara, and J. C. Magee, *Compartmentalized Dendritic Plasticity and Input Feature Storage in Neurons*, Nature (London) **452**, 436 (2008).

[21] B. W. Mel, *NMDA-Based Pattern Discrimination in a Modeled Cortical Neuron*, Neural Comput. **4**, 502 (1992).

[22] P. Poirazi and B. W. Mel, *Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue*, Neuron **29**, 779 (2001).

[23] P. A. Rhodes, *Recoding Patterns of Sensory Input: Higher-Order Features and the Function of Nonlinear Dendritic Trees*, Neural Comput. **20**, 2000 (2008).

[24] M. Schiess, R. Urbanczik, and W. Senn, *Gradient Estimation in Dendritic Reinforcement Learning*, J. Math. Neurosci. **2**, 1 (2012).

[25] J. J. Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[26] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Spin-Glass Models of Neural Networks*, Phys. Rev. A **32**, 1007 (1985).

[27] K. Morita, M. Okada, and K. Aihara, *Selectivity and Stability via Dendritic Nonlinearity*, Neural Comput. **19**, 1798 (2007).

[28] K. Morita, *Possible Role of Dendritic Compartmentalization in the Spatial Working Memory Circuit*, J. Neurosci. **28**, 7699 (2008).

[29] J. E. Lisman, J. M. Fellous, and X. J. Wang, *A Role for NMDA-Receptor Channels in Working Memory*, Nat. Neurosci. **1**, 273 (1998).

[30] X. J. Wang, *Synaptic Reverberation Underlying Mnemonic Persistent Activity*, Trends Neurosci. **24**, 455 (2001).

[31] D. Zhang, Y. Li, M. J. Rasch, and S. Wu, *Nonlinear Multiplicative Dendritic Integration in Neuron and Network Models*, Front. Comput. Neurosci. **7** (2013).

[32] Y. Katz, W. L. Kath, N. Spruston, and M. E. Hasselmo, *Coincidence Detection of Place and Temporal Context in a Network Model of Spiking Hippocampal Neurons*, PLoS Comput. Biol. **3**, e234 (2007).

[33] G. J. Stuart and M. Häusser, *Dendritic Coincidence Detection of EPSPs and Action Potentials*, Nat. Neurosci. **4**, 63 (2001).

[34] M. A. Long, D. Z. Jin, and M. S. Fee, *Support for a Synaptic Chain Model of Neuronal Sequence Generation*, Nature (London) **468**, 394 (2010).

[35] R. D. Traub and R. K. S. Wong, *Cellular Mechanism of Neuronal Synchronization in Epilepsy*, Science **216**, 745 (1982).

[36] S. Jahnke, M. Timme, and R. M. Memmesheimer, *Guiding Synchrony through Random Networks*, Phys. Rev. X **2**, 041016 (2012).

[37] R. M. Memmesheimer and M. Timme, *Non-additive Coupling Enables Propagation of Synchronous Spiking Activity in Purely Random Networks*, PLoS Comput. Biol. **8**, e1002384 (2012).

[38] R. M. Memmesheimer, *Quantitative Prediction of Intermittent High-Frequency Oscillations in Neural Networks with Supralinear Dendritic Interactions*, Proc. Natl. Acad. Sci. U.S.A. **107**, 11 092 (2010).

[39] J. J. Hopfield, *Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons*, Proc. Natl. Acad. Sci. U.S.A. **81**, 3088 (1984).

[40] M. Shiino, *Stochastic Analyses of the Dynamics of Generalized Little-Hopfield-Hemmen Type Neural Networks*, J. Stat. Phys. **59**, 1051 (1990).

[41] M. Morita, *Associative Memory with Nonmonotone Dynamics*, Neural Netw. **6**, 115 (1993).

[42] J. Inoue, *Retrieval Phase Diagrams of Non-monotonic Hopfield Networks*, J. Phys. A **29**, 4815 (1996).

[43] H. Qiao, J. Peng, and Z. B. Xu, *Nonlinear Measures: A New Approach to Exponential Stability Analysis for Hopfield-Type Neural Networks*, Neural Netw. **12**, 360 (2001).

