Spike-Time Reliability of
Layered Neural Oscillator Networks

Kevin K. Lin[1], Eric Shea-Brown[2], and Lai-Sang Young[3]

[1] Department of Mathematics, University of Arizona,
[2] Department of Applied Mathematics, University of Washington
[3] Courant Institute of Mathematical Sciences, New York University

January 9, 2009

Abstract

We study the *reliability* of layered networks of coupled "type I" neural oscillators in response to fluctuating input signals. Reliability means that a signal elicits essentially identical responses upon repeated presentations, regardless of the network's initial condition. We study reliability on two distinct scales: *neuronal reliability*, which concerns the repeatability of spike times of individual neurons embedded within a network, and *pooled-response reliability*, which concerns the repeatability of total synaptic outputs from a subpopulation of the neurons in a network. We find that neuronal reliability depends strongly both on the overall architecture of a network, such as whether it is arranged into one or two layers, and on the strengths of the synaptic connections. Specifically, for the type of single-neuron dynamics and coupling considered, single-layer networks are found to be very reliable, while two-layer networks lose their reliability with the introduction of even a small amount of feedback. As expected, pooled responses for large enough populations become more reliable, even when individual neurons are not. We also study the effects of noise on reliability, and find that noise that affects all neurons similarly has much greater impact on reliability than noise that affects each neuron differently. Qualitative explanations are proposed for the phenomena observed.

Introduction

If the same sensory stimulus is presented multiple times to a neuronal network, how similar are the spike trains that it evokes? The answer impacts the fidelity of neural codes that are based on temporal patterns of spikes [44, 41]. This question of the reproducibility, or *reliability,* of stimulus-induced spike times has a long history in neuroscience.

Electrophysiology experiments have investigated the reliability of neural systems on different scales. *In vitro* experiments have established that single, synaptically isolated neurons are reliable under a broad range of conditions, *i.e.*, the spike times of an isolated neuron in response to

1

repeated injections of a fixed, fluctuating current signal tend to be repeatable across multiple trials [13, 37, 22]. At the network or systems level, experiments have been carried out in various preparations in which sensory stimuli (*e.g.*, aperiodic visual or auditory signals) are repeatedly presented, and recordings from single units repeatedly made. While many such studies find considerable reliability *in vitro* and *in vivo*, spike trains drawn from intact networks do appear less reproducible than for isolated neurons [9, 39, 14, 4, 35]. Moreover, repeatability has been shown to diminish for neurons that are sampled from areas more and more distant from sensory inputs [25].

Theoretical studies have found that models of isolated neurons tend to be reliable [57, 46, 51, 40, 22]; here, we undertake a corresponding study for large networks. Two of the known reasons that networks might display unreliability are *noise* and *dynamical instability*. Biological networks tend to be noisy. Specifically, neural networks experience synaptic variability and failures (synaptic noise) [20, 23]; these effects do not arise for direct current injections, as is often done in single-cell studies. Second, network-level interactions can lead to dynamical instabilities which are not present in isolated neurons, causing a network's response to depend sensitively on its internal state at the time of the stimulus is presented [53, 50, 33, 55]. We will attempt to address both of these issues in this paper.

From a theoretical standpoint, under what conditions is a network reliable? We answer this question for a class of neural oscillator networks that idealize a commonly occurring situation in neuroscience: *layered* networks [49, 16]. Specifically we consider *single-layer networks*, in which a group of similar, sparsely coupled neurons all directly receive the same fluctuating stimulus, and *two-layer networks*, in which the signal is received only by an initial "input layer" and transmitted to a second layer via recurrent connections. Individual cells are modeled as "Theta neurons," one-dimensional phase reductions of Type I neurons operating in mean-driven regimes [18]; this model is chosen primarily to facilitate numerical computation. Reliability both of individual neurons within a network and of their pooled responses are studied. We find that individual neurons can be reliable or unreliable depending on certain features of network architecture, such as their layered structure and the strengths of inter-layer and intra-layer connections. When the responses of large enough populations are pooled together, unreliability diminishes considerably. To make transparent the mechanisms involved in the determination of neuronal reliability, we first neglect the effects of noise, introducing it only later on. We propose to separate noise into two types, which we call "global" and "local" noise, and show that global noise is far more detrimental to reliability.

While we try to incorporate features that reflect the complexities of neurobiological networks, we have also made enough idealizations so that the entire model is characterized by a dozen or so parameters — stimulus amplitude, mean synaptic strengths, degree of heterogeneity *etc.* — putting it in a range where it is feasible to attempt to systematically relate reliability to network structure and system parameters. We view our neural oscillator network as a driven dynamical system. Since large networks give rise to systems with many degrees of freedom, a statistical approach is taken. We cast the problem in the framework of random dynamical systems theory, and seek to develop a qualitative understanding of how networks respond to fluctuating stimuli.

In terms of methodology, numerical simulations are used extensively throughout: for discovery purposes, to test mathematical conjectures, and to confirm the results of theoretical deductions based on earlier steps. Whenever we can, we provide theoretical explanations for the phenomena observed. Our arguments are, for the most part, not intended as rigorous proofs. While state-of-the-art dynamical systems theory provides many insights, the types of results discussed here lie considerably outside of the reach of rigorous mathematics.

Previous theoretical studies have addressed the question of reliability on the level of networks: see, *e.g.*, [53, 10, 36, 50, 33, 5, 6]. These and other references on topics related to our work will be discussed in the text as we go along. Some of the results of the present paper have been announced in a brief note [34].

1 Description of Model

In Sect. 1.1, we introduce the class of systems studied in this paper. In Sect. 1.2, we discuss how various model details are motivated by neuroscience.

1.1 Model details

We study pulse-coupled networks of phase oscillators described by equations of the general form [56, 19, 18]

$$\dot{\theta}_i = \omega_i + z(\theta_i) \cdot \left[\sum_{j \neq i} a_{ji}\, g(\theta_j) + \varepsilon_i I(t) \right], \qquad i = 1, \cdots, N \ . \tag{1}$$

The variables θ_i are the states of the phase oscillators, *i.e.*, they are angles parametrized by numbers in the interval $[0, 1]$ with periodic boundary conditions. The coupling matrix (a_{ji}) describes the network structure: for each (i, j), a_{ji} can be positive, negative or zero; $a_{ji} \neq 0$ means that oscillator i receives input directly from oscillator j. The phase response curve $z(\theta)$ measures how big an effect an input has on a neuron in state θ, be it an external stimulus or another neuron's synaptic output. Here, we take $z(\theta) = \frac{1}{2\pi}[1 - \cos(2\pi\theta)]$. This characterizes general oscillators near a saddle-node bifurcation on a periodic orbit, and corresponds to a generic "Type I" or "theta neuron" model (see, *e.g.*, [18]). Coupling between oscillators is mediated by a smooth pulsatile function g with the property that $g(\theta) > 0$ for $\theta \in (-\frac{1}{20}, \frac{1}{20})$, $g(\theta) = 0$ elsewhere, and $\int_0^1 g(\theta)\, d\theta = 1$. (The specific function we use is $g(\theta) \propto (1 - 400 \cdot \theta^2)^3$ for $|\theta| \leq \frac{1}{20}$.) The number N is arbitrary, and is taken to be $100 - 200$ in most of our simulations.

We restrict our attention to networks receiving a single input signal $I(t)$, which for simplicity we take to be a ("frozen") realization of white noise, *i.e.*, $I(t)\, dt = dW_t$ where W_t denotes standard Brownian motion. (We have found numerically that the addition of low-frequency components does not substantially change the outcome; see Sect. 3.1.)

An important aspect of this model is that there is a degree of *cell heterogeneity*, by which we mean variations in the parameters a_{ji} and ω_i, represented by $\rho \in [0, 1]$. Specifically, we assume the intrinsic frequencies ω_i in the network are chosen independently from the uniform distribution on the interval $\omega \cdot [1 - \rho, 1 + \rho]$ for some number ω. The coupling constants a_{ji}, which we discuss below, are also heterogeneous.

Two types of network structures are considered in this paper:

Single-layer networks. We set $\varepsilon_i \equiv \varepsilon$ for all i, so that all neurons receive the same input $I(t)$ at the same amplitude ε. The network topology is generated by (i) making each neuron receive input from exactly κ other randomly-chosen neurons, *i.e.* all vertices in the connection graph have in-degree κ; and (ii) requiring that the resulting network have no disconnected neurons or subnetworks. All the coupling constants a_{ji} are taken to be near a number a, which we call the *synaptic strength*;
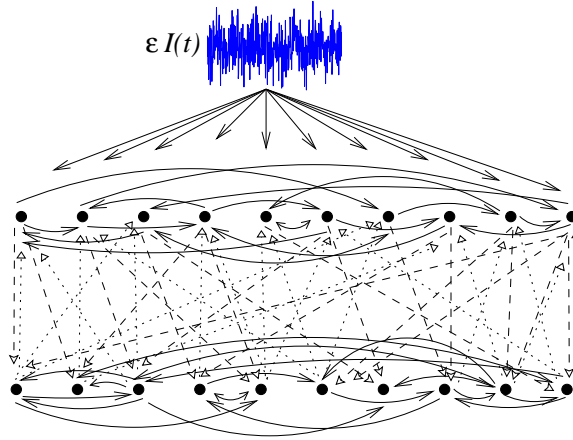
Figure 1: Schematic of the two-layer network, with $N = 20$ and $\kappa_1 = \kappa_2 = \kappa_{\text{ff}} = \kappa_{\text{fb}} = 2$.

specifically we take the a_{ji} to be independent samples from the uniform distribution on $a \cdot [1 - \rho, 1 + \rho]$ where ρ is as above.

Two-layer networks. We begin with two groups of neurons of size $\frac{N}{2}$ each; see Fig. 1. For neuron i in the first group, we set $\varepsilon_i \equiv \varepsilon$; this is layer 1. For each i in the second group, we set $\varepsilon_i \equiv 0$; this is layer 2. We first describe the dynamics within individual layers: Each neuron in layer 1 hears from exactly κ_1 neurons from within layer 1, with synaptic strengths $\approx a_1$. Layer 2 has a similar structure, the corresponding parameters being κ_2 and a_2, respectively. As for inter-layer activity, each neuron in layer 2 receives input from exactly κ_{ff} neurons in layer 1, with synaptic strengths $\approx a_{\text{ff}}$, while each neuron in layer 1 receives input from exactly κ_{fb} neurons from layer 2, with synaptic strengths $\approx a_{\text{fb}}$. (Here "ff" and "fb" refer to "feedforward" and "feedback".) As always, the actual coupling constants are randomly chosen to lie within $1 \pm \rho$ of the numbers given. Finally, we assume as before that the network has no disconnected components.

Notice that each choice of connection graph and system parameters, *i.e.*, a_{ji}, ω_i, and ε_i, determines a network. When studying the reliability of this network, these parameters will remain fixed, as will $I(t)$, and each *trial* corresponds to a randomly chosen initial condition in the system defined by (1). So far the effects of noise have not been taken into account. When we wish to simulate the effects of random noise on the system, separate white-noise terms that vary from trial to trial will be added; this is done in Sect. 6.

Glossary of notation and default values

"Variable" below means it is a quantity we intend to vary.

- N is system size or total number of neurons; variable, usually $100 - 200$
- ε is the amplitude of the input stimulus; variable
- ρ is the degree of cell heterogeneity; default value $= 0.1$
- ω is the mean frequency of the neurons; default value $= 1$

For single-layer systems:
- a is the synaptic strength; variable

4

- κ is the in-degree of the network as a directed graph, *i.e.* each neuron hears from κ others; default value $= 20\%$ of N

For two-layer systems:
- a_1, a_2 are synaptic strengths within layers 1 and 2; variable
- $a_{\text{ff}}, a_{\text{fb}}$ are ff/fb synaptic strengths; variable
- κ_1, κ_2 are numbers of cells from the from same layer that provide input to each neuron; default value $= 20\%$ of $N/2$
- $\kappa_{\text{ff}}, \kappa_{\text{fb}}$ are numbers of cells from the other layer that provide input to each neuron within a given layer; default value $= 20\%$ of $N/2$

1.2   Model assumptions in relation to neurobiological networks

Individual cells in our network are modeled by Theta neurons, which represent Type I neurons in intrinsically active, "mean-driven" firing regimes. This model assumes that in the absence of any input, each neuron possesses sufficient inward current to cause periodic firing with frequency ω_i [21, 45, 18]. The Theta neuron model used here can be derived from the quadratic integrate and fire model via a coordinate change (*e.g.*, [30, 18]).

We comment below on other model assumptions:

Some idealizations are made solely to limit the scope of our mathematical study and to keep the number of defining parameters small. These include:

(i) We consider isolated patches of neurons that form either one or two layers. Coupling architectures are idealized and not intended to model any specific part of the nervous system. Properties of all cells within a layer are assumed to be similar; in particular, they are either all excitatory or all inhibitory (to simulate the net effect of both types of neurons within a layer).

(ii) The network *topologies* we consider are relatively homogeneous: we have fixed the in-degree of each node to give better control of the input received by each neuron; see Sect. 3.1.

(iii) We consider a single sensory stimulus $I(t)$, which is received by all cells in layer 1 in the same way. This is to simulate a situation where the first layer is a small patch of cells "close" to the stimulus and with nearly identical receptive fields. Our assumption that $I(t)$ is a "frozen" white noise is a simplification justified by the fact that low frequency components have negligible effects; see Sect. 3.1.

(iv) The functions $g(\theta_j)$ represent coupling via pulsatile synaptic currents. The fact that $g(\cdot)$ is written as a function of phase rather than as a function of time following spikes is a simplification, but the approximation is good for brief impulses of synaptic current, as is sometimes done for medium-to-fast synapses. We neglect any effects due to conductance-based coupling, which can be incorporated by modifying the $z(\cdot)$ function. We also neglect asymmetries in rise and fall times of synaptic currents (but do not believe that this will make a major difference for the brief synaptic events considered here).

