

Renewal theory of coupled neuronal pools: Stable states and slow trajectories

Christian Leibold

Department Biologie II, LMU Munich, Großhadernerstrasse 2, D-82152 Planegg, Germany and
Bernstein Center for Computational Neuroscience Munich, Großhadernerstrasse 2, D-82152 Planegg, Germany
(Received 15 April 2011; revised manuscript received 24 July 2011; published 30 September 2011)

A theory is provided to analyze the dynamics of delay-coupled pools of spiking neurons based on stability analysis of stationary firing. Transitions between stable and unstable regimes can be predicted by bifurcation analysis of the underlying integral dynamics. Close to the bifurcation point the network exhibits slowly changing activities and allows for slow collective phenomena like continuous attractors.

DOI: [10.1103/PhysRevE.84.031935](https://doi.org/10.1103/PhysRevE.84.031935)

PACS number(s): 87.19.lj, 87.18.Sn, 02.30.Ks

I. INTRODUCTION

Although information processing in biological neuronal networks is often linked to exquisitely timed individual spikes [1–3], cognitive phenomena, such as short-term memory and decision making, are also reflected by slowly changing neuronal firing rates [4,5]. To act as such rate encoders, neurons have to be able to gradually and persistently elevate their firing rate over some period of time. The mechanisms that were identified to underlie such ongoing firing are manifold. They reach from specific ion channels in the cell membrane (e.g., [6]) to the wiring up of respective neuronal circuits (e.g., [7]). Since the physiological impact of the neuronal circuitry is more difficult to track down experimentally than that of cellular mechanisms, theoretical models have to be set up to validate network-based hypothesis of persistent firing and to study the interplay between cellular and circuit mechanisms.

A theoretical approach to study the rate encoding of functionally equivalent subsets of neurons is to group them into pools, thereby translating the description from the single cell to the population level. Classically, pool dynamics are described by a set of coupled ordinary differential equations [8–10]. These approaches have been, and still are, very successfully applied to model neuronal processing in both cognitive and sensory cortical areas (e.g., [11,12]). Such a temporally local description, however, cannot account for intrinsic delays and the dependence of spike probability on the spiking history. Therefore relating simulations of spiking neurons to such classical pool models is often very difficult.

These drawbacks are partially amended by pool dynamics considering neuronal spiking as a renewal process [13–16]. These theories provide exact descriptions of the pool firing rate if the number N of neurons in a pool tends to infinity. A disadvantage of these models is that they are rather difficult to analyze. In particular, a theory of how to analytically treat the dynamics, if several of such renewal pools are coupled, is not available so far. Introducing couplings is an important next step since it opens the way to investigate more complex network topologies in the future. In this paper, I propose an extension of the stability analysis of stationary firing to multiple coupled pools also allowing coupling delays. The theory is semianalytically treatable and asymptotically correct for large pool sizes $N \rightarrow \infty$, which distinguishes it from most other approaches that are based on simulations of spiking neurons. Finally, to demonstrate the inner working of the

theory, it is applied to three well-studied example cases in which collective dynamical phenomena are known to arise.

II. MODEL

Let us consider d pools of neurons. Each pool $i = 1, \dots, d$ is fully and homogeneously connected and consists of a homogeneous population of neurons receiving identical input. The synaptic connections between the pools are determined by the synaptic weights W_{ij} . The neuronal properties across pools may be different, although this degree of freedom will not be made use of in this paper.

Following [13,16], the pool dynamics is formulated on the level of a renewal process in which a *hazard function* $p_i(t|t')$ determines the conditional probability $p_i(t|t') \Delta t$ that a neuron in pool i fires in the infinitesimally short interval $[t, t + \Delta t)$, if its last spike occurred at time t' . This conditional probability allows one to construct the survivor function

$$S_i(t|t') = \Theta(t - t') \exp \left[- \int_{t'}^t ds p_i(s|t') \right] \quad (1)$$

that provides the probability that a neuron in pool i does not fire in the interval $[t', t)$ if its last spike occurred at time t' . Here, and elsewhere, Θ denotes the Heaviside step function $\Theta(t) = 1$ for $t \geq 0$, and $\Theta(t) = 0$ otherwise.

By splitting up the joint probability $\rho_i(t, t')$ that a neuron fired its last spike at time t' in a conditional and a marginal distribution, one defines the firing probability density (or firing rate) $A_i(t')$ as

$$\rho_i(t, t') = S_i(t|t') A_i(t'). \quad (2)$$

In what follows, the firing rate vector $\vec{A} = (A_1, \dots, A_d)^T$ will be considered as the macroscopic dynamical variable of interest (although strictly speaking it is no dynamical variable).

Since ρ_i as a probability is normalized, the integral equation

$$\mathcal{N}_i(t) = \int_{-\infty}^t dt' A_i(t') S_i(t|t') \equiv 1 \quad (3)$$

is valid for all times t , and so defines the dynamics of the firing rates $A_i(t)$ via a self-consistency equation [13,16]. For the purpose of numerical simulation, the dynamics can be made explicit by taking the derivative of Eq. (3) with respect to time t .

To facilitate the analysis of the integral dynamics from Eq. (3) for d coupled pools, I consider a specific model for the conditional firing probability density $p_i(t|t')$ that realizes the

renewal property in the simplest nontrivial way. This model consists of three constituents: a synaptic coupling kernel, a gain function, and refractoriness.

First the activities A_j of the pools are assumed to be convolved (*) by a synaptic coupling kernel

$$\kappa * A_j(t) = \int_{-\infty}^t ds A_j(s) \kappa(t-s). \quad (4)$$

If multiplied with the synaptic weight W_{ij} the convolution in Eq. (4) models the contribution of the activity of pool j to the postsynaptic potential of neuron i in the spirit of spike response models [14]. Without loss of generality, synaptic kernels are assumed to be normalized, $\int_0^\infty ds \kappa(s) = 1$. All examples in this paper are derived with an exponential kernel

$$\kappa(s) = \Theta(s - \Delta) \beta e^{-\beta(s-\Delta)}, \quad \beta > 0. \quad (5)$$

The two parameters β and Δ are interpreted as the inverse time constant of the postsynaptic potential and the synaptic delay, respectively.