[44] J. Schiller, G. Major, H. J. Koester, and Y. Schiller, *NMDA Spikes in Basal Dendrites of Cortical Pyramidal Neurons*, Nature (London) **404**, 285 (2000).

[45] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, Cambridge, England, 1991).

[46] A. Wald, *On Cumulative Sums of Random Variables*, Ann. Math. Stat. **15**, 283 (1944).

[47] D. Blackwell and M. A. Girshick, *On Functions of Sequences of Independent Chance Vectors with Applications to the Problem of the "Random Walk" in k Dimensions*, Ann. Math. Stat. **17**, 310 (1946).

[48] R. Sheldon, *A First Course in Probability* (Pearson Education, Delhi, 2002).

[49] H. G. Kim, M. Beierlein, and B. W. Connors, *Inhibitory Control of Excitable Dendrites in Neocortex*, J. Neurophysiol. **74**, 1810 (1995).

[50] A. I. Gulyás, M. Megias, Z. Emri, and T. F. Freund, *Total Number and Ratio of Excitatory and Inhibitory Synapses Converging onto Single Interneurons of Different Types in the CA1 Area of the Rat Hippocampus*, J. Neurosci. **19**, 10 082 (1999).

[51] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

[52] M. V. Tsodyks and M. V. Feigel'man, *The Enhanced Storage Capacity in Neural Networks with Low Activity Level*, Europhys. Lett. **6**, 101 (1988).

[53] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Storing Infinite Numbers of Patterns in a Spin Glass Model of Neural Networks*, Phys. Rev. Lett. **55**, 1530 (1985).

[54] M. Graupner and N. Brunel, *Calcium-Based Plasticity Model Explains Sensitivity of Synaptic Changes to Spike Pattern, Rate, and Dendritic Location*, Proc. Natl. Acad. Sci. U.S.A. **109**, 3991 (2012).

[55] M. W. Hirsch, *Convergent Activation Dynamics in Continuous Time Networks*, Neural Netw. **2**, 331 (1989).

[56] D. K. Smetters and A. Zador, *Synaptic Transmission: Noisy Synapses and Noisy Neurons*, Curr. Biol. **6**, 1217 (1996).

[57] R. J. Glauber, *Time-Dependent Statistics of the Ising Model*, J. Math. Phys. (N.Y.) **4**, 294 (1963).

[58] M. Häusser and B. W. Mel, *Dendrites: Bug or Feature?*, Curr. Opin. Neurobiol. **13**, 372 (2003).

[59] G. N. Elston and M. G. Rosa, *Morphological Variation of Layer III Pyramidal Neurones in the Occipitotemporal Pathway of the Macaque Monkey Visual Cortex*, Cereb. Cortex **8**, 278 (1998).

[60] F. T. Sommer and T. Wennekers, *Associative Memory in Networks of Spiking Neurons*, Neural Netw. **14**, 825 (2001).

[61] Y. Aviel, D. Horn, and M. Abeles, *Memory Capacity of Balanced Networks*, Neural Comput. **17**, 691 (2005).

[62] N. L. Golding, N. P. Staff, and N. Spruston, *Dendritic Spikes as a Mechanism for Cooperative Long-Term Potentiation*, Nature (London) **418**, 326 (2002).

[63] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, *Noise Enhancement of Information Transfer in Crayfish Mechanoreceptors by Stochastic Resonance*, Nature (London) **365**, 337 (1993).

[64] S. Kadar, J. Wang, and K. Showalter, *Noise-Supported Travelling Waves in Sub-excitable Media*, Nature (London) **391**, 770 (1998).

[65] B. Crespi, *Storage Capacity of Non-monotonic Neurons*, Neural Netw. **12**, 1377 (1999).

[66] R. Legenstein and W. Maass, *Branch-Specific Plasticity Enables Self-Organization of Nonlinear Computation in Single Neurons*, J. Neurosci. **31**, 10 787 (2011).

[67] T. Geszti, *Physical Models of Neural Networks* (World Scientific, Singapore, 1990).