We have, on the other hand, attempted to incorporate various features that are motivated by the complexities of neurobiological networks. These include:

(i) The layered structure of our models reflects the layering that is typical both within and between different areas of the nervous system [49, 16]. Layered networks are also common in computational and *in vitro* (dynamic clamp) studies; see, *e.g.*, [43, 11, 52, 15].
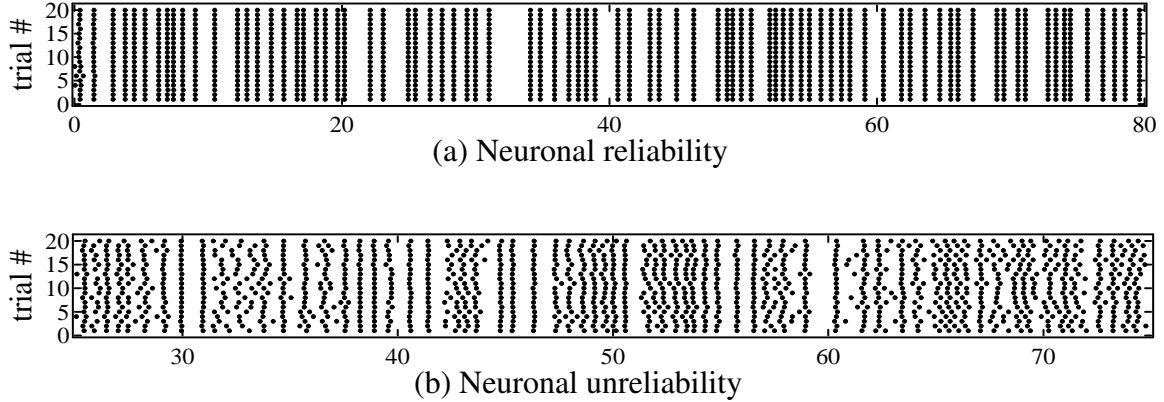
Figure 2: Raster plots showing the spike times of two neurons across 20 trials in response to a fixed stimulus $\varepsilon I(t)$. In (a), the neuron is chosen at random from a single-layer network with $N = 100$, $\kappa = 20$, $a = 0.05$, and $\varepsilon = 2.5$. In (b), the neuron comes from a two-layer network with $N = 100$, $\kappa_1 = \kappa_2 = \kappa_{\text{ff}} = \kappa_{\text{fb}} = 10$, $a_1 = a_2 = 0.1$, $a_{\text{ff}} = 0.28$, $a_{\text{fb}} = 0.25$, and $\varepsilon = 2.5$.

    (ii) As is typical in many nervous system areas, we consider sparsely coupled networks, typically with each neuron receiving inputs from 20% of the network.

    (iii) We allow up to 20% ($\rho = 0.1$) heterogeneity in the parameters describing individual neurons and their couplings.

    (iv) We study coupling strengths that are in biologically plausible ranges (*cf.* [27, 24, 49]). The following is a rough translation of the parameters used here to those that characterize the quadratic integrate-and-fire model of [30]: if individual frequencies are ≈ 10 Hz, then the duration of coupling impulses is roughly ≈ 10 ms, and a coupling strength of $a_{ij} = 0.3$ (the largest we consider for our typical case of $N = 100$) corresponds to a post-synaptic potential of roughly 1.5 mV.

    (v) In Sect. 6 we will study reliability in the presence of trial-to-trial and neuron-to-neuron fluctuations, in an attempt to simulate the noisy environment of biological neural networks.

2   Neuronal and Pooled-Response Reliability

In this paper, we treat the reliability of network responses to the stimulus $\varepsilon I(t)$ on two different scales: on the neuronal level, and on the level of pooled responses. We make precise here what is meant by these two kinds of reliability, and discuss how they will be measured.

2.1   Neuronal reliability and Lyapunov exponents

Neuronal reliability refers to the repeatability of spike times for individual neurons embedded in a network. Fig. 2 shows raster plots of spike times of two neurons drawn randomly from two networks with different parameters and configurations, for repeated presentations of a fixed stimulus $\varepsilon I(t)$. The plot in Fig. 2(a) shows *neuronal reliability*, *i.e.*, repeatable performance from trial to trial, following a brief transient. For the network in Fig. 2(b), neuronal reliability is not present: spike times differ persistently from trial to trial. We emphasize that in both cases, plotted are responses of a neuron embedded in a network. Single Theta neurons *in isolation* will not produce

raster plots like that in Fig. 2(b), for they are always reliable [46, 51].

Neuronal reliability is closely related to stability properties of the dynamical system defined by Eq. (1). Recall that given a dynamical system, be it autonomous or driven, the rates of divergence of nearby orbits are measured by numbers called *Lyapunov exponents*. Let $\lambda_{\max}(x)$ denote the largest Lyapunov exponent at a phase point x. Then typical pairs of orbits with initial conditions near x will tend to converge or diverge in time depending on whether $\lambda_{\max}(x) < 0$ or > 0. That is to say, $\lambda_{\max} < 0$ can be interpreted as a notion of stability, while $\lambda_{\max} > 0$ implies sensitive dependence on initial conditions and chaos.

A priori, Lyapunov exponents provide information only on nearby orbits, as they are defined in terms of growth rates of tangent vectors. But much more is true for stochastically driven systems such as that defined by Eq. (1), thanks to some very general results from random dynamical systems theory. First, for stochastically driven systems, under fairly general conditions, the numerical value of $\lambda_{\max}$ is independent of initial network condition or the particular realization of white noise (or stimulus). Thus it is truly a signature of the system, defined entirely by system parameters and stimulus amplitude.

Below we discuss in nontechnical terms some results from this theory that are directly relevant for us, referring the reader to Appendix A for a more technical mathematical discussion.

The sign of $\lambda_{\max}$ as test of reliability

It has been proved [31, 7] that under mild conditions, if $\lambda_{\max} < 0$, then trajectories starting from any ensemble of initial conditions will, after a transient, coalesce to a small blob which continues to evolve as effectively a single trajectory. That is to say, independent of the initial state of the system at the moment when the stimulus is turned on, the system will, after a period of adaptation, produce essentially identical spike patterns. This phenomenon, referred to as *random sinks*, represents the complete entrainment of the network dynamics to the stimulus, and hence the reliable response of every neuron in the network.

Conversely, if $\lambda_{\max} > 0$, then by a theorem in [32], trajectories of the system organize themselves around a messy and complicated-looking object in the phase space called a *random strange attractor*. These objects evolve with time, retaining their complex structures; in particular, they do not collapse to a point. This means that at any given point in time, the network may be in many different states depending on its initial condition, *i.e.*, spike-time recordings will differ depending on the system's initial state. This is exactly what it means for a system to be unreliable.

To summarize, the dichotomy in dynamical picture discussed above leads to the following *Lyapunov exponent test*: If $\lambda_{\max} < 0$, then the system is reliable. If $\lambda_{\max} > 0$, then it is unreliable. In the case $\lambda_{\max} = 0$, this test is inconclusive. Similar ideas have been used in, *e.g.*, [50, 33, 40, 42].

The magnitude of $\lambda_{\max}$ as measure of degree of reliability

Reliability, however, is generally viewed as a relative notion, *i.e.*, some systems are "more reliable" or "less reliable" than others. We claim that the *magnitude* of $\lambda_{\max}$ carries useful information. There are no theorems to cite here, but ideas underlying the results discussed above tell us that other things being equal, the more negative $\lambda_{\max}$, the stronger the tendencies of trajectories to coalesce, hence the more reliable the system. Conversely, for $\lambda_{\max} > 0$, the larger its magnitude, the greater the instability, which usually translates into greater unreliability. There are two caveats, however, when interpreting the magnitude of $\lambda_{\max}$: Since $\lambda_{\max}$ is a quantity that involves averaging

over time, it may not reflect the initial response of the network upon presentation of the stimulus; and second, it is not clear how one should compare the magnitudes of $\lambda_{\max}$ for two systems that are very disparate in size (nor is it clear how to compare their reliability).

For purposes of comparing systems with similar geometry and size, and for stimuli that are presented for a reasonable duration (depending on the system), we have found the magnitude of $\lambda_{\max}$ to be a very good indicator of the degree of reliability.

Lyapunov exponents as a theoretical tool

A very obvious question is: Since reliability measures across-trial variability, why does one not simply perform a number of trials (using the same input), and compute their across-trial variances? We give three reasons in favor of $\lambda_{\max}$ as a tool in *theoretical investigations* — without asserting that it is better for all circumstances:

(1) Consider a network of size N. Are we to carry out the above procedure for a single neuron, a few neurons, or for all N of them? Keeping track of n neurons will require treating variances of n-component functions. Also, since these variances vary with time, do we plot their time courses, or average them? A virtue of using $\lambda_{\max}$ is that it sums up the stability property of a system in a very compact way. When plotted as a function of parameters it enables us to view at a glance the entire landscape, and to identify emerging trends.

(2) For reliable systems, a variance computation will yield 0, while the magnitude of $\lambda_{\max}$ tells us how reliable the system is. This information can be useful: for example, systems that are more reliable are likely to have greater tolerance for system noise.

(3) A third advantage of using $\lambda_{\max}$ is that known mathematical properties of Lyapunov exponents can be leveraged. Under quite general conditions, $\lambda_{\max}$ varies continuously, even smoothly, with parameters. For example, if $\lambda_{\max}$ is found to be very negative for a system, then it is likely to remain negative for a set of nearby parameters; the size of this parameter region can sometimes be estimated with knowledge of how fast $\lambda_{\max}$ is changing. Knowing that a system has zero across-trial variance alone will not yield this kind of information.

2.2 Pooled-response reliability

Let $\mathcal{C}$ denote a collection of some or all of the N neurons in the network. To measure the total output of the subpopulation defined by $\mathcal{C}$ in response to $\varepsilon I(t)$, we introduce the "synaptic output function"

$$S_{\mathcal{C}}(t) = \sum_i f_{syn}(t - T_i) \,.$$

Here T_i are the spike times of *any* neuron in $\mathcal{C}$, and f_{syn} is a postsynaptic current modeled by the one-sided exponential

$$f_{syn}(t) = \begin{cases} \tau^{-1} \exp(-t/\tau) \,, & t \geq 0 \,; \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{2}$$

Similar output functions are commonly studied – see, *e.g.*, [38, 54]. Note that $\int_0^\infty f_{syn}(t)\, dt = 1$. The synaptic time constant τ is set to $\frac{1}{15}$ in this paper; if neurons have a natural frequency of ≈ 10

Hz, this corresponds to $\approx$ 7 ms. *Pooled-response reliability* assesses how repeatable the time-course of $S_{\mathcal{C}}(t)$ is from trial to trial. A measure of this repeatability is the variance of $S_{\mathcal{C}}(t)$ across a large number of trials.

In Sects. 3 and 4 below, we will study, via Lyapunov exponents, neuronal reliability for one- and two-layer networks. Pooled-response reliability is discussed in Sect. 5.

3 Neuronal Reliability of Single-layer Networks

This section is devoted to a numerical investigation of the Lyapunov exponents $\lambda_{\max}$ of single-layer networks as defined in Sect. 1, specifically the dependence of $\lambda_{\max}$ on network parameters. The purpose of Sect. 3.1 is to identify which features of the system most seriously affect neuronal reliability. We will see that single-layer networks are highly reliable; qualitative explanations are given in Sect. 3.2.

3.1 Dependence on system parameters

As we will show, it is productive to view $\lambda_{\max}$ as a function of the quantity $A = \kappa a$, which has the interpretation of being the *a priori synaptic kick amplitude* experienced by each neuron.[1] To see this, we focus on an arbitrary neuron. Since $\omega = 1$, we expect each of the κ neurons that provide input to it to spike once per unit time, with average synaptic strength a. The impact of these spikes depends on the phase of the receiving neuron via the phase response curve z, but without further information about the network's dynamics, we take the average $\langle z \rangle = \frac{1}{2\pi}$ as a representative value. Therefore, in each unit of time, we expect the phase of each neuron to be perturbed by $\frac{\kappa a}{2\pi}$ due to the coupling terms, the effect of which is to cause it to travel (forwards if $A > 0$ and backwards if $A < 0$) an *extra* $\frac{A}{2\pi}$ fraction of a lap per unit time.

Four different aspects of how $\lambda_{\max}$ varies with system parameters are discussed in paragraphs A–D below.

A. Basic relationship between $\lambda_{\max}, \varepsilon$ and A. Recall that ε is the stimulus amplitude. The qualitative relation between these 3 quantities is illustrated in Fig. 3. Notice that A and ε alone do not determine a system; there are many other variables (such as specific choices of connection graphs, a_{ji}, *etc.*) The different curves in Fig. 3 correspond in fact to different choices of these parameters. For the moment, however, let us suppress these differences, and focus on the strikingly similar features of the three plots.

In each of the 3 figures, the curves can be divided in an obvious way into 2–3 groups. For example, in Fig. 3(a), there are 3 groups with 2 curves in each. The curves within each group share the same value of ε but differ in other details. Let us refer to each group as a single curve for now. Viewed this way, Fig. 3(a) shows 3 curves with $\varepsilon = 0.5, 1.5, 2.5$; all other system parameters are fixed except for a, which is chosen to give the values of A in the x-axis. Similarly, Fig. 3(b),(c) each

---

[1] "*A priori*" here refers to one's best guess based on system parameters alone without further knowledge of the dynamics. For example, as the system evolves, each neuron will, in time, acquire a mean frequency, which is likely to be different than its intrinsic frequency. But before studying the dynamics of the system – that is, *a priori* – there is no information on how the two will differ. So we take its *a priori* value to be equal to its intrinsic frequency.
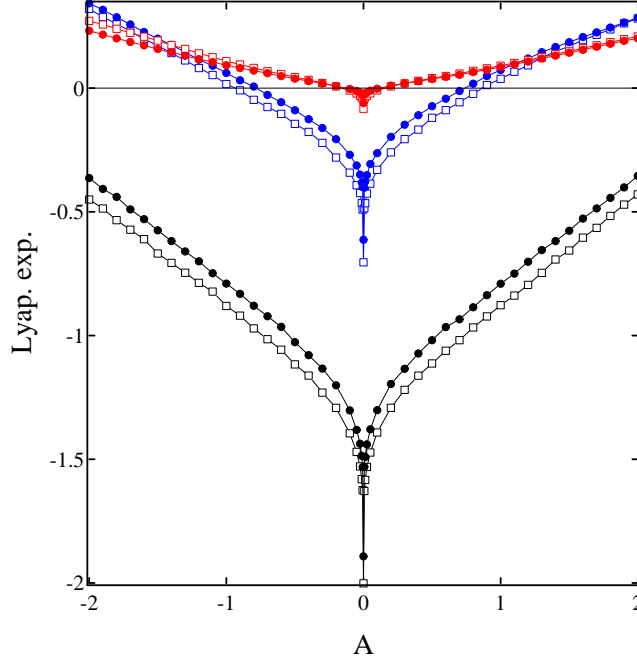
(a) $\varepsilon = 0.5$ *(top), 1.5 (middle), 2.5 (bottom)*
$N = 100, \kappa = 20$
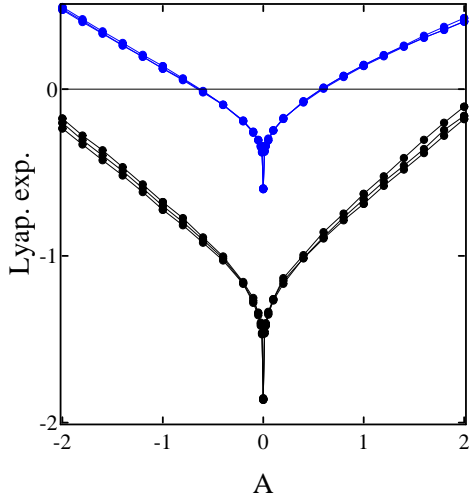
(b) $\varepsilon = 1.5$ *(top), 2.5 (bottom)*
$N = 100, \kappa = 10$
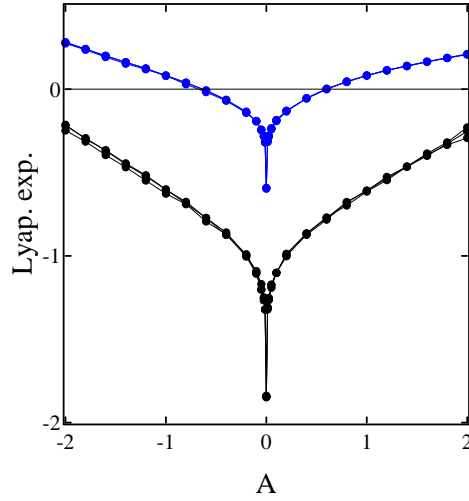
(c) $\varepsilon = 1.5$ *(top), 2.5 (bottom)*
$N = 400, \kappa = 80$