Second, the filter output is assumed to instantaneously influence the firing rate of the postsynaptic neuron. This instantaneous action is modeled by an exponential gain function

$$f_i[\kappa * \vec{A}(t)] = v_0 \exp[\vec{w}_i \cdot (\kappa * \vec{A}(t)) + I_i(t)], \quad (6)$$

with I_i quantifying the external inputs and v_0 taking the role of the spontaneous firing rate. The vectors \vec{w}_i describe the synaptic strength with which the other pools are connected to pool i . The corresponding synaptic weight matrix W_{ij} can thus be derived from the elements of the individual weight vectors as $W_{ij} = (\vec{w}_i)_j$.

Third, refractoriness is modeled by a multiplicative probability term

$$g(t) = \Theta(t - \tau) \quad (7)$$

that incorporates an absolute refractory time τ , implying that two consecutive spikes of a neuron have to be separated at least by this interval.

In summary, these ideas lead to the following model for the firing probability density:

$$\begin{aligned} p_i(t|t') &= f_i(\kappa * A_1(t), \dots, \kappa * A_d(t)) g(t - t') \\ &=: f_i(\kappa * \vec{A}(t)) g(t - t'). \end{aligned} \quad (8)$$

III. THEORY OF LINEAR STABILITY

The pool dynamics defined by the integral equation (3) can generally only be treated numerically. To also allow analytical treatment, Eq. (3) is linearized around stationary states.

A. Stationary states

In this paper only constant stationary states are considered, i.e., the firing rate solution $\vec{A}(t) = \vec{A}^* = \text{const}$ is constant in time. This of course also restricts the study to constant inputs $I_i(t) = I_i = \text{const}$ (which can be relaxed a bit by considering piecewise constant inputs). In such a regime, according to Eq. (8), the firing probability density p_i and hence the survivor function only depend on the time difference $t - t'$ to the last

spike. Particularly, one finds

$$S_i^*(t - t') = \Theta(t - t') \exp[-f_i(\vec{A}^*) G(t - t')] \quad (9)$$

with

$$G(s) = \Theta(s - \tau)(s - \tau). \quad (10)$$

From the dynamical equation (3), one readily obtains the fixed-point condition

$$(A_i^*)^{-1} = \int_0^\infty ds \exp[-f_i(\vec{A}^*) G(s)] = \tau + (f_i(\vec{A}^*))^{-1}, \quad (11)$$

which is solved by

$$A_i^* = \frac{f_i(\vec{A}^*)}{1 + \tau f_i(\vec{A}^*)}. \quad (12)$$

In Fig. 1, the fixed-point states from Eq. (12) are compared to simulations of neurons firing according to the renewal process from Eq. (8). For the specified model, excitatory-coupled pools exhibit bistability, in which the two stable fixed points are separated by an unstable one [Figs. 1(A)–1(C)]. However, if several pools are coupled, the stability properties of the fixed points are not readily available from inspecting the isoclines [Figs. 1(D)–1(F)]. Thus it is necessary to derive a formal stability criterion based on linear perturbation theory.

B. Perturbations

Stability analysis of the stationary states \vec{A}^* is done by linearizing the integral equation (3). For small perturbations

$$\vec{\alpha}(t) = \vec{A}(t) - \vec{A}^* \quad (13)$$

one obtains the linear relation

$$0 = \sum_{j=1}^d \int_{-\infty}^t d\tilde{t} \frac{\delta \mathcal{N}_i(t)}{\delta A_j(\tilde{t})} \alpha_j(\tilde{t}), \quad i = 1, \dots, d. \quad (14)$$

Equation (14) thus imposes a condition for permitted perturbations $\vec{\alpha}(t)$. Since \vec{A}^* is a stationary solution, $\vec{\alpha}(t) = 0$ is always permitted, so the search has to be restricted to nontrivial perturbations $\vec{\alpha}(t) \neq 0$.

To allow further analysis of Eq. (14), one requires an explicit expression of the functional derivative, which, for the present model, is

$$\begin{aligned} \frac{\delta \mathcal{N}_i(t)}{\delta A_j(t-x)} &= S_i^*(x) \delta_{ij} - A_i^* \int_0^\infty du S_i^*(u) \int_{t-u}^t ds \frac{\delta p_i(s|t-u)}{\delta A_j(t-x)} \\ &= S_i^*(x) \delta_{ij} - A_i^* W_{ij} f_i(\vec{A}^*) \int_0^\infty du S_i^*(u) \\ &\quad \times \int_0^u ds g(s) \kappa(x-u+s). \end{aligned} \quad (15)$$

This expression reveals that the functional derivative has no explicit dependence on time t , which justifies to introduce the abbreviation $\Lambda_{ij}(x) := \frac{\delta \mathcal{N}_i(t)}{\delta A_j(t-x)}$. The integrals in Eq. (14) are thus convolutions and the problem can be analyzed more easily in the Laplace domain.

To this end, one assumes that the perturbations have exponential time dependence and, without loss of generality,

