APS physics

CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Synchrony in stochastically driven neuronal networks with complex topologies

Katherine A. Newhall, Maxim S. Shkarayev, Peter R. Kramer, Gregor Kovačič, and David Cai

Phys. Rev. E **91**, 052806 — Published 11 May 2015

DOI: 10.1103/PhysRevE.91.052806

Synchrony in Stochastically Driven Neuronal Networks with Complex Topologies

Katherine A. Newhall,[1] Maxim S. Shkarayev,[2] Peter R. Kramer,[3] Gregor Kovačič,[3] and David Cai[4, 5, 6, *]

[1] *Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3250, USA*
[2] *Department of Physics and Astronomy, Iowa State University, 12 Physics Hall, Ames, IA 50011-3160, USA*
[3] *Mathematical Sciences Department, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA*
[4] *Courant Institute of Mathematical Sciences and Center for Neural Science, New York University, 251 Mercer Street, New York, NY 10012, USA*
[5] *Department of Mathematics, MOE-LSC and Institute of Natural Sciences, Shanghai Jiao Tong University, Dong Chuan Road 800, Shanghai 200240, China*
[6] *NYUAD Institute, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, United Arab Emirates*

(Dated: April 20, 2015)

We study the synchronization of a stochastically-driven, current-based, integrate-and-fire neuronal model on a preferential-attachment network with scale-free characteristics and high clustering. The synchrony is induced by cascading total firing events where every neuron in the network fires at the same instant of time. We show that in the regime where the system remains in this highly synchronous state, the firing rate of the network is completely independent of the synaptic coupling, and depends solely on the external drive. On the other hand, the ability for the network to maintain synchrony depends on a balance between the fluctuations of the external input and the synaptic coupling strength. In order to accurately predict the probability of repeated cascading total firing events we go beyond mean-field and tree-like approximations and conduct a detailed second order calculation taking into account local clustering. Our explicit analytical results are shown to give excellent agreement with direct numerical simulations for the particular preferential-attachment network model investigated.

PACS numbers: 87.19.lj, 05.20.Dd, 05.10.Gg, 87.19.lp
Keywords: neuronal networks, integrate-and-fire, synchrony, first passage time

I. INTRODUCTION

Pulse-coupled network models have recently proved their utility in branches of science and engineering ranging from internet traffic regulation [1, 2] to power grid management [3] to epidemic spreading [4–9] in addition to their well-established role in neuroscience [10–13]. In neuroscience applications the node dynamics in the network are often described by the integrate-and-fire (IF) oscillator due to its simplicity [14, 15]. On the one hand, networks of IF oscillators frequently capture *network mechanisms* as accurately as the more complicated models such as Hodgkin-Huxley, while on the other hand, they are sufficiently simple to be simulated at a fraction of the cost or even to be treated analytically.

The architectural connectivity influences the nature of the dynamics exhibited by a given pulse-coupled network. Naturally, the topology of such a network can be represented as a directed graph, with the network's IF dynamical units represented by the graph's nodes and the connections between pairs of units by directed edges between pairs of nodes. In particular, each node is characterized by the number of directed edges emanating from it (*outgoing degree*) and terminating at it (*incoming degree*). Frequently, only the statistics of these degrees are known, or even just the asymptotic behavior of these statistics for large values of the outgoing and incoming node-degrees. For example, *scale-free* networks are characterized by the degrees being asymptotically distributed according to decaying power laws. We are therefore interested in predicting the behavior of the dynamics based on a few key statistical descriptors of the network topology.

A commonly studied dynamical regime of pulse-coupled networks is synchronous oscillations. Within the brain, there are rich oscillatory phenomena [16–21]. A great deal of theoretical work [22–28] has explored mechanisms of synchrony in well-defined networks, often with deterministic driving. Particular attention was frequently paid to which features of the phase response curve are conducive to synchrony [27]. We will be concerned with developing an inherently statistical approach applicable to all-excitatory networks [23, 24, 29–32] that have stochastic driving and only a statistical description of their connectivity architecture.

To highlight our focus on these statistical aspects, we will select a simple IF model for our investigation. In IF networks with instantaneous pulse coupling, synchronous oscillations can take the form of (statistically) periodic *total firing events* during which all the neurons in the network fire at once [24, 29–31]. While total firing events are a highly idealized form of synchrony, their analysis may provide insight into more realistic correlated firing events in neuronal networks. For the case of all-to-all connected networks, analytical methods were developed in [29] (and in [24] for a discrete state IF model) to quantify the strength of deterministic driving needed to

* cai@cims.nyu.edu

completely synchronize the neuronal membrane-potential trajectories, and in [30, 31] to characterize the parameter regimes in which a network of neurons driven by independent, stochastic Poisson pulse-trains has high probability to sustain total firing events. Note that under independent Poisson pulse-train drive for each neuron, while the neurons *fire* all at once as a result of *cascading* total firing events, the trajectories of their membrane potentials between two cascading events evolve separately. The emergence of synchrony in this case depends strongly on the interplay between the spreading of the membrane-potential trajectories induced by the independent Poisson-train drive to each neuron and their synchronization due to their reset after every total firing cascade [30, 31].

During a cascading total firing event in an all-to-all, instantaneously-coupled, IF neuronal network, when the first neuron fires, the membrane potentials of all the remaining neurons jump instantaneously by the same amount proportional to the coupling strength. If one or more of these potentials exceed the firing threshold, the corresponding neurons fire, and again instantaneously increase the membrane potentials of all the remaining neurons. This cascading process repeats until all the neurons in the network have fired. The all-to-all connectivity of the network is crucial in this process: if certain interneuronal connections are missing, the firing of one neuron does not necessarily increase the membrane potentials of all the remaining neurons, but just of those connected to it. A central advance presented in the present work therefore addresses the interplay between the network topology and the time-evolving ensemble of the neuronal membrane potentials that enables cascading total firing events to still take place with high probability in the appropriate parameter regimes. We are thereby attempting to contribute insight from another perspective to the broad question of how the statistical properties of the network are reflected in the statistical properties of the network dynamics [10, 32–37]. Moreover, we are here building a connection between our previous results concerning the statistical description of cascading total firing events in all-to-all coupled IF networks [30, 31] with those addressing steady-state statistics of firing rates and membrane-potential correlations in the asynchronous operating regime of IF networks with specific nontrivial architectural connectivity [33].

The statistics we derived in [30, 31] for the time intervals between repeated cascading total firing events in all-to-all connected networks carries over essentially unchanged for nontrivial network topology, as we shall discuss in Subsection III A. We must however substantially rework the derivation for the probability that such statistically synchronized firing dynamics is self-sustaining, and the bulk of our effort in the current work is to develop and apply a new theoretical approach for computing this probability. We improve upon mean-field [28, 38–44] and tree-like [45–48] approximations for network dynamics by explicitly accounting for clustering. The derivation we employ relies crucially on our empirical observation that, in the parameter regimes exhibiting at least a partial degree of synchrony, once two or three neurons have fired in succession, the probability that the rest of the cascade would fail becomes negligible. While our simulations and specific calculations are carried out for a specific, scale-free network, we emphasize that our general framework for computing the probability of repeated cascading total firing events is applicable to IF networks with general complex topology.

We employ two main ingredients in the derivation of the probability of repeated cascading total firing events in complex IF networks. The first is an exit-time problem used for computing probability distribution of the time when the first neuronal membrane potential reaches firing threshold after a total firing event, which is obtained by solving a Fokker-Planck equation. The second is the Gaussian approximation to the membrane-potential distribution for the neurons that have not yet fired, found without consideration of the reset mechanism and using the exact evaluation of the membrane-potential cumulants. Since between pairs of cascading total firing events the neurons are effectively uncoupled, these two ingredients are independent of the network architecture, and can in fact be calculated for individual neurons with feedforward Poisson-train drive. Therefore, we were able to use the results of [30, 31] for both, and we here only describe them briefly in order to complete the description of our new combinatorial method for finding the probability of repeated cascading total firing events in systems with complex network topology.

The remainder of the paper is organized as follows. In Sec. II, we discuss the current-based integrate-and-fire network model and define the network terminology we use throughout this paper. We begin our theoretical characterization of the perfectly synchronous firing state in Sec. III. This begins with a derivation of the time between total firing events in Sec. III A, and continues in Sec. III B with the calculation of the probability to see repeated cascading total firing events for a general network topology. We complete this description for a specific scale-free network, described in the beginning of Sec. IV, by presenting distributions related to the outgoing degrees of one, two and three connected nodes in Sec. IV A. In Sec. IV B, we obtain excellent agreement between our theory and results of numerical simulation, thus confirming the validity of our approximations. We close with the conclusions in Sec. V.

Derivations of the analytical approximations are presented in the appendices. In Appendix A, we explain how the distribution of the neuronal voltages can be approximated in terms of a Gaussian distribution. Appendix B complements the results of [33] to calculate the one-, two-, and three-node statistics of the scale-free network model needed in the analytical expressions for the probability to see repeated total firing events. The development of our precise second-order approximations for this probability is presented in Appendix C, and a useful asymptotic sim-

plification of one of these approximations is computed in Appendix D. Some technical results needed for these asymptotic calculations are derived in Appendix E.

II. NEURONAL NETWORK MODEL

The network model we consider consists of N coupled, current-based, excitatory integrate-and-fire point neurons [15, 49], which are governed by the system of equations

$$\frac{dv_j}{dt} = -g_L(v_j - V_R) + I_j(t), \quad j = 1, \ldots, N, \tag{1a}$$

where v_j is the membrane potential of the jth neuron, g_L is the leakage conductance, and V_R is the reset voltage. The injected current, $I_j(t)$, is modeled by the pulse train

$$I_j(t) = f \sum_l \delta(t - s_{jl}) + S \sum_{i=0}^{N-1} \sum_k D_{ji}\delta(t - \tau_{ik}), \tag{1b}$$

where $\delta(\cdot)$ is the Dirac delta function. Once the voltage, v_j, surpasses the firing threshold, V_T, its evolution by Eq. (1) is interrupted. It is reset to the reset voltage, V_R, and then resumes its evolution under Eq. (1) again.

The external input to each neuron, the first term in Eq. (1b), is modeled by an independent Poisson train of current spikes with rate ν. At time $t = s_{jl}$, when the lth of these spikes arrives at the jth neuron, this neuron's voltage instantaneously increases by an amount f. The input current spikes arriving from other neurons in the network make up the second term in Eq. (1b). At time τ_{ik}, when the ith neuron fires its kth spike, the jth neuron's voltage will instantaneously increase by an amount S precisely when $D_{ji} = 1$. The $N \times N$ connectivity matrix, D, encodes the network architecture with $D_{ji} = 1$ if the jth neuron has a directed connection from the ith neuron and $D_{ji} = 0$ otherwise. Due to the instantaneous jumps in voltage at each spike time, and an analytical solution to Eq. (1a) between arriving spike times, we are able to efficiently simulate the system (1) without discretization errors using an event-driven algorithm similar to the one reviewed in Sec. 2.4.1 of [50]. For the purpose of numerical simulations in this paper, we use the values $g_L = 1$, $V_R = 0$ and $V_T = 1$. This nondimentionalization can be derived by suitable scaling of realistic physiological neuronal values [51].

In this work we will assume that D represents a graph with only one *strongly connected component*, meaning that between any pair of nodes in the graph there exists a directed path from one to the other and back again. (A *directed path* is a list of nodes where each node has an outgoing directed edge emanating from it received by the next node in the list. The *path length* is one less than the number of nodes in the list.) If part of the network were disconnected from the remaining network, then it would be impossible for a cascading total firing event to occur regardless of the synaptic coupling strength.

Later, in Sec. IV, we will look specifically at a directed version of the network model with significant clustering discussed in [52]. Clustering (also referred to as *transitivity* [48]) refers to a relative frequency of closed triangles – three nodes in which each has an edge connecting it to the other two . We believe that the clustering is an important network statistic affecting the synchronizability of the of the system (1), and requires developing techniques that go beyond existing "tree-like" approximations that ignore clustering.

III. THEORETICAL CHARACTERIZATION OF SYNCHRONY

As discussed in the introduction, the stochastically-driven model (1) maintains synchrony through cascading total firing events. Between such events the neuronal voltage trajectories separate due to the independent noise driving each neuron, in contrast to those in the deterministic version of system (1). We characterize the propensity of a network to sustain synchrony through cascading total firing events by the probability to see these events in succession. In particular, we define the state of the N neuronal voltages, at the moment of time some neuronal voltage first reaches threshold following a cascading total firing event, to be *cascade-susceptible* if the firing of that neuron induces the entire network to fire, thus giving rise to another such event (we define cascade-susceptible precisely in Sec. III B). There is a well defined random time between these two events which depends solely on the external drive and not the network topology. In Sec. III A we summarize the calculation of the probability density function (pdf) for this random time between pairs of total firing events; the firing rate of the synchronous network is the inverse of its mean. We then proceed in Sec. III B to calculate the probability the network is cascade-susceptible when the first neuron fires since the previous cascading total firing event, using knowledge about the neuronal voltage distribution at the time the first neuron fires and the local network topology.

A. Time between synchronous events

In the introduction we pointed out that between cascading total firing events, the neurons are effectively uncoupled as no neurons are firing. The random time between total firing events in a synchronous system is given by the time $T^{(1)}$ at which the first neuron after a total firing event reaches threshold voltage V_T. This is simply the minimum of the N independent times that each neuron would take to reach threshold voltage V_T, if influenced only by its own external drive. The network topology plays no role in this random time $T^{(1)}$, so its pdf (needed in Sec. III B) will be the same as for the all-to-all network previously considered in [30, 31]. Note that the firing rate

of the synchronous network is given by the statistical average of the inverse of $T^{(1)}$, and is consequently indeed independent of network topology. We summarize next the results from [30, 31] for the pdf of the random time between total firing events in a synchronous network.

The pdf, $p_T^{(1)}(t)$, of the minimum exit time, $T^{(1)}$, of the N neurons is related to the pdf, $p_T(t)$, of a single neuron's exit time through the formula

$$p_T^{(1)}(t) = N p_T(t)\Big(1 - F_T(t)\Big)^{N-1}, \tag{2}$$

where $F_T(t) = \int_0^t p_T(t')\,\mathrm{d}t'$ is the cumulative distribution function (cdf) of the exit time for a single neuron [53]. We compute the single-neuron exit time distribution, $p_T(t)$, using an equation describing the evolution of the function $G(x,t)$, the probability that a neuron's voltage has not yet crossed the threshold given that it started at the voltage x at the time $t = 0$. The cdf for the first exit time, $F_T(t)$, is expressed as $F_T(t) = 1 - G(V_R, t)$, as the neuronal voltage always evolves from the reversal potential after a cascading total firing event. Under a diffusion approximation of the external driving [54], the function $G(x,t)$ satisfies the equation

$$\frac{\partial}{\partial t}G(x,t) = \Big[-g_L(x - V_R) + f\nu\Big]\frac{\partial}{\partial x}G(x,t) + \frac{f^2\nu}{2}\frac{\partial^2}{\partial x^2}G(x,t) \tag{3a}$$

with the boundary conditions

$$\frac{\partial}{\partial x}G(x,t)\Big|_{x=V_R} = 0 \tag{3b}$$

and

$$G(V_T, t) = 0, \tag{3c}$$

and the initial condition

$$G(x, 0) = 1. \tag{3d}$$

The system (3) gives a valid description provided the external spike strength, f, is small, at least $f \ll V_T - V_R$.

We find the cdf $F_T(t) = 1 - G(V_R, t)$ by computing the solution of the parabolic partial differential equation for $G(x,t)$, Eq. (3), using the Crank-Nicolson scheme [55, Sec. 2.6.3]. We find the pdf $p_T(t) = dF_T(t)/dt$ by numerically differentiating the function $F_T(t)$. We use both $F_T(t)$ and $p_T(t)$ to numerically evaluate the desired pdf, $p_T^{(1)}(t)$, for the minimum exit time of all the N voltages via Eq. (2). Alternatively, the pdf $p_T^{(1)}(t)$ can be found exactly from a related Fokker-Planck Equation which describes the evolution of the neuronal voltage pdf in terms of an eigenfunction expansion involving confluent hypergeometric functions, as discussed in [30].

We compare the theoretically obtained firing rate, the inverse of the expected time between cascading total firing events,

$$\left\langle T^{(1)} \right\rangle = \int_0^\infty t p_T^{(1)}(t) dt, \tag{4}$$

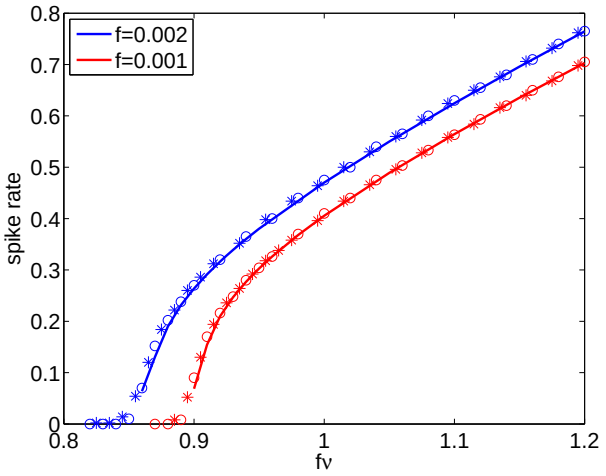

FIG. 1. (color online) Firing rate of a neuronal network in a perfectly synchronous parameter regime as computed as the inverse of Eq. (4) (lines) and from direct numerical simulation of system (1) (symbols) for the directed version of the scale-free network in [52] (and presented in Section IV) with $N = 4000$ neurons, network parameter $m = 50$, (circles) $S = 0.075$ (asterisks) $S = 0.150$ and for the indicated values of input spike strength, f and mean external driving strength $f\nu$.

found by integrating over the pdf in Eq. (2), with direct numerical simulations of the system (1), for the network model to be defined in Section IV, as a function of the mean external driving strength, $f\nu$, in Fig. 1. There is excellent agreement in the superthreshold regime ($f\nu > 1$) and well into the subthreshold regime ($f\nu < 1$). Also, we point out the dependence of the firing rate on the size of the fluctuations. As f, the strength of the external driving spikes, increases, it becomes more likely to find one neuronal voltage further from the mean; this voltage reaches threshold faster and causes a total firing event. Therefore we see an increase in the mean firing rate of the network with an increase of f, with external driving strength $f\nu$ held fixed. The exact scaling of the firing rate in the superthreshold ($f\nu > (V_T - V_R)/g_L$) region, proportional to $\sqrt{f}$ for fixed $f\nu$, was obtained in [31].

Finally, we note for the purposes of the calculation of synchronizability, to which we turn next, that the above analysis for the statistics of the time $T^{(1)}$ until some neuron first crosses threshold after a total firing event applies even if that neuron does not trigger a total firing event (i.e., the network is not actually in a synchronized state.).

B. Synchronizability of a network

In this section, we investigate in which parameter regimes the neuronal network (1) exhibits synchronous behavior through total firing events by calculating the probability that the neuronal voltages are cascade-

susceptible when the first neuron fires after the previous total firing event. We employ the same strategy for this analysis as in [30], but need to take into account that, in contrast to the all-to-all network considered in that work, neurons receive different numbers of spikes depending on their connections in the more complex network we will consider here. The neuron that initiates a cascading firing event instantaneously sends a spike to all its *neighbors*, the neurons connected to it via outgoing edges in the network. For the cascading firing event to continue, at least one of these neighboring neurons must fire, instantaneously sending out another spike to all of its neighbors. In this fashion, the cascading firing event continues until all of the neurons in the network have fired, or the spikes from the neurons that have fired fail to excite any other neuronal voltages to threshold.