Figure 3: Largest Lyapunov exponent $\lambda_{\max}$ versus $A = \kappa \cdot a$ for single-layer networks. In (a), the input $I(t)$ is white noise plus a sinusoidal component of amplitude c and frequency $w = 2$; see Eq. (3). Filled circles: $c = 0$; open squares, $c = 2$. All six curves in (a) use the same network connectivity graph. Panels (b) and (c) show the corresponding results for $c = 0$. Three different realizations of connection graphs are used for each choice of ε in (b) and (c).

shows 2 curves with $\varepsilon = 1.5, 2.5$. Plots for other values of ε are not shown, but they interpolate between these graphs in a straightforward way. Our main findings are:

(1) *When $A = 0$, $\lambda_{\max} < 0$.*
(2) *In general, $\lambda_{\max}$ can be positive or negative. When ε is fixed, $\lambda_{\max}$ increases with $|A|$. For fixed A, $\lambda_{\max}$ decreases with ε (provided ε is not too small).*

The first observation is entirely expected, as the situation is identical to that of single phase oscillators, for which it is well known that $\lambda_{\max} < 0$ [51, 46, 26, 31]. Observation (2) suggests a competition between the entraining effects of the stimulus and the destabilizing effects of the coupling: increasing stimulus amplitude (with other parameters fixed) leads to greater reliability, while increasing coupling strength leads to less reliable responses. The system will be reliable or unreliable depending on the outcome of this competition. Moreover, all seems quite independent of the excitatory or inhibitory nature of the neurons.

This competition idea appears to be valid for as long as the stimulus is not too weak (relative to A). We do not discuss the case of weak stimuli in this paper. Reliability properties of such systems are strongly influenced by intrinsic network properties, *i.e.* properties of the network when $\varepsilon = 0$. This is an interesting topic in its own right, one that would, however, take us too far afield (see, *e.g.*, Sect. 6 of [33]). The $\varepsilon = 0.5$ curve in Fig. 3(a) is in fact there to illustrate a different phenomenon (see paragraph B); we will generally consider only $\varepsilon > 1$.

B. Fluctuating nature of the input $I(t)$. Plot (a) sums up the results of a study to determine the degree to which the nature of $I(t)$ matters. For example, do high or low frequency stimuli have stronger entraining effects, and how important is the fact that it is *rapidly fluctuating* (as opposed to being a smooth function)? To study this, we set

$$I(t)\,dt = \varepsilon dW + c\sin(wt)\,dt\,, \tag{3}$$

taking $w = 2$ and letting c vary between 0 and 2. When $c \neq 0$, the second term is a stimulus component that varies on a similar timescale as the interval between a neuron's spikes. The graphs marked by open squares in Fig. 3(a) show that this second component has little effect on $\lambda_{\max}$ even for $c = 2$, which is significant relative to the white noise amplitudes ε used in this figure. Moreover, the low frequency component alone does not have any entraining effects on the system, as evidenced by the case where $\varepsilon = 0.5, c = 2$. These results are representative of a number of different parameter combinations (including different drive frequencies w) investigated. We summarize the preceding discussion as

Rapidly fluctuating components in the stimulus have far greater effects on neuronal reliability.

For simplicity, we set $c = 0$ in the rest of this paper, taking $I(t)$ to be a white-noise stimulus.

Remark: We also investigated the effects of other fluctuating inputs such as Poisson processes of impulses at various rates. The results can be summarized as follows: On a purely qualitative level, the graphs of $\lambda_{\max}$ as functions of A are similar to those in Fig. 3; quantitative values of $\lambda_{\max}$ depend strongly on the specific type and parametrization of the input; a wide range of numerical values of $\lambda_{\max}$ was observed. We refrain from comparisons, since there is no reasonable way to match the magnitudes of different types of forcing.

11

C. Scaling of N, κ **and** a**.** These 3 parameters must be considered together. First, κ should be seen as a fraction of N; we will refer to this fraction as the *connectivity* of the network. In this paper, we focus on sparse connectivity (as explained in Sect. 1.2), using a default value of 20% in simulations. For our results to be meaningful, we need to confirm that the basic relationship in paragraph A is robust for a range of connectivity. Our simulations confirm this and more: they show that even the numerical values of $\lambda_{\max}$ remain fairly stable for connectivity in the range $10 - 30\%$, increasing only slightly toward the 10% end. A sample of our results is shown in Fig. 3(b).

Next we fix a value of connectivity, say at 20%. As N increases, the number of neurons whose synaptic outputs directly impact a specific neuron increases. If the value of a is kept fixed, the total perturbation from these couplings will be larger and larger. If our network dynamics are to approach a meaningful limit as $N \to \infty$, it is necessary to stabilize the total input received by a typical neuron, and the only way to do this is to compensate by decreasing a. This explains our choice of $A = \kappa a$ as scaling parameter. A sample of our results for larger N is shown in Fig. 3(c).

Exact numerical values of $\lambda_{\max}$ may vary, but the basic relationship between $\lambda_{\max}, \varepsilon$ and A do not depend on connectivity or system size.

These results imply that other things being equal, many weak synaptic inputs have roughly the same effect as fewer inputs with strengths that compensate for the numbers – provided the numbers are not too small. Our intuition (supported by limited simulations) is that keeping connectivity fixed, a limiting behavior will be reached as $N \to \infty$. This, however, is beyond the scope of the present paper.

D. Detailed network architectures. Even as we fix N, κ and a, there are many possibilities for admissible connection graphs (we require only that the network not separate into disjoint components), and for each connection graph, values of a_{ji} (as well as ω_i) may vary up to 20% at $\rho = 0.1$. The dynamics in systems resulting from these choices are not equivalent. The question is: to what degree do they affect $\lambda_{\max}$? The answer is: *very little*. In Fig. 3(b) and (c), three $\lambda_{\max}$ curves are plotted for each choice of ε, with each curve corresponding to a distinct connection graph. There are small differences in $\lambda_{\max}$, especially for larger ε, but the variations are negligible. These results are consistent with those of many other simulations not shown. We summarize as follows:

Once κ, a and N are fixed, other aspects of detailed network architecture do not have a significant effect on neuronal reliability.

This completes our discussion of dependence on parameters.

We finish with an observation that begs for an explanation: for $1 \leq A \leq 2$, for instance, the *a priori* kick amplitude estimate at the beginning of this subsection tells us that each neuron is expected to go an extra $\frac{1}{6}$ to $\frac{1}{3}$ of a lap per unit time due to excitation by other neurons. This is a very large perturbation, one that should lead to a great deal of instability in the dynamics. Yet, $\lambda_{\max}$ is robustly negative at $\varepsilon = 2.5$ and $A = 2$. That is to say, the systems under consideration are somehow more reliable than expected.

3.2 Why are single-layer networks so reliable?

The following simple reasoning may give some insight into the situation:

Start with the situation where $A = 0$ (zero coupling) and $\rho = 0$ (zero cell heterogeneity). In other words, we have a collection of identical neurons that are not aware of each other's presence, all responding to the same input. The system is clearly reliable, as the situation is equivalent to that of a single neuron and single neurons are reliable. Moreover, the neurons are *synchronized* after some period of time, since in the dynamical system corresponding to a single neuron, trajectories starting from all initial conditions coalesce into one.

Now allow ρ and $|A|$ to be slightly positive. From Sect. 3.1, we see that the system will be reliable, but perfect synchrony is lost: perfect synchrony is not possible no matter what A is for $\rho > 0$. Since this is a small perturbation of the $\rho = A = 0$ case, it is not hard to imagine that the neurons will remain *nearly synchronized* most of the time.

Continuing to keep ρ very small, we now increase $|A|$ gradually, and argue that the effects of the coupling cannot be felt until $|A|$ becomes quite large. The reasons are as follows: In near-synchronous states, the effects of synaptic inputs to a neuron are severely diminished. This is because a neuron spikes when its phase θ is near 0, and the phase response curve z has the property that $z(0) = z'(0) = 0$. Thus, if neuron i hears from neuron j and $\theta_j = 0$ at approximately the same time that $\theta_i = 0$ (as is typically the case for networks in near-synchronous states), then the kick from neuron j to neuron i will have little impact. Whatever destabilizing effects couplings may have on a network's dynamics is thus attenuated in near-synchronous networks, and once the system comes close to synchrony, the tendency is to remain near-synchronous.

For larger ρ, one would expect the system to wander farther away from the near-synchronous states, and to do so more often. Hand in hand with that is the increased impact of synaptic events. When ρ and A become large enough, this will lead to unreliability as the results in Sect. 3.1 imply.

We now use numerics to confirm that at $\rho = 0.1$ and not-so-small $|A|$, the neurons remain nearly synchronized, as one might extrapolate from our theoretical understanding of the situation for small ρ and $|A|$. Recall that ρ measures heterogeneity in ω_i and a_{ji} (but not in κ, the in-degree to a neuron in the connection graph), and $\rho = 0.1$ means that ω_i and a_{ji} may vary up to 20%. For definiteness, we focus on $A = 1$.

To quantify the degree of synchrony of a state, we introduce a measure d_{sync} defined to be the minimum mean-square distance of the network state $\Theta = (\theta_1, \cdots, \theta_N)$ from a completely-synchronized state. That is, d_{sync} is the distance from Θ to the nearest state on the "diagonal" $\{(\theta_1, \cdots, \theta_N) : \theta_1 = \cdots = \theta_N\}$. Thus, $d_{\text{sync}} = 0$ means complete synchrony, and the larger this number, the more spread out the phases of the neurons in the system. (In practice, we discard 5% of the outliers before computing mean-square distances, because these values tend to unduly bias the outcome.)

Fig. 4 shows snapshots of the phases in a single-layer network for different values of d_{sync}; the phases are plotted relative to a mean phase $\langle \theta \rangle$. The d_{sync} values shown correspond approximately to the 10th, 30th, ..., 90th percentiles of computed d_{sync} values for this network. These snapshots show a picture of partial synchrony: the phases of the neurons are localized to bands of varying widths much of the time. By examining d_{sync} as a function of time (data not shown), one can see that the system synchronizes, desynchronizes, and regroups, over and over in a process that is stationary in time.

Table 1 shows the median value of d_{sync} and the corresponding value of $\lambda_{\max}$ for a selection of cell heterogeneities ρ.[2] (The snapshots in Fig. 4 can be used to calibrate d_{sync}.) As predicted earlier,

---

[2]The d_{sync} values in Table 1 are computed over 1.6×10^4 time units. This is the reason for the discrepancy with
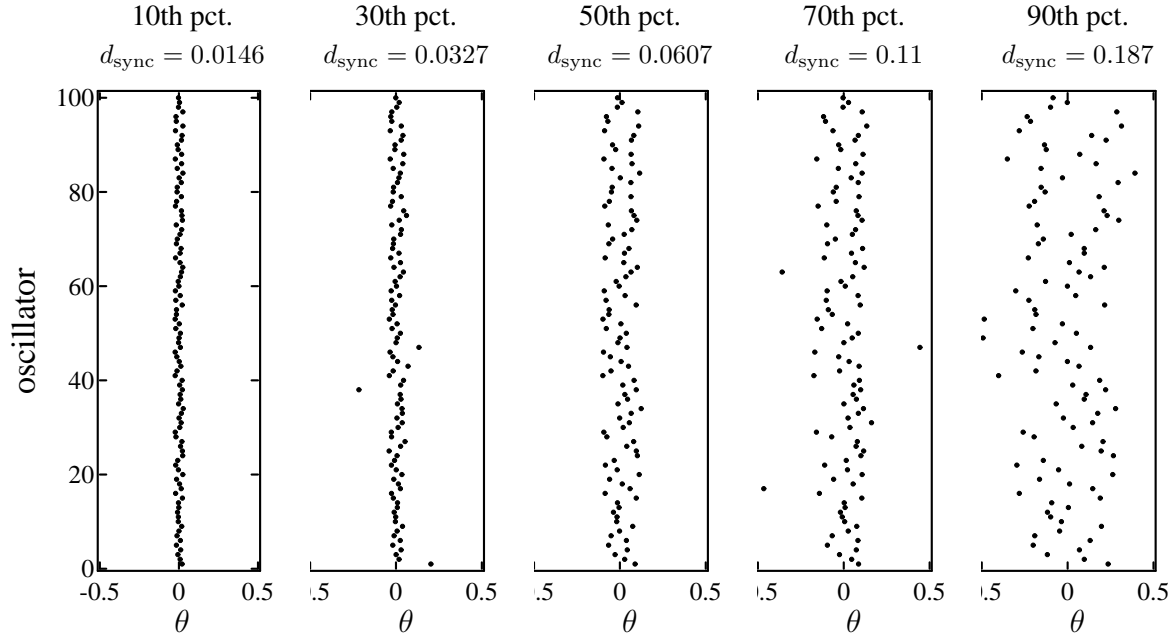
Figure 4: Snapshots of phase distributions illustrating different values of d_{sync}. We plot the phases of neurons in a single-layer network relative to a mean phase $\langle \theta \rangle$. The d_{sync} values correspond to the indicated percentiles for this network computed over 200 time units. Here, $N = 100$, $\kappa = 20$, $\rho = 0.1$, $A = 1$, and $\varepsilon = 2.5$.

the system is practically synchronous for small ρ, and the degree of synchrony steadily decreases as ρ increases. As a result of the increased sensitivity, the degree of neuronal reliability as measured by λ_{max} decreases simultaneously. However, we see that the system retains partial synchrony as well as substantial reliability even for $\rho = 0.3$, where intrinsic frequencies and coupling constants can vary as much as 60%. The data shown are for a fixed value of A. We remark that unlike λ_{max}, d_{sync} values do not vary a great deal with A; they are mostly determined by ρ.

	Heterogeneity ρ			
	0	0.01	0.1	0.3
Median d_{sync}	0	0.0046	0.050	0.12
λ_{max}	-1.9	-1.7	-0.70	-0.18

Table 1: Median value of d_{sync} versus the degree of cell heterogeneity ρ. Here, $N = 100$, $\kappa = 20$, $A = 1$, and $\varepsilon = 2.5$.