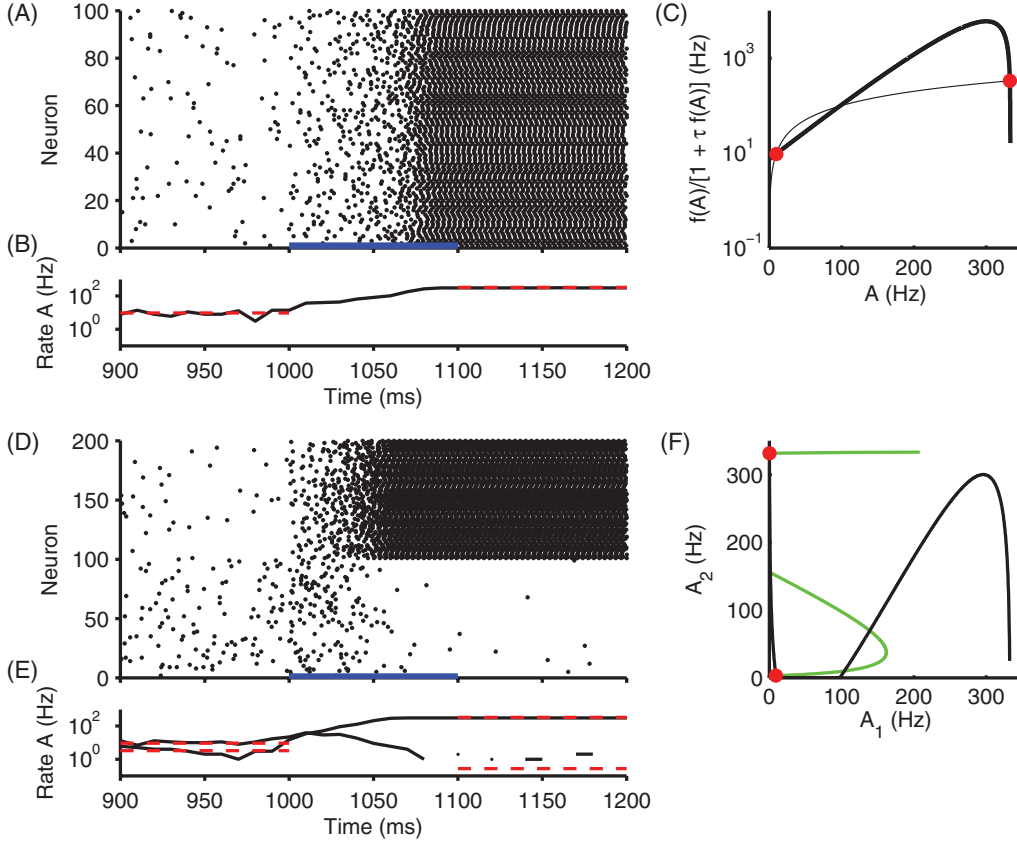


FIG. 1. (Color online) Multistability of excitatory-coupled pools. (A) Spike raster plot of one simulated pool with 100 neurons ($W = 30$, $I = 2$). The thick blue bar on the time axis indicates the time interval at which the neurons received an additional stimulus. (B) Rate $A(t)$ derived from simulation in (A). Red dashed lines indicate the predicted stationary states, which are (C) the points of intersection between the identity (thin line) and the right hand side of Eq. (12) (thick line). Red discs correspond to the stable states from B. (D) Same as (A) for two pools with $W = \begin{pmatrix} 30 & -10 \\ -10 & 30 \end{pmatrix}$, and $I = (2, 1)^T$. (E) Same as (B) for $\bar{A}(t)$ from (D). (F) Isoclines for A_1 (black) and A_2 [green (gray)]. Red discs indicate stationary solutions from (E).

start at time $t = 0$;

$$\vec{\alpha}(t) = \Theta(t) \exp(-zt) \hat{\vec{\alpha}}(z). \quad (16)$$

Note that according to this definition, and though in contrast to most other approaches on linear stability analysis, a positive real part of z indicates a stable perturbation, whereas a negative real part of z indicates instability. Using Eq. (16), the convolution integrals in Eq. (14) read

$$\begin{aligned} \int_{-\infty}^t d\tilde{t} \frac{\delta \mathcal{N}_i(t)}{\delta A_j(\tilde{t})} \alpha_j(\tilde{t}) &= \int_0^t d\tilde{t} \Lambda_{ij}(t - \tilde{t}) \alpha_j(\tilde{t}) \\ &= \int_0^t dt' \Lambda_{ij}(t') \alpha_j(t - t') \\ &= \hat{\alpha}_j(z) e^{-zt} \int_0^t dt' \Lambda_{ij}(t') e^{zt'}. \end{aligned} \quad (17)$$

As a consequence, if Λ_{ij} is integrable, the convolution integral on the left-hand side exists independently of z and t . The integrability conditions for the Λ_{ij} s imply that for $t \rightarrow \infty$, the

integral over such Λ is bounded by

$$\left| \int_0^t dt' \Lambda(t') - L_1 \right| < L_2 e^{-\lambda t}, \quad (18)$$

for some maximal constant $\lambda > 0$. Stability of the stationary solution is then determined by the integral on the right-hand side of Eq. (17) for $t \rightarrow \infty$. Two cases have to be discerned:

(1) If $\text{Re}(z) - \lambda \geq 0$, the integral diverges like $L_2 e^{t(\text{Re}(z) - \lambda)}$. This divergence is balanced by the exponential prefactor e^{-zt} , such that the overall convolution integral converges like $e^{-\lambda t}$ independently of z . As a result, Eq. (14) does not impose any constraint on z and all perturbations with $\text{Re}(z) - \lambda \geq 0$ are permitted. Since in particular $\text{Re}(z) > 0$, all those perturbations are stable. As an intuitive interpretation, $\text{Re}(z) \geq \lambda$ means that the perturbations die out faster [at rate $\text{Re}(z) > 0$] than the integral (at rate $\lambda > 0$) and thus do not contribute in the limit $t \rightarrow \infty$.

(2) If $\text{Re}(z) - \lambda < 0$, the integral on the right-hand side exists for all t . One thus can properly define the Laplace transform $\hat{\Lambda}(z) = \int_0^\infty dx \Lambda(x) e^{xz}$, and, requiring independence of time (i.e., disregarding the prefactor e^{zt}), Eq. (14) turns into a