We approximate the probability that the cascading firing event includes all neurons by subtracting from unity the probability that the cascade fails after one or two neurons fire. Numerical simulations shown in Fig. 2 give strong evidence that this approximation holds with a high degree of accuracy, at least for the statistical network model we consider in Section IV. Indeed, studies of other IF neuronal network models [32, 56], Watts models [35, 46, 47, 57], and epidemic models [58] often find a markedly bimodal distribution of cascade sizes, i.e., that if the cascade does not terminate after a few neurons have fired in succession, then most or all of the network will be drawn into a "giant" cascade or "big burst." Here, of course, we are making the stronger assumption, supported by the numerical simulations in Fig. 2, that two neurons is an adequate cutoff to account for cascades that fail to entrain the whole network. For the calculation of the probability of cascade failure on a network with complex topology, it is important to keep track of the number of spikes different neurons in the network have received at various stages of the firing cascade. We include the effects of the network topology by including distributions for the numbers of neighbors of the first and second neuron to fire, as well as the number of neurons that will receive two spikes if both neurons fire. In Sec. IV, we derive these distributions for a specific scale-free network.

To calculate the probability, $P(C)$, that the system is cascade-susceptible, we first precisely formulate the notion of cascade-susceptibility discussed previously in the introduction and in the beginning of Sec. III. We define the event C to consist of all arrangements of neuronal voltages at the time when the first neuron fires so that a cascading total firing event ensues. The time is measured from the previous total firing event, giving the condition that all neuronal voltages equal V_R at time zero. We calculate the probability of the event C by first integrating over the conditional probability of the random time $T^{(1)}$ at which the first neuron fires:

$$P(C) = \int_0^\infty P(C \mid T^{(1)} = t) p_T^{(1)}(t) dt. \qquad (5)$$

Here, the pdf for the first exit time of the N neurons,

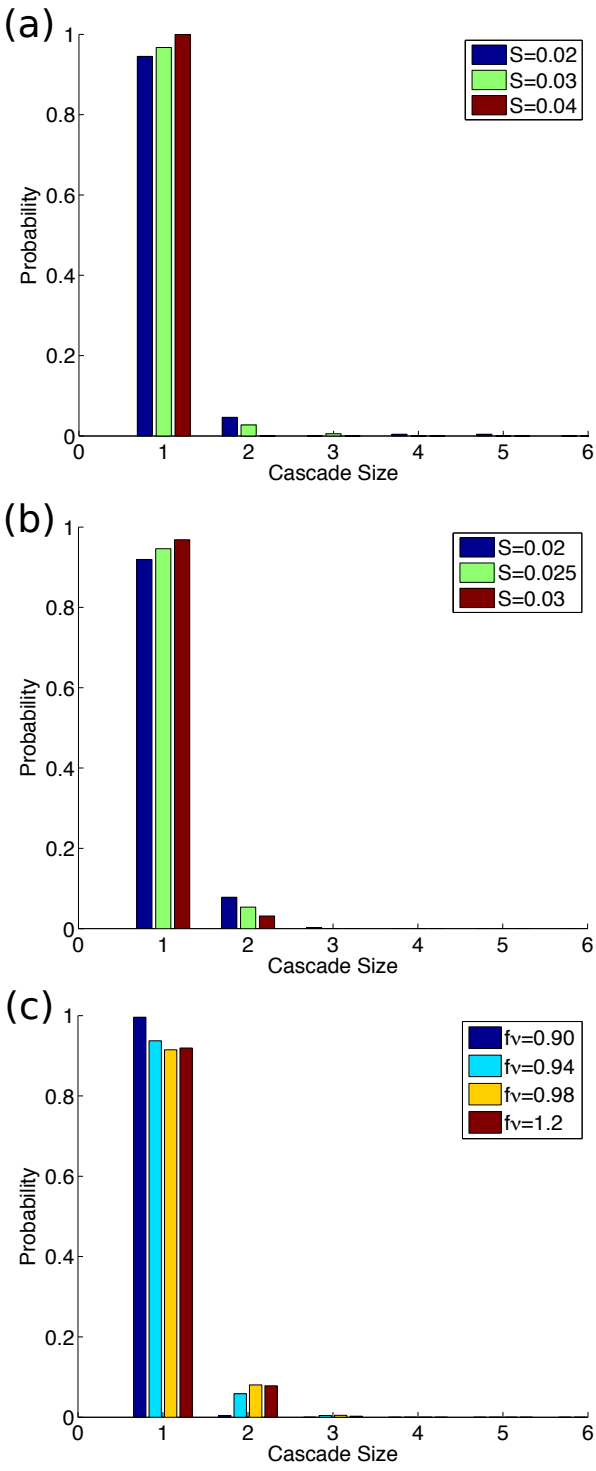

FIG. 2. (color online) Probability distributions of the number of neurons firing in a cascading firing event conditioned on the cascade not including all N neurons, obtained from numerical simulation. All three cases use the particular scale-free network model (described in Section IV) with network parameter $m = 50$, and external spike size $f = 0.001$. The parameters (a) $f\nu = 1.2$, $N = 4000$ (b) $f\nu = 1.2$, $N = 1000$ and (c) $S = 0.02$, $N = 1000$, are just three sets of examples indicating that the probability of a cascade failing after two neurons have fired is negligible.

$p_T^{(1)}(t)$, is given in Eq. (2).

We will now proceed to evaluate the conditional probability in Eq. (5) by subtracting from unity the probability of the complement of the event C, i.e. the probability the cascade fails to include all neurons. The complement, C^C, is divided into the mutually exclusive events A_j, corresponding to exactly j of the N neurons firing in the cascading event. The total probability of cascade failure is the sum over failure at each possible step, and the probability of the cascade succeeding is

$$P(C \mid T^{(1)} = t) = 1 - \sum_{j=1}^{N-1} P_t(A_j), \tag{6}$$

where we use the notation $P_t(\cdot)$ to indicate the probability of an event given the condition $T^{(1)} = t$. We do not compute the probability of the events A_j in general, but instead approximate Eq. (6) with the first two terms. As stated previously, this approximation is empirically justified, at least for the network model to be studied in Sec. IV, in Fig. 2.

The calculation of the failure probabilities, $P_t(A_1)$ and $P_t(A_2)$, depends on the voltages of the neurons and whether or not they themselves fire when receiving spikes. We therefore must first consider the voltage distribution for a typical neuron in the network other than the neuron that initiates the cascading firing event at time $T^{(1)} = t$. Furthermore, as each incoming network spike increases the voltages of the neurons at which it arrives by an amount S, we divide the voltage interval $V_R \le x \le V_T$ into bins of width S starting at V_T, so that the first bin is $V_T - S < x \le V_T$. The probability, $p_k(t)$, for a neuron's voltage to be in the kth bin, given that a different neuron fired at time $T^{(1)} = t$, is described by the formula

$$p_k(t) = \int_{V_T - kS}^{V_T - (k-1)S} \tilde{p}_v(x,t)dx, \tag{7}$$

where $\tilde{p}_v(x,t)$ is the pdf for voltage of a neuron that did not fire up to time t. As described in [30, 31], we approximate the conditioning on $T^{(1)} = t$ by truncating the freely evolving single-neuronal voltage pdf, $p_v(x,t)$, to the interval $V_R \le x \le V_T$ and normalizing it to unit integral:

$$\tilde{p}_v(x,t) = \frac{p_v(x,t)}{\int_{V_R}^{V_T} p_v(x',t)\,dx'}, \qquad V_R \le x \le V_T. \tag{8}$$

For $p_v(x,t)$ we use the Gaussian approximation

$$p_v(x,t) \sim \frac{1}{\sqrt{2\pi}\sigma(t)} \exp\left(-\frac{(x-\mu(t))^2}{2\sigma^2(t)}\right), \tag{9a}$$

with the average voltage

$$\mu(t) = V_R + \frac{f\nu}{g_L}\left(1 - e^{-g_L t}\right) \tag{9b}$$

and the voltage variance

$$\sigma^2(t) = \frac{f^2\nu}{2g_L}\left(1 - e^{-2g_L t}\right), \tag{9c}$$

derived in App. A following the standard technique of computing cumulants from the characteristic function [54] in the limit of small f while $f\nu = O(1)$.

Returning to Eq. (6), the first term, $P_t(A_1)$, is the probability that only the neuron initiating the cascading firing event spikes. We include the network topology by conditioning on the number of outgoing connections (the number of neighbors), K_1, of the first neuron to fire. Then we require that all K_1 neurons receiving a spike from it have voltage lower than $V_T - S$, so that

$$P_t(A_1) = \sum_{k_1=0}^{N-1} \Big[1 - p_1(t)\Big]^{k_1} P_K(k_1), \tag{10}$$

where $p_1(t)$ is given by Eq. (7). Note that the distribution for the number of outgoing connections, $P_K(k)$, is independent of neuronal voltages. The one-term approximation of $P(C)$, from Eq. (5), is

$$P(C) \approx 1 - P(A_1) \approx 1 - \int_0^\infty P_t(A_1) p_T^{(1)}(t)dt \tag{11}$$

where $p_T^{(1)}(t)$ is the pdf of the minimum first exit time, given by Eq. (2).

We now consider $P_t(A_2)$, the probability that exactly two neurons fire in the cascade. As in the previous paragraph, we condition on the number of outgoing connections, K_1, of the first neuron to fire,

$$P_t(A_2) = \sum_{k_1=0}^{N-1} P_t(A_2 \mid K_1 = k_1) P_K(k_1). \tag{12}$$

We note immediately that $P_t(A_2 \mid K_1 = 0) = 0$ because there is no way a second neuron can fire if the initiating neuron has no outgoing connections. We therefore can drop the $k_1 = 0$ term from the sum in Eq. (12), and will implicitly assume $k_1 \ge 1$ in all further considerations. We define F to be the event that exactly one of the K_1 neurons connected to the first neuron fires. Because $A_2 \subseteq F$,

$$P_t(A_2|K_1 = k_1) = P_t(A_2|F, K_1 = k_1) P_t(F|K_1 = k_1). \tag{13}$$

To compute the probability of the event A_2 we also condition upon the second neuron to fire having K_2 outgoing connections,

$$\begin{aligned} P_t(A_2|F,K_1 = k_1) = \sum_{k_2=0}^{N-2} P_t(K_2 = k_2|F, K_1 = k_1) \\ \times P_t(A_2|F, K_2 = k_2, K_1 = k_1). \end{aligned} \tag{14}$$

Note that we have assumed pairs of connected neurons to share only one edge. Were this not the case, we would

have to ignore the edge from the second neuron to fire received by the first neuron to fire as this first neuron cannot fire a second time. The first probability in Eq. (14) is independent of the event F (as well as the first passage time $T^{(1)}$), and is given by $P_{K_2|K_1}(k_2|k_1)$, the distribution of the number of outgoing connections from the second neuron, given that it is connected by an edge directed away from the first neuron, which has k_1 outgoing connections. Now, $P_t(A_2|F, K_2 = k_2, K_1 = k_1)$ is just the probability that no other neurons connected to the second firing neuron themselves fire a spike, conditioned on the numbers of neurons receiving outgoing connections from the first and second firing neurons. Note that the total number of neurons under consideration may not be $k_1 + k_2 - 1$.

In the all-to-all coupled network, all neurons connected to the first neuron would also be connected to the second neuron, and therefore would receive two spikes [30, 31]. For a general network, we condition on the number, L, of neurons that will receive spikes from both neurons that fire, and which we shall call *doubly-excited neurons*:

$$\begin{aligned} P_t(A_2|F, K_2 = k_2, K_1 = k_1) = \\ \sum_{l=0}^{\kappa} P_t(A_2|F, L = l, K_2 = k_2, K_1 = k_1) \\ \times P_t(L = l|F, K_2 = k_2, K_1 = k_1), \end{aligned} \tag{15}$$

where $\kappa = \min(k_1 - 1, k_2)$. The second factor in Eq. (15) is independent of the event F (as well as the first passage time $T^{(1)}$); it is given by $P_{L|K_1,K_2}(l|k_1, k_2)$, the distribution of the number of neurons in the network receiving connections from both the first neuron and the second neuron to fire, given these two neurons are connected and have k_1 and k_2 outgoing connections respectively.

The probability, $P_t(A_2|F, L = l, K_2 = k_2, K_1 = k_1)$, of event A_2, given the first neuron to fire has k_1 outgoing connections, exactly one other neuron fires which has k_2 outgoing connections, and l neurons receive spikes from both neurons that fire, is the probability that the neurons receiving spikes from the second firing neuron do not themselves spike,

$$\begin{aligned} P_t(A_2|F, L = l, K_2 = k_2, K_1 = k_1) \\ = \Big[1 - q_2(t)\Big]^l \Big[1 - p_1(t)\Big]^{k_2 - l}, \end{aligned} \tag{16}$$

where $p_1(t)$ is the probability the neuronal voltage is in the bin $B_1 = [V_T - S, V_T)$, defined in Eq. (7). The quantity $q_2(t)$ is the probability a neuron's voltage does not lie in the bin $B_2 = [V_T - 2S, V_T - S)$, given that it does not lie in the bin $B_1 = [V_T - S, V_T)$, but in this bin's complement, $B_1^C = [V_R, V_T - S)$. (This is known from event F; this neuron did not fire when it received the first spike.) In terms of the probabilities defined in Eq. (7),

$$q_2(t) = P_t(B_2|B_1^C) = \frac{P_t(B_2)}{P_t(B_1^C)} = \frac{p_2(t)}{1 - p_1(t)}.$$

We now proceed to the probability $P_t(F|K_1 = k_1)$ in Eq. (13). This event is that exactly one of the k_1 neurons connected to the first neuron fires due to the incoming spike. This is equivalent to exactly one success out of k_1 Bernoulli trials, each with a success probability of $p_1(t)$:

$$P_t(F|K_1 = k_1) = k_1 p_1(t)\big[1 - p_1(t)\big]^{k_1 - 1}. \tag{17}$$

Combining the derived quantities in Eqs. (13), (14), (15), (16), and (17), we obtain $P_t(A_2)$ from Eq. (12) as

$$P_t(A_2) = \sum_{k_1=1}^{N-1} \sum_{k_2=0}^{N-2} k_1 p_1(t)\Big(1 - p_1(t)\Big)^{k_1 - 1 + k_2} \left[\sum_{l=0}^{\kappa} \left(\frac{1 - p_1(t) - p_2(t)}{\big(1 - p_1(t)\big)^2}\right)^l P_{L|K_1,K_2}(l|k_1, k_2)\right] P_{K_2|K_1}(k_2|k_1) P_K(k_1), \tag{18}$$

where $\kappa = \min(k_1 - 1, k_2)$. Here and throughout this paper we use the notation $P_{X|Y_1,\ldots,Y_n}(x|y_1, \ldots, y_n)$ for the probability $P(X = x|Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ of any random variable X taking the value x given the values of the random variables $\{Y_j\}_{j=1}^n$ are $\{y_j\}_{j=1}^n$.

The approximation for the probability that a network is cascade-susceptible is obtained by approximating the integrand in Eq. (5), i.e., the sum in Eq. (6), by its first two terms:

$$\begin{aligned} P(C) \approx 1 - \big[P(A_1) + P(A_2)\big] = \\ 1 - \int_0^{\infty} \big[P_t(A_1) + P_t(A_2)\big] p_T^{(1)}(t) dt \end{aligned} \tag{19}$$

where $P_t(A_1)$ is given in Eq. (10), $P_t(A_2)$ in Eq. (18) and $p_T^{(1)}(t)$, the pdf of the minimum first exit time, in Eq. (2). What remains to be defined are the network specific distributions, $P_{L|K_1,K_2}(l|k_1, k_2)$, $P_{K_2|K_1}(k_2|k_1)$ and $P_K(k_1)$, which we will develop for a specific scale-free network model next in Sec. IV A. Note that for the all-to-all coupled network with bidirectional edges, we can take $K_1 = N - 1, K_2 = N - 2$ and $L = N - 2$ with probability one, and both Eq. (10) and (18) reduce to the formula presented in [30, 31].

The computation of the term $P_{L|K_1,K_2}$, which is essentially a measure of clustering, can be rather challenging, and we will for comparison consider a "tree-like" approximation in which such clustering effects are ignored altogether, i.e., L is simply assumed to be exactly zero. Such assumptions are made frequently in the literature on net-

work dynamics [46–48], and can often yield surprisingly effective approximations even in clustered networks [45]. Our tree-like approximation for $P_t(A_2)$ reads:

$$P_t^{\text{tree}}(A_2) \equiv \sum_{k_1=1}^{N-1} \sum_{k_2=0}^{N-2} k_1 p_1(t) \Big(1-p_1(t)\Big)^{k_1-1+k_2} \times P_{K_2|K_1}(k_2|k_1) P_K(k_1), \tag{20}$$

where $\kappa = \min(k_1 - 1, k_2)$, in place of $P_t(A_2)$ in Eq. (19).

Before proceeding to the analysis of a specific network model, we note that the formula (18) is surprisingly ambiguous about whether clustering, defined at the end of Sec. II, should generally enhance or suppress the synchronizability of a network. The expression does not directly involve the conventional measures of clustering [48], but we can think of the distribution of the random variable L as characterizing the effects of clustering for our purposes, in that it describes how many neurons form the third edge of an appropriately directed triangle involving the first two nodes to fire. In particular, we would expect a positive association between the mean value of L and the clustering coefficient, which assigns a numerical value to the amount of clustering in the network (see Sec. III B of [48] for various definitions of the clustering coefficient). Note however that our approximations refer to the full probability distribution of L, which should be somewhat more comprehensive than a summary statistic like the clustering coefficient. If the following inequality holds:

$$\rho(t) \equiv \frac{1-p_1(t)-p_2(t)}{(1-p_1(t))^2} < 1, \tag{21}$$

then we would (not quite rigorously) conclude that increasing clustering would, by generally increasing the size of L, decrease $P_t(A_2)$ in Eq. (18) and therefore increase $P(C)$ in Eq. (19). That is, we would say more highly clustered networks would be more synchronizable under our IF model, in agreement with the findings of [35] for the Watts model [57]. Indeed, recalling the definitions (7) of $p_j(t)$ as the probability for a neuron (other than the neuron initiating the potential cascade) to require exactly j input spikes from other neurons to participate in the cascade, we can interpret inequality (21) to say that it is more likely for the first two neurons to excite a third neuron by directing their outputs to a common neuron, rather than to two distinct ones. However, we have not found an argument for why Eq. (21) should generally be true, other than some rough argument that, in most cases, $p_2(t)$ will be substantially larger than $p_1(t)$ because the probability distribution of the other neuronal voltages (8) at the first neuron's firing time is usually unimodal and peaked at more than two spike amplitudes, $2S$, from threshold. And indeed, we do find through numerical simulations that Eq. (21) does seem to hold for the particular model on which we focus, and to which we turn next.