Summary of Sect. 3: *For models of the type described in Sect. 1, we find that in single-layer systems driven by a single stimulus, the quantity A (representing the a priori synaptic kick amplitude received by a neuron) is a natural scaling parameter, and neuronal reliability is largely determined by the competition between ε and A. A salient feature of these systems is that they are considerably more reliable than the parameter A would indicate. This is explained by the tendency of the neurons to synchronize in response to a common stimulus. Partial synchrony is observed even for*

---

the median d_{sync} value for $\rho = 0.1$ in Fig. 4, which were computed over 200 time units.

14

significant cell heterogeneity ρ. Two factors that contribute directly to these findings are (i) the form of the PRC we use, and (ii) the fact that the inputs received by all neurons are identical.

Related works. Several other studies have found that coupling can lead to unreliability in randomly connected networks, for different neural models or different types of inputs from those studied here. Perhaps the closest study to ours is [50], which also discusses relations of the type in paragraphs A and C of Sect. 3, including competition between coupling and stimulus effects. This study also uses Lyapunov exponents to assess reliability of neural oscillator networks. The neural oscillators in that paper are driven by independent stimuli, however, producing quite different dynamics from those in the present setting. For discrete-state networks, [53, 54, 10, 36] also describe how recurrent connectivity can create instabilities related to those observed here in large, randomly connected networks. Additionally, an abstract modeling framework is used in [5] to study the effects of coupling on reliability, and is applied to models of spike-response and conductance-based integrate-and-fire neurons receiving statistically independent inputs in [6]. Specifically, unreliable dynamics were found for networks of excitatory and inhibitory cells designed to represent firing regimes in the cortex.

Another related work is [8], which studies how network interactions can improve the reliability of entrainment to purely periodic stimuli in certain frequency ranges. This result, therefore, is opposite to the trend found here; the difference could be due in part to the periodicity of the stimuli, which can interact in a "resonant" manner with the subthreshold dynamics of single cells in a manner that differs from the white noise stimuli we consider here. The role of the frequency content of the stimulus $I(t)$ on neuronal reliability – paragraph B above – was also studied in the context of isolated cells by [22], where it was found that stimulus components at a neuron's firing frequency tended to produce reliability.

Finally, we note that synchronization is a much studied topic with too large a literature to systematically acknowledge. We simply reference [8], which finds that increased reliability is accompanied by increased synchrony in network responses to periodic inputs, as well as the book [42] and references therein.

4 Neuronal reliability of two-layer networks

This section contains a study of neuronal reliability via $\lambda_{\max}$ for two-layer systems. Specifically, we will study feedback from a second layer and how this affects reliability. Sect. 4.1 is an overview of the landscape (in a manner analogous to Sect. 3.1), and in Sect. 4.2, we attempt to explain what we regard as the most striking feature for these networks.

4.1 Dependence on system parameters

As with single-layer networks, we will express $\lambda_{\max}$ in terms of

$$A_1 = \kappa_1 a_1, \quad A_2 = \kappa_2 a_2, \quad A_{\text{ff}} = \kappa_{\text{ff}} a_{\text{ff}}, \quad \text{and} \quad A_{\text{fb}} = \kappa_{\text{fb}} a_{\text{fb}},$$

with the same interpretations as before. For example, A_{ff} is the *a priori* total kick per unit time received by each neuron in layer 2 from neurons in layer 1. In the absence of any knowledge of network dynamics, these kicks alone (not counting kicks from its own layer) cause the neuron in

15

layer 2 to travel an additional fraction $\frac{A_{\text{ff}}}{2\pi}$ of a lap per unit time. Similarly, A_{fb} is the *a priori* total kick per unit time received by each neuron in layer 1 from neurons in layer 2, and so on.

For two-layer networks, we will generally fix $\varepsilon = 2.5$ and set $|A_1| = |A_2| = 1$, with A_1 and A_2 having the same signs as A_{ff} and A_{fb}, respectively. The numerical values chosen are somewhat arbitrary, but well within the range that gives neuronal reliability in single-layer systems (see Fig. 3). Notice that whether a connection $j \to i$ is excitatory or inhibitory is determined solely by j, *i.e.* neurons are always excitatory or always inhibitory, as they generally must be.

We start by plotting $\lambda_{\max}$ as a function of A_{ff} and A_{fb} for a prototypical network. The results are shown in Fig. 5(a). We limit our discussion to the region $|A_{\text{ff}}| \gtrsim 1.5$, where the signals received by layer 1 are transmitted to layer 2 with some strength. To us, by far the most striking observation from Fig. 5(a) is

Two-layer networks with any appreciable feedback are neuronally unreliable.

At $A_{\text{fb}} = 0$ (and taking A_2 to be either 1 or -1), the system is purely feedforward, so the behavior of layer 1 is identical to that of a single-layer system at $|A| = 1$. At such values of A, the single layer system has been shown to be quite robustly reliable (Fig. 3, bottom curves). For $A_{\text{fb}} = 0$, Layer 2 in itself also behaves like a single-layer system, except that unlike layer 1, the input it receives is not white noise but the outputs of layer 1. Evidently, when $|A_{\text{ff}}|$ is large enough, the second layer is reliable with this input and $|A_2| = 1$, *i.e.*, $\lambda_{\max} < 0$ along the $A_{\text{fb}} = 0$ axis in the panel in Fig. 5(a).

As one moves away from the $A_{\text{fb}} = 0$ axis, $\lambda_{\max}$ increases sharply. Details do not show clearly in Fig. 5(a); they are given in Fig. 5(b). There, we plot $\lambda_{\max}$ as a function of A_{fb} for $A_{\text{ff}} = 2.8$, representing excitatory feedforward synapses between the layers. (When $A_{\text{fb}} = 0$, we plot $\lambda_{\max}$ for both $A_2 = 1$ and $A_2 = -1$.) Notice that for $A_{\text{ff}} > 0$, the system is somewhat more unreliable for $A_{\text{fb}} > 0$ than $A_{\text{fb}} < 0$. A possible explanation might be that for the neurons in layer 1, the effects of kicks from layer 2, which are in the opposite direction, may cancel some of the effects of kicks from its own layer.

Fig. 5(b) confirms the neuronal unreliability of the network for all parameters considered except for a very small range of A_{fb} when $A_{\text{ff}} = 2.8$. The plot contains the results for multiple choices of connection graphs, demonstrating again that $\lambda_{\max}$ is almost completely independent of details at this level.

Finally, to confirm again that the scalings in paragraph C of Sect. 3.1 make sense, we plot $\lambda_{\max}$ as a function of N for $A_{\text{ff}} = 2.8$, $A_{\text{ff}} = 2.5$ in Fig. 5(c). As noted in Sect. 2, we do not know how to interpret the magnitude of $\lambda_{\max}$ in terms of levels of spike time variability across such a wide spectrum of network sizes, but these results show that the system remains unreliable for a wide range of N.[3]

4.2 Why can't two-layer networks tolerate feedback?

Comparing the data presented in Sects. 3.1 and 4.1, what stands out the most to us is the contrast between the *reliability* of the single-layer system and the *unreliability* of the two-layer system with

---

[3]Inspection of raster plots at $N = 1000$, for example, clearly shows substantial trial-to-trial variability, with an interesting effect: the variability of single neurons tends to wax and wane over time, and this waxing and waning is differently timed for different neurons.
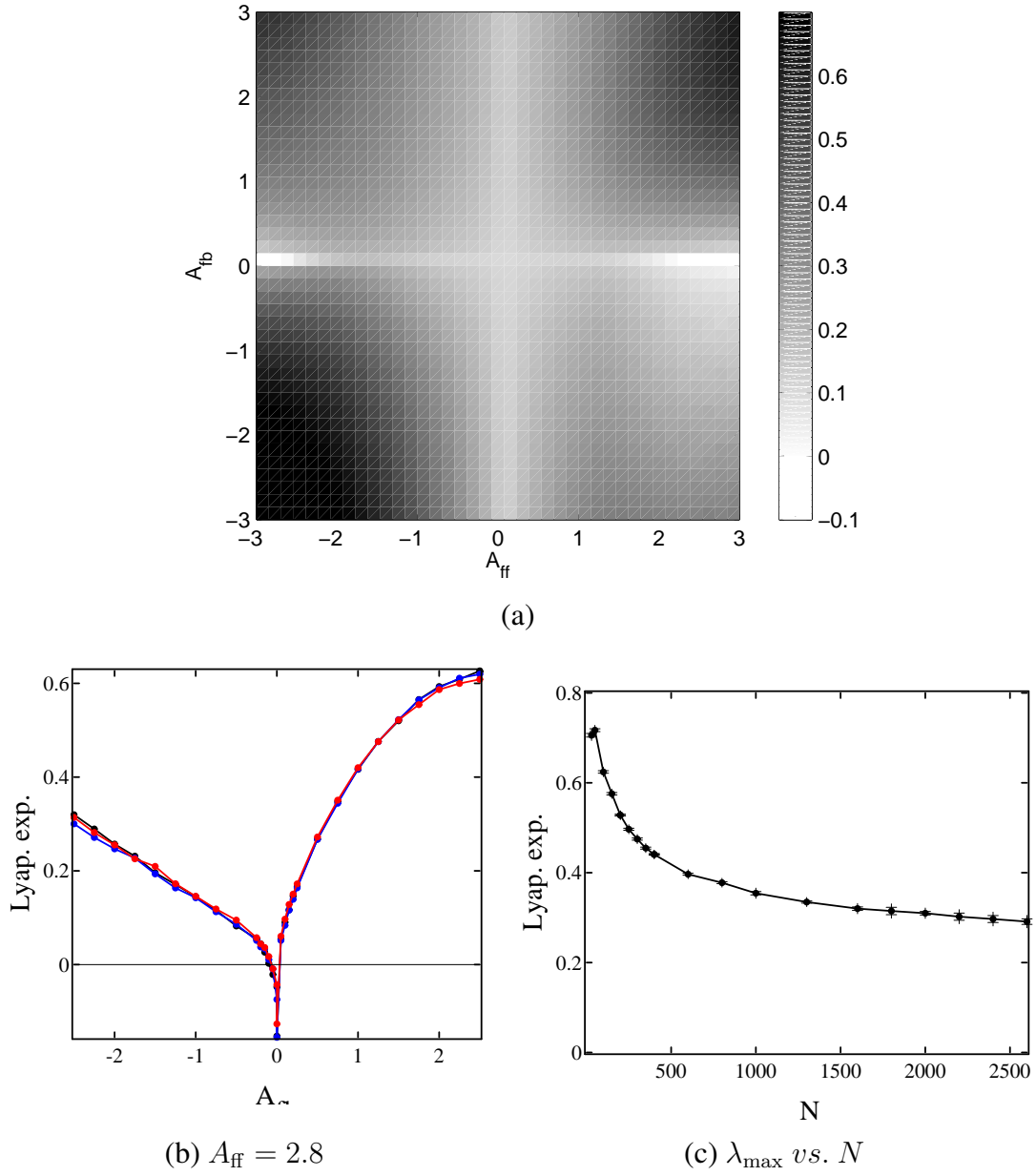
(a)

(b) $A_{\mathrm{ff}} = 2.8$

(c) $\lambda_{\max}$ *vs.* N

Figure 5: The Lyapunov exponent $\lambda_{\max}$ of the two-layer network. Panel (a) shows $\lambda_{\max}$ *vs.* A_{ff} and A_{fb} for $N = 100$; panel (b) displays the slice $A_{\mathrm{ff}} = 2.8$. Panel (c) shows $\lambda_{\max}$ *vs.* N for $5 \leq N \leq 2600$, with $A_{\mathrm{ff}} = 2.8$ and $A_{\mathrm{fb}} = 2.5$. In all plots, $\varepsilon = 2.5$.
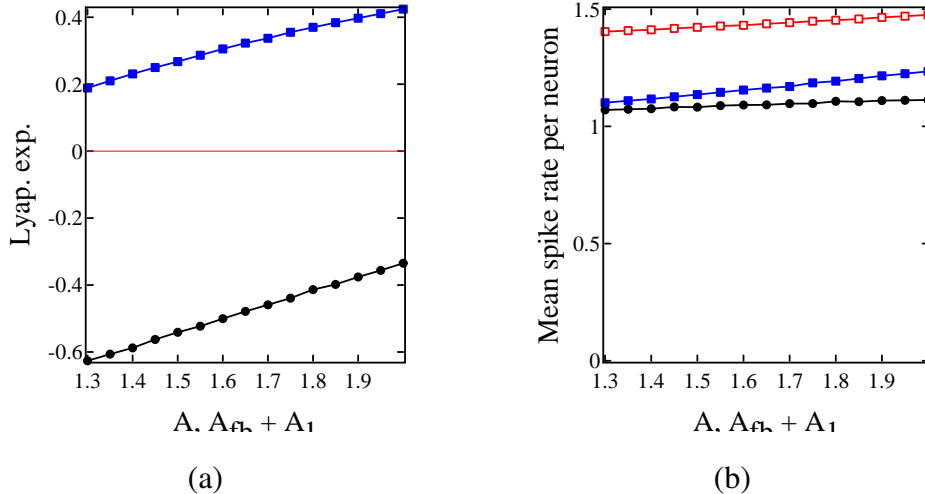
Figure 6: Comparison of single-layer network with layer 1 in a two-layer network. In (a), we plot λ_{max} versus A for a single layer network *(discs)* and λ_{max} versus $A_1 + A_{\text{fb}}$ *(squares)* for a two-layer network. In (b), we plot the mean firing rate r for neurons in the single-layer network *(solid circles)* against the firing rates r_i for layer i neurons $(i = 1, 2)$ in the two-layer network *(solid and open squares, respectively)*. Here, we set $N = 100$ and $\varepsilon = 2.5$ for both networks. For the single layer, $\kappa = 20$. For the two-layer, $\kappa_1 = \kappa_2 = \kappa_{\text{ff}} = \kappa_{\text{fb}} = 10$.

feedback at comparable values of ε and summed synaptic strengths. The purpose of this subsection is to propose a systematic explanation of this phenomenon based partly on theory and partly on numerical evidence.

For concreteness, we fix a family of single-layer networks and a family of two-layer networks, with parameters chosen to be comparable in a sense to be explained. For single-layer networks, we let $A \in [1.3, 2]$, and for two-layer networks, we fix $A_1 = A_2 = 1$, $A_{\text{ff}} = 2.8$, and let $A_{\text{fb}} \in [0.3, 1]$; we use $\varepsilon = 2.5$ for both networks. We will focus on comparing the single-layer network to layer 1 of the two-layer network (sometimes referred to in abbreviated form as "single layer" and "layer 1," respectively). The choices above give the same range of summed synaptic strengths for neurons in both cases, *i.e.*, A and $A_1 + A_{\text{fb}}$ both lie in the range $[1.3, 2]$, and they both receive directly a stimulus of the same amplitude; this is what we meant by the two systems being "comparable".