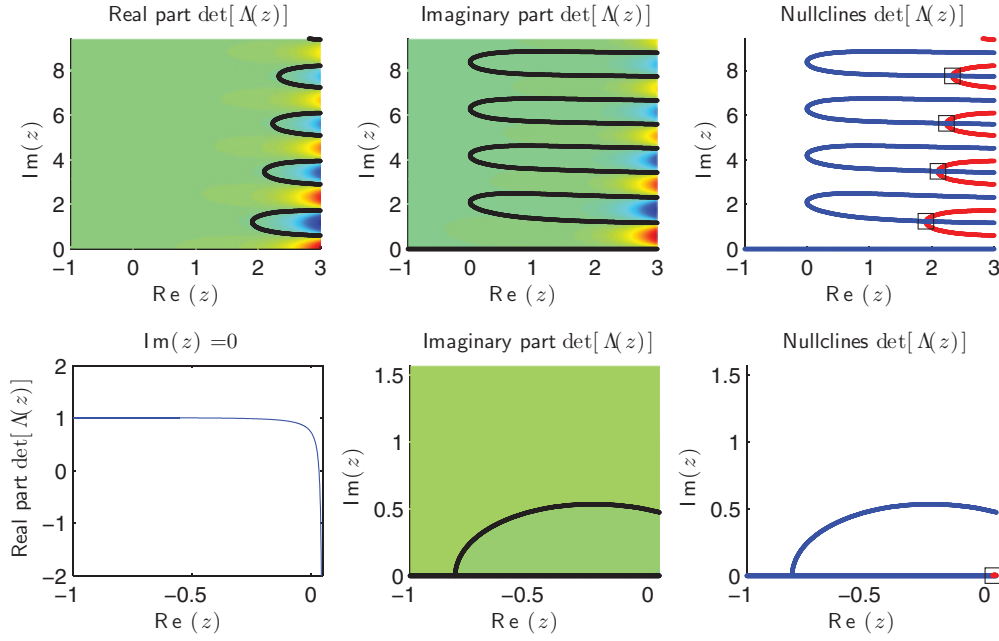


FIG. 2. (Color online) Complex roots. (Top) Real and imaginary part [color (gray) coded] of $\det \hat{\Lambda}(z)$ for uncoupled pools $W = 0$. Thick black lines indicate respective nullclines. Right: roots of $\det \hat{\Lambda}(z)$ (black squares) are determined by the intersection of the nullclines (red for real part, blue for imaginary part). (Bottom) Introducing synaptic self-coupling $W = 30$ limits the range of slow real parts to $\text{Re}(z) < \beta = 0.05$. Left: The singularity in $\hat{\kappa}(z)$ guarantees a zero crossing of the real part at $\text{Im}(z) = 0$ (the two right-most panels are equivalent to top).

linear matrix equation,

$$0 = \hat{\Lambda}(z) \hat{\alpha}(z). \quad (19)$$

Being a homogeneous linear equation for $\hat{\alpha}(z)$, the existence of nonzero perturbations imposes the condition

$$\det[\hat{\Lambda}(z)] = 0. \quad (20)$$

The roots z of Eq. (20) determine the possible perturbation modes $\hat{\alpha}(z) \neq 0$ at the respective fixed point. The real part $\text{Re}(z)$ determines the stability of the mode: $\text{Re}(z) > 0$ accounts for a stable mode, whereas $\text{Re}(z) < 0$ identifies an unstable mode. A nonvanishing imaginary part $\text{Im}(z) \neq 0$ indicates an oscillatory perturbation, with period $2\pi/\text{Im}(z)$. Since $\text{Re}(z) - \lambda < 0$, stable perturbations decay more slowly than the convolution integrals in Eq. (14). I thus call them *slow* modes.

Using the specific model assumptions for the functions f and g one finds explicit forms for the survivor function,

$$\hat{S}_i^*(z) = \frac{1}{f_i(\vec{A}^*) - z} \left(f_i(\vec{A}^*) \frac{e^{z\tau} - 1}{z} + 1 \right) \quad (21)$$

and the matrix

$$\hat{\Lambda}_{ij}(z) = \delta_{ij} \hat{S}_i^*(z) - \frac{A_i^* W_{ij} \hat{\kappa}(z)}{f_i(\vec{A}^*) - z}. \quad (22)$$

The feasibility condition from Eq. (20) is hence equivalent to

$$\det \left[\left(f_i(\vec{A}^*) \frac{e^{z\tau} - 1}{z} + 1 \right) \delta_{ij} - \hat{\kappa}(z) M_{ij} \right] = 0 \quad (23)$$

with $(M)_{ij} = M_{ij} = A_i^* W_{ij}$. Let us next consider two special cases in which Eq. (23) becomes particularly simple.

No refractoriness. If the refractory time is chosen $\tau = 0$, the neurons are allowed to fire with infinite rate. In that case, the first term in Eq. (23) simplifies to the unit matrix $\mathbb{1}$ and the feasibility condition can be translated to

$$\det(\hat{\kappa}(z)^{-1} \mathbb{1} - M) = 0. \quad (24)$$

This equation can be readily solved for those values of z at which $\hat{\kappa}^{-1}(z)$ equals an eigenvalue μ of M . For the specific choice of an exponential coupling kernel, one finds

$$\hat{\kappa}(z) = \frac{e^{\Delta z}}{1 - z/\beta} \quad (25)$$

and thus the feasibility condition reads

$$e^{-\Delta z} (1 - z/\beta) = \lambda \in \sigma(M). \quad (26)$$

It is shown in the Appendix that (1) Eq. (26) has exactly one root for real positive eigenvalues μ of M . Moreover, it is shown that (2) if the modulus $|\mu| < 1$, a root z has always positive real part and the perturbation is stable, and (3) z has a nonzero imaginary part only if also μ has a nonzero imaginary part. Imaginary parts of the eigenvalues thus directly indicate oscillatory behavior of the perturbations.