IV. APPLICATION TO A SCALE-FREE NETWORK MODEL

Here we investigate the interplay of the network topology with the synchronizability of the system (1) for the scale-free network with clustering presented in [52]. The reasons for choosing this network are two-fold. First, as shown in [52], this is a highly clustered network (as defined in the last paragraph of Sec. II), which, in its directed variant, provides high likelihood that between each pair of nodes there exists a relatively short directed path from one to the other (as defined in the last paragraph of Sec. II). In terms of modeling neuronal activity, this allows the spikes generated from the firing neurons to quickly traverse the network and cause a cascading total firing event. Second, this network has non-trivial local topology around each node, yet not so complex that we would be unable to obtain relatively accurate descriptions of the probability distributions for the outgoing degree of one node and two connected nodes, as well as the number of nodes receiving outgoing edges from these two connected nodes, which are necessary to evaluate the probability of a cascading total firing event.

We begin our investigation in Sec. IV A by completing the theoretical characterization of the synchronizability of the system (1) from Sec. III B. To this end, we calculate the required distributions, i.e. the number of outgoing connections, $P_K(k)$, the conditional probability distribution for outgoing connections of two connected neurons, $P_{K_2|K_1}(k_2|k_1)$ and the number of neurons to receive two spikes, $P_{L|K_1,K_2}(l|k_1,k_2)$. We do not find the last of these three distributions exactly, but instead compute very accurate approximations which moreover serve as rigorous lower and upper bounds. When used to compute the probability, $P_t(A_2)$, that the cascade fails after exactly two neurons fire in Eq. (18), and then the probability, $P(C)$, that the system is cascade-susceptible in Eq. (5), the bounds result in approximations for $P(C)$. Then, in Sec. IV B, we compare the analytical calculation of $P(C)$ to direct numerical simulations of system (1), describe when the higher-order network statistics are important for the description of synchrony, and discuss the qualitative role of the clustering in the network.

A. Network construction and statistics

We create a realization of the undirected scale-free network we have chosen to use as an example in this paper by growing it in stages according to the algorithm described in [52]. We begin with m all-to-all connected nodes that are deemed *active*. Each stage has two steps: First, a new active node is added to the network and undirected edges are created to connect it with each of the m existing active nodes. Second, one of the $m+1$ active nodes is randomly chosen to be permanently deactivated with probability inversely proportional to its current degree (the total number of edges). The stage ends as it began,

with m active nodes. We repeat this procedure until the network has grown to include N nodes, and then convert it into a directed network by randomly assigning a direction to each edge. We note that "active" is only a term used for the growing of the network and has nothing to do with the numerical simulations of integrate-and-fire neurons.

To compute the expressions in Sec. III B in our various approximations of the cascade-susceptible probability $P(C)$, we need first of all the probability distributions $P_K(k)$ and $P_{K_2|K_1}(k_2|k_1)$ for the outgoing degrees of the first two nodes to fire. Precise formulas are presented and derived in Appendix B, with some results originating in [52] and [33]. Here we will simply state some simpler but useful approximations based on some more accurate asymptotic expressions valid when the active cluster size is large, and the network size is much larger than the active cluster size ($N \gg m \gg 1$). First, we have the following rough approximation for the outgoing degree distribution of a single node based on the more precise asymptotic formula (B4):

$$P_K(k) \approx \begin{cases} \frac{m^2}{2k^3} & \text{if } k \geq m/2, \\ 0 & \text{if } k < m/2. \end{cases} \tag{22}$$

Second, we have the following rough approximation for the conditional outgoing degree distribution for the K_2 outgoing connections of a second node which receives a directed connection from an initially specified node with K_1 outgoing connections, based on the more precise asymptotic formula (B7):

$$P_{K_2|K_1}(k_2|k_1) \approx \begin{cases} \frac{m^2(k_1+k_2-m)}{2k_2^3} & \text{if } k_1, k_2 \geq m/2, \\ 0 & \text{else.} \end{cases} \tag{23}$$

The one term (essentially mean field) approximation can then be written explicitly for our scale-free network model by substituting the approximation (22) (or its more precise expression (B2)) for the single-node outgoing edge distribution $P_K(k)$ into the formula (10) for $P_t(A_1)$, which appears in the one-term expansion (11) for $P(C)$. To proceed to a two-term expansion Eq. (19) for $P(C)$, we need to compute the expression (18) for $P_t(A_2)$. First of all, this requires the conditional distribution of the number of outgoing edges of two connected nodes, $P_{K_2|K_1}(k_2|k_1)$, given as a rough approximation in Eq. (23) (and more precisely by the expression (B5)). A much more challenging term in the expression (18) for $P_t(A_2)$ is the distribution, $P_{L|K_1,K_2}(l|k_1,k_2)$, which describes the probability that two connected neurons with outgoing degrees k_1 and k_2 both have outgoing edges to L common neurons, where k_1 is the outgoing degree of the first neuron to fire, and k_2 is the outgoing degree of the neuron that fires as a result of the first neuron's firing. The value of L determines the number of neurons that will receive two spikes if both these neurons fire. The determination of a precise expression for $P_{L|K_1,K_2}$ appears to involve a rather unwieldy calculation, so we will satisfy ourselves with a range of approximations. The simplest approximation of all is the tree-like approximation (20) in which we simply take $P_{L|K_1,K_2}(l|k_1,k_2) = \delta_{l,0}$, that is, $L \equiv 0$. Another simple approximation would be to take $P_{L|K_1,K_2}(l|k_1,k_2) = \delta_{l,\bar{l}}$, that is, $L \equiv \bar{l}$, a nonzero constant which would crudely characterize the clustering without any measure of its variability or dependence on outgoing degrees. Then the approximation for Eq. (18) would read:

$$P_t^{(L=\bar{l})}(A_2) = P_t^{\text{tree}}(A_2) \left(\frac{1 - p_1(t) - p_2(t)}{\big(1 - p_1(t)\big)^2} \right)^{\bar{l}}, \tag{24}$$

where $P_t^{\text{tree}}(A_2)$ is given in Eq. (20). We will consider the particular choices $\bar{l} = (m-1)/4$ and $\bar{l} = (13m-9)/36$, to be motivated as good estimates below.

Our most detailed estimation of $P_{L|K_1,K_2}(l|k_1,k_2)$ will use the network generation algorithm to define simpler random variables $\tilde{L}$ and $\hat{L}$ which we take as approximate proxies of L, but for which $P_{\tilde{L}|K_1,K_2}(l|k_1,k_2)$ and $P_{\hat{L}|K_1,K_2}(l|k_1,k_2)$ can be explicitly computed. The resulting precise formulas are somewhat cumbersome and thus presented in detail in Appendix C. We display here only a systematic but not quite rigorous approximation based on an asymptotic simplification for the statistics of $\tilde{L}$:

$$\tilde{P}_t(A_2) \lesssim P_t^{\text{tree}}(A_2) \left(\frac{1 - p_1(t) - p_2(t)}{(1 - p_1(t))^2} \right)^{m/4} \exp\left[\frac{3m}{32} \left(\ln \frac{1 - p_1(t) - p_2(t)}{\big(1 - p_1(t)\big)^2} \right)^2 \right] \text{ for } N \gg m \gg 1. \tag{25}$$

where $P_t^{\text{tree}}(A_2)$ is given in Eq. (20). This approximation is formally applicable when $N \gg m \gg 1$, that is, when the size of the active cluster used to generate the network is large, and the size of the network considerably larger than the active cluster. We use the notation $\tilde{P}$ to stress that this expression for the probability of A_2 is an approximation based on using the simpler random variable $\tilde{L}$ in place of L.

Our proxy random variables are in fact not only approximations, but rigorous bounds that satisfy $\tilde{L} \leq L \leq \hat{L}$ (in every realization of the network). As explained in Appendix C 1, the approximations based on these

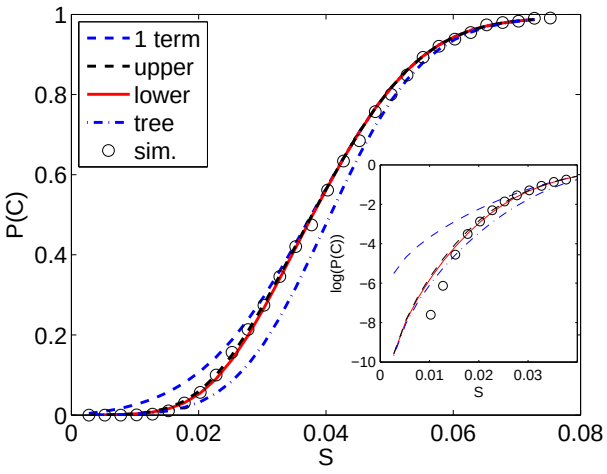

FIG. 3. (color online) The probability, $P(C)$, that the system (1) becomes cascade-susceptible again after a total firing event as a function of the coupling strength, S, for 4 different approximations of $P(C)$ in Eq. (6) using $f = 0.001$, $\nu = 1.2/f$, $m = 50$ and $N = 4,000$. (blue dash — 1 term) Eq. (6) truncated after one term; (blue dot dash — tree) Eq. (6) truncated after two terms with the network assumed to have a tree-like structure ($L = 0$ with probability 1); (dashed black — upper) an upper bound on the two term approximation by using the distribution for $\hat{L}$, the upper bound on L, Eq. (C4); (solid red — lower) a lower bound on the two term approximation by using the distribution for $\tilde{L}$, the lower bound on L, Eq. (C3); (black circles — sim.) results from numerical simulations averaged over 10 network realizations. Note that the upper (dashed black) and lower (solid red) bounds almost coincide with each other, and they are in excellent agreement with the numerical results. (inset) The same data plotted on a semi-log plot.

proxy random variables therefore produce explicit upper and lower bounds on $P_t(A_2)$, and therefore, respectively, lower and upper bounds for the two-term approximation (19) for $P(C)$. This is why we have represented the approximation in Eq. (25) using the notation $\lesssim$.

B. Results

Our theoretical characterization of synchronizability is completed by combining the network specific distributions derived in Sec. IV A with the theoretical calculation in Sec. III B of the probability, $P(C)$, that the system (1) returns to a cascade-susceptible state after a total firing event. Comparison of this theoretical characterization with the results from numerical simulations allows us to discuss the validity of our approximations, and to investigate how the network synchrony depends on the model parameters: f, the external input strength, S, the synaptic input strength, and $f\nu$, the mean external drive strength, as well as the network parameters: N, the number of nodes and m, the parameter that controls the total number of edges in the network. The numerically simulated values for $P(C)$ are obtained by repeatedly starting all neurons at reset voltage, which is the state after a previous cascading total firing event, and simulating the network dynamics until the first neuron fires; the probability to be cascade-susceptible is the fraction of the total number of simulations represented by those that lead to the firing of all N neurons in the network at that time.

We investigate the approximations in the theoretical calculation of the cascade-susceptible probability $P(C)$ by plotting in Fig. 3 their predictions along with the results of numerical simulations of the system (1). The one-term approximation, which assumes that either only one neuron fires or all neurons fire in a cascade, $P(C) \approx 1-P(A_1)$ defined in Eq. (11), is determined by the probability of cascade failure after the first neuron fires and the outgoing degree distribution of a single node, $P_K(k)$. More sophisticated approximations can be obtained by including the second term, $P(A_2)$, the probability that the cascade fails after exactly two neurons fire, as in Eq. (19). The tree-like approximation, using $P_t^{\text{tree}}(A_2)$ (Eq. (20)) in place of $P_t(A_2)$, just involves the distribution relating the outgoing degrees of two connected nodes, $P_{K_2|K_1}(k_2|k_1)$. More systematic two-term approximations can be taken from replacing the distribution of L with that of its lower bounding random variable $\tilde{L}$ from Eq. (C3), or of its upper bounding random variable $\hat{L}$ from Eq. (C4). These latter approximations, based on these bounding random variables, are based on the statistical rules for generating the network, and are not simply obtained from joint statistics of two nodes. That is, the form of these upper and lower bounds, developed in Appendix C, would have to be rederived for different statistical network models. We will also explore an intermediate approximation $P_t^{(L=\bar{l})}(A_2)$ (Eq. (24)) for $P_t(A_2)$ in which L is replaced by a constant $\bar{l}$. We will consider the two particular values $\bar{l} = (m-1)/4$ and $\bar{l} = (13m-9)/36$ which correspond to the means of the lower and upper bounding random variables, $\tilde{L}$ and $\hat{L}$, respectively, when $N \gg m \gg 1$ (see Appendix C).

We see from Fig. 3 that the one-term approximation (dashed blue line) gives excellent agreement with the simulations for sufficiently large coupling strengths, but deviates significantly in characterizing less synchronizable networks ($P(C) \lesssim 0.3$). The overall quality of the one-term approximation reflects the idea that the mean of the degree distribution of the nodes is the most important determinant of synchrony, which was emphasized by [59] for a different (Hindmarsh-Rose) neuronal model on k-regular networks. The tree-like two-term approximation (dot-dashed blue line) actually gives a worse approximation than the one-term approximation for all but quite small coupling strengths. We remark, though, that the tree-like approximation may still be considered somewhat "unreasonably effective," despite the highly clustered character of our statistical network model, for reasons expounded in [45]. The lower (red solid line) and

upper (black dashed line) bounds for the two-term approximation not only serve as bounds, as they should, but also as accurate approximations and substantial improvements to the one-term approximation for networks with low synchronizability ($10^{-4} \lesssim P(C) \lesssim 0.3$). Moreover, in Fig. 4, we see that the simplified asymptotic approximation (25) for the lower bound is excellent for all $S < 0.06$, though breaks down for $S \sim 0.06$, presumably because $1-\rho(t)$ does not remain suitably small, as implicitly assumed in the approximation (see the discussion at the end of Appendix D). The simple approximation (24) obtained by treating the number of doubly-excited neurons L as a constant (given by the mean of the corresponding bound on L) are seen to work remarkably well, though their relative error deteriorates considerably more rapidly with coupling strength than the more systematic asymptotic approximation (25). Figure 4 shows, therefore, that the plots of the upper and lower bounds in Fig. 3, which refer to the precise but complicated expressions (C3) and (C4), can be very well approximated by using the much simpler heuristic expression (24), and the lower bound can be even better approximated by the explicit expression (25) obtained from a systematic asymptotic calculation.

The good performance of the simple approximation (24), applied pointwise for all possible times $t > 0$ of the firing of the first neuron to the approximation of the term $P_t(A_2)$ in Eq. (18), raises a question of whether a similar scaling approximation could work for the time-integrated quantity, namely:

$$P^{(L=\bar{l})}(A_2) = P^{\text{tree}}(A_2)\big[F(S,f,\nu)\big]^{\bar{l}} \qquad (26)$$

where $P^{\text{tree}}(A_2)$ is the value obtained by substituting the tree-like ($L \equiv 0$) approximation $P_t^{\text{tree}}(A_2)$ in place of $P_t(A_2)$ in Eq. (19), $\bar{l}$ is some constant exponent, and $F(S,f,\nu)$ is some suitable scaling factor. Because the effect of clustering in the two-term approximation (19) of the cascade-susceptible probability, $P(C)$, is contained entirely in the term $P(A_2)$, $P^{\text{tree}}(A_2)$ neglects clustering all together, and the scaling factor $F(S,f,\nu)$ is a function only of neuronal parameters but not the network, the hypothesis (26) implies the effects of clustering can be captured entirely by some single summary statistic $\bar{l}$. Because of the way in which the random variable L appears in the expression for $P_t(A_2)$ in Eq. (18), we would expect that the best choice for $\bar{l}$ would be the mean of the random variable L. As we do not know how to compute this directly, we will instead, in analogy with our studies of the local-in-time approximations Eq. (24), consider the values $\bar{l} = (m-1)/4$ and $\bar{l} = (13m-9)/36$, the means of the lower and upper bounding random variables for L (derived in Appendix C). We stress that even if the approximation (24) were valid for all t, the relation (26) does not necessarily follow because the time integrals don't commute with exponentiation, unless the integrand in Eq. (19) were tightly concentrated near a single point in time.

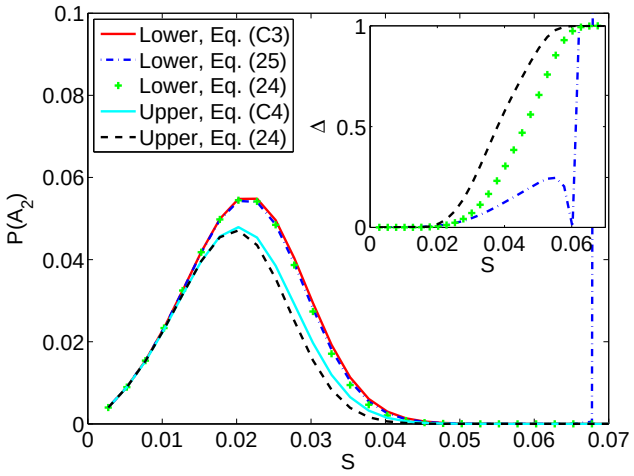

FIG. 4. (color online) Comparison of approximate expressions for $P(A_2)$, the probability for a cascade to terminate with the second neuron. We display the following expressions involving the lower bounding random variable $\tilde{L}$: (solid red) precise evaluation (Eq. (C3)); (blue dot-dash) asymptotic approximation (Eq. (25)); (green stars) simple heuristic approximation (Eq. (24)) using the value $\bar{l} = (m-1)/4$ (mean of lower bound $\tilde{L}$). We display the following expressions involving the upper bounding random variable $\hat{L}$: (cyan solid) precise evaluation (Eq. (C4)); (black dash) simple heuristic approximation (Eq. (24)) using the values $\bar{l} = (13m-9)/36$ (approximate mean of upper bound $\hat{L}$) Note that, as explained in Appendix C 1, the lower (respectively upper) bound $\tilde{L}$ ($\hat{L}$) gives rise to an upper (lower) bound on $P(A_2)$ and consequently a lower (upper) bound on $P(C)$ via Eq. (19). The inset Δ displays the relative errors to the corresponding precise expressions (Eq. (C3) or (C4)), with the same colors and linestyles as in the main plot.

To explore the validity of the hypothesized simple correction Eq. (26) to a tree-like approximation, we plot in Fig. 5 the quantity

$$\left(\frac{P(A_2)}{P^{\text{tree}}(A_2)}\right)^{1/\bar{l}} \qquad (27)$$

for the two values $\bar{l} = (m-1)/4$ and $\bar{l} = (13m-9)/36$, for three different choices of network parameters. $P(A_2)$ is computed by integrating over $P_t(A_2)$ in Eq. (18) using, respectively, the lower and upper bound distributions (C3) and (C4), whose means (for $N \gg m \gg 1$) are $(m-1)/4$ and $(13m-9)/36$; we recall from Fig. 3 that these theoretical expressions give very good representations of the simulated values of $P(A_2)$. If the hypothesis (26) were true, then the plots in Fig. 5 for different network parameters (but the same neuronal parameters f, S, and ν) should fall on top of each other on a curve, which would then represent the function $F(f,S,\nu)$.