Before proceeding further, observe that the neuronal reliability of layer 1 is captured in *projected* Lyapunov exponents, *i.e.*, the rates of separation of typical pairs of trajectories in the two-layer system, measuring their distances only after projecting onto the coordinates corresponding to layer 1. It can be shown that the positivity of projected Lyapunov exponents corresponds to neurons in layer 1 producing unrepeatable responses depending on their initial states. We find (numerically) that for the parameters in question, these numbers are positive and equal to λ_{max}, the usual Lyapunov exponent of the two-layer system. That is to say, the unreliability of the two-layer system manifests itself already in layer 1.

For easy viewing, we collect in Fig. 6(a) the relevant λ_{max} values from Sects. 3.1 and 4.1, plotting those of the single layer network and of layer 1 as functions of A and $A_1 + A_{\text{fb}}$, respectively. One sees that in this range, λ_{max} is definitively negative for the single layer, and definitively positive for layer 1. The goal of the rest of this subsection is to provide an explanation for this.

Effective synaptic inputs

18

The numbers A and $A_1 + A_{\text{fb}}$ are *a priori* estimates on the total synaptic input received by a neuron in the two families of networks; they are based on zero knowledge of network dynamics. Our plan is to show that when these "zeroth order" estimates are replaced by "first-order" approximations (these terms are in quotation as they are being used loosely), we will see that the systems thought to be comparable earlier are in fact not comparable. The quantity in question is how the *actual* impact due to coupling compares to its *a priori* values represented by A and $A_1 + A_{\text{fb}}$.

For the single-layer system, a better approximation of this quantity than A is

$$|A| \cdot r \cdot \bar{z} \ . \tag{4}$$

Here, r is the average firing rate per neuron in the network, and $\bar{z}$ is the average of the phase response curve $z(\cdot)$ for a receiving neuron, at the actual times when it receives a synaptic input. For layer 1 of the two-layer system, the corresponding quantity is

$$|A_1| \cdot r_1 \cdot \bar{z}_1 \ + \ |A_{\text{fb}}| \cdot r_2 \cdot \bar{z}_{\text{fb}} \tag{5}$$

where r_1 and r_2 are the average firing rates for neurons in layers 1 and 2, and $\bar{z}_1$ and $\bar{z}_{\text{fb}}$ are the mean values of $z(\cdot)$ when synaptic events initiated by neurons in layer 1 and layer 2, respectively, are received. These expressions are oversimplified, in that we have ignored details and correlations among spike times; that is why we refer to them as "first order approximations".

We now proceed to compare the quantities in (4) and (5).

Firing rates. As discussed in Sect. 3.2, the coupling in the single-layer system has diminished impact due to its tendency to synchronize. Consequently, we do not expect r to increase with A by a great deal in spite of the fact that all the neurons in the system are excitatory. This is confirmed in Fig. 6(b), which shows r as having increased by < 0.15 even at $A = 2$. Comparing the single-layer system with layer 2, we expect r_2 to be substantially larger, because the intra-layer dynamics of the two are similar, but the external driving for the single layer system is white noise (which in principle accelerates and decelerates with equal probability), whereas layer 2 is driven by layer 1, which is purely excitatory. This also is confirmed in Fig. 6(b). Finally, we see from Fig. 6(b) that r_1 is slightly greater than r; this may be the result of layer 2 firing more frequently. Nevertheless, to fully understand the effect of the input from layer 2, one must also factor in the role of the phase response curve z, which we discuss next.

Distribution of phases and sensitivity to inputs. Let P be the phase distribution for the single-layer network on the arrival of kicks, *i.e.*, whenever a neuron in the network spikes, we record the phase of the receiving neuron. Let P_{21} denote the corresponding phase distribution for a receiving neuron in layer 1 when synaptic inputs from layer 2 arrives, *i.e.*, we record θ_i whenever a synaptic event occurs along an edge $j \to i$, where neuron j belongs to layer 2 and neuron i to layer 1; the distributions P_{12}, P_{11}, and P_{22} are similarly defined.

For the single-layer network: From Sect. 3.2, we expect P to be centered at $\theta = 0$ and sharply peaked; this is confirmed again in Fig. 7(a).

Insight into the workings of a two-layer network can be obtained from the corresponding two-neuron system: while this is an obvious oversimplification, partial justification is provided by the tendency toward synchronization within layers. The phase space of the dynamical system describing two neurons is the 2-dimensional torus. This toy model is simple enough that much can be deduced (if not rigorously proved). In particular, it is not hard to see that the distributions
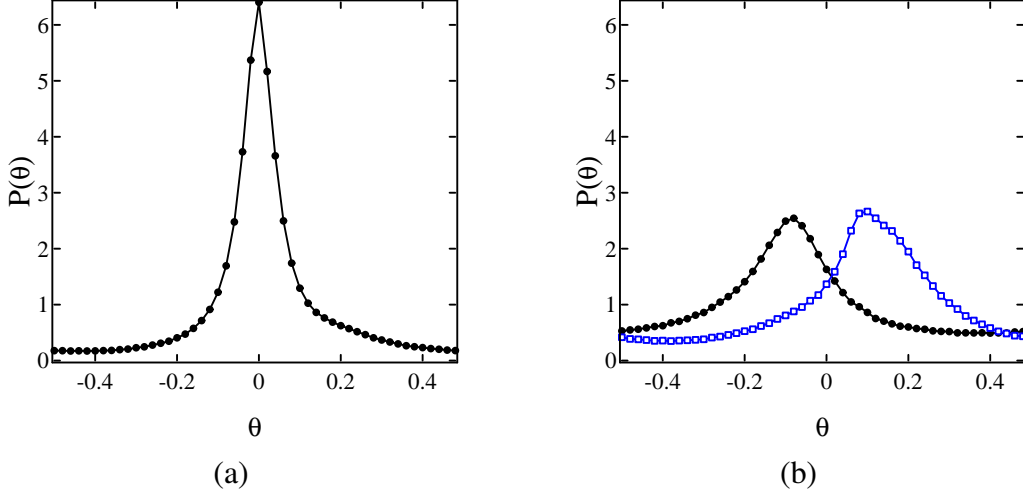
Figure 7: Phase distribution of neurons at the arrival of synaptic impulses for networks with $N = 100$. In (a), we plot the phase distribution P for a single-layer network with $A = 1.8$ and $\kappa = 20$. In (b), we plot the phase distributions P_{21} *(right)* and P_{12} *(left)* for a two-layer network with $A_1 = 1$ and $A_{\text{fb}} = 0.8$. All other parameters are the same as in Fig. 6

P_{12} and P_{21} are considerably more spread out than P, and that their peaks are off-center, the one for P_{21} being to the right of $\theta = 0$ and the one for P_{12} to the left. These predictions for large networks are confirmed in Fig. 7(b), which shows the corresponding distributions for a network with $N = 100$. (P_{11}, not shown, is qualitatively similar to P.)

A more detailed discussion of the 2D toy model is given in Appendix B. We note here that the distributions P_{21} and P_{12} lead to larger $\bar{z}_{\text{ff}}$ and $\bar{z}_{\text{fb}}$, so that inter-layer kicks have significantly greater effects than intra-layer events.

As a concrete example, we reconsider two sets of parameters previously taken to be comparable: for the single-layer network at $A = 1.8$, we have

$$A \cdot r \cdot \bar{z} \;=\; 1.8 \times 1.1 \times 0.049 \;=\; 0.097 \,,$$

while for layer 1 of the two-layer network with $A_1 = 1$ and $A_{\text{fb}} = 0.8$, we have

$$A_1 \cdot r_1 \cdot \bar{z}_1 + A_{\text{fb}} \cdot r_2 \cdot \bar{z}_{\text{fb}} \;=\; 1 \times 1.19 \times 0.059 + 0.8 \times 1.46 \times 0.107 \;=\; 0.195 \,.$$

Thus we see that their *effective* synaptic inputs really differ by a factor of 2. Not to suggest that these two systems are directly comparable, but the single-layer network with $A = 2 \times 1.8 = 3.6$ and the two-layer system with $A_1 = 1, A_{\text{fb}} = 0.8$ both have definitively positive and roughly comparable Lyapunov exponents.

Finally, we must emphasize that although the explanation above does address some of the most important contributing factors, it is only a partial explanation for this very intriguing phenomenon. More than averaged coupling strength alone enters in determining whether the largest Lyapunov exponent λ_{max} will be positive. Aside from correlations among different neurons' spike times, the nature and geometry of the coupling also play important roles. For large dynamical systems, such geometric and analytic considerations are out of reach at the present time.

Summary of Sect. 4: *To us, the most striking feature of the two-layer networks studied is that they become unreliable as soon as a small amount of feedback (less than $\frac{1}{10}$ the value of A_{ff} in the*

cases studied) is introduced. We believe an important contributing factor is that layers driven by different inputs are generally not synchronized to each other. Such is the case for the two layers here, and this asynchrony leads to increased sensitivity to inter-layer couplings and a decrease in reliability. These findings are likely to be applicable to networks with more than two layers.

Related works. Here, we mention prior studies on the effects of feedback and network architecture on neural reliability. The simplest setting in which this was addressed is that of [33], in which reliability was studied for a two-cell analog of the present two-layer system. There, it was found that feedback can lead to unreliability even for this small system; nevertheless, greater feedback strengths could be tolerated in the two-cell network before unreliability was produced. Additionally, reliability is studied for a variety of different network architectures in [5], in the context of an alternative, abstracted formalism for spiking networks. Different architectures, i.e. ring- vs. random-connectivity, did produce differences in reliability, but the distinction between these settings and those studied here precludes a direct comparison with our findings. Finally, a method of dissecting larger networks into interacting "modules" within and among which unreliability could be generated and propagated was studied in [33].

We note that, the effect of recurrent connections on the feedforward propagation of impulsive or more general inputs is analyzed, in a different spirit from reliability studies, in [3, 55].

5   Reliability of Pooled Responses

We now shift our attention from the behavior of individual neurons to bulk measurements, to study the reliability of responses of populations of neurons pooled together. Dependence on population size and relation to neuronal reliability are discussed.

Two representative networks are used a number of times for illustration in this section and the next. They are the networks for which rasters are plotted in Fig. 2 – except that we will use different N for different purposes:

Case A. Single-layer network with $A = 1$ and $\varepsilon = 2.5$. Shown to be neuronally reliable ($\lambda_{\max} = -0.77$ at $N = 100$).

Case B. Two-layer network with $A_1 = A_2 = 1$, $A_{\mathrm{ff}} = 2.8$, $A_{\mathrm{fb}} = 2.5$ and $\varepsilon = 2.5$. Shown to be neuronally unreliable ($\lambda_{\max} = 0.53$ at $N = 100$).

5.1   Measuring pooled-response reliability

As defined in Sect. 2, *pooled-response reliability* for a collection $\mathcal{C}$ of neurons in the network refers to the repeatability of the time-course $S_{\mathcal{C}}(t)$ representing the total synaptic output at time t of all the neurons in $\mathcal{C}$. See Sect. 2 for the precise definition of $S_{\mathcal{C}}(t)$.

Our first order of business is to propose a reasonable way to quantify the reliability of $S_{\mathcal{C}}(t)$. Let us suppress mention of $\mathcal{C}$ when it is understood. For each (fixed) t, a natural measure of the repeatability of $S(t)$ is by its across-trial variance, *i.e.*, the variance of the distribution of $S(t)$ computed over a very large number of trials (equivalently over a very large number of initial conditions of the network). We denote this number by $V(t)$. The size of $V(t)$ will depend on the number of neurons in $\mathcal{C}$. To put it on a scale that is independent of pool size, we consider $V(t)/n^2$ where n is the number of neurons in $\mathcal{C}$. Notice that this is equivalent to the variance of $\frac{1}{n}S(t)$. The

fact that $V(t)$ is time-dependent makes it a little unwieldy to work with. We therefore distill this information further to give a single number by time averaging $V(t)$. More precisely, define

$$\bar{V} = \lim_{T \to \infty} \frac{1}{T} \int_0^T V(t) dt.$$

The number $\bar{V}/n^2$ will be our measure of pooled-response reliability.

Fig. 8 shows plots of $S(t)$ for Cases A and B and various choices of populations $\mathcal{C}$. Some of these plots will be used to calibrate $\bar{V}/n^2$, *i.e.*, to give some idea what values of $\bar{V}/n^2$ correspond to high and low unreliability.

Fig. 8(a) shows $S(t)$ for Case A with $N = 100$ and $\mathcal{C}$ consisting of all 100 neurons. The results of only a single trial are shown, as $S(t)$ is identical across all trials given the same stimulus: it is easy to see that *neuronal reliability implies reliability of $S(t)$*. Notice that perfect synchrony would lead to $S(t)$ plots comprising of purely vertical bars. Here, we see tall, well defined peaks which nevertheless have some width. This corresponds to the near-synchronous behavior described in Sect. 3.2, where the phases of neurons concentrate in narrow bands. We also see stretches of time where the spiking shows greater disorder, corresponding to the more asynchronous configurations seen in Fig. 4.

In Fig. 8(b), plots of $S(t)$ for Case B are shown. The network used in these simulations has 1000 neurons, and the panels on the left and right correspond to two different choices of $\mathcal{C}$. The panels on the left represent 5 different trials for a collection of 5 neurons chosen randomly from layer 1. The value of $\bar{V}/n^2$ is computed to be 0.9 (over 200 trials). This is a case of high unreliability, as is evident from comparison of the 5 plots. The panels on the right are trials for a $\mathcal{C}$ consisting of 100 neurons also randomly chosen from layer 1. For this $\mathcal{C}$, $\bar{V}/n^2$ is computed to be about 0.2. These 5 plots may appear nearly identical, but they have obvious degradations upon closer inspection; we regard this as an instance of low unreliability. Compared to the reliable case, there are more peaks per unit time in these plots ($n = 100$ in both cases), and they are not as tall or as well defined, reflecting the less synchronous behavior.

We end this subsection by pointing out a potential defect of $\bar{V}/n^2$ as a measure of unreliability: since $V(t)$ measures the across-trials variance at each instant in time and does not take into account jittering, *i.e.* small shifts in spike times across trials, deviations in the *shapes* of the $S(t)$ – if that is what matters – are likely to be smaller than suggested by $\bar{V}/n^2$.

5.2 Dependence on population size

Intuitively, effects of averaging over progressively larger populations suggest that *pooled* responses may become more reliable as population size n increases. We investigate this systematically as follows: Continuing to use the network in Case B but with $N = 3000$, we compute $\bar{V}/n^2$ for various randomly chosen collections $\mathcal{C}_n$ of neurons of size n, always taking them from layer 1 for consistency and averaging the computed values over the different choices of $\mathcal{C}_n$ if they differ. (These values in fact do not differ appreciably for $n > 20$ or so.) The results are plotted in Fig. 9.