No coupling. If pools are considered as uncoupled, i.e., $W = 0$, then the feasibility conditions from Eq. (23) maps to

$$\prod_{i=1}^d \left(f_i(\vec{A}^*) \frac{e^{z\tau} - 1}{z} + 1 \right) = 0. \quad (27)$$

Hence the roots z of this condition are given by

$$\frac{e^{z\tau} - 1}{z} = -f_i^{-1}(\vec{A}^*), \quad i = 1, \dots, d. \quad (28)$$

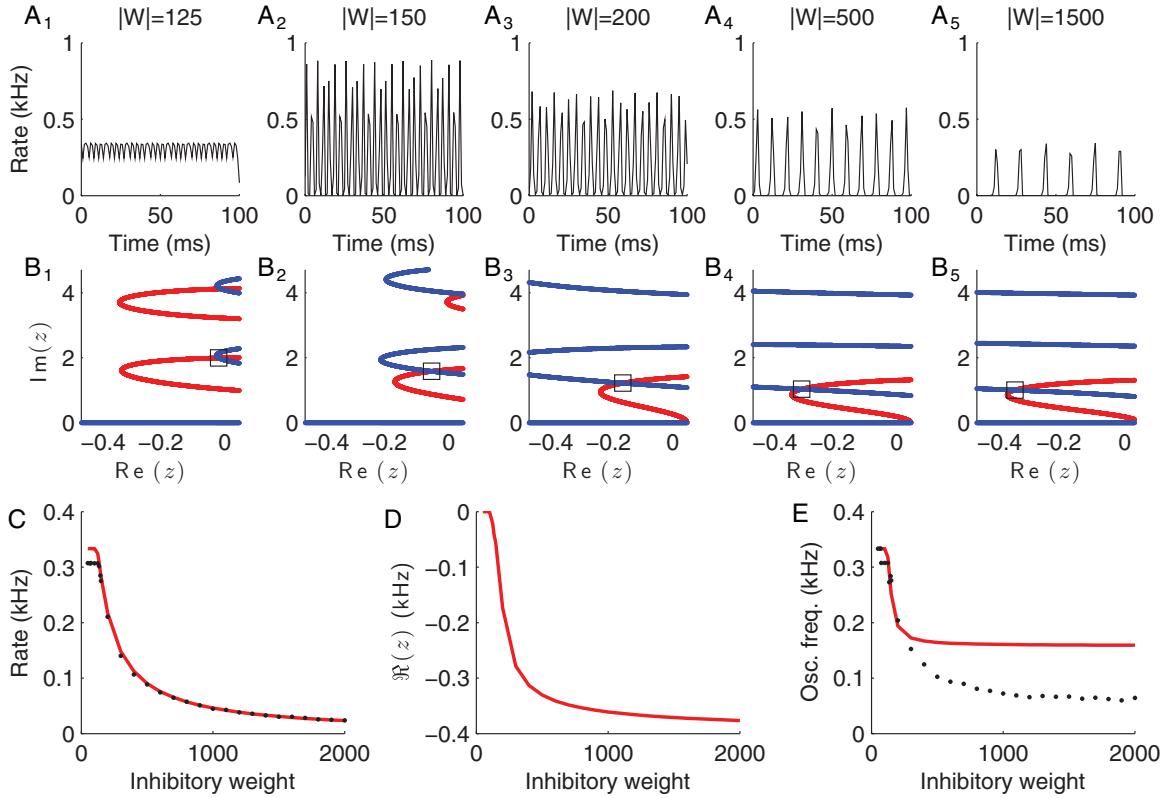


FIG. 3. (Color online) Population oscillations. (A_{1–5}) Population rate from network simulations for five different inhibitory synaptic coupling strengths $|W|$. (B_{1–5}) Roots (squares) of the feasibility conditions for the weights from A. Color (gray) code corresponds to Fig. 2, top right. (C) Empirical firing rate (black dots) and stationary state (red solid line) from the fixed point equation (12). (D) Real part of the root of Eq. (23). (E) Imaginary part (divided by 2π , red) of the root of Eq. (23) and empirical oscillation frequency (black dots).

For each value of $f_i(\vec{A}^*)$ Eq. (28) can have multiple complex roots, as shown in the upper panel of Fig. 2.

General case. In general, neurons are refractory $\tau > 0$ and coupled $W \neq 0$ and hence the feasibility condition can only be evaluated numerically. Most notably also for arbitrarily small positive weights $W_{ij} > 0$ the pole of \hat{k} at $z = \beta$ [Eq. (25)] introduces a slow root with $\text{Re}(z) < \beta$ (Fig. 2, bottom). This shows that positive feedback, no matter how small, always introduces a slow time scale. Since the integrability of the Laplace transform of Λ requires $\text{Re}(z) < \beta$, the roots of the uncoupled network are no longer effective in that case and the network dynamics permits all (stable) perturbations at time scales faster than $1/\beta$.

Here and unless otherwise mentioned the following parameters are used: the decay rate of the coupling kernel $\beta = 1/(20 \text{ ms})$, indicating a neuronal integration time constant of 20 ms; the refractory time $\tau = 3 \text{ ms}$, imposing a maximal firing rate of 333 Hz; the synaptic delay $\Delta = 2 \text{ ms}$.

IV. APPLICATIONS

Our understanding of dynamical systems is dominated by concepts derived from studying ordinary differential equations, such as, e.g., phase space trajectories and fixed points. Integral dynamics such as the one presented here do not allow using these intuitive ideas, at least in the strict sense. In this section, I will introduce three examples of collective neuronal

phenomena that have been (partly extensively) studied in the past to illustrate how the theory presented here applies to them. Interestingly, the slow stable modes turn out to be quite well reflected by the classical concept of a phase space.

A. Gamma rhythm $d = 1$

As an example for a homogeneously coupled population of neurons (a single pool, $d = 1$), I reconsider inhibitory networks that generate oscillations in the gamma range (30–100 Hz). All neurons receive the same positive drive $I = 50$, and the synaptic weight W is used as a parameter to control the population oscillation. In line with a multitude of previous studies (e.g., [16–19]), Fig. 3(A) shows that oscillation frequency of a simulated pool of $N = 4,500$ neurons decreases with increasing inhibitory weight $|W|$.