For small values of the coupling strength, S, the scaling approximation Eq. (26) gives an excellent description of

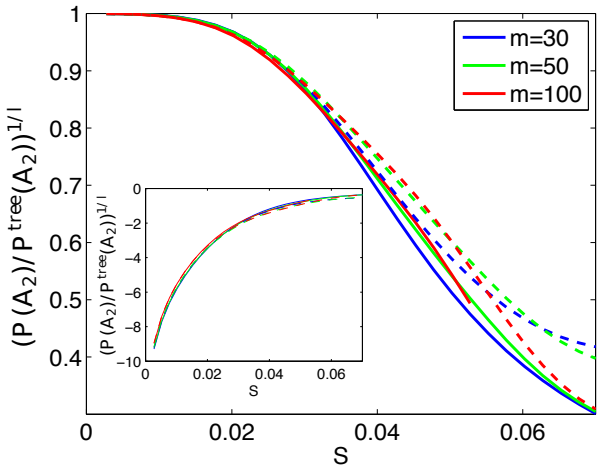

FIG. 5. (color online) The quantity $(P(A_2)/P^{\rm tree}(A_2))^{1/\bar{l}}$ from Eq. (27) plotted for the indicated values of m and $f = 0.001$, $f\nu = 1.2$ and $N = 4,000$. (solid lines) The lower bound distribution for L, Eq. (C3), or the (dashed lines) upper bound distribution for L, Eq. (C4), is used for both the computation of $P_t(A_2)$ in Eq. (18) and for choosing the value of $\bar{l}$ as the respective means, $(m-1)/4$ and $(13m-9)/36$, of these distributions. (inset) The same data plotted as $1-(P(A_2)/P^{\rm tree}(A_2))^{1/\bar{l}}$ with a logarithmic scale.

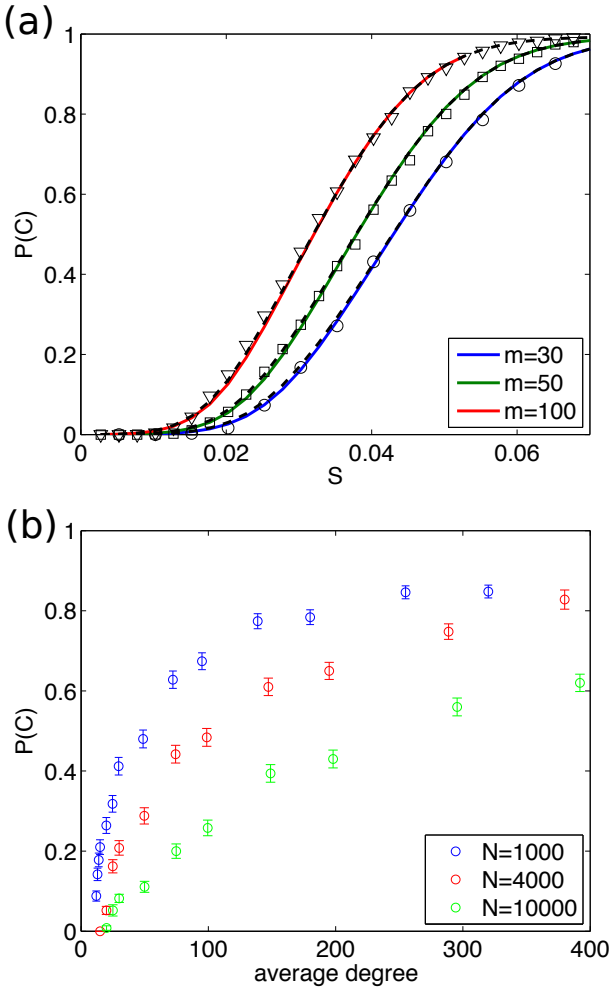

FIG. 6. (color online) (a) The probability, $P(C)$, that the system (1) is again cascade-susceptible after a total firing event, as a function of the coupling strength, S, for the three indicated values of network parameter m, $f = 0.001$, $\nu = 1.2/f$, and $N = 4,000$. Analytical computation from Sec. III B (solid lines) using the lower bound on L, Eq. (C3), and (dashed lines) the upper bound on L, Eq. (C4), are compared to (symbols) results from direct numerical simulations. Note that the solid and dashed lines representing the lower and upper bound calculations almost coincide. (b) The probability, $P(C)$, that the system (1) is cascade-susceptible after a total firing event, as a function of average degree of a node, $\frac{m(m-1)}{2N} + \frac{(N-m)m}{N}$, for the indicated values of network size N, $f = 0.001$, $f\nu = 1.2$, and $S = 0.0301$, averaged over 500 Monte Carlo simulations and 20 networks.

the dependence on clustering, as can be most clearly seen form the logarithmic plot in the inset. For larger values of S, we see some divergence of the plots across networks, more so for the value $\bar{l} = (13m-9)/36$ corresponding to the upper bound bounding random variable for L. The value $\bar{l} = (m-1)/4$, corresponding to the mean of the lower bounding random variable for L , however, does give quite good collapse and suggests that the simple representation Eq. (26) for clustering may work well with $\bar{l} = (m-1)/4$.

We also investigate the effect of the network parameter m on $P(C)$, the probability to return to a cascade-susceptible state after a total firing event. Note that m governs the mean degree of the nodes in the network, but changing it also affects other statistical network quantities. The leading order effect of increasing m amounts to increasing, on average, the degree of the first neuron to fire, leading to smaller probabilities of cascade failure. This statement is confirmed over a broad range of synaptic coupling strength, S, in Fig. 6 (top). Nonetheless, there is still a significant higher-order effect. If we compare scale-free networks with the same average degree, as in Fig. 6 (bottom), a noticeable dependence on network size – another network property – still remains.

The synchronizability of the system (1) has a direct dependence on the size of the external input fluctuations. While keeping the mean external input, $f\nu$, constant, we plot in Fig. 7(a) the probability, $P(C)$, that the systems is cascade-susceptible as a function of coupling strength, S, for four different value of external input spike strength, f. Decreasing f (while keeping $f\nu$, constant) allows the system to maintain a higher level of synchrony at a lower value of coupling strength, S. This is partially explained by the fact that smaller fluctuations in the external input naturally narrow the distribution of neuronal voltages, as the voltage variance in Eq. (9c) scales like f while keeping $f\nu$ constant. But our characterization of synchronizability depends on the voltage distribution at

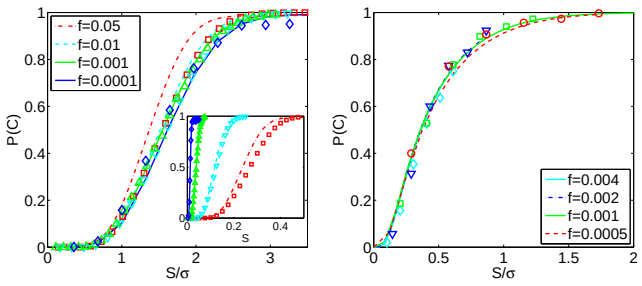

FIG. 7. (color online) The probability, $P(C)$, the system (1) is again cascade-susceptible after a total firing event, as a function of the coupling strength, S, scaled by the magnitude of the standard deviation of a single neuronal voltage, $\sigma \sim f\sqrt{\nu/2}$, which removes much of the dependence on the external spike strength, f. (solid lines) Analytical expressions using the lower bound distribution (C3) for L to compute $P_t(A_2)$ in (18) are compared to results from direct numerical simulations (symbols). (left) The scale-free network model with $N = 4,000$, $m = 50$ and $\nu = 1.2/f$ and (inset) same data without the scaling. (right) Reproduction of the data for an all-to-all network model explored in [31] with $N = 100$ and $\nu = 1.2/f$.

the first exit time, $T^{(1)}$; for which the pdf in Eq. (2) depends separately (through F_T in Eq. (3a)) on $f\nu$ and f. As seen in Fig. 1 in Sec. III A, decreasing the external input spike strength, f (for fixed $f\nu$), increases the time between total firing events and therefore the variability of the neuronal voltages.

We find empirically that we can account for these various effects of the external input spike strength f by simply rescaling the coupling strength by the magnitude, $f\sqrt{\nu/2}$, of the standard deviation of a single neuronal voltage (see Eq. (9c)), as seen in Fig. 7(b). That is, the synchronizability of the network seems to depend primarily on the ratio of the coupling strength S to the standard deviation of the neuronal voltages. The deviations of the results of direct numerical simulations when $f = 0.05$ and 0.1 from the theoretical expressions can be readily understood from the known deterioration of the diffusion approximation behind Eq. (3), which defines the first exit time pdf, and the Gaussian approximation of the voltage distribution, Eq. (9) as f increases. A rather systematic exploration of the dependence of the synchronizability (characterized differently than here) on the governing parameters of another scale-free network model with discrete IF dynamics can be found in [32]. They found it difficult to achieve data collapse with respect to primitive quantities, such as our combination $S/(f\sqrt{\nu/2})$, but their discrete model had no simple way to characterize the standard deviation of voltages of neurons at the time of cascade initiation. The work [32] also finds substantial variability, across realizations of scale-free networks with the same parameter values, of the synchronizability of the network under their dynamical model.

V. CONCLUSIONS

Neurons exhibit synchronous firing not only in a number of model networks [16–21], but in experimental measurements as well [60–64]. Their underlying mechanisms and possible functions may not be fully understood, but there is certainly a link between the network representing the neuronal connections and the appearance of synchronous firing. In fact, some network topologies may enhance a network's ability to synchronize [65], or increase the speed at which it is attracted to the synchronous state [10]. In this work, we demonstrate a means for characterizing the ability of a stochastically-driven, pulse-coupled, current-based integrate-and-fire model to maintain a synchronous firing state, and present an analytical calculation for the probability to see repeated cascading total firing events based on a statistical representation of the local network topology around a single firing neuron.

Although this definition of synchrony is rather stringent, it offers a way to quantitatively compare the synchronous states in a probabilistic setting across different networks, and its calculation involves techniques which may prove useful to more general studies of dynamics on networks. This framework for describing synchrony was employed in [66] for all-to-all networks of both inhibitory and excitatory neurons, leading to a reduced description similar to a firing rate model, but with the addition of a stochastic jump process to account for small synchronous bursts [67]. While the relaxation of instantaneous synapses to more realistic time courses does not allow for instantaneous bursts, we expect the bursts to be spread over a characteristic time-window width, leaving the underlying mechanism of competition between noise and excitation; see for example, Sec. VC in [31].

The calculation of how a global property of a coupled system depends upon its network topology is a broad question being pursued by many researchers other than neuroscientists. For example, those interested in the spread of epidemics [4–9, 68–70] are also investigating techniques for representing dynamical disease spread over a network. One important question is how much information about the network is needed to accurately describe the global property. In epidemic modeling, accurate predictions for the infected fraction of the population was carried out by writing differential equations for interacting nodes, and using a moment-closure for interacting triples of nodes [4, 5]. Another useful technique for dealing with random graphs is to assume the network has a tree-like structure, and apply results from branching process and bond-percolation theory [71, 72], which allows for the calculation of epidemic size [6–8]. In this work, we found that a tree-like approximation significantly underestimates the network's ability to maintain a synchronous firing state (see for example Fig 3). Such a tree-like approximation was also abandoned for this scale-free network in [52] and instead the epidemic threshold was calculated in terms of the average degree of the neighbors of the largest degree nodes (a.k.a. hubs) in the net-

work. Our approach also differs conceptually from pair approximation procedures, such as those reviewed in [68] and [73].

The techniques in this paper, accounting explicitly for the statistical network topology including clustering effects, may apply to studying other coupled dynamical systems on networks. For example, numerical simulation studies in [35] of "information cascades" (the Watts model [57]) on a different random network model found that global cascades occur more easily as the network becomes more clustered, and our results agree with this conclusion for our network dynamics as well. Some recent theories [74, 75] have extended tree-like approximations to account for triangles, but these are generally applied to special random configuration models [74] for which triangles are introduced explicitly in the generation of the network, and therefore relatively easy to account for. Here we have computed clustering effects for a random directed scale-free network model [52], which to be sure is idealized, but for which the statistics of appropriately directed triangles required a nontrivial calculation (App. C). These calculations have been shown to substantially improve upon the accuracy of tree-like or mean field approximations and give excellent agreement with the direct numerical simulations.

ACKNOWLEDGMENTS

K.A.N. was supported by an NSF Graduate Research Fellowship and by NSF grant DMS-0636358. M.S. was partly supported by NSF grant DMS-0636358. P.R.K. was partly supported by NSF grant DMS-0449717. G.K. was supported by NSF grant DMS-0506287. D.C. was partially supported by NSF DMS-1009575, Shanghai No. 14JC1403800 and the NYU Abu Dhabi Institute No. G1301.

Appendix A: Voltage pdf for freely evolving neuron

In this appendix, we derive the Gaussian approximation, Eq. (9), of the pdf for a typical freely evolving neuronal voltage. By freely evolving we mean that it is not reset to V_R when it reaches V_T and it is not subject to any injected current spikes from other neurons in the network. If at $t = 0$ we set $v_j(0) = V_R$, $j = 1, \ldots, N$ then the solution to Eq. (1) under these conditions is

$$v_j(t) = V_R + f \sum_{l=1}^{M_j(t)} e^{-g_L(t-s_{jl})}. \tag{A1}$$

The number $M_j(t)$ of the external spikes arriving at the jth neuron before the time t is random and Poisson-distributed with mean νt. After the random sum in (A1) is shifted by its mean and scaled by its standard deviation, the cumulants of this rescaled voltage are calculated using successive derivatives of the logarithm of its characteristic function. As $f \to 0$ while $f\nu = O(1)$, cumulants of order three and higher can be neglected. Therefore, in the small-fluctuation regime, when f is small while $f\nu = O(1)$, we can reasonably approximate the distribution of the neuronal voltage, $v_j(t)$, which is not reset to V_R upon reaching V_T, by the Gaussian distribution [54] in Eq. (9). More details concerning this argument can be found in Newhall *et al.* [30, 31].

Appendix B: Statistical Distributions of Network Model

We collect here various formulas for probability distributions which are used in approximation formulas in the main text and Appendix C. We present in turn, the needed statistical distributions for the properties of a single node (Appendix B 1) and two nodes (Appendix B 2). Typically these distributions, which involve reference to as many as three random variables, will be composed of distributions involving fewer random variables. The fundamental building blocks of our calculations are the prior results from Klemm and Eguíluz [52] and Shkarayev *et al.* [33] which are derived under the assumption $N \gg m \gg 1$, i.e., that the network size is large relative to the size of the active cluster which generated the network, which is itself large. All our numerical examples work at least plausibly in such a regime. The derivation of the probability distributions presented in Appendices B 1 and B 2 are collected in Appendix B 3. We will also report some asymptotic simplifications formaly valid for $N \gg m \gg 1$; these are derived in Shkarayev *et al.* [33] as well as within Appendix D.

1. Single-node Distribution

In order to obtain the distribution of a node's outgoing degree, $P_K(k)$, we recall that the direction of each undirected edge is assigned randomly, i.e. for a given node, the probability that an undirected adjoining edge becomes an outgoing edge is 1/2. Therefore, the probability a node with degree E (the sum of its outgoing degree and its incoming degree) becomes a node with outgoing degree K when direction is assigned is given by the binomial distribution:

$$P_{K|E}(k|\epsilon) = \binom{\epsilon}{k} \left(\frac{1}{2}\right)^{\epsilon}, \tag{B1}$$

when $\epsilon \geq k$, and vanishes otherwise. We determine the distribution $P_K(k)$, by considering all possible values for the undirected degree, E, that could lead to a node with outgoing degree K:

$$P_K(k) = \sum_{\epsilon=\hat{m}}^{N-1} P_{K|E}(k|\epsilon) P_E(\epsilon) \tag{B2}$$

where $\hat{m} = \max(m, k)$. The analytical expression for the distribution of a node's undirected degree, $P_E(\epsilon)$, in the limit of large N and m, was found in [52] to be

$$P_E(\epsilon) \sim \begin{cases} \frac{2m^2}{\epsilon^3} & \text{for } \epsilon \geq m, \\ 0 & \text{else,} \end{cases} \text{ for } N \gg m \gg 1. \quad \text{(B3)}$$

We can also obtain the following simplified partial description of the outgoing edge distribution using the derivation in Shkarayev *et al.* [33]:

$$P_K(k) \sim \begin{cases} \mathrm{TST}(m^{-1}) & \text{if } \frac{k-m/2}{\sqrt{m}} \ll -1, \\ \frac{m^2}{2k^3} & \text{if } \frac{k-m/2}{\sqrt{m}} \gg 1. \end{cases} \text{ for } N \gg m \gg 1. \quad \text{(B4)}$$

$\mathrm{TST}(m^{-1})$ refers to a quantity transcendentally small with respect to the small parameter m^{-1} (and which we neglect in practice).

2. Two-node Distribution

We proceed next to joint network statistics of two nodes, where node 2 receives an outgoing edge from node 1. We stress that in this context we pick pairs of connected nodes in the manner neurons fire in cascading firing events, as opposed to selecting an edge at random. That is, we select node 1 at random (as node 1 represents the first neuron to fire), and then select node 2 randomly from one of the nodes that are at the other end of one of the K_1 outgoing edges from node 1. We are always implicitly assuming $K_1 \geq 1$ when we refer to a two-node distribution, otherwise node 2 would be undefined.

We begin with the conditional distribution $P_{K_2|K_1}(k_2|k_1)$ that node 2 has k_2 outgoing edges, given that node 1 has k_1 outgoing edges. We determine $P_{K_2|K_1}(k_2|k_1)$ by considering all possible combinations of E_1 and E_2, the numbers of undirected edges, that could lead to the just-described arrangement of outgoing directed edges:

$$P_{K_2|K_1}(k_2|k_1) = \sum_{\epsilon_1=\hat{m}_1}^{N-1} \sum_{\epsilon_2=\hat{m}_2}^{N-1} \binom{\epsilon_2 - 1}{k_2} \binom{\epsilon_1}{k_1} \frac{P_{E_2|E_1}(\epsilon_2|\epsilon_1) P_E(\epsilon_1)}{2^{\epsilon_1+\epsilon_2-1} P_K(k_1)}, \quad \text{(B5)}$$

where $\hat{m}_1 = \max(m, k_1)$ and $\hat{m}_2 = \max(m, k_2)$. A detailed derivation of Eq. (B5) is presented in Appendix B 3 b. The conditional distribution, $P_{E_2|E_1}(\epsilon_2|\epsilon_1)$, of the number of undirected edges of node 2, E_2, given node 2 is connected to node 1 which has E_1 undirected edges,

$$P_{E_2|E_1}(\epsilon_2|\epsilon_1) = \epsilon_1^{-1} P_E(\epsilon_2)(\epsilon_1 + \epsilon_2 - 2m), \text{ for } N \gg m \gg 1. \quad \text{(B6)}$$

is derived in Appendix B 3 a. We can also state a simplified partial description of the conditional outgoing degree distribution by substituting Eq. (B6) and (B3) into Eq. (D3), derived in Appendix D:

$$P_{K_2|K_1}(k_2|k_1) \sim \begin{cases} \mathrm{TST}(m^{-1}) & \text{if } \{\frac{k_i - m/2}{\sqrt{m}}\}_{i=1,2} \ll -1, \\ \frac{m^2(k_1+k_2-m)}{2k_2^3} & \text{if } \{\frac{k_i - m/2}{\sqrt{m}}\}_{i=1,2} \gg 1. \end{cases} \text{ for } N \gg m \gg 1. \quad \text{(B7)}$$

We also will need to modify the conditional probability distribution (B1) for the number of outgoing edges, given the total number of incident edges, when referring to node 2, because it is known by definition to have an incoming edge (from node 1):

$$P_{K_2|E_2}(k_2|\epsilon_2) = \binom{\epsilon_2 - 1}{k_2} \left(\frac{1}{2}\right)^{\epsilon_2 - 1} \quad \text{(B8)}$$

Eq. (B1) does apply to node 1 ($P_{K_1|E_1}(k_1|\epsilon_1) = P_{K|E}(k_1|\epsilon_1)$) since it is indeed selected completely randomly from the network.