For $n = 1$, $\bar{V}$ is well over 1. We have not included this data point so as not to skew the scale of our plot. An observation that stands out immediately is that pooled responses show a dramatic decline in unreliability as n increases from 5 to 25. This means that even for a network that is quite unreliable on the neuronal level, reliability can be gained by pooling together the responses of a
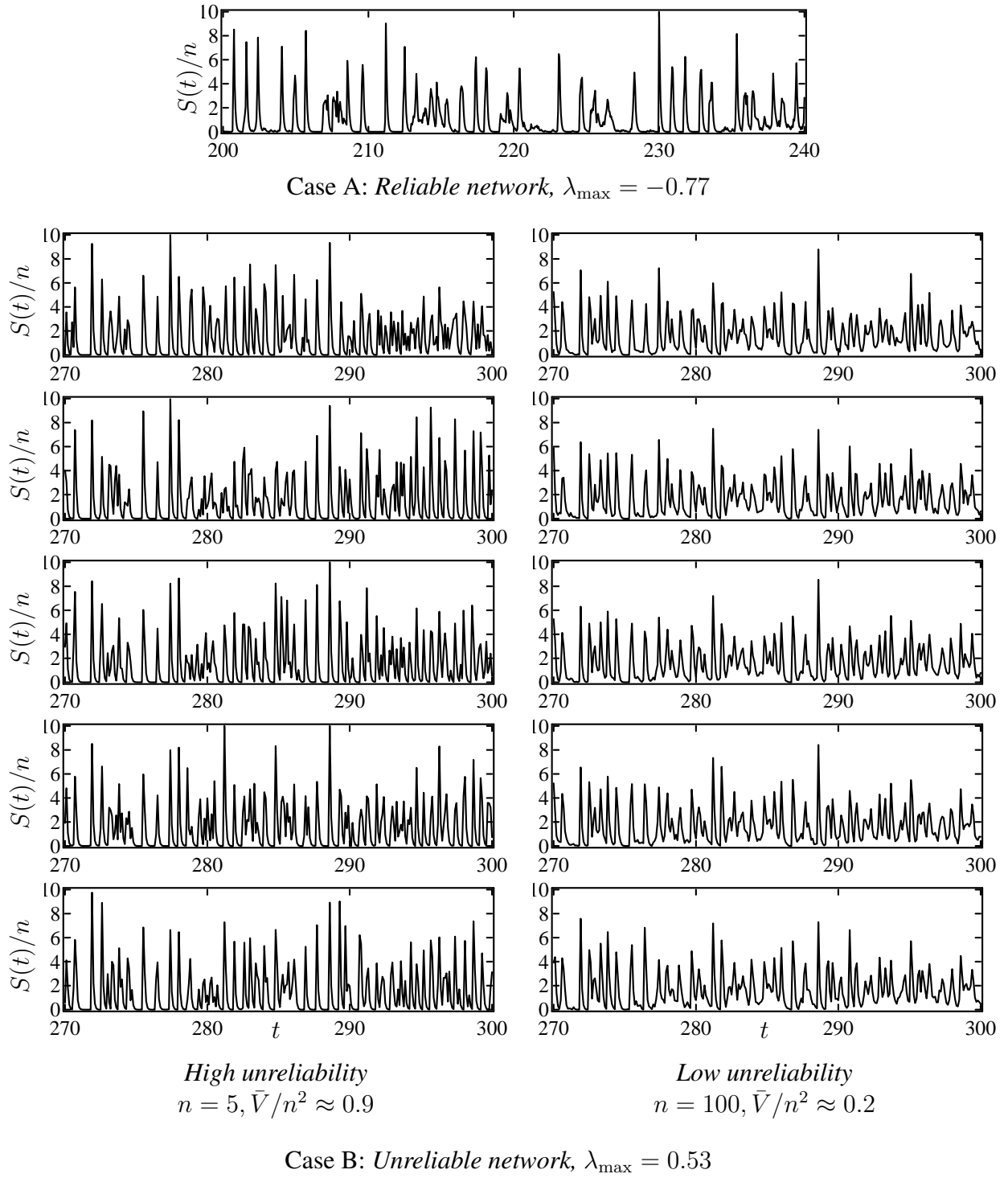
Case A: *Reliable network,* $\lambda_{\max} = -0.77$

High unreliability
$n = 5, \bar{V}/n^2 \approx 0.9$

Low unreliability
$n = 100, \bar{V}/n^2 \approx 0.2$

Case B: *Unreliable network,* $\lambda_{\max} = 0.53$

Figure 8: The pooled response $S(t)$ in two representative cases. *Case A:* Single-layer network, $N = 100$, $\kappa = 10$, $A = 1$, $\varepsilon = 2.5$. *Case B:* Two-layer network, $N = 1000$, $\kappa_1 = \kappa_2 = \kappa_{\text{ff}} = \kappa_{\text{fb}} = 10$, $A_{\text{ff}} = 2.8$, $A_{\text{fb}} = 2.5$, $A_1 = A_2 = 1$, $\varepsilon = 2.5$.
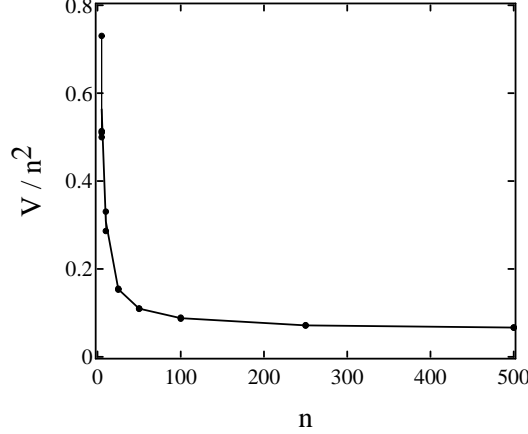
23

Figure 9: The scaled variance $\bar{V}/n^2$ as a function of subpopulation size n. Here, we use a two-layer network with $N = 3000$, $\kappa_1 = \kappa_2 = \kappa_{\text{ff}} = \kappa_{\text{fb}} = 0.1 \cdot N$, $A_{\text{ff}} = 2.8$, $A_{\text{fb}} = 2.5$, $A_1 = A_2 = 1$, and $\varepsilon = 2.5$.

small population. If carried over to real-life networks, this phenomenon may have ramifications. For example, by pooling together the spike times of a relatively small number of neurons, one may be able to dramatically improve the precision of neural codes based on temporal spike patterns (see, *e.g.*, [58, 2] and references therein). As n increases, $\bar{V}/n^2$ continues to decrease in the range tested. For $n \geq 100$, $\bar{V}/n^2$ dips below 0.1. Extrapolating from Fig. 8(b)(right), one concludes that pooled responses of these larger populations are, for practical purposes, reliable. We do not have enough data to determine if $\bar{V}/n^2$ will go to zero as $n \to \infty$.

The data we have shown are for neurons sampled from layer 1. Their pooled responses are slightly more unreliable than those of similar size populations from layer 2. For mixed samples comprised of some neurons from layer 1 and some from layer 2, $\bar{V}/n^2$ is somewhat lower; we attribute that to the fact that spike times from the two layers are less correlated (see Sect. 4.2).

With regard to the theoretical question on whether $\bar{V}/n^2 \to 0$ as $n \to \infty$ (which necessitates N tending to infinity as well, obviously), we conjecture an affirmative answer but do not know for certain. For fixed $\mathcal{C}$ and t, if we let $X_j = f_{syn}(t - t_j)$ where t_j is the last spike time before time t for neuron j (see Sect. 1.2 for the definition of $f_{syn}(t)$), then

$$V(t) = \text{Var}\left(\frac{1}{n}\sum_{j \in \mathcal{C}} X_j\right) .$$

Whether or not this variance drops off to zero as $n \to \infty$ depends on the asymptotic value of

$$\frac{1}{n^2}\sum_{i \neq j; i,j \in \mathcal{C}} \text{Cov}(X_i, X_j) . \tag{6}$$

In our model, each neuron receives input from a fixed fraction (20%) of all neurons in the network. This in itself may suggest that (6) is likely to remain bounded away from zero. On the other hand, there is a corresponding decrease in coupling strengths (recall that we use $A = \kappa a$ as our scaling parameter; see Sect. 3.1), which will decrease the size of each term in Eq. (6). The balance of these two aspects of the modeling makes it difficult to control the large n limit of (6).[4]

---

[4] As explained in Appendix A, the distribution of states at t units of time after the initial presentation of the stimulus
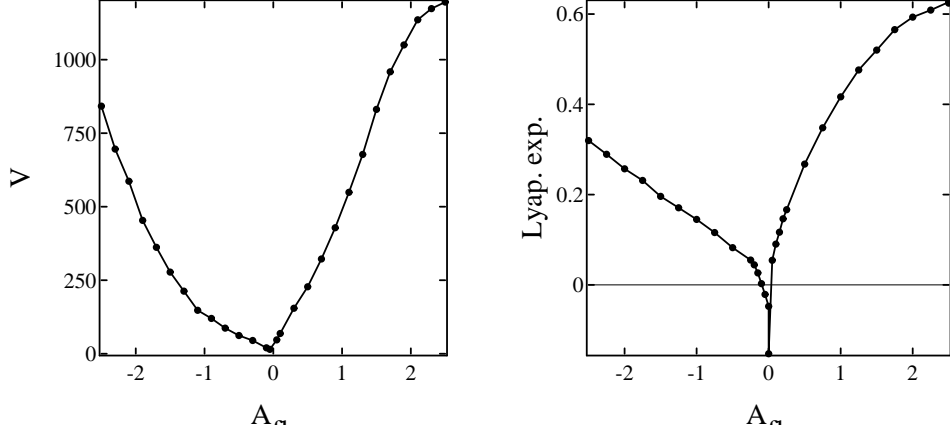
24

Figure 10: The mean synaptic variance $\bar{V}$ and Lyapunov exponent $\lambda_{\max}$ versus A_{fb}. Here, we use a two-layer network with $N = 100$, $\kappa_1 = \kappa_2 = \kappa_{\mathrm{ff}} = \kappa_{\mathrm{fb}} = 10$, $A_{\mathrm{ff}} = 2.8$, $A_{\mathrm{fb}} = 2.5$, $A_1 = A_2 = 1$, $\varepsilon = 2.5$. Note that the computed $\bar{V}$ are slightly positive even when $\lambda_{\max} < 0$ because computing necessarily relies on finite-time averages.

Similar effects of correlations increasing the variability of summed responses have been observed experimentally (for example, see [58]) as well as in modeling studies. In many of these studies, the source of correlations is fluctuating inputs that are common to different neurons (*e.g.*, [48, 38]). A modeling study closer to the setting of this section is [47], in which correlations arise through complex network interactions, although different questions about stimulus encoding are addressed there.

5.3 Relation to neuronal reliability

We have introduced two ways of measuring the reliability of a network: neuronal reliability as measured by $\lambda_{\max}$, and pooled-response reliability (taking $\mathcal{C}$ to be all neurons in the network for definiteness) as given by $\bar{V}/N^2$. Mathematically, these two quantities measure two different properties related to reliability. We now ask if they agree qualitatively even though we know the latter decreases with N. To that end, we consider the system studied in Sect. 4, namely the two-layer system with feedforward strength fixed at $A_{\mathrm{ff}} = 2.8$ and system size fixed at $N = 100$. We vary A_{fb} as a parameter, and compare $\lambda_{\max}$ and $\bar{V}/N^2$. A plot of the latter is shown in Fig. 10 (left), with the $\lambda_{\max}$ plot recalled from Sect. 4.1 shown alongside it to facilitate comparison. These plots show a strong correlation between the two measurements of unreliability we have introduced, a fact corroborated by other plots not shown.

6 Effects of Noise on Reliability

Up until now, we have considered reliability in noise-free settings. That is, we have neglected synaptic noise and other sources of trial-to-trial variability and fluctuations. For definiteness, we

is given by one of the sample measures μ_ω, which are singular and quite complicated. Estimating (6) will require knowledge of these measures.
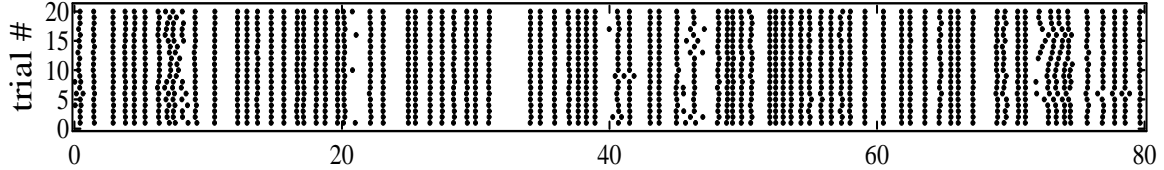
Figure 11: Response of a single-layer network in the presence of noise. Raster plot shows the spike times of a neuron in a single-layer network over 20 trials, with noise amplitudes $\sigma_{\text{local}} = \sigma_{\text{global}} = 0.1$. The parameters are those of Case A from Sect. 5 (see Fig. 8).

define noise here to be trial-to-trial fluctuations not modeled by Eq. (1). The purpose of this section is to investigate whether networks that produce repeatable responses in the absence of noise continue to (approximately) do so when noise is introduced, and whether some types of noise have a greater impact on repeatability than others.

For a conceptual understanding of its impact on reliability, we find it useful to distinguish between (i) noise that affects each neuron differently (*e.g.* synaptic release and membrane noise), and (ii) noise that affects the entire population in roughly the same way at each point in time (*e.g.*, noise associated with the stimulus $I(t)$) [38, 20, 23]. As an idealization, we add to Eq. (1) two noise terms:

$$\dot{\theta}_i = \omega_i + z(\theta_i) \cdot \left[\sum_{j \neq i} a_{ji} \, g(\theta_j) + \varepsilon_i I(t) + \sigma_{\text{local}} \eta_i(t) + \sigma_{\text{global}} \zeta(t) \right].$$

Here $\zeta(t)$ and $\eta_i(t)$ are white noise realizations which vary *independently from trial to trial*; additionally, the $\eta_i(t)$ are *independent for each i*. We refer to $\eta_i(t)$ and $\zeta(t)$ as "local" and "global" noise; their respective amplitudes are denoted by σ_{local} and σ_{global}.

Our simulations show that neuronal reliability persists under weak local and global noise – although a gradual degradation of spike time precision from trial to trial is unavoidable. See Fig. 11, which shows a raster plot of 20 trials for an arbitrarily chosen neuron in the network in Case A (see beginning of Sect. 5) at $\sigma_{\text{local}} = \sigma_{\text{global}} = 0.1$. The response is, for the most part, reliable, but there are some time intervals during which unreliable behavior is clearly visible. As expected, pooled responses can tolerate higher-amplitude noise terms.

Because the local noise terms $\eta_i(t)$ are statistically independent and identically distributed with mean 0, when n is large we expect their impact to be attenuated when we average over neurons in $S(t)$. In contrast, there can be no such attenuation for the global noise term $\zeta(t)$, since it is the same for all neurons. This suggests that global noise may have a more severe effect on pooled-response reliability than local noise. We confirm this thinking by computing $\bar{V}/n^2$ for the networks in both Case A and Case B. The results are shown in Table 2 below.