In the language of the present theory, the dependence of oscillation frequency on synaptic coupling is reflected by the corresponding roots of the feasibility condition from Eq. (23). All roots have a negative real part and a nonvanishing imaginary part [Fig. 3(B)] indicating that the stationary states are unstable and that the unstable perturbations exhibit oscillatory nature. A comparison between the stationary rate A^* from Eq. (12) and the mean spike rate of the simulations shows excellent agreement [Fig. 3(C)] despite the fast speeds of divergence from the asynchronous state, as indicated by the strongly negative real parts of the roots in Fig. 3(D). Only for

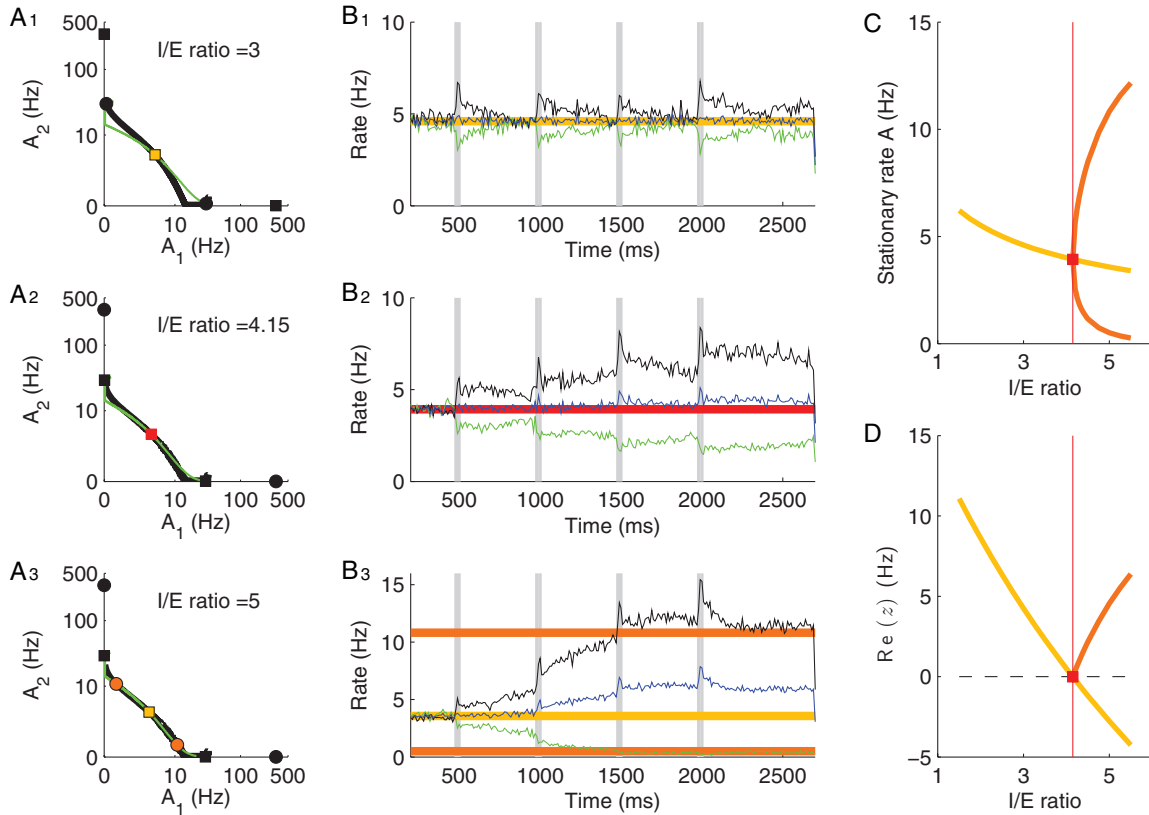


FIG. 4. (Color online) Line attractor. (A_{1-3}) State spaces for three different I/E ratios (as indicated). Squares indicate stable stationary solutions, discs indicate unstable stationary solutions. Black thick and green thin lines mark the isoclines (below 33 Hz); cf. Fig. 2. Note that axes are logarithmically scaled. (B_{1-3}) Simulations corresponding to the respective situations in (A). Black (upper) and green (lower) traces represent the rates derived from the two individual pools. The blue (middle) trace is the average activity of the system. Gray vertical bars mark the times of the stimuli. The colors (gray) of levels of the horizontal lines correspond to the stationary states from (A). (C),(D) Bifurcation diagrams. (C) Stationary states as a function of the I/E ratio [colors (gray) of levels from (A) and (B)]. (D) Roots of the feasibility condition from Eq. (23).

very low values of $|W|$ does the stationary rate overestimate the simulation results, as the strict upper bound of the firing rate induces a bias.

The $|W|$ dependence of the mean rate suggests that the dynamics is still governed by the asynchronous (unstable) stationary state. It is thus justified to also compare the oscillation frequency of the simulated network with the imaginary part of the root of Eq. (23): For small modulus of the weight $|W|$ (fast oscillations) the imaginary part turns out to be a good predictor [Fig. 3(E)]. However, as the modulus $|W|$ increases, the empirical oscillation frequency becomes only about half of the predicted one, indicating that the stability analysis only qualitatively captures the network dynamics in this regime.

B. Line attractors $d = 2$

Irregular-asynchronously firing pools of neurons can only encode inputs by their mean firing rate. It is therefore interesting to study if the linear perturbation theory can predict whether two pools can fire with a continuum of firing rates. A dynamical system that exhibits such a continuum of stable states is called a line attractor. In neurobiology, line

attractors are considered to be necessary for integrating stimuli and keeping the integral value in short-term memory for a sufficiently long period of time. Line attractors have been suggested to stabilize eye position [7] and to perform path integration [20].

Here a line attractor is considered to consist of two pools with positive self-coupling and negative cross coupling. The specific choice of the coupling matrix is $W = 50 \begin{pmatrix} 1 & \\ -\rho & -1 \end{pmatrix}$, in which ρ measures the ratio between inhibitory and excitatory coupling strength (I/E ratio). In addition, the pools receive homogeneous excitatory drives $\vec{I} = (2, 2)^T$.