Finally, the formulas for the approximations (C3) and (C4) to $P_{L|K_1,K_2}$ using the lower bounding random variable $\tilde{L}$ and upper bounding random variable $\hat{L}$ make reference to certain conditional probability distributions involving combinations of three random variables associated to the two nodes. Here we present the formulas relating these somewhat more complex probability distributions to the more elementary conditional probability distributions involving one or two random variables listed above:

$$P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2) = \frac{P_{K_2|E_2}(k_2|\epsilon_2)}{P_{K_2|K_1}(k_2|k_1) P_K(k_1)} \times \sum_{\epsilon_1=\hat{m}_1}^{N-1} P_{K|E}(k_1|\epsilon_1) P_{E_2|E_1}(\epsilon_2|\epsilon_1) P_E(\epsilon_1), \quad \text{(B9)}$$

where $\hat{m}_1 = \max(m, k_1)$.

$$P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2) = \frac{P_{K|E}(k_1|\epsilon_1) P_E(\epsilon_1)}{P_{K_2|K_1}(k_2|k_1) P_K(k_1)} \times \sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{K_2|E_2}(k_2|\epsilon_2) P_{E_2|E_1}(\epsilon_2|\epsilon_1), \quad \text{(B10)}$$

where $\hat{m}_2 = \max(m, k_2)$.

$$P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1) = \frac{P_{K|E}(k_1|\epsilon_1) P_{E_2|E_1}(\epsilon_2|\epsilon_1) P_E(\epsilon_1)}{\sum_{\epsilon=\hat{m}_1}^{N-1} P_{K|E}(k_1|\epsilon) P_{E_2|E_1}(\epsilon_2|\epsilon) P_E(\epsilon)}. \quad \text{(B11)}$$

3. Derivation of Key Multi-Node Statistical Distributions

We present here the derivations of the formulas listed previously in this appendix.

a. Conditional Distribution for the the Degrees of Two Connected Nodes

For two connected nodes 1 and 2, we derive the conditional distribution, $P_{E_2|E_1}(\epsilon_2|\epsilon_1)$ in Eq. (B6), which encodes the correlation between the degrees, E_1 and E_2, of these two nodes. We begin with the formula from [33] for the "edge-distribution function" $T(x,y)$ describing the probability that a randomly chosen edge with randomly chosen direction would originate in a node of degree x and terminate in a node of degree y:

$$T(x,y) = \begin{cases} P_E(x)P_E(y)\frac{x+y-2m}{2m} & x,y \geq m \\ 0 & \text{else} \end{cases}, \quad \text{(B12)}$$

where $P_E(\cdot)$ is the distribution in Eq. (B3). This formula for $T(x,y)$ was obtained by writing a set of recursive equations for how pairs involving active and inactive nodes gain edges during the network generation process, and solving these equations under the assumption of large network size and large active cluster size, $N \gg m \gg 1$.

We next recall that the (marginal) degree distribution of a randomly chosen node from a randomly chosen edge is given in terms of the usual degree distribution $P_E(\epsilon)$ of a randomly chosen node by

$$\frac{xP_E(x)}{\sum_{\epsilon=m}^{N-1}\epsilon P_E(\epsilon)} \sim \frac{xP_E(x)}{2m} \text{ for } N \gg m \gg 1,$$

because there are x as many edges incident to a node of degree x as there are nodes of degree x. The second expression follows from using the asymptotic form (B3) for the degree distribution $P_E(\epsilon)$. Consequently, we have that the conditional distribution for the degree y of the second node on a randomly chosen edge, given the degree x of the other node on the edge, can be expressed as:

$$\frac{T(x,y)}{xP_E(x)/2m} = P_E(y)\frac{x+y-2m}{x}.$$

But conditioning a random edge on the degree of a specified node is equivalent to choosing a random edge of a node with that degree, so the last expression must be equal to $P_{E_2|E_1}(y|x)$, yielding the result in Eq. (B6).

b. Conditional Distribution for the the Outgoing Degrees of Two Connected Nodes

For two connected nodes, we derive the conditional distribution, $P_{K_2|K_1}(k_2|k_1)$ in Eq. (B5), which encodes the correlation between the outgoing degrees, K_1 and K_2, of these two nodes. We derive this distribution from the joint distribution, $P_{K_2,K_1}(k_2,k_1)$, of the outgoing degrees K_1 and K_2 using the formula

$$P_{K_2|K_1}(k_2|k_1) = \frac{P_{K_2,K_1}(k_2,k_1)}{P_K(k_1)}, \quad \text{(B13)}$$

where $P_K(k_1)$ is computed from Eq. (B2). By summing over all possible numbers, E_1 and E_2, of edges the two nodes had before direction was assigned, we obtain the following equation for the joint distribution:

$$\begin{aligned} P_{K_2,K_1}(k_2,k_1) = &\sum_{\epsilon_1=\hat{m}_1}^{N-1}\sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{E_2|E_1}(\epsilon_2|\epsilon_1) \\ &\times P_{K_2,K_1|E_1,E_2}(k_2,k_1|\epsilon_1,\epsilon_2)P_E(\epsilon_1), \end{aligned} \quad \text{(B14)}$$

where $\hat{m}_1 = \max(m,k_1)$, $\hat{m}_2 = \max(m,k_2)$ and $P_E(\epsilon_1)$ is given in Eq. (B3). As the outgoing degrees, K_1 and K_2, of the connected nodes under consideration are conditionally independent given their undirected degrees, the probability $P_{K_2,K_1|E_1,E_2}(k_2,k_1|\epsilon_1,\epsilon_2)$ can be factored into

$$P_{K_2,K_1|E_1,E_2}(k_2,k_1|\epsilon_1,\epsilon_2) = P_{K_2|E_2}(k_2|\epsilon_2)P_{K|E}(k_1|\epsilon_1), \quad \text{(B15)}$$

where $P_{K_2|E_2}$ is given by the expression (B8). Combining Eq. (B15) and Eq. (B14) with Eqs. (B1), (B6), (B8), and (B13), we obtain Eq. (B5).

c. Conditional distribution for the degree of node 2 given its outgoing degree and the outgoing degree of node 1

We derive in terms of known distributions the distribution $P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2)$ (Eq. (B9)) of the degree, E_2, of node 2, given that node 1 has K_1 outgoing connections, one of which is received by node 2 which has K_2 outgoing connections. We employ Bayes' Law and obtain

$$P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2) = \frac{P_{E_2|K_1}(\epsilon_2|k_1)P_{K_2|E_2}(k_2|\epsilon_2)}{P_{K_2|K_1}(k_2|k_1)}, \quad \text{(B16)}$$

where we have used conditional independence to write $P_{K_2|E_2,K_1}(k_2|\epsilon_2,k_1)$ as $P_{K_2|E_2}(k_2|\epsilon_2)$. We determine the probability $P_{E_2|K_1}(\epsilon_2|k_1)$ by summing over all possible numbers of connections, E_1, node 1 had,

$$P_{E_2|K_1}(\epsilon_2|k_1) = \sum_{\epsilon_1=\hat{m}_1}^{N-1} P_{E_2|E_1,K_1}(\epsilon_2|\epsilon_1,k_1)P_{E|K}(\epsilon_1|k_1), \quad \text{(B17)}$$

where $\hat{m}_1 = \max(m,k_1)$. To determine the second probability, $P_{E|K}(\epsilon_1|k_1)$, in Eq. (B17), we apply Bayes' Law and obtain,

$$P_{E|K}(\epsilon_1|k_1) = \frac{P_E(\epsilon_1)P_{K|E}(k_1|\epsilon_1)}{P_K(k_1)} \quad \text{(B18)}$$

Equation (B17) reduces to

$$P_{E_2|K_1}(\epsilon_2|k_1) = \sum_{\epsilon_1=\hat{m}_1}^{N-1} P_{K|E}(k_1|\epsilon_1)\frac{P_{E_2|E_1}(\epsilon_2|\epsilon_1)P_E(\epsilon_1)}{P_K(k_1)}. \quad \text{(B19)}$$

This equation, when substituted back into Eq. (B16) yields Eq. (B9).

d. Conditional distribution for the degree of node 1 given its outgoing degree and the outgoing degree of node 2

In a similar manner to Appendix B 3 c, we derive in terms of known distributions the distribution $P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2)$ (Eq. (B10)) of the degree, E_1, of node 1, given that node 1 has K_1 outgoing connections, one of which is received by node 2 which has K_2 outgoing connections. We employ Bayes' Law and obtain

$$P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2) = \frac{P_{K_2|E_1}(k_2|\epsilon_1)P_{E|K}(\epsilon_1|k_1)}{P_{K_2|K_1}(k_2|k_1)}, \tag{B20}$$

where we have used conditional independence to write $P_{K_2|E_1,K_1}(k_2|\epsilon_1,k_1)$ as $P_{K_2|E_1}(k_2|\epsilon_1)$. We determine the probability $P_{K_2|E_1}(k_2|\epsilon_1)$ by summing over all possible number of edges, E_2, node 2 has,

$$P_{K_2|E_1}(k_2|\epsilon_1) = \sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{K_2|E_1,E_2}(k_2|\epsilon_1,\epsilon_2)P_{E_2|E_1}(\epsilon_2|\epsilon_1) \tag{B21}$$

where $\hat{m}_2 = \max(m,k_2)$. The first probability, $P_{K_2|E_1,E_2}(k_2|\epsilon_1,\epsilon_2)$, in Eq. (B21) is equivalent to $P_{K_2|E_2}(k_2|\epsilon_2)$ as K_2 is conditionally independent of E_1 once the value of E_2 is known. Substituting Eq. (B21) into Eq. (B20), and using Bayes' law to express

$$P_{E|K}(\epsilon_1|k_1) = \frac{P_{K|E}(k_1|\epsilon_1)P_E(\epsilon_1)}{P_K(k_1)}, \tag{B22}$$

we arrive at the desired Eq. (B10).

e. Conditional distribution for the degree of node 1 given its outgoing degree and the outgoing degree of node 2

We derive in terms of known distributions, the distribution $P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1)$ for the degree, E_1 of node 1 given that it has k_1 outgoing connections, one of which is received by node 2, with degree ϵ_2. Applying Bayes' Law we obtain

$$P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1) = \frac{P_{E|K}(\epsilon_1|k_1)P_{E_2|E_1,K_1}(\epsilon_2|\epsilon_1,k_1)}{P_{E_2|K_1}(\epsilon_2|k_1)}. \tag{B23}$$

The probability $P_{E|K}(\epsilon_1|k_1)$ is given in Eq. (B18) and the probability $P_{E_2|E_1,K_1}(\epsilon_2|\epsilon_1,k_1)$ is independent of K_1 and equals $P_{E_2|E_1}(\epsilon_2|\epsilon_1)$. Together with $P_{E_2|K_1}(\epsilon_2|k_1)$ in Eq. (B19), Eq. (B23) reduces to Eq. (B11).

Appendix C: Derivation of Analytical Expressions for Upper and Lower Bounds for Second Order Approximation of Cascade-Susceptible Probability

We describe here how we obtain computable two-term approximations (19) of the cascade-susceptible probability $P(C)$ through the explicit construction of lower and upper bounding random variables, $\tilde{L} \le L \le \hat{L}$, for the key random variable L representing the number of doubly-excited nodes receiving inputs from the two nodes defined in the discussion in Subsection III B, and which we label as "1" and "2" in the same way. First in Appendix C 1, we outline the general strategy. In Appendix C 2, we define the explicit lower and upper bounding random variables, $\tilde{L}$ and $\hat{L}$, that we will use. The resulting approximations for $P(C)$ are presented in Appendix C 3. These results are derived in precise form through purely probabilistic manipulations in Appendix C 4.

1. Approximation and Bounding Strategy

Our introduction of computable bounding random variables for L not only gives us an approximation of $P(C)$ but also bounds on the two-term approximation (19). This follows from the fact that L only affects the two-term approximation (19) in the term $P_t(A_2)$ in Eq. (18), and this term is monotonic with respect to L if the inequality Eq. (21) is satisfied. As noted in the main text, we will not attempt to derive the conditions under which this inequality holds, but will simply accept it as an empirical observation. Proceeding under this assumption, we note that the square bracket $[\cdot]$ in Eq. (18) can be written as $\mathbb{E}[(\rho(t))^L|K_1=k_1,K_2=k_2]$, a conditional expectation of a function $g(l)=(\rho(t))^l$ of L. Taking Eq. (21) as granted, we have that $g(l)=(\rho(t))^l$ is a monotonically decreasing function of l, so

$$\begin{aligned}
&\mathbb{E}\big[(\rho(t))^{\hat{L}}\big|K_1=k_1,K_2=k_2\big] \\
&\quad\le \mathbb{E}\big[(\rho(t))^{L}\big|K_1=k_1,K_2=k_2\big] \\
&\quad\le \mathbb{E}\big[(\rho(t))^{\tilde{L}}\big|K_1=k_1,K_2=k_2\big]
\end{aligned}$$

In other words, by replacing $P_{L|K_{1,2}}(l|k_1,k_2)$ in Eq. (18) by $P_{\tilde{L}|K_1,K_2}(l|k_1,k_2)$, we obtain an upper bound on $P_t(A_2)$ (and a lower bound on the approximation for $P(C)$ in Eq. (19)). Similarly, by replacing $P_{L|K_{1,2}}(l|k_1,k_2)$ in Eq. (18) by $P_{\hat{L}|K_1,K_2}(l|k_1,k_2)$, we obtain a lower bound on $P_t(A_2)$ (and therefore an upper bound on the approximation for $P(C)$ in Eq. (19)).

If Eq. (21) fails to hold, then we no longer can claim this strategy will lead necessarily to bounds, but we can still study the results as approximations, which presumably have more to do with how well the bounding random variables $\tilde{L}$ and $\hat{L}$ actually approximate L.

2. Definition of Bounding Random Variables

We now proceed to define our lower and upper bound variables, $\tilde{L}$ and $\hat{L}$ respectively, by (possibly) undercounting or overcounting, respectively the set of nodes that are connected to both nodes 1 and 2 by inwardly directed edges. Returning to the construction of the undirected

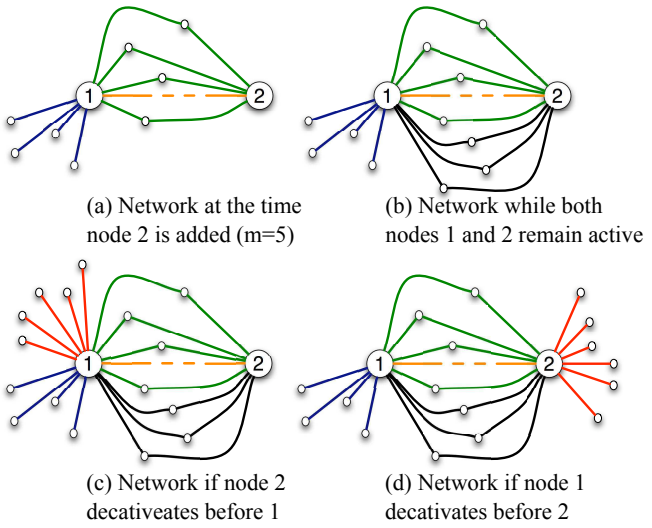

FIG. 8. (color online) Depiction of the undirected network around two connected nodes as it is being grown. Blue: existing connections from node 1 to deactivated nodes at the time node 2 joined the network. Orange dash: the known connection between node 1 and node 2. Green: primal connections to the $m-1$ other active nodes when node 2 joined the network. Black: connections made while both nodes 1 and 2 remain active. Red: connections made to the remaining active node once one of the two nodes has been deactivated.

network (described at the beginning of Sec. (IV A)), we note that at the step the later of these two nodes entered the network, the earlier node was active, and therefore both of these nodes were connected at this step to the other $m-1$ active nodes (Fig. 8a). We call the connections to these $m-1$ nodes *primal*, noting that the edge connecting node 1 to node 2 is not counted as primal. We will define $\tilde{L}$ to be the number of nodes in the network that are connected to nodes 1 and 2 by primal connections that are both inwardly directed toward that node. Since the set of primal connections is a subset of the set of all connections, the result is a lower bound of L in each realization of the network: $\tilde{L} \leq L$.

We define the upper bound random variable $\hat{L}$ in each realization by the following process: 1) determining which node, 1 or 2, has the lesser degree (and calling it node A and the other node B), 2) successively replacing each edge between node A and a non-neighbor of node B with an edge between node A and a randomly chosen neighbor of node B (that does not already have a connection to node A), preserving directionality of the edge with respect to node A, and 3) defining $\hat{L}$ to be the number of nodes in this rewired network that are connected to both nodes 1 and 2 by inwardly directed edges. If nodes 1 and 2 have the same degree, then in step 1, we assign one of them to be node A (and the other node B) with equal probability. Clearly $L \leq \hat{L}$ because the rewiring moves do not disrupt any nodes that were connected, via appropriate directions, to nodes 1 and 2, and only possibly make some new nodes fall into this category.

In Fig. 8 we provide a visualization for these upper and lower bounds in terms of various stages in the generation of the undirected network, taking (without any real loss of generality) node 1 to be the first of the two nodes to enter the network. Panel a) shows a generic picture of the part of the network directly connected to nodes 1 or 2 immediately after the second node enters the network. The orange edge is the connection between nodes 1 and 2, green edges denote the primal connections created at this step; and the blue edges portray the other edges adjacent to node 1 that were generated at previous steps. While nodes 1 and 2 remain active, they are simultaneously connected jointly to further nodes (denoted with black edges in panel b). At some point, either node 1 or node 2 will deactivate. If node 2 deactivates first, then node 1 will continue to accrue new connections (red edges in panel c) until it also deactivates. If on the other hand, node 1 deactivates first, then node 2 will accrue more connections (red edges in panel d) until it also deactivates. If the network generation terminates before either node deactivates, we would simply have the special case of panel c) or d) in which no red edges appear. Now, we can interpret L as the number of nodes connected to both nodes 1 and 2 by either green or black edges, with appropriate directions. Similarly, the lower bound $\tilde{L}$ is the number of nodes connected to both nodes 1 and 2 by green (not black) edges, with appropriate directions. If we have the situation of panel c), so that node 2 plays the role of node A, then in fact the upper bound is precise ($\hat{L} = L$) because no rewiring is done in the construction of $\hat{L}$. In the situation of panel d), some edges of the node of lower degree (which here happen to be the blue edges incident to node 1, playing the role of node A) are redrawn to connect to neighbors of the other node (here those dangling nodes incident to the red edges), so $\hat{L}$ would overcount by the number of those neighboring nodes whose edges in the rewired network were appropriately directed. The fact that the number of rewired edges drawn is determined by the node with smaller degree may give us some *a priori* confidence that the upper bounding random variable $\hat{L}$ may be fairly close to L. This is further bolstered by the observation that the situation in panel c) (with no overcounting at all) will occur more often that the situation in panel d) because the deactivation rate of a node is inversely proportional to its current degree. We will see in fact in Subsection IV B that both the lower and upper bounds based on the random variables $\hat{L}$ and $\tilde{L}$ are sharp in practice.