	Noise amplitudes $(\sigma_{\text{local}}, \sigma_{\text{global}})$			
	$(0, 0)$	$(0.5, 0)$	$(0.25, 0.25)$	$(0, 0.5)$
Case A: *Reliable*	0.0	0.04	0.27	0.84
Case B: *Unreliable*	0.22	0.13	0.56	1.1

Table 2: The (scaled) synaptic variance $\bar{V}/n^2$ in response to various combinations of noise parameters. In both cases, $N = 200$ and $n = 100$; in the unreliable case, $\mathcal{C}$ is all of layer 1. Similar results (not shown) are found for layer 2 neurons.

26

The results shown in this table are self-explanatory and, we think, illuminating: The first column, which is noise-free, is as in the previous section; unreliability is low in Case B given the pool size considered ($n = 100$). The second column confirms that local noise alone has little impact on pooled-response reliability due to its averaging effects: the reliable case remains, for all practical purposes, reliable, and the randomizing effects of the local noise have, in fact, *diminished* the unreliability in Case B. This observation is clearly valid beyond the settings considered here. It raises questions on the practical level: For example, it suggests that those types of noise in the nervous system (such as synaptic noise) which are relatively close to what we call "local noise" here may have relatively little impact on reliability; they may even serve to mollify the uncertainty created by other aspects of network dynamics. Moving on to the rightmost column, we notice the large values of $\bar{V}/n^2$ resulting from global noise alone, in both the reliable and unreliable cases. A comparison with Fig. 8(b) convinces one easily that at $\sigma_{\text{global}} = 0.5$, there is little reliability to speak of. We expect that this finding will also have practical implications. The third column shows results somewhere between the second and the 4th as expected.

The loss of reliability in a neuronally reliable system due to global noise can be understood as follows: Recall from Sect. 2 that reliability means all trajectories independent of initial conditions coalesce to a "random sink" for each t. Within each trial, since $\zeta(t)$ is a term of the same type as $I(t)$, its presence strengthens the effects of the stimulus, leading to more robust entrainment (see Sect. 3.1). Recall, however, that $\zeta(t)$ varies from trial to trial, so the trajectories entrain to a *different* stimulus, and therefore coalesce to a different state on each trial. When σ_{global} is large enough, this provides a mechanism for destroying reliability.

We finish with the following findings on how the effects of local and global noise vary with system size. In the case of local noise, due to the averaging effects discussed above, one expects the impact of such noise to diminish with increasing N. This is confirmed in the first row of Table 3, where we give $\bar{V}/n^2$ for the pooled response of layer 1 neurons in Case B (here $n = N/2$).

The situation differs dramatically in the presence of global noise. Here, no such averaging occurs, and the second row of Table 3 shows that unreliability in fact *increases* with N. While do not have a mathematical explanation of this numerical observation, its practical implications are clear: if sufficient global noise is present, the networks studied here cannot produce reliable responses, regardless of system size.

	System size N		
	200	600	1000
Local noise only, $\sigma_{\text{local}} = 0.5$	0.13	0.06	0.04
Global noise only, $\sigma_{\text{global}} = 0.5$	1.08	1.28	1.32

Table 3: The synaptic variance $\bar{V}/n^2$ for the pooled response of layer 1 neurons in Case B (see Sect. 5). Note that $n = \frac{N}{2}$. Similar results (not shown) are found for layer 2 neurons.

Needless to say, in realistic situations, the noise that impacts a neuronal population is correlated, *i.e.*, it corresponds to a combination of local and global noise in our modeling. We hope that by separating it into these two diametrically opposite types, we have provided some insight into the effect of noise on reliability.

Related works. Correlated noise as common input fluctuations to different cells has been directly

measured experimentally (*e.g.*, [12, 29]), and has also been studied theoretically in a wide variety of settings. One of the studies that is closest to the present section is [38], which examines the variability of pooled responses to time-dependent stimuli. There, in the setting of random-walk type neuron models without recurrent connections, a strong effect of correlated inputs on increasing this variability is also identified.

7   Discussion

In this section, we discuss the implications of our findings for *neural coding*.

Neural coding of time-dependent signals by spike trains has been widely studied (see [44] and references therein). If patterns of spike times are to carry information about such a signal, then these patterns must be at least partially repeatable each time the signal is presented. In other words, the system must possess some degree of *reliability*. It is commonly accepted that single cells in isolation display such repeatable responses, but many questions pertaining to the reliability of spiking networks remain (see Introduction). These questions motivated the present study.

The setting of our study is that of layered networks of Type I neural oscillators. One of our main findings is that when a group of neural oscillators all receive very similar inputs, individual neurons respond reliably, even in the presence of strong coupling. However, reliability is rapidly degraded by feedback from a downstream layer. We emphasize that this unreliability is caused entirely by dynamical instability: it occurs even in the absence of noise.

Given the ubiquity of strong feedback in the multilayered networks that occur in neurobiology, this immediately raises the question of whether biological networks can ever be expected to carry temporal stimuli. In other words, has temporal reliability been "sacrificed" in networks with strong feedback, perhaps to transmit information by modes other than temporal spike patterns? While our findings concern a specific class of models and cannot address these questions in general, they do suggest some important mitigating factors:

First, for the neural oscillator networks considered here, some degree of repeatability is preserved even for systems that are not neuronally reliable (*i.e.*, those with $\lambda_{\max} > 0$); see Fig. 2(b). That is to say, while feedback diminishes the repeatability of spike times, it does not eliminate temporal information completely but rather "blurs" it.

More importantly, we demonstrated that repeatability can be recovered, even in the presence of strong feedback, by pooling over ensembles of cells. With the simplest population coding scheme, *i.e.*, averaging, highly repeatable temporal information about the signal can be obtained by pooling together relatively few (*e.g.*, 100) neurons. This remains true in the presence of certain types of network fluctuations. Specifically, pooling remains effective in the presence of "local" noise, and was found to break down only when there is a significant component to the added noise that is globally correlated across network. This finding is in agreement with known results in other situations, *e.g.*, [38, 58].

We emphasize that the networks considered here are highly idealized. Furthermore, biological neurons often operate in the excitable, rather than oscillatory, regime. It is therefore difficult to compare our findings to the *in vivo* recordings discussed in the Introduction. Moreover, the generality of our findings beyond networks of Theta neural oscillators, *i.e.*, to systems with other phase response curves, remains to be explored. Nevertheless, we anticipate that some of our basic findings, especially on the role of feedback, the potential of population averaging to improve reli-

ability, and the effect of global noise, will be useful in considering network effects on encoding in more realistic situations.

Summary and Conclusions

We have introduced a class of coupled-oscillator networks (intended as models of idealized biological networks) for which we have carried out a systematic study of stimulus-response reliability. Networks in this class have layered structures, are sparsely coupled and heterogeneous, and their constituent components are phase oscillators modeled on type 1 neurons. Our findings, all of which are new in the present context, can be summarized as follows:

(1) *On the neuronal level, single-layer networks are fairly reliable even with substantial coupling, due to a tendency to synchronize. However, recurrent connections – even a small amount of feedback – can be strongly destabilizing in two-layer systems.* In general, individual neurons can behave reliably or unreliably as a result of the competition between entrainment to the stimulus or upstream layer and the perturbative effects of other synaptic events. Other aspects of architecture have less effect on reliability. Lyapunov exponents are used to measure neuronal reliability.

(2) *Pooled responses of even relatively small subpopulations are essentially reliable*, independent of whether individual neurons within the network are reliable or not. This provides a means to stabilize instabilities due to network dynamics, such as those caused by inter-layer connections. In a fixed-size network, pooled responses have similar reliability properties as individual neurons but with lower volatility.

(3) *Global noise, i.e., noise that affects the entire population in roughly the same way, can seriously jeopardize even pooled-response reliability, while local noise is less damaging and can in fact have mollifying effects.*

Implications for neural coding are straightforward and briefly discussed. While the findings above concern a specific class of models, we anticipate that some of the issues raised, such as diminished neuronal reliability due to feedback, recovery of reliability by pooling responses, and the detrimental effects of global noise, may be relevant more generally.

Appendix A. Review of random dynamical systems theory

In this appendix, we review some relevant mathematical theory that justifies the use of Lyapunov exponents in determining the reliability of a system. "Reliability" here refers exclusively to "neuronal reliability". As these results are very general and can potentially be used elsewhere, we will present them in a context considerably more general than the system defined by Eq. (1).

Consider a stochastic differential equation (SDE) of the form

$$dx_t = a(x_t)\, dt + \sum_{i=1}^{k} b_i(x_t) \circ dW_t^i \,. \tag{7}$$

Here $x_t \in M$ where M is a compact Riemannian manifold of any dimension $d \geq 1$, $a(\cdot)$ and $b(\cdot)$ are smooth functions on M, and $(W_t^1, \cdots, W_t^k)$ is a k-dimensional standard Brownian motion. In general, the equation is assumed to be of Stratonovich type, but when $M = \mathbb{T}^N \equiv \mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$, we have the choice between the Itô and Stratonovich integrals.[5] Eq. (1) is a special case of this setting with $M = \mathbb{T}^N$.

A. Stochastic flows associated with SDEs (see *e.g.* [28, 7])

In general, one fixes an initial x_0, and looks at the distribution of x_t for $t > 0$. Under fairly general conditions, these distributions converge to the unique stationary measure μ, the density of which is given by the Fokker-Planck equation. For our purposes, however, this is not the most relevant viewpoint. Since reliability is about a system's reaction to a single realization of Brownian motion at a time, and concerns the simultaneous evolution of all or large ensembles of initial conditions, of relevance to us are not the distributions of x_t but *flow-maps* of the form $F_{t_1,t_2;\omega}$. Here $t_1 < t_2$ are two points in time, ω is a sample Brownian path, and $F_{t_1,t_2;\omega}(x_{t_1}) = x_{t_2}$ where x_t is the solution of (7) corresponding to ω. A well known theorem states that such *stochastic flows of diffeomorphisms* are well defined if the functions $a(x)$ and $b(x)$ in Eq. (7) are sufficiently smooth (see [28]). More precisely, the maps $F_{t_1,t_2;\omega}$ are well defined for almost every ω, and they are invertible, smooth transformations with smooth inverses. Moreover, $F_{t_1,t_2;\omega}$ and $F_{t_3,t_4;\omega}$ are independent for $t_1 < t_2 < t_3 < t_4$. These results allow us to treat the evolution of systems described by (7) as compositions of random, *i.i.d.*, smooth maps.

Since reliability questions involve one ω at a time, the stationary measure μ, which gives the steady-state distribution averaged over all ω, is not the object of direct interest. Of relevance are the *sample measures* $\{\mu_\omega\}$, which are the conditional measures of μ given the past. More precisely, we think of ω as defined for all $t \in (-\infty, \infty)$ and not just for $t > 0$. Then μ_ω describes what one sees at $t = 0$ given that the system has experienced the input defined by ω for all $t < 0$. Two useful facts about these sample measures are

(a) $(F_{-t,0;\omega})_*\mu \to \mu_\omega$ as $t \to \infty$, where $(F_{-t,0;\omega})_*\mu$ is the measure obtained by transporting μ forward by $F_{-t,0;\omega}$, and

(b) the family $\{\mu_\omega\}$ is invariant in the sense that $(F_{0,t;\omega})_*(\mu_\omega) = \mu_{\sigma_t(\omega)}$ where $\sigma_t(\omega)$ is the time-shift of the sample path ω by t.

Thus in the context of a reliability study, if our initial distribution is given by a probability density ρ and we apply the stimulus corresponding to ω, then the distribution at time t is $(F_{0,t;\omega})_*\rho$. For t sufficiently large, one expects in most situations that $(F_{0,t;\omega})_*\rho$ is very close to $(F_{0,t;\omega})_*\mu$, which by (a) above is essentially given by $\mu_{\sigma_t(\omega)}$. The time-shift by t of ω is necessary because by definition, μ_ω is the conditional distribution of μ at time 0.

---

[5] In this paper, we use the Itô interpretation throughout.

B. Lyapunov exponents of random dynamical systems (see *e.g.* [1])

The fact that the evolution of systems described by (7) can be represented as compositions of random, *i.i.d.*, smooth maps allows us to tap into a large part of dynamical systems theory, namely the theory of *random dynamical systems* (RDS). Many of the techniques for analyzing smooth deterministic systems have been extended to this random setting, including the notion of Lyapunov exponents. For the stochastic flows above, the largest Lyapunov exponent is defined to be

$$\lambda_{\max}(x, \omega) = \lim_{t \to \infty} \frac{1}{t} \log \|DF_{0,t;\omega}(x)\| .$$

These numbers are known to be defined for μ-a.e. $x \in M$ and a.e. ω. Moreover, they are nonrandom, *i.e.*, they do not depend on ω, and when μ is ergodic, $\lambda_{\max}$ does not depend on x either, *i.e.*, $\lambda_{\max}$ is equal to a single (fixed) number for almost every initial condition in the phase space and for almost every sample path.

Numerical calculation of Lyapunov exponents. Lyapunov exponents can be computed numerically by solving the variational equations associated with Eq. (7); the largest Lyapunov $\lambda_{\max}$ is given by the logarithmic growth rate of a typical tangent vector. This is what we have done in this paper, using the Euler method for SDEs to solve the variational equations.

C. Implications of the sign of $\lambda_{\max}$

The two results below are in fact valid in greater generality, but let us restrict ourselves to the SDE setting at the beginning of this subsection.

Theorem 1 *Let μ be an ergodic stationary measure of the RDS defined by Eq. (7).*

(1) (Random sinks) [31] *If $\lambda_{\max} < 0$, then for a.e. ω, μ_ω is supported on a finite set of points.*

(2) (Random strange attractors) [32] *If μ has a density and $\lambda_{\max} > 0$, then for a.e. ω, μ_ω is a random SRB measure.*

In Part (1) of the theorem, if in addition to $\lambda_{\max} < 0$, mild conditions (on the relative motions of two points) are assumed, then almost surely μ_ω is supported on a single point [7]. From the discussion above, μ_ω being supported on a single point corresponds to the collapse of trajectories starting from almost all initial conditions to a single trajectory. In the context of Eq.(1), this is exactly what it means for the system to be neuronally reliable as explained in Sect. 2.1.