The stationary states are then analyzed in dependence on the I/E ratio ρ (Fig. 4). For weak inhibition (small ρ) the system exhibits five stationary solutions: two stable oscillatory solutions for which one of the pools is silent and the other pool fires at maximal rate, a symmetric stable solution in which both pools fire at the same low rate, and two unstable solutions that separate the basins of attraction of the stable solutions [Fig. 4(A₁)]. At some critical I/E ratio ρ_c the symmetric stationary solution loses stability, i.e., the real part of the root of Eq. (23) becomes zero and the isoclines become parallel at the stationary point [Fig. 4(A₂)]. For $\rho > \rho_c$ the symmetric

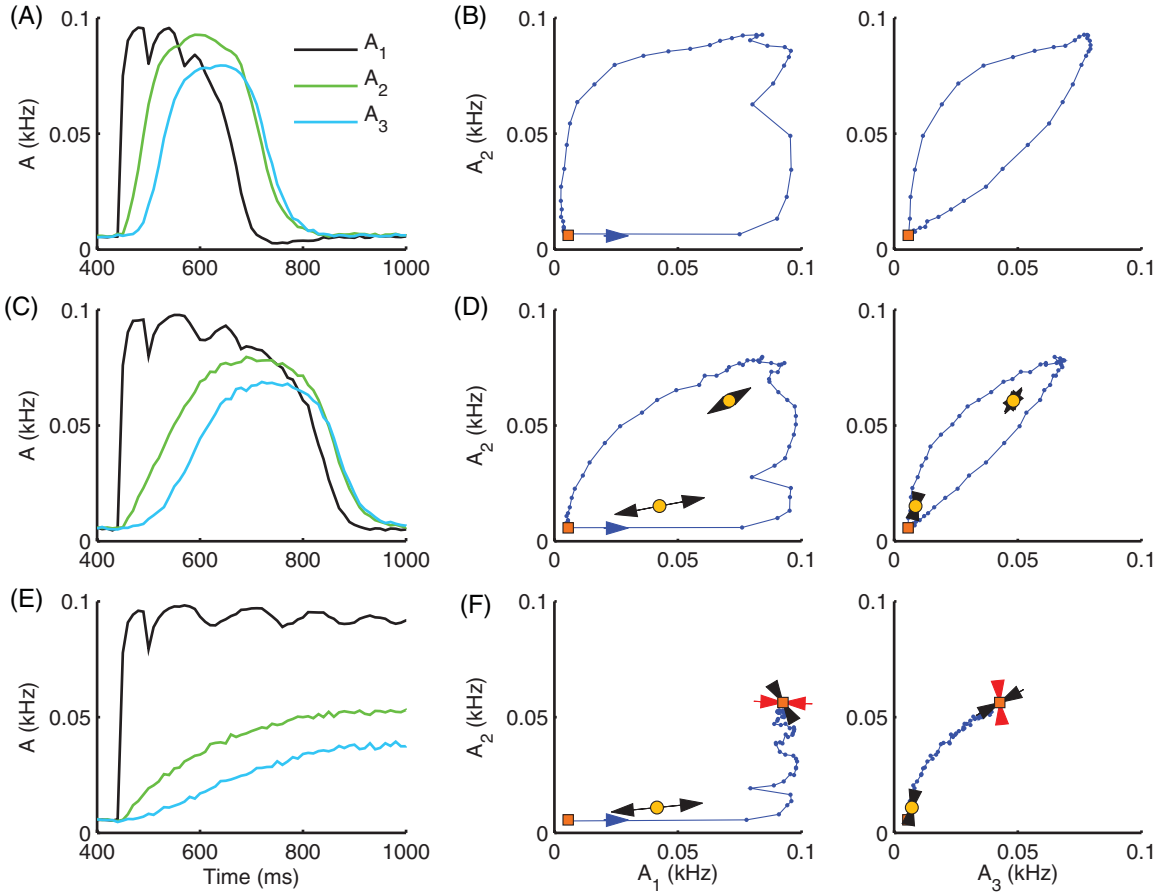


FIG. 5. (Color online) Slow transients. (A),(C),(E) Firing rates of three coupled pools from stochastic simulations. Pool 1 received a strong stimulus of amplitude $I_1 = 6.5$ between 450 and 500 ms. The coupling strength k between the first and the second pool is $k = 30$ [in (A)], $k = 20$ [in (C)], and $k = 15$ [in (E)]. The rates are averaged over 10 ms. (B),(D),(F) Two different projections (blue lines with dots) of the activity variables from the simulations in (A),(C),(E). The arrow in the left graphs indicate the time course. The temporal separation of the dots is 10 ms. Theoretically obtained fixed points are plotted as orange squares (stable) and yellow discs (unstable). The arrows indicate the direction and magnitude of the slow stable and unstable modes. Arrows are drawn black if the imaginary part of the corresponding root of Eq. (23) is zero. Red (gray) arrows indicate oscillatory modes with nonzero imaginary part of the root. Here, the refractory time is $\tau = 10$ ms, and the coupling kernel has a time constant of $1/\beta = 30$ ms.

stationary state is unstable and spawns two new neighboring stable solutions [Fig. 4(A₃)]. The resulting bifurcation diagram [Figs. 4(C) and 4(D)] thereby reminds one of a pitch fork bifurcation.

Linear perturbation analysis thus predicts that at the critical I/E ratio ρ_c the system should behave as a line attractor. This prediction was tested by three simulations for $\rho < \rho_c$, $\rho \approx \rho_c$, and $\rho > \rho_c$. The setting was such that after an initialization phase of 500 ms, the pools received four additional current stimuli $\vec{I}^{(k)} = a(1, -1)^T, k = 1, \dots, 4$, with amplitude $a = 0.25$ for small time intervals of 20 ms at times $t^{\text{stim}(k)} = k \cdot 500$ ms. For $\rho < \rho_c$ (where the symmetric stationary state is stable), the firing rates of the pools always decays back to the symmetric state [yellow in Fig. 4(B₁)], i.e., the system did not show considerable short-term memory. For $\rho > \rho_c$, where the symmetric stationary state is unstable, the firing rates drift away from the symmetric stationary state and approach one of the stable asymmetric states [orange in Fig. 4(B₃)]. Only at the critical I/E ratio $\rho = \rho_c$, the firing rates remain at different levels for a duration that considerably exceeds

the coupling time constant of 20 ms: In the vicinity of the symmetric stationary state, the system acts as a continuous attractor [Fig. 4(B₂)].

To conclude, linear perturbation theory is able to also quantitatively identify network states in which the system integrates a stimulus. These continuous attractor states are particularly sensitive to noise. The pools thus have to be either large (as in the present example), or the intrinsic time constants (e.g., $1/\beta$) have to be long.