3. Statistics of Random Variables Bounding the Number L of Doubly-Excited Nodes

For each of the bounding random variables, $\tilde{L}$ and $\hat{L}$ defined in Appendix C 2 we will need their (unconditional) mean as well as their probability distribution,

conditioned upon the number of outgoing edges of each of the two neurons under consideration. Here we present the formulas, and derive them in Appendix C 4. Beyond the one-node and two-node probability distributions presented in Appendix B, the results will repeatedly involve the hypergeometric distribution [53, pg. 68]

$$H(k;n,r,s) = \begin{cases} \dfrac{\binom{r}{k}\binom{n-r}{s-k}}{\binom{n}{s}} & \text{for } 0 \le k \le r, s \le n, \\ 0 & \text{else,} \end{cases} \tag{C1}$$

which describes the probability of choosing exactly k special objects when s objects are chosen without replacement from among n objects, of which r are special objects.

The (unconditional) means of the bounding random variables are:

$$\mathbb{E}\tilde{L} = (m-1)/4, \qquad \mathbb{E}\hat{L} = (13m-9)/36. \tag{C2}$$

The conditional probability distribution for the lower bounding random variable $\tilde{L}$, given the number of outgoing edges of the two neurons under consideration is:

$$\begin{aligned} P_{\tilde{L}|K_1,K_2}(l|k_1,k_2) &= \sum_{g_1=0}^{m-1}\sum_{g_2=0}^{m-1} H(l;m-1,g_1,g_2) \\ &\times \left[\sum_{\epsilon_1=\hat{m}_1}^{N-1} H(g_1;\epsilon_1-1,m-1,k_1-1) P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2)\right] \\ &\times \left[\sum_{\epsilon_2=\hat{m}_2}^{N-1} H(g_2;\epsilon_2-1,m-1,k_2) P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2)\right], \end{aligned} \tag{C3}$$

where $\hat{m}_1 = \max(m,k_1)$ and $\hat{m}_2 = \max(m,k_2)$. This expression becomes directly computable in terms of the specified random network statistics when we use the expressions (B10) and (B9) for $P_{E_1|K_1,K_2}$ and $P_{E_2|K_1,K_2}$, in conjunction with the explicit expressions for $P_{K|E}$ (Eq. (B1)), $P_{K_2|E_2}$ (Eq. (B8)), $P_K(k_1)$ (Eq. (B2)), $P_E(\epsilon_1)$ (Eq. (B3)), $P_{K_2|K_1}(k_2|k_1)$ (Eq. (B5)), and $P_{E_2|E_1}$ (Eq. (B6)). Substituting this result into Eq. (18) gives an explicit, albeit admittedly cumbersome, theoretical lower bound on the two-term approximation for $P(C)$ presented in Eq. (19). We show in Appendix D how the simpler expression (25) can be derived semi-rigorously from Eq. (C3).

The conditional probability distribution for the upper bounding random variable $\tilde{L}$, given the number of outgoing edges of the two neurons under consideration is:

$$\begin{aligned} P_{\hat{L}|K_1,K_2}(l|k_1,k_2) &= \sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2) \\ &\times \sum_{\epsilon_1=\hat{m}_1}^{N-1} H(l;\epsilon_M,k_1-1,k_2) P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1), \end{aligned} \tag{C4}$$

where $\epsilon_M = \max(\epsilon_1,\epsilon_2) - 1$. We obtain an explicitly computable formula for this conditional probability $P_{\hat{L}|K_1,K_2}$ by combining this expression with the conditional probability distributions, $P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2)$, given in Eq. (B9), and $P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1)$, given in Eq. (B11), in conjunction with the explicit expressions for $P_{K|E}$ (Eq. (B1)), $P_{K_2|E_2}$ (Eq. (B8)), $P_{E_2|E_1}$ (Eq. (B6)), and P_E (Eq. (B3)).

Though the upper bound estimate for $P(C)$ resulting from substituting Eq. (C4) into Eq. (18) in place of $P_{L|K_1,K_2}$ involves one fewer summation than the lower bound estimate proceeding from Eq. (C3), we were not successful in obtaining a useful asymptotic simplification parallel to that in Eq. (25). The culprit is the appearance of $\epsilon_M = \max(\epsilon_1,\epsilon_2)$ as an argument deep in the expression (C4), which spoils our procedure in Appendix D relying on successive approximations, for $N \gg m \gg 1$, of the sums by integrals of Gaussian functions over the whole real line.

4. Derivation of Precise Statistical Formulas for Bounding Random Variables

a. Probability distribution for lower bound

We proceed to consider how to compute the distribution, $P_{\tilde{L}|K_1,K_2}(l|k_1,k_2)$ (to be substituted for $P_{L|K_1,K_2}(l|k_1,k_2)$ in Eq. (18) to evaluate a lower bound for $P(C)$), for the lower bound, $\tilde{L}$, on the number of neurons receiving two spikes when both the neurons with k_1 outgoing edges and the neuron with k_2 outgoing edges fire. The random variable $\tilde{L}$ considers only those neurons with primal connections from both node 1 and node 2. With no other information, $\tilde{L}$ would be binomially distributed with $m-1$ trials, each with probability $1/4$, which comes from picking the correct direction for both primal edges emanating from each of the two nodes. From this, we immediately compute that the mean value $\mathbb{E}\tilde{L} = (m-1)/4$.

For the conditional probability distribution, the fact that the information about k_1 and k_2 is given, complicates the situation. For example, if the total number of undirected connections, E_1, is only slightly larger than k_1, then we know that almost all the connections must have been picked to be outgoing, including those that are primal. The conditional probability that any given primal connection is outgoing from node 1 would not simply be $1/2$ in general. To address this issue, we take into account the number of outgoing primal connections, G_1 and G_2, from node 1 and node 2. Note in particular the value of $\tilde{L}$ cannot be larger than either G_1 or G_2.

The number of primal connections, without regard to direction, is $m-1$ from each node. If we knew that there were g_1 and g_2 outgoing primal connections emanating from node 1 and node 2 respectively, then the lower bound, $\tilde{L}$, would satisfy the hypergeometric distri-

bution, $H(l; m-1, g_1, g_2)$, defined in Eq. (C1). This is because $\tilde{L}$ is the number of the $m-1$ nodes, with primal connections to both nodes 1 and 2, that are both among the subset of g_1 nodes receiving outgoing primal connections from node 1 and the subset of g_2 nodes receiving outgoing primal connections from node 2, with both subsets determined randomly and independently.

In order to determine the desired distribution, $P_{\tilde{L}|K_1,K_2}(l|k_1,k_2)$, we apply the law of total probability [53] by summing over all possible values for G_1 and G_2, the numbers of outgoing primal connections,

$$\begin{aligned} &P_{\tilde{L}|K_1,K_2}(l|k_1,k_2) = \\ &\sum_{g_1=0}^{m-1}\sum_{g_2=0}^{m-1} P_{\tilde{L}|K_1,K_2,G_1,G_2}(l|k_1,k_2,g_1,g_2) \\ &\times P_{G_1|K_1,K_2}(g_1|k_1,k_2)P_{G_2|K_1,K_2}(g_2|k_1,k_2). \end{aligned} \tag{C5}$$

Note that G_1 and G_2 are conditionally independent of each other given the numbers of outgoing connections, K_1 and K_2. Moreover, the value of $\tilde{L}$ is conditionally independent of K_1 and K_2 once the values of $G_1 = g_1$ and $G_2 = g_2$ are known, and cannot exceed either g_1 or g_2. Therefore, the distribution $P_{\tilde{L}|K_1,K_2,G_1,G_2}(l|k_1,k_2,g_1,g_2)$ is equivalent to $P_{\tilde{L}|G_1,G_2}(l|g_1,g_2)$, which in turn is $H(l; m-1, g_1, g_2)$ as discussed in the previous paragraph.

If the degree, $E_1 = \epsilon_1$, of node 1 were also known, the number of its outgoing primal connections would obey the hypergeometric distribution $H(g_1; \epsilon_1 - 1, m-1, k_1 - 1)$. This can be understood as G_1 is the number of the $m-1$ primal edges that are selected to be outgoing when $k_1 - 1$ connections (since we do not count the outgoing connection from node 1 to node 2) out of the total $\epsilon_1 - 1$ are selected to be outgoing. Therefore, we evaluate $P_{G_1|K_1,K_2}(g_1|k_1,k_2)$ by considering all possible values for E_1,

$$\begin{aligned} P_{G_1|K_1,K_2}(g_1|k_1,k_2) = &\sum_{\epsilon_1=\hat{m}_1}^{N-1} P_{G_1|E_1,K_1,K_2}(g_1|\epsilon_1,k_1,k_2) \\ &\times P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2), \end{aligned} \tag{C6}$$

where $\hat{m}_1 = \max(m, k_1)$. With this information G_1 is conditionally independent of K_2, and the first distribution $P_{G_1|E_1,K_1,K_2}(g_1|\epsilon_1,k_1,k_2)$ is equivalent to $P_{G_1|E_1,K_1}(g_1|\epsilon_1,k_1)$, which in turn is $H(g_1; \epsilon_1 - 1, m - 1, k_1 - 1)$.

The distribution $P_{G_2|K_1,K_2}(g_2|k_1,k_2)$ in Eq. (C5) is found in an almost identical fashion to $P_{G_1|K_1,K_2}(g_1|k_1,k_2)$, except that the known edge to node 1 is incoming rather than outgoing:

$$\begin{aligned} P_{G_2|K_1,K_2}(g_2|k_1,k_2) = &\sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2) \\ &\times H(g_2; \epsilon_2 - 1, m-1, k_2) \end{aligned} \tag{C7}$$

where $\hat{m}_2 = \max(m, k_2)$. Assembling the above results, we obtain Eq. (C3).

b. *Probability distribution for upper bound*

Here we derive the mean and the distribution in Eq. (C4) for the upper bound, $\hat{L}$, of the random variable L. We begin by noting that, after the rewiring procedure described in Subsection C 2, the number of nodes connected by edges, without regard to direction, to both nodes 1 and 2 is $\min(E_1, E_2) - 1$, where E_1 is the degree of node 1 and E_2 is the degree of node 2.

From this observation, together with the assignment of each possible direction to each edge independently with probability 1/2, we find that the number $\hat{L}$ of nodes receiving outgoing connections from both nodes 1 and 2 in this rewired network must satisfy:

$$\mathbb{E}(\hat{L}|E_1, E_2) = \frac{1}{4}(\min(E_1, E_2) - 1). \tag{C8}$$

The unconditional mean of $\hat{L}$ is then computed by applying the law of total expectation [53]:

$$\mathbb{E}\hat{L} = \sum_{\epsilon_1=1}^{N-1}\sum_{\epsilon_2=1}^{N-1} \mathbb{E}(\hat{L}|E_1 = \epsilon_1, E_2 = \epsilon_2) P_{E_1,E_2}(\epsilon_1, \epsilon_2) \tag{C9}$$

For $1 \ll m \ll N$, we have an explicit formula for $P_{E_1,E_2}(\epsilon_1,\epsilon_2) = P_{E_2|E_1}(\epsilon_2|\epsilon_1)P_E(\epsilon_1)$ by appeal to Eqs. (B6) and (B3). Substituting this expression and Eq. (C8) into Eq. (C9), and approximating the sums by semidefinite integrals (consistent with the assumption $1 \ll m \ll N$), we have

$$\begin{aligned} \mathbb{E}\hat{L} &\sim \int_m^\infty \int_m^\infty \frac{1}{4}(\min(\epsilon_1,\epsilon_2) - 1)\frac{\epsilon_1 + \epsilon_2 - 2m}{\epsilon_1}\left(\frac{2m^2}{\epsilon_1^3}\right) \\ &\quad \times \left(\frac{2m^2}{\epsilon_2^3}\right) \mathrm{d}\epsilon_1\, \mathrm{d}\epsilon_2 \\ &= -\frac{1}{4} + m^4 \int_m^\infty \left[\int_m^{\epsilon_1} \frac{\epsilon_1 + \epsilon_2 - 2m}{\epsilon_1^4\epsilon_2^2}\, \mathrm{d}\epsilon_2 \right. \\ &\quad \left. + \int_{\epsilon_1}^\infty \frac{\epsilon_1 + \epsilon_2 - 2m}{\epsilon_1^3\epsilon_2^3}\, \mathrm{d}\epsilon_2\right] \\ &= \frac{13m - 9}{36}. \end{aligned}$$

To determine the full conditional probability distribution, $P_{\hat{L}|K_1,K_2}(l|k_1,k_2)$, we sum over all possible values for E_1 and E_2 using the law of total probability:

$$\begin{aligned} &P_{\hat{L}|K_1,K_2}(l|k_1,k_2) = \\ &\sum_{\epsilon_1=\hat{m}_1}^{N-1}\sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{\hat{L}|E_1,E_2,K_1,K_2}(l|\epsilon_1,\epsilon_2,k_1,k_2) \\ &\times P_{E_1|E_2,K_1,K_2}(\epsilon_1|\epsilon_2,k_1,k_2)P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2), \end{aligned} \tag{C10}$$

where $\hat{m}_1 = \max(m, k_1)$ and $\hat{m}_2 = \max(m, k_2)$. We obtain the distribution $P_{E_1|E_2,K_1,K_2}(\epsilon_1|\epsilon_2,k_1,k_2)$ by first noting that it equals $P_{E_1|E_2,K_1}(\epsilon_1|\epsilon_2,k_1)$ as E_1 is conditionally independent of K_2 once the value e_2 is known.

Next, we consider first the case where node 1 plays the role of node A in generating the rewired network (so in particular $E_1 \leq E_2$). Noting that in the rewired network, all neighbors of node 1 are also neighbors of node 2, we see that the value of $\hat{L}$ is the number of the $k_1 - 1$ nodes receiving outgoing connections from node 1 that are among the $\epsilon_2 - 1$ other neighbors of node 2 that receive one of the k_2 outgoing connections from node 2 (the minus one is because we exclude the known outgoing connection from node 1 to node 2). The random variable $\hat{L}$ therefore satisfies the hypergeometric distribution, defined in Eq. (C1),

$$P_{\hat{L}|E_1,E_2,K_1,K_2}(l|\epsilon_1,\epsilon_2,k_1,k_2) = H(l;\epsilon_2-1,k_1-1,k_2).$$

When node 2 plays the role of node A in generating the rewired network (so that $E_2 \leq E_1$, then all neighbors of node 2 are also neighbors of node 1, and the value of $\hat{L}$ is the number of the k_2 nodes receiving outgoing connections from node 2 that are among the $e_1 - 1$ other neighbors of node 1 that receive one of the $k_1 - 1$ outgoing connections from node 1. The random variable $\hat{L}$ therefore satisfies the hypergeometric distribution,

$$P_{\hat{L}|E_1,E_2,K_1,K_2}(l|\epsilon_1,\epsilon_2,k_1,k_2) = H(l;\epsilon_1-1,k_2,k_1-1).$$

The hypergeometric distribution is symmetric under interchange of the last two arguments [53], so we may combine the above results to write the distribution for the upper bound $\hat{L}$ as

$$P_{\hat{L}|E_1,E_2,K_1,K_2}(l|\epsilon_1,\epsilon_2,k_1,k_2) = H(l;\epsilon_M,k_1-1,k_2) \tag{C11}$$

where $\epsilon_M = \max(\epsilon_1,\epsilon_2)-1$. By substituting the distribution in Eq. (C11) into Eq. (C10) we obtain the expression for $\hat{L}$ that appears in Eq. (C4).

Appendix D: Central Limit Approximations for Large m

Explicit formulas for some of the scale-free network statistics presented in Appendix B generally relied on the assumption $N \gg m \gg 1$. Those results followed from standard asymptotic results in probability theory and the approximation of the discrete probability distributions by continuous ones [33], and are quite rigorously derived. We now show how we obtain the approximate formula (25) for the probability of cascade failure at the second step from the precise expression (C3) for the conditional probability distribution for the lower bound $\tilde{L}$ on the number of doubly-excited nodes in the same asymptotic regime $N \gg m \gg 1$. This calculation will be mostly systematic, but will encounter some steps where we make uncontrolled approximations. For this reason, we separately report and consider the more precise formula (C3) as well as the simpler explicit approximation in Eq. (25), as the asymptotic simplification is theoretically less secure than those reported in Appendix B.

We proceed essentially through successive applications of Laplace's method [76], which can be interpreted as approximations akin to the central limit theorem. The main complication in this asymptotic calculation is the propagation of the Gaussian approximations through the successive levels of conditioning.

We begin with a standard Gaussian approximation [53, 77, 78]) to the binomial distribution (B1), recalling that $\epsilon \geq m$ so that large m implies large ϵ:

$$P_{K|E}(k|\epsilon) \sim \phi(k;\epsilon/2,\epsilon/4) \text{ for } m \gg 1 \tag{D1}$$

where

$$\phi(x;\mu,\sigma^2) \equiv \frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

denotes a Gaussian with mean μ and variance σ^2. More precisely, the DeMoivre-Laplace theorem [77, 78] proves that such approximation (D1) is only valid in the core of the probability distribution (i,e., within a few standard deviations of the mean), but not in the tails. But one can check that the statistics of interest are all dominated for large m by the core of the probability distributions, so we will not state explicitly this restriction of validity.