The conclusion of Part (2) requires clarification: In deterministic dynamical systems theory, SRB measures are natural invariant measures that describe the asymptotic dynamics of chaotic dissipative systems (in the same way that Liouville measures are the natural invariant measures for Hamiltonian systems). SRB measures are typically singular. They are concentrated on unstable manifolds, which are families of curves, surfaces etc. that wind around in a complicated way in the phase space [17]. Part (2) of Theorem 1 generalizes these ideas to random dynamical systems. Here, random (meaning ω-dependent) SRB measures live on random unstable manifolds, which are complicated families of curves, surfaces, *etc.* that evolve with time. In particular, in a system with random SRB measures, different initial conditions lead to very different outcomes at time t when acted on by the same stimulus; this is true for all $t > 0$, however large. In the context of a reliability study, therefore, it is natural to regard the distinctive geometry of random SRB measures as a signature of unreliability.
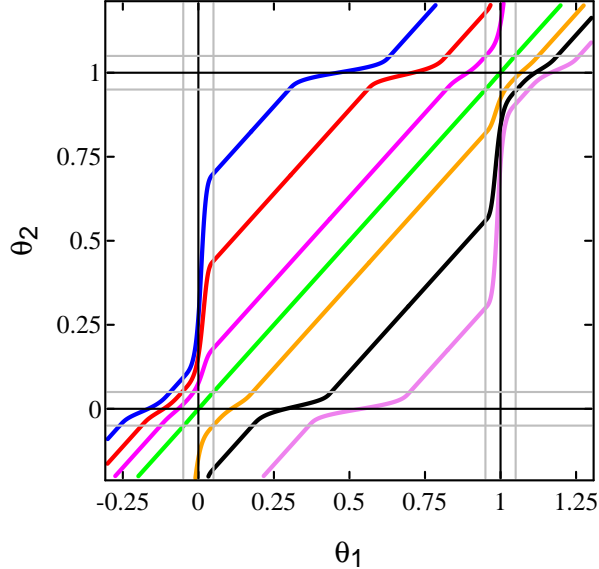
31

Figure 12: A few trajectories for a two-oscillator toy model with $\varepsilon = 0$. The trajectories are drawn on a "lift" of the 2-torus to the plane. The parameters are $\omega_1 = \omega_2 = 1$, $A_{\text{ff}} = 2.8$, and $A_{\text{fb}} = 0.8$.

We do not claim here that mathematically, the results in Theorem 1 apply to Eq.(1). To formally apply these results, conditions of ergodicity, invariant density *etc.* have to be verified. Evidence – both analytic and numerical – point to an affirmative answer when the coupling constants a, a_i, a_{ff} *etc.* are nonzero.

Appendix B. A 2-D toy model of two-layer networks

We provide here more detail on how two-neuron models can be used to shed light on two-layer networks as suggested in Sect. 4.2. Specifically, we will explain how the shapes of the phase distributions P_{12} and P_{21} are predicted.

Consider a system comprised of *two neurons* whose dynamics obey Eq. (1). For definiteness, we set $\omega_1 = \omega_2 = 1$, $A_{\text{ff}} = 2.8$, and $A_{\text{fb}} = 0.8$ to mimic the parameters in the two-layer networks considered in Sect. 4.2, with neurons 1 and 2 representing layers 1 and 2 in the two-layer system. The phase space of this system is the 2-torus, which we identify with the square $[0.1]^2$ with periodic boundary conditions; the coordinates are denoted by (θ_1, θ_2). In Fig. 12, we show a few trajectories of the undriven system, *i.e.*, with $\varepsilon = 0$. Away from the edges, they are northeasterly with slope 1; near the edges, they are bent due to the coupling. We now turn on the stimulus, setting $\varepsilon = 2.5$ as in Sect. 4.2. Because only neuron 1 hears the stimulus, it perturbs trajectories only in the θ_1 direction. When the stimulus is turned on, trajectories will, for the most part, continue to go in roughly the same directions as those shown in Fig. 12, but they become "wriggly", being driven randomly to the left and right by the white-noise stimulus.

We will refer to the top and bottom edges of the square (which are identified with each other) as Σ. Every time a trajectory crosses Σ, neuron 2 spikes, and the location in Σ tells us the phase of neuron 1 when this spiking occurs. We view Σ as a cross-section to the flow, and consider the induced return map $\Phi : \Sigma \to \Sigma$. In the case of two neurons with feedback, the distribution of

trajectories of Φ on Σ tells us the phase distribution of neuron 1 when it receives a synaptic input from neuron 2. In our analogy with the two-layer system, this is the analog of P_{21}. Similarly, the distribution of returns to the left and right edges identified (we call that Σ') represents the phases of neuron 2 when it receives an input from neuron 1., *i.e.*, the distribution analogous to P_{12}.

To understand these distributions, let us view the return dynamics to Σ as the result of two "moves" in succession (this is not entirely accurate but will suffice for present purposes): The first is the return map for the flow with $\varepsilon = 0$, and the second is a "smearing", obtained by, *e.g.*, taking a convolution with a Gaussian, to simulate the perturbations experienced by the trajectory between returns to Σ.

The return dynamics of the undriven flow are very simple: From the geometry of the flowlines, one sees that starting from any point in Σ, there is a leftward displacement due to the fact that the upward kick along the vertical edges are stronger than the rightward kicks along the horizontal edges (*i.e.* $A_{\mathrm{ff}} > A_{\mathrm{fb}}$). This leftward displacement is quite substantial away from $\theta_1 = 0$, reaching a maximum at $\theta_1 \approx 0.75$. Because of the character of the phase response function, this displacement is very small (but strictly positive) near $\theta_1 \approx 0$. It is so weak there that with $\varepsilon = 0$, all trajectories spend most of their time near the diagonal, with only brief excursions in between. In other words, when $\varepsilon = 0$, the phase distributions on Σ peak sharply at $\theta_1 = 0$.

With $\varepsilon = 2.5$, the "smearing" is nontrivial. Immediately, one sees that it causes the distribution to be more spread out. It is also easy to see that some concentration near $\theta_1 = 0$ will be retained, only that the peak will be more rounded. We now explain why one should expect the peak to be shifted to the right: Suppose we start with a roughly constant distribution on an interval centered at $\theta_1 = 0$. Since the return map is nearly the identity in this region, we may assume it does not change this distribution substantially. Next we take a convolution, which causes the distribution to have wider support. Now the part of the distribution that is pushed to the right of the original interval will get pushed back in when we apply the return map of the undriven flow again, due to the leftward displacement discussed earlier, whereas the part that is pushed to the left will be pushed further away from $\theta_1 = 0$ by the $\varepsilon = 0$ dynamics. The result is an obvious asymmetry in the distribution, one that is reinforced in subsequent iterations. (The argument we have presented does not constitute a proof, but a proof is probably not out of reach.)

Once we have convinced ourselves of the geometry of P_{21}, it is easy to see from the northeasterly direction of the flowlines that if P_{21} peaks to the right of $\theta_1 = 0$ on Σ, then P_{12} must peak just below $\theta_2 = 0$ on Σ'. This completes the explanation promised in Sect. 4.2.

We finish with a couple of remarks that may be illuminating:

(1) The phenomenon described above occurs for both excitatory and inhibitory couplings: with $A_{\mathrm{ff}} < 0$ and $|A_{\mathrm{fb}}| < |A_{\mathrm{ff}}|$, the displacement of the return map $\Phi : \Sigma \to \Sigma$ is to the right. But in the inhibitory situation, there are other forces shaping the phase distribution, making the picture there more complicated. (In case the reader wonders how to interpret our reasoning in the pure-feedforward case: statements about phase distributions on Σ are valid, except that spikings of neuron 2 do not impact neuron 1!)

(2) We may also represent the *single-layer* system by two neurons. Here by far the biggest difference is that both neurons receive the stimulus in the same way, and that translates into perturbations that are *in the direction of the diagonal*. Such perturbations are not very effective in spreading out distributions, especially when the trajectories are concentrated near the diagonal. These observations provide a geometric understanding for the material in Sect. 3.2.

A detailed analysis of the two-neuron model with different emphasis is carried out in [33].

References

[1] L. Arnold. *Random Dynamical Systems*. Springer, New York, 2003.

[2] B. Averbeck, P.E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nat Rev Neurosci*, 7(5):358–366, May 2006.

[3] Y. Aviel, C. Mehring, M. Abeles, and D. Horn. On embedding synfire chains in a balanced network. *Neural Comp.*, 15:1321–1340, 2003.

[4] W. Bair, E. Zohary, and W.T. Newsome. Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior. *J. Neurosci.*, 21(5):1676–1697, 2001.

[5] A. Banerjee. On the sensitive dependence on initial conditions of the dynamics of networks of spiking neurons. *J. Comput. Neurosci.*, 20:321–348, 2006.

[6] A. Banerjee, P. Seriès, and A. Pouget. Dynamical constraints on using precise spike timing to compute in recurrent cortical networks. *Neural Computation*, 20:974–993, 2008.

[7] P.H. Baxendale. Stability and equilibrium properties of stochastic flows of diffeomorphisms. In *Progr. Probab. 27*. Birkhauser, 1992.

[8] M. Bazhenov, N. Rulkov, J. Fellous, and I. Timofeev. Role of network dynamics in shaping spike timing reliability. *Phys. Rev. E*, 72:041903, 2005.

[9] M. Berry, D. Warland, and M. Meister. The structure and precision of retinal spike trains. *PNAS*, 94:5411–5416, 1997.

[10] N. Bertschlinger and T. Natschlager. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comp.*, 16:1413–1436, 2004.

[11] C. Borgers, C. Epstein, and N. Kopell. Background gamma rhythmicity and attention in cortical local circuits: A computational study. *J. Neurosci.*, 102:7002–7007, 2005.

[12] R. M. Bruno and B Sakmann. Cortex is driven by weak but synchronously active thalamo-cortical synapses. *Nature*, 312:16221627, 2006.

[13] H.L. Bryant and J.P. Segundo. Spike initiation by transmembrane current: a white-noise analysis. *Journal of Physiology*, 260:279–314, 1976.

[14] R. de Reuter van Steveninck, R. Lewen, S. Strong, R. Koberle, and W. Bialek. Reproducibility and variability in neuronal spike trains. *Science*, 275:1805–1808, 1997.

[15] B. Doiron, M.J. Chacron, L. Maler, A. Longtin, and J. Bastian. Inhibitory feedback required for network burst responses to communication but not to prey stimuli. *Nature*, 421:539–543, 2003.

[16] E. Douglas and K. Martin. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451, 2004.

[17] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617–656, 1985.

[18] G. B. Ermentrout. Type I membranes, phase resetting curves, and synchrony. *Neural Comp.*, 8:979–1001, 1996.

[19] G.B. Ermentrout and N. Kopell. Frequency plateaus in a chain of weakly coupled oscillators, I. *SIAM J. Math. Anal.*, 15:215–237, 1984.

[20] A. Aldo Faisal, Luc P. J. Selen, and Daniel M. Wolpert. Noise in the nervous system. *Nat. Rev. Neurosci.*, 9:292–303, 2008.

[21] A. Hodgkin. The local electric changes associated with repetitive action in a non-medulated axon. *J. Physiol.*, 117:500–544, 1948.

[22] J. Hunter, J. Milton, P. Thomas, and J. Cowan. Resonance effect for neural spike time reliability. *J. Neurophysiol.*, 80:1427–1438, 1998.

[23] D. Johnston and S. Wu. *Foundations of Cellular Neurophysiology*. MIT Press, Cambridge, MA, 1997.

[24] E. Kandel, J. Schwartz, and T. Jessell. *Principles of Neural Science*. McGraw-Hill, 4 edition, 1991.

[25] P. Kara, P. Reinagel, and R.C. Reid. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27:636–646, 2000.

[26] Yu. Kifer. *Ergodic Theory of Random Transformations*. Birkhauser, 1986.

[27] C. Koch. *Biophysics of computation: information processing in single neurons*. Oxford University Press, 1999.

[28] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[29] I. Lampl, I. Reichova, and D. S. Ferster. Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron*, 22:361374, 1999.

[30] P. E. Latham, B. J. Richmond, P. G. Nelson, and S. Nirenberg. Intrinsic dynamics in neuronal networks. I. theory. *J. Neurophysiol.*, 83:808–827, 2000.

[31] Y. Le Jan. Équilibre statistique pour les produits de difféomorphismes aléatoires indépendants. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(1):111–120, 1987.

[32] F. Ledrappier and L.-S. Young. Entropy formula for random transformations. *Probab. Th. and Rel. Fields*, 80:217–240, 1988.

[33] K. K. Lin, E. Shea-Brown, and L.-S. Young. Reliability of coupled oscillators. *J. Nonlin. Sci. (accepted subject to final revisions)*, 2008.

[34] K. K. Lin, E. Shea-Brown, and L.-S. Young. Reliability of layered neural oscillator networks. *submitted*, 2008.

[35] T. Lu, L. Liang, and X. Wang. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.*, 4:1131–1138, 2001.

[36] H.R. Maei and P.E. Latham. Can randomly connected networks exhibit long memories? *Preprint, Gatsby Computational Neuroscience Unit*, 2005.

[37] Z. Mainen and T. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268:1503–1506, 1995.

[38] M. Mazurek and M. Shadlen. Limits to the temporal fidelity of cortical spike rate signals. *Nature Neurosci.*, 5:463–471, 2002.

[39] G. Murphy and F. Rieke. Network variability limits stimulus-evoked spike timing precision in retinal ganglion cells. *Neuron*, 52:511–524, 2007.

[40] K. Pakdaman and D. Mestivier. External noise synchronizes forced oscillators. *Phys. Rev. E*, 64:030901–030904, 2001.

[41] D. Perkel and T. Bullock. Neural coding. *Neurosci. Res. Program Bull.*, 6:221–344, 1968.

[42] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, Cambridge, 2001.

[43] A. Reyes. Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. *Nature Neuroscience*, 6:593–599, 2003.

[44] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA, 1996.

[45] J. Rinzel and G.B. Ermentrout. Analysis of neural excitability and oscillations. In C. Koch and I. Segev, editors, *Methods in Neuronal Modeling*, pages 251–291. MIT Press, 1998.

[46] J. Ritt. Evaluation of entrainment of a nonlinear neural oscillator to white noise. *Phys. Rev. E*, 68:041915–041921, 2003.

[47] P Seriès, Peter E Latham, and Alexandre Pouget. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat Neurosci*, 7:11291135, 2004.

[48] M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.*, 18:3870–3896, 1998.

[49] G. Shepard. *The Synaptic Organization of the Brain*. Oxford Univ. Press, 2004.

[50] J. Teramae and T. Fukai. Reliability of temporal coding on pulse-coupled networks of oscillators. *arXiv:0708.0862v1* [nlin.AO], 2007.

[51] J. Teramae and D. Tanaka. Robustness of the noise-induced phase synchronization in a general class of limit cycle oscillators. *Phys. Rev. Lett.*, 93:204103–204106, 2004.

[52] D. Terman, J. Rubin, A. Yew, and C.J. Wilson. Activity patterns in a model for the subthalamopallidal network of the basal ganglia. *J. Neurosci.*, 22:2963–2976, 2002.

[53] C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726, 1996.

[54] C. van Vreeswijk and H Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Comp.*, 10:1321–1371, 1998.

[55] T. Vogels and L. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.*, 25:10786–10795, 2005.

[56] A. Winfree. *The Geometry of Biological Time*. Springer, New York, 2001.

[57] C. Zhou and J. Kurths. Noise-induced synchronization and coherence resonance of a hodgkin-huxley model of thermally sensitive neurons. *Chaos*, 13:401–409, 2003.

[58] E. Zohary, M. N. Shadlen, and W. T. Newsome. Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature*, 370:140–143, 1994.