Next, we apply the same concepts from the proof of the DeMoivre-Laplace theorem successively to other formulas building on this conditional probability. The relevant technical results regarding asymptotic evaluation of integrals are collected in Appendix E. We begin with Eq. (B2), where the fact that m is large, P_E has support only on $\epsilon \geq m \gg 1$ and varies on a scale $m \gg 1$ (Eq. (B3)), and, from Eq. (D1), $P_{K|E}$ varies on a length scale $\gtrsim \sqrt{m} \gg 1$ suggests that we can approximate the sum with unit step by an integral:

$$P_K(k) \sim \int_m^\infty \phi(k;\epsilon/2,\epsilon/4) P_E(e)\,\mathrm{d}\epsilon$$

Next we rescale the integration variable $\epsilon = m\tilde{\epsilon}$ to obtain:

$$P_K(k) \sim \int_1^\infty \phi(k;m\tilde{\epsilon}/2,m\tilde{\epsilon}/4) P_{\tilde{E}}(\tilde{\epsilon})\,\mathrm{d}\tilde{\epsilon}$$

where $P_{\tilde{E}}(\tilde{\epsilon}) = mP_E(m\tilde{\epsilon})$ is a rescaled probability distribution for $\tilde{E} = E/m$. From Eq. (B3) we see that the function $P_{\tilde{E}}$ is independent of m for large m. We now apply Lemma 1 from Appendix E concerning the integral of a slowly varying function against a sharply peaked Gaussian (standard variation small compared to mean) to obtain:

$$\begin{aligned} P_K(k) &\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k-m/2}{\sqrt{m}} \ll -1, \\ \frac{2}{m} P_{\tilde{E}}(2k/m) & \text{if } \frac{k-m/2}{\sqrt{m}} \gg 1 \end{cases} \\ &= \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k-m/2}{\sqrt{m}} \ll -1, \\ 2P_E(2k) & \text{if } \frac{k-m/2}{\sqrt{m}} \gg 1. \end{cases} \end{aligned} \tag{D2}$$

That is, the degree distribution of the outgoing edges is, for large m, well approximated by a simple rescaling of

the total edge distribution for a node (similar results were derived in somewhat less precise form in Shkarayev *et al.* [33]). This could have been argued also using simpler asymptotic arguments on probability generating functions, but we used the above argument because we will need a generalization of it later in the calculation. Similar arguments show:

$$P_{K_1,K_2}(k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ 4P_{E_1,E_2}(2k_1,2k_2) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases}$$

and therefore

$$P_{K_2|K_1}(k_2|k_1) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ 2P_{E_2|E_1}(2k_2|2k_1) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases} \tag{D3}$$

We now apply these approximations to simplify the expressions for $P_{E_2|K_1,K_2}$ in Eq. (B9) and $P_{E_1|K_1,K_2}$ in Eq. (B10). We will proceed using formal asymptotic analysis, without attempting to be fully rigorous. To this end, note from Eq. (B6) that when m is large, $P_{E_2|E_1}(\epsilon_2|\epsilon_1)$ has support only on $\epsilon_1, \epsilon_2 \geq m \gg 1$ and varies on a scale $m \gg 1$, and, from Eq. (D1), $P_{K|E}$ varies on a length scale $\gtrsim \sqrt{m} \gg 1$. This suggests that we can approximate the sum with unit step by an integral:

$$\begin{aligned} &\sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{K_2|E_2}(k_2|\epsilon_2) P_{E_2|E_1}(\epsilon_2|\epsilon_1) \\ &\quad \sim \int_{\hat{m}_2}^{\infty} P_{K_2|E_2}(k_2|\epsilon_2) P_{E_2|E_1}(\epsilon_2|\epsilon_1)\,\mathrm{d}\epsilon_2 \\ &\quad \sim \int_{\hat{m}_2}^{\infty} \phi(k_2;\epsilon_2/2,\epsilon_2/4) P_{E_2|E_1}(\epsilon_2|\epsilon_1)\,\mathrm{d}\epsilon_2 \\ &\quad \sim \int_{\hat{m}_2/m}^{\infty} \phi(k_2;m\tilde{\epsilon}_2/2,m\tilde{\epsilon}_2/4) P_{\tilde{E}_2|\tilde{E}_1}(\tilde{\epsilon}_2|\epsilon_1/m)\,\mathrm{d}\tilde{\epsilon}_2 \end{aligned} \tag{D4}$$

where $P_{\tilde{E}_2|\tilde{E}_1}(\tilde{\epsilon}_2|\tilde{\epsilon}_1) = mP_{E_2|E_1}(m\tilde{\epsilon}_2|m\tilde{\epsilon}_1)$ is a rescaled probability distribution. From Eq. (B6) we see that the function $P_{\tilde{E}_2|\tilde{E}_1}$ becomes independent of m for large m. To Eq. (D4), we now apply Lemma 1 from Appendix E:

$$\begin{aligned} &\sum_{\epsilon_2=\hat{m}_2}^{N-1} P_{K_2|E_2}(k_2|\epsilon_2) P_{E_2|E_1}(\epsilon_2|\epsilon_1) \\ &\quad \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_2-\hat{m}_2/2}{\sqrt{\hat{m}_2}} \ll -1, \\ \frac{2}{m} P_{\tilde{E}_2|\tilde{E}_1}(2k_2/m|\epsilon_1/m) & \text{if } \frac{k_2-\hat{m}_2/2}{\sqrt{\hat{m}_2}} \gg 1. \end{cases} \\ &\quad \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_2-\hat{m}_2/2}{\sqrt{\hat{m}_2}} \ll -1, \\ 2P_{E_2|E_1}(2k_2|\epsilon_1) & \text{if } \frac{k_2-\hat{m}_2/2}{\sqrt{\hat{m}_2}} \gg 1. \end{cases} \end{aligned}$$

Recalling that $\hat{m}_2 = \max(m,k_2)$, applying the Gaussian approximation (D1) to the factor $P_{K|E}(k_1|\epsilon_1)$ in Eq. (B10), we deduce:

$$P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \frac{2P_{E_2|E_1}(2k_2|\epsilon_1)\phi(k_1;\epsilon_1/2,\epsilon_1/4)P_E(\epsilon_1)}{P_{K_2|K_1}(k_2|k_1)P_{K_1}(k_1)} & \text{if } \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases} \tag{D5}$$

Arguing similarly regarding the expression Eq. (B9), we can simplify it for large m to:

$$P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}} \ll -1, \\ \frac{2P_{E_2|E_1}(\epsilon_2|2k_1)\phi(k_2;\epsilon_2/2,\epsilon_2/4)P_E(2k_1)}{P_{K_2|K_1}(k_2|k_1)P_{K_1}(k_1)} & \text{if } \frac{k_1-m/2}{\sqrt{m}} \gg 1. \end{cases} \tag{D6}$$

Having used the large m asymptotics to remove the summations in these conditional probabilities, we turn next to approximating the summations in the $[\cdot]$ brackets in Eq. (C3).

The hypergeometric functions appearing here can also be approximated for large m using the following asymptotic result [77]:

$$H(k;n,r,s) \sim \phi(k;rs/n, rs(n-r)(n-s)/n^3) \quad \text{for } n \gg 1; r/n, s/n, k/n \sim \mathrm{ord}\,(1). \tag{D7}$$

This distinguished asymptotic limit is applicable to all hypergeometric functions appearing in Eq. (C3), with m playing the role of large parameter, because the probability distribution for the edges is concentrated about $\epsilon_1, \epsilon_2 \sim \mathrm{ord}\,(m)$, and from Eqs. (D5) and (D6), the contributions from $k_1, k_2 \not\sim \mathrm{ord}\,(m)$ are transcendentally small. (To be more rigorous, we might concern ourselves with situations where these parameters are in fact much larger than m, but given the decay of the edge distribution $P_E(\epsilon)$ and the fact that in our simulations the network size N is not orders of magnitude larger than m make it plausible that we can neglect the contribution from this regime, at least to leading order.) Also, if g_1 or

g_2 are much smaller or larger than m, the true hypergeometric distribution as well as the approximation (D7) will both give transcendentally small weight so we will simply extend the approximation (D7) to those regimes with no loss to leading order accuracy. Applying these observations to the first $[\cdot]$ bracket in Eq. (C3), approximating the sum by an integral for similar reasons as above, and rescaling the integration variable $\tilde{\epsilon}_1 = \epsilon_1/m$, we obtain:

$$\sum_{\epsilon_1=\hat{m}_1}^{N-1} H(g_1;\epsilon_1-1,m-1,k_1-1)P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2)$$
$$\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \int_{\hat{m}_1/m}^{\infty} \frac{2P_{E_2|\tilde{E}_1}(2k_2|m\tilde{\epsilon}_1)P_{\tilde{E}}(\tilde{\epsilon}_1)\phi(g_1;k_1/\tilde{\epsilon}_1,m(k_1/m)(\tilde{\epsilon}_1-1)(\tilde{\epsilon}_1-k_1/m)/\tilde{\epsilon}_1^3)\phi(k_1;m\tilde{\epsilon}_1/2,m\tilde{\epsilon}_1/4)}{P_{K_2|K_1}(k_2|k_1)P_{K_1}(k_1)}\,\mathrm{d}\tilde{\epsilon}_1 & \text{if } \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases}$$

Now we *cannot* apply Lemma 1 from Appendix E to this integral because each of the Gaussians fails to have logarithmically bounded derivative, and more fundamentally, are sharply peaked functions (standard deviation $\sim \mathrm{ord}\,(\sqrt{m})$ much smaller than their mean $\sim \mathrm{ord}\,(m)$.) We therefore apply Lemma 2 from Appendix E, adapted to this calculation, which yields:

$$\begin{aligned} &\sum_{\epsilon_1=\hat{m}_1}^{N-1} H(g_1;\epsilon_1-1,m-1,k_1-1)P_{E_1|K_1,K_2}(\epsilon_1|k_1,k_2) \\ &\quad\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \frac{4P_{E_2|E_1}(2k_2|2k_1)P_E(2k_1)}{P_{K_2|K_1}(k_2|k_1)P_{K_1}(k_1)}\phi\left(g_1;\frac{m}{2},\frac{m}{4}\right) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1, \end{cases} \\ &\quad\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \phi\left(g_1;\frac{m}{2},\frac{m}{4}\right) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1, \end{cases} \end{aligned} \tag{D8}$$

The fraction was removed in the last expression through the use of Eqs. (D2) and (D3).

A similar calculation shows

$$\sum_{\epsilon_2=\hat{m}_2}^{N-1} H(g_2;\epsilon_2-1,m-1,k_2)P_{E_2|K_1,K_2}(\epsilon_2|k_1,k_2)$$
$$\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \phi\left(g_2;\frac{m}{2},\frac{m}{4}\right) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1, \end{cases} \tag{D9}$$

We next turn to the summation over g_1 and g_2 in Eq. (C3). Applying the asymptotic approximation Eq. (D7) of the hypergeometric distribution $H(l;m-1,g_1,g_2)$ and the approximations (D8) and (D9) to the other factors, and then approximating the sums over g_1 and g_2 by integrals, and then rescaling the integration variables to be order unity, we have:

$$P_{\tilde{L}|K_1,K_2}(l|k_1,k_2)$$
$$\sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \int_0^1\int_0^1 \phi(l;m\tilde{g}_1\tilde{g}_2, m\tilde{g}_1\tilde{g}_2(1-\tilde{g}_1)(1-\tilde{g}_2))\phi\left(\tilde{g}_1;\frac{1}{2},\frac{1}{4m}\right)\phi\left(\tilde{g}_2;\frac{1}{2},\frac{1}{4m}\right)\,\mathrm{d}\tilde{g}_1\,\mathrm{d}\tilde{g}_2 & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases}$$

The integrals here essentially amount to computing the unconditional distributions of random variables $\tilde{L}$, which is Gaussian when conditioned on two independent Gaussian random variables G_1 and G_2. This can be eval-

uated asymptotically for large m using Lemma 3 from Appendix E separately for each integral to obtain:

$$P_{\tilde{L}|K_1,K_2}(l|k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \phi\left(l; \frac{m}{4}, \frac{3m}{16}\right) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases} \tag{D10}$$

Admittedly, this is such a simple result that it seems like a more direct argument should have sufficed to achieve it. We have however not been able to find a more elegant derivation. The Gaussian structure could presumably be argued informally through central limit theorem ideas, and its mean is easy to see. But the value of the variance is important, and we have found no more efficient method to compute its value correctly than chasing through the above sequence of calculations. In essence, one could say the above efforts have amounted to more or less computing the large m asymptotics for the conditional variance of $\tilde{L}$ on K_1 and K_2 by averaging its variance conditioned on the other random variables whose probability distributions we have integrated out.

Anyway, the last step in the derivation of the lower bound approximation to $P_t(A_2)$ is to approximate the summation over l in Eq. (18), when the approximation (D10) for $P_{\tilde{L}|K_1,K_2}$ is substituted in place of $P_{L|K_1,K_2}$, and the summation approximated by an integral:

$$\sum_{l=0}^{\kappa} \left(\frac{1-p_1(t)-p_2(t)}{\left(1-p_1(t)\right)^2}\right)^l P_{\tilde{L}|K_1,K_2}(l|k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ \int_0^\kappa \phi\left(l; \frac{m}{4}, \frac{3m}{16}\right) (\rho(t))^l \, \mathrm{d}l & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1, \end{cases}$$

where $\rho(t)$ is defined in Eq. (21). Now we rescale the integration variable and complete the square in the exponent:

$$\begin{aligned} &\int_0^\kappa \phi\left(l; \frac{m}{4}, \frac{3m}{16}\right) (\rho(t))^l \, \mathrm{d}l \\ &\quad = \int_0^{\kappa/m} \phi\left(\tilde{l}; \frac{1}{4}, \frac{3}{16m}\right) (\rho(t))^{(m\tilde{l})} \, \mathrm{d}\tilde{l} \\ &\quad = \frac{1}{\sqrt{3\pi/(8m)}} \int_0^{\kappa/m} \mathrm{e}^{-\left(\frac{(\tilde{l}-1/4)^2}{3/(8m)}\right) + m\tilde{l}\ln\rho(t)} \, \mathrm{d}\tilde{l} \\ &\quad = \frac{1}{\sqrt{3\pi/(8m)}} \int_0^{\kappa/m} \mathrm{e}^{-\frac{8m}{3}\left(\tilde{l}-1/4-\frac{3}{16}\ln\rho(t)\right)^2} \\ &\qquad\qquad \times \mathrm{e}^{\frac{m}{4}\ln\rho(t)+\frac{3m}{32}(\ln\rho(t))^2} \, \mathrm{d}\tilde{l}. \end{aligned} \tag{D11}$$

We will now use the observation that $\rho(t)$ is generally fairly close to 1 in our numerical examples, so that $\ln\rho(t)$ is small. Then, since the upper integration limit $\kappa/m \gtrsim 1/2$ over the range of k_1, k_2 values for which the integral approximation is appropriate, the integral in the last expression in Eq. (D11) involves a Gaussian centered near $1/4$ with small $O(1/m)$ variance, integrated over at least the interval $[0, 1/2]$. This justifies replacing the integral over the finite domain by an integral over the whole real line, yielding:

$$\begin{aligned} &\int_0^\kappa \phi\left(l; \frac{m}{4}, \frac{3m}{16}\right) (\rho(t))^l \, \mathrm{d}l \\ &\qquad \sim \exp\left[\frac{m}{4}\ln\rho(t) + \frac{3m}{32}(\ln\rho(t))^2\right] \end{aligned}$$

So we have argued that:

$$\sum_{l=0}^{\kappa} \left(\frac{1-p_1(t)-p_2(t)}{\left(1-p_1(t)\right)^2}\right)^l P_{\tilde{L}|K_1,K_2}(l|k_1,k_2) \sim \begin{cases} \mathrm{TST}(1/m) & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \ll -1, \\ (\rho(t))^{m/4} \exp\left[\frac{3m}{32}(\ln\rho(t))^2\right] & \text{if } \frac{k_1-m/2}{\sqrt{m}}, \frac{k_2-m/2}{\sqrt{m}} \gg 1. \end{cases} \tag{D12}$$

Thus far, we have attempted to be precise and systematic, if not fully mathematically rigorous, in our asymptotic derivations. But continuing in the same vein beyond this point becomes much more arduous so we will proceed the rest of the way without the same level of systematic argument, justifying our arguments *post hoc* through agreement with numerical computations (Fig. 5).

First of all, $\rho(t)$ (21) is independent of m, so the expression (D12) may appear to be transcendentally small with respect to $1/m$ if $\rho(t) < 1$, and one may ask why we don't neglect this expression in the same way as we have other transcendentally small terms. The reason is that we can see from the definition of $\rho(t)$ that it should be close to unity if the spike size S is small enough, and indeed in our numerical experiments, we find this to be at least roughly true. Now, to be fully precise, we should connect the closeness of $\rho(t)$ to unity in terms of the small parameter $1/m$ through certain restrictions on the neuronal parameters, but we will not attempt this. Rather we will simply take the expression (D12) as a proposed approximation based on the above mostly systematic arguments. We pause only to point out that if indeed we were to assume $1-\rho(t)$ is really $\mathrm{ord}\,(1/m)$ in some formal sense, then $\ln\rho(t) \sim \mathrm{ord}\,(1/m)$ and the expression (D12) actually simplifies asymptotically (for $N \gg m \gg 1$) to the approximation Eq. (24) where L is simply replaced by the constant $L = \bar{l} = \mathbb{E}\tilde{L} = m/4$. This could explain the very good performance of this approximation in Fig. 5. We note though the improved accuracy (emphasized in the inset figure) of the expression (25), which we next show to follow from the more precise approximation (D12).

We proceed then, with these caveats, to take Eq. (D12) as a proposed approximation based on our mostly systematic arguments from above. Now when we attempt to substitute this approximation into Eq. (18), we encounter the fact that we must sum over all nonnegative integer values of k_1 and k_2, and we only have asymptotics for limited parameter regimes. One expects though that the sums will be concentrated at values of k_1 and

k_2 for which the nontrivial approximation in Eq. (D12) applies. Indeed the factors $P_{K_2|K_1}(k_2|k_1)P_K(k_1)$ would, according to Eqs. (D2) and (D3), seem to be primarly concentrated on these values. The other factors in the sum, which depend exponentially on k_1 and k_2 will skew the weight somewhat to smaller values, but we will not attempt a precise argument here. Rather, we will simply posit that we can simply replace the sum over l in Eq. (18) *for all* k_1 and k_2 by the nontrivial expression in Eq. (D12) that was derived only for the regime $(k_1 - m/2)/\sqrt{m}, (k_2 - m/2)/\sqrt{m} \gg 1$. This is how we obtain the simplified expression Eq. (25) for the upper bound for $P_t(A_2)$ in the main text.

Appendix E: Asymptotic Results for Gaussian Integrals

We collect in this appendix a couple of key results, and their derivations, that are repeatedly applied in the large m asymptotic simplifications in Appendix D.

Lemma 1 *For smooth functions g with bounded logarithmic derivative $\frac{\mathrm{d}\ln g}{\mathrm{d}y}$ on the positive real axis and positive constants c_1 , c_2, and c_3 independent of λ (at least asymptotically),*

$$\int_{c_3}^{\infty} \phi(x;\lambda c_1 y, \lambda c_2 y) g(y)\,\mathrm{d}y \sim \begin{cases} \frac{1}{\lambda c_1} g(x/c_1\lambda) & \text{if } \frac{x-\lambda c_1 c_3}{\sqrt{\lambda c_2 c_3}} \gg 1, \\ \mathrm{TST}(1/\lambda) & \text{if } \frac{x-\lambda c_1 c_3}{\sqrt{\lambda c_2 c_3}} \ll -1 \text{ and } x>0, \end{cases} \quad \text{for } \lambda \gg 1. \tag{E1}$$

where $\mathrm{TST}(1/\lambda)$ *denotes transcendentally small terms with respect to the small parameter* $1/\lambda$

Note we will want to apply this lemma to a case in which $g(y)$ diverges at $y = 0$ so we don't want to replace the right hand side simply by $g(0)$. Intuitively, the lemma says the integral is negligible when the core of the Gaussian is outside the integration domain, but when it is inside the integration domain, it is thin enough relative to its mean that g can be approximated as constant over this Gaussian core. The extra factor of $1/(\lambda c_1)$ arises because of the need to express the Gaussian in terms of the integration variable y rather than its more natural argument x.

To be more precise, the validity of this lemma can be established for the case $x - \lambda c_1 c_3 \ll -\sqrt{\lambda c_2 c_3}$ by noting that for all y' in the integration domain, the Gaussian factor is evaluated many standard deviations away from its mean ($\frac{|x-\lambda c_1 y'|}{\sqrt{\lambda c_2 y'}} \gg 1$), giving a transcendentally small contribution. To see why, note that the function

$$\begin{aligned} h(y') &= \frac{|x-\lambda c_1 y'|}{\sqrt{\lambda c_2 y'}} = \frac{\lambda c_1 y' - x}{\sqrt{\lambda c_2 y'}} \\ &= \sqrt{\lambda c_1 c_2^{-1}}(y'^{1/2} - x y'^{-1/2}) \end{aligned}$$

has positive derivative for $y' > 0$ and $x > 0$, and is therefore minimized on the integration domain at $y' = c_3$, and $h(c_3) = |x - \lambda c_1 c_3|/\sqrt{\lambda c_2 c_3} \gg 1$ by hypothesis.

To treat the case $(x - \lambda c_1 c_3)/\sqrt{\lambda c_2 c_3} \gg 1$, we first note that this hypothesis implies the Gaussian in the integrand is many standard deviations away from the lower limit. Together with the boundedness of the logarithmic derivative of g, this implies that the integral will be dominated by the core of the Gaussian factor (again with transcendentally small error). We therefore make a nonlinear change of variable to bring the Gaussian into focus:

$$\eta(x,y) = \frac{c_1\lambda y - x}{\sqrt{c_2 \lambda y}}. \tag{E2}$$

The slight complication here is that the variance of the Gaussian in Eq. (E1) is set by the integration variable y rather than the external parameter x, so the change of variable $\eta(x, y)$ is nonlinear in the integration variable. Proceeding, we have:

$$\begin{aligned} &\int_{c_3}^{\infty} \phi(x;\lambda c_1 y, \lambda c_2 y) g(y)\,\mathrm{d}y \\ &\qquad = \frac{1}{\sqrt{2\pi c_2 \lambda}} \int_{\eta(x,c_3)}^{\infty} \mathrm{e}^{-\eta^2/2} \frac{g(y(\eta,x))}{\sqrt{y(\eta,x)}} \left|\frac{\partial y(\eta,x)}{\partial \eta}\right| \mathrm{d}\eta, \end{aligned}$$

where $y(\eta, x)$ is the inverse function obtained by solving Eq. (E2) for y. This can be done with the quadratic formula, but we can take a more direct approach to its asymptotic approximation by rewriting Eq. (E2) as:

$$y = \frac{x}{c_1\lambda} + \frac{\sqrt{c_2 y}\eta}{c_1\sqrt{\lambda}}.$$

As noted above, only order unity values of η are relevant, and by hypothesis x is at least $\mathrm{ord}\,(\lambda)$ in magnitude. The only self-consistent dominant balance in this relationship is therefore $y \sim \frac{x}{c_1\lambda}$, from which iteration shows:

$$y = \frac{x}{c_1\lambda} + \frac{\sqrt{c_2 x}\eta}{c_1^{3/2}\lambda} + O(\lambda^{-1}). \tag{E3}$$

Since $g(y)/y$ has bounded logarithmic derivative, it can be well approximated by the leading term in this iterated

asymptotic expansion, and we have:

$$\int_{c_3}^{\infty} \phi(x;\lambda c_1 y,\lambda c_2 y) g(y)\,\mathrm{d}y \sim \frac{1}{\sqrt{2\pi c_2\lambda}} \int_{\eta(x,c_3)}^{\infty} \mathrm{e}^{-\eta^2/2} \frac{g(x/(c_1\lambda))}{\sqrt{x/(c_1\lambda)}} \frac{\sqrt{c_2 x}}{c_1^{3/2}\lambda}\,\mathrm{d}\eta.$$

The hypothesis $(x-\lambda c_1 c_3)/\sqrt{\lambda c_2 c_3} \gg 1$ implies $\eta(x,c_3) \ll -1$, so we may replace the lower limit by $-\infty$, perform the Gaussian integral over η, leaving the result in Eq. (E1).

Lemma 2 *For smooth functions g, h with bounded logarithmic derivatives over positive values of their arguments, and positive constants c_1 , c_2, and c_3 independent of λ (at least asymptotically),*

$$\int_{c_3}^{\infty} \phi(x;\lambda c_1 y,\lambda c_2 y)\phi(z;c_4 x/y,\lambda h(x/\lambda,y)) g(y)\,\mathrm{d}y \sim \begin{cases} \mathrm{TST}(1/\lambda) & \text{if } \frac{x-\lambda c_1 c_3}{\sqrt{\lambda c_2 c_3}} \ll -1 \text{ and } x>0, \\ \frac{1}{c_1\lambda} g(x/(c_1\lambda))\phi(z;c_1 c_4\lambda, \lambda h(\frac{x}{\lambda},\frac{x}{c_1\lambda}) + c_1 c_2 c_4^2 \lambda^2 x^{-1}) & \text{if } \frac{x-\lambda c_1 c_3}{\sqrt{\lambda c_2 c_3}} \gg 1 \end{cases} \quad \text{for } \lambda \gg 1. \tag{E4}$$

We note that a naive application of Lemma 1 to the expression Eq. (E4), i.e., simply evaluating the second Gaussian factor at the peak value $y = x/(\lambda c_1)$ of the first Gaussian, yields an incorrect result. This is an invalid application of Lemma 1 because the logarithmic derivative of the second factor is actually $\mathrm{ord}\,(\lambda)$, and therefore surely not bounded as $\lambda \to \infty$.

To establish this lemma, we first note that the integral is transcendentally small for $x - \lambda c_1 c_3 \ll -\sqrt{\lambda c_2 c_3}$ for the same reasons as in Lemma 1. For $\frac{x-\lambda c_1 c_3}{\sqrt{\lambda c_2 c_3}} \gg 1$, we make the same change of variable (E2) as in Lemma 1, yielding now:

$$\int_{c_3}^{\infty} \phi(x;\lambda c_1 y,\lambda c_2 y)\phi(z;c_4 x/y),\lambda h(x/\lambda,y)) g(y)\,\mathrm{d}y = \frac{1}{\sqrt{2\pi c_2\lambda}} \int_{\eta(x,c_3)}^{\infty} \mathrm{e}^{-\eta^2/2} \phi(z;c_4 x/y(\eta,x)),\lambda h(x/\lambda,y(\eta,x))) \frac{g(y(\eta,x))}{\sqrt{y(\eta,x)}} \left|\frac{\partial y(\eta,x)}{\partial \eta}\right| \mathrm{d}\eta.$$

Substituting now the asymptotic expansion (E3) for $y(\eta,x)$, we obtain:

$$\int_{c_3}^{\infty} \phi(x;\lambda c_1 y,\lambda c_2 y)\phi(z;c_4 x/y),\lambda h(x/\lambda,y)) g(y)\,\mathrm{d}y \sim \frac{1}{\sqrt{2\pi c_2\lambda}} \int_{\eta(x,c_3)}^{\infty} \mathrm{e}^{-\eta^2/2} \phi\left(z;c_1 c_4\lambda\left(1-\eta\sqrt{c_2/(c_1 x)}+O(x^{-1})\right),\lambda h(x/\lambda,(x/(c_1\lambda))(1+O(x^{-1/2}))\right) \times \frac{g(x/(c_1\lambda)(1+O(x^{-1/2})))}{\sqrt{x/(c_1\lambda)(1+O(x^{-1/2}))}} \frac{\sqrt{c_2 x}}{c_1^{3/2}\lambda}(1+O(x^{-1/2}))\,\mathrm{d}\eta.$$

Recalling that $x \gtrsim \mathrm{ord}\,(\lambda)$ by assumption, dropping all terms that don't contribute to leading order (which requires some care in the Gaussian factor), and noting for the reasons as in Lemma 1 we can take the lower integration limit to $-\infty$, we find:

$$\begin{aligned}\int_{c_3}^{\infty} &\phi(x;\lambda c_1 y,\lambda c_2 y)\phi(z;c_4 x/y),\lambda h(x/\lambda,y)) g(y)\,\mathrm{d}y \\ &\sim \frac{1}{\sqrt{2\pi}c_1\lambda} \int_{-\infty}^{\infty} \mathrm{e}^{-\eta^2/2} \phi\left(z;c_1 c_4\lambda\left(1-\eta\sqrt{c_2/(c_1 x)}\right),\lambda h(x/\lambda,x/(c_1\lambda))\right) g(x/(c_1\lambda))\,\mathrm{d}\eta \\ &= \frac{1}{c_1\lambda} g(x/(c_1\lambda)) \int_{-\infty}^{\infty} \phi(\eta;0,1)\phi\left(z;c_1 c_4\lambda\left(1-\eta\sqrt{c_2/(c_1 x)}\right),\lambda h(x/\lambda,x/(c_1\lambda))\right)\,\mathrm{d}\eta\end{aligned}$$

This last integral can be evaluated tediously by combining the Gaussians and completing the square with respect to the integration variable η in the exponential. Alter-

natively and more simply, one can interpret the integral, through the law of total probability for continuous random variables [53], as giving the probability density of a random variable Z, which is Gaussian with mean and variance given by the arguments of the second Gaussian factor, when conditioned on a standard Gaussian random variable $\mathcal{N}$. Because $\mathcal{N}$ only influences Z through the conditional mean, the random variable Z can be seen to be unconditionally Gaussian, with mean and variance given in terms of those of the component random variables $\mathcal{N}$ and Z by the formulas for total mean and total variance [53]. We thereby obtain directly the expression in Eq. (E4).

Lemma 3 *For a smooth function h with bounded logarithmic derivatives over positive values of its argument and positive constants $c_1 < c_3$ and c_2 independent of λ (at least asymptotically),*

$$\int_0^{c_3} \phi(y; c_1, \lambda^{-1}c_2)\phi(z; \lambda c_4 y, \lambda h(y))\,\mathrm{d}y \sim \phi\left(z; \lambda c_4 c_1, \lambda(h(c_1) + c_2 c_4^2)\right). \text{ for } \lambda \gg 1.$$

The derivation is just an easier application of the techniques used for the previous lemma, now using the simpler change of variables:

$$\xi(y) = \sqrt{\lambda}(y - c_1)/\sqrt{c_2}, \tag{E5}$$

leading to:

$$\begin{aligned}\int_0^{c_3} \phi(y; c_1, \lambda^{-1}c_2)\phi(z; c_4\lambda y, \lambda h(y))\,\mathrm{d}y &= \int_{-\lambda^{1/2}c_1/c_2}^{\lambda^{1/2}(c_3-c_1)/c_2} \phi(\xi; 0, 1)\phi(z; \lambda c_4(c_1 + \sqrt{c_2/\lambda}\xi), \lambda h(c_1 + \sqrt{c_2/\lambda}\xi))\,\mathrm{d}\xi \\ &\sim \int_{-\infty}^{\infty} \phi(\xi; 0, 1)\phi(z; \lambda c_4 c_1 + \lambda^{1/2} c_4 c_2^{1/2}\xi, \lambda h(c_1))\,\mathrm{d}\xi \\ &= \phi\left(z; \lambda c_4 c_1, \lambda(h(c_1) + c_4^2 c_2)\right).\end{aligned}$$

[1] P. Holme, Adv Complex Systems **6**, 163 (2003).
[2] W. Ben-Ameur and E. Gourdin, SIAM J Discrete Math **17**, 18 (2003).
[3] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, Proceedings of the National Academy of Sciences **109**, E680 (2012).
[4] M. J. Keeling, Proc. R. Soc. Lond. B **266**, 859 (1999).
[5] K. T. D. Earmes and M. J. Keeling, PNAS **99**, 13330 (2002).
[6] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, J Theoretical Biology **240**, 400 (2006).
[7] M. E. J. Newman, Phys Rev Lett **95**, 108701 (2005).
[8] M. E. J. Newman, Phys Rev E **76**, 045101(R) (2007).
[9] V. M. Egu??luz and K. Klemm, Phys Rev Let **89** (2002).
[10] M. Timme, F. Wolf, and T. Geisel, Physical Reveiw Letters **92** (2004).
[11] M. Timme, Europhysics Letters **76**, 367 (2006).
[12] D. Guo and C. Li, Phys Rev E **79** (2009).
[13] S. Lu, J. Fang, A. Guo, and Y. Peng, Neural Networks **22**, 30 (2009).
[14] A. N. Burkitt, Biol Cybern **95**, 97 (2006 Aug).
[15] A. N. Burkitt, Biol Cybern **95**, 1 (2006 Jul).
[16] R. D. Traub, J. Jeffreys, and M. Whittington, *Fast Oscillations in Cortical Circuits* (MIT press, Cambridge,MA, 1999).
[17] G. Buzsaki and A. Draguhn, Science **304**, 1926 (2004 Jun 25).
[18] T. J. Sejnowski and O. Paulsen, J Neurosci **26**, 1673 (2006 Feb 8).
[19] M. Steriade, Neuroscience **137**, 1087 (2006).
[20] A. K. Roopun, M. A. Kramer, L. M. Carracedo, M. Kaiser, C. H. Davies, R. D. Traub, N. J. Kopell, and M. A. Whittington, Front Neurosci **2**, 145 (2008).
[21] P. Fries, Annu Rev Neurosci **32**, 209 (2009).
[22] S. Achuthan and C. C. Canavier, Journal of Neuroscience **29**, 5218 (2009).
[23] D. Hansel, G. Mato, and C. Meunier, Neural Comput **7**, 307 (1995).
[24] C. Vanvreeswijk and L. F. Abbott, SIAM Journal on Applied Mathematics **53**, 253 (1993), ArticleType: research-article / Full publication date: Feb., 1993 / Copyright © 1993 Society for Industrial and Applied Mathematics.
[25] I. Belykh, E. de Lange, and M. Hasler, Phys Rev Lett **94** (2005).
[26] S. Ostojic, N. Brunel, and V. Hakim, Journal of Computational Neuroscience **26**, 369 (2009).
[27] R. M. Smeal, G. B. Ermentrout, and J. A. White, Philosophical Transactions of the Royal Society B: Biological Sciences **365**, 2407 (2010).
[28] W. Gerstner and J. L. van Hemmen, Physical Review Letters **71**, 312 (1993).
[29] R. E. Mirollo and S. H. Strogatz, SIAM Journal on Applied Mathematics **50**, 1645 (1990).
[30] K. A. Newhall, G. Kovačič, P. R. Kramer, D. Zhou, A. V. Rangan, and D. Cai, Commun. Math. Sci. **8**, 541 (2010).
[31] K. A. Newhall, G. Kovačič, P. R. Kramer, and D. Cai, Phys. Rev. E **82** (2010).
[32] R. DeVille and C. Peskin, Bulletin of Mathematical Biology **74**, 769 (2012).

[33] M. S. Shkarayev, G. Kovačič, and D. Cai, Phys. Rev. E **85**, 036104 (2012).

[34] D. Schmidt, J. Best, and M. S. Blumberg, in *Dynamical Systems and Differential Equations, DCDS Supplement 2011 Proceedings of the 8th AIMS International Conference (Dresden , Germany)*, edited by W. Feng, Z. Feng, M. Grasselli, A. Ibragimov, X. Lu, S. Siegmund, and J. Voigt (American Institute of Mathematical Sciences, 2011) pp. 1279–1288.

[35] Y. Ikeda, T. Hasegawa, and K. Nemoto, Journal of Physics: Conference Series **221**, 012005 (2010).

[36] M. D. LaMar and G. D. Smith, Physical Review E **81**, 046206 (2010).

[37] C. Li and Q. Zheng, Physical Biology **7**, 036010 (2010).

[38] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, Physical Review E **85**, 026106 (2012).

[39] J. Touboul and G. Ermentrout, Journal of Computational Neuroscience **31**, 453 (2011).

[40] A. Treves, Network **4**, 259 (1993).

[41] D. J. Amit and N. Brunel, Cereb Cortex **7**, 237 (1997 Apr-May).

[42] N. Brunel and V. Hakim, Neural Comp. **11**, 1621 (1999).

[43] D. Cai, L. Tao, A. V. Rangan, and D. W. McLaughlin, Commun. Math. Sci. **4**, 97 (2006).

[44] G. Kovačič, L. Tao, A. V. Rangan, and D. Cai, Phys. Rev. E **80**, 021904 (2009).

[45] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Physical Review E **83**, 036112 (2011).

[46] J. P. Gleeson, Physical Review E **77**, 046117 (2008).

[47] J. L. Payne, K. D. Harris, and P. S. Dodds, Physical Review E **84**, 016110 (2011).

[48] M. E. J. Newman, SIAM Review **45**, 167 (2003).

[49] H. C. Tuckwell, *Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories*, Cambridge Studies in Mathematical Biology (Cambridge University Press, 1988).

[50] R. Brette, M. Rudolph, T. Carnevale, M. Hines, D. Beeman, J. M. Bower, M. Diesmann, A. Morrison, P. H. Goodman, F. C. Harris Jr, M. Zirpe, T. Natschlager, D. Pecevski, B. Ermentrout, M. Djurfeldt, A. Lansner, O. Rochel, T. Vieville, E. Muller, A. P. Davison, S. E. Boustani, and A. Destexhe, J. Comput. Neurosci. **23**, 349 (2007).

[51] A. Rangan, D. Cai, and D. McLaughlin, PNAS **102** (2005).

[52] K. Klemm and V. Eguíluz, Physical Review E **65** (2002).

[53] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability* (Athena Scientific, 2002).

[54] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer, 2004).

[55] J. Thomas, *Numerical Parital Differential Equations: Finite Difference Methods* (Springer, 1998).

[56] R. E. L. DeVille and C. S. Peskin, Bull. Math. Biol. **70**, 1608 (2008).

[57] D. J. Watts, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).

[58] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, Lecture Notes in Statistics, Vol. 151 (Springer-Verlag, New York, 2000) pp. x+137.

[59] I. Belykh, E. de Lange, and M. Hasler, Physical Review Letters **94**, 188101 (2005).

[60] G. Buzsaki, Z. Horvath, R. Urioste, J. Hetke, and K. Wise, Science **256**, 1025 (1992).

[61] A. Bragin, G. Jando, Z. Nadasdy, J. Hetke, K. Wise, and G. Buzsaki, J Neurosci **15**, 47 (1995).

[62] J. Csicsvari, H. Hirase, A. Czurko, and G. Buzsaki, Neuron **21**, 179 (1998).

[63] J. Csicsvari, H. Hirase, A. Czurko, A. Mamiya, and G. Buzsaki, J Neurosci **19**, 274 (1999).

[64] J. A. Henrie and R. Shapley, J Neurophysiol **94**, 479 (2005).

[65] N. Masuda and K. Aihara, Biol Cybern **90**, 302 (2004).

[66] J. Zhang and et al, J. Comp. Neurosci. **36**, 279 (2014).

[67] J. Zhang and A. V. Rangan, J. Comput. Neurosci. **38**, 355 (2015).

[68] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, Interdisciplinary Perspectives on Infectious Diseases **2011**, 1 (2011).

[69] E. M. Volz, J. C. Miller, A. Galvani, and L. Ancel Meyers, PLoS Comput Biol **7**, e1002042 (2011).

[70] E. Kenah and J. C. Miller, Interdisciplinary Perspectives on Infectious Diseases **2011**, 1 (2011).

[71] R. Durrett, *Random Graph Dynamics* (Cambridge University Press, New York, 2007).

[72] D. S. Callaway, M. E. J. Newman, S. Strogatz, and D. J. Watts, Phys Rev Lett **85** (2000).

[73] A. Hackett, S. Melnik, and J. P. Gleeson, Physical Review E **83**, 056107 (2011).

[74] M. E. J. Newman, Physical Review Letters **103**, 058701 (2009).

[75] A. Hackett, S. Melnik, and J. P. Gleeson, Physical Review E **83**, 056107 (2011).

[76] A. H. Nayfeh, *Introduction to perturbation techniques*, Wiley Classics Library (John Wiley & Sons, New York, 1981).

[77] W. Feller, *An introduction to probability theory and its applications*, 3rd ed., Vol. 1 (John Wiley & Sons, New York, London, Sydney, 1968).

[78] Y. G. Sinai, *Probability theory* (Springer-Verlag, Berlin, 1992).