C. Trajectories in activity space $d = 3$

The example of the line attractor shows that pool dynamics can act on time scales that are much slower than the intrinsic cellular processes such that the temporal evolution of pool firing rates is a true network phenomenon. One thus may ask whether such coupled pools may also be able to represent long time intervals themselves, making use of the time it takes to follow a specific trajectory, such as a limit cycle or a slow transient sequence of activity states. Such activity

sequences have been investigated in various flavors in the past, e.g., [21–23]. These models consider sequential activation of neuron populations either as a complex spatiotemporal memory trace that is stored in the network, or as a short-term memory buffer for sensory stimuli. To the best of my knowledge, no model so far was able to provide a macroscopic description of activation sequences of recurrently coupled pools that is rigorously derived from microscopic networks of spiking neurons in continuous time.

In this section, an implementation of such a slow trajectory is presented for the example of three coupled pools. The first pool thereby represents a bistable switch (cf. Fig. 1) that acts as a short-term memory buffer for a brief (50 ms) input pulse. Once activated, the first pool excites a second pool, which exhibits a slow buildup in rate that acts as a time delay. Finally, a third inhibitory pool is driven by the second pool. The third pool inhibits the first pool and stops the short-term memory buffer. As a result, the dynamical system is set back into a low-activity stable state. The dynamics is realized using the weight matrix

$$W = \begin{pmatrix} 70 & 0 & -20 \\ k & 35 & 0 \\ 0 & 50 & 0 \end{pmatrix}, \quad (29)$$

in which we use the weight k between the first and the second pool as parameter to analyze the system behavior. Additionally, all pools receive a constant drive $\vec{I} = \frac{3}{2}(1, 1, 1)^T$. Typical trajectories from cellular simulation are shown in Fig. 5.

The duration of the transient burst of activity can be adjusted by the coupling strength k . If k is large the delay between activation of the first and the second pool is relatively short and so is the burst duration. The “trajectory” in the \vec{A} space is dominated by two stationary states with high rates. For large $k > k_c^{(1)}$ these stationary states do not actually exist, although their “ghosts” still mediate slow trajectories [Figs. 5(A) and 5(B)]. At some critical $k = k_c^{(1)}$ the high-rate stationary states appear [Figs. 5(C) and 5(D)] and the dynamics slows down further for $k < k_c^{(1)}$. Below a second critical value $k < k_c^{(2)} < k_c^{(1)}$, one of the stationary states becomes stable and thus the transient behavior disappears [Figs. 5(E) and 5(F)]. In this regime, the dynamics is dominated by the bistability of the first pool.

To conclude, linear stability analysis of the normalization condition from Eq. (3) allows qualitative statements about the dynamics of the high-dimensional dynamical system, similar to classical systems of ordinary differential equations [24], although, strictly speaking, the notion of phase spaces and trajectories is incorrect for an integral dynamics. For example, trajectories may intersect as the dynamics is not local in time and, similarly, the arrows in Fig. 5 do not represent a flow field but the direction of slow stable and unstable modes.

V. DISCUSSION

This paper presents a theoretical approach to analyze the firing of coupled pools of neurons. A stability criterion is derived for stationary states, which requires finding the complex roots of a matrix determinant. These roots indicate

slow stable and unstable modes, i.e., activity features that go beyond the single-cell time scales and as such represent collective dynamical phenomena.

The theory is applied to well-studied examples, which particularly reveal slow dynamical features. It is important to understand which collective mechanisms can underly such slow time scales, because they are not readily accessible to experimental measurements. Slow *cellular* time scales, however, are much easier to track down experimentally, and thus possible additional network mechanisms are often not recognized.

So far, the presented theory only applies to simple (one-dimensional) neuron models, that cannot extend to relevant intrinsic behaviors such as adaptation or slow positive self-feedback. Particularly, such interplay between cellular and network mechanisms may be important to understand slow neuronal processing [7]. An extension to high-dimensional neuron models has been achieved for a pool model without explicit renewal property [25]. For the present theory such extensions are not yet available as they require extending the renewal assumption, one of the corner stones of the approach. Yet it seems possible but tedious to generalize the dynamical equation (3) to many previous spike times.

Slow time scales of neuronal network activity have been described in many contexts. Aside from the ones mentioned (integration of retinal slip and path integration), such activity patterns are generally assumed to occur for most short-term memory tasks. A well-studied example is decision making, where a slow time evolution of firing rates has been observed in the prefrontal cortex preceding the decision [4,26]. More recently, short-term memory has also been connected to short-term synaptic plasticity [27,28], which means that there the memory is not represented by long-lasting firing, but by cell- (or synapse)-intrinsic dynamical processes. In order to study the interaction of the two mechanisms (synaptic and network), future work requires also incorporating synaptic short-term plasticity into the renewal pool description presented in this paper.

ACKNOWLEDGMENTS

The author is grateful to Chun-Wei Yuan for comments on the manuscript. This work was funded by the German Ministry of Education and Research (BMBF) Bernstein Fokus Neuronal Basis of Learning (Grant No. 01GQ0981) and Bernstein Center for Computational Neuroscience Munich (Grant No. 01GQ0440).

APPENDIX

In the case of nonrefractory neurons $\tau = 0$ the permitted slow stable and unstable perturbations are determined as the roots $z = x + i y$ of

$$\left(1 - \frac{z}{\beta}\right) e^{-\Delta z} = \mu = |\mu| e^{i \arg(\mu)} \quad (A1)$$

with μ being an eigenvalue of the matrix $M_{ij} = A_i W_{ij}$. Moreover, integrability of the Laplace transform requires

$\text{Re}(z) < \beta$. Equation (A1) can be split up into modulus and phase. For the modulus one obtains

$$y_{\pm}(x) = \pm\beta\sqrt{\left[|\mu|^2 e^{2x\Delta} - \left(1 - \frac{x}{\beta}\right)^2\right]}, \quad (\text{A2})$$

which defines two functions $y_+(x) > 0$ and $y_-(x) = -y_+(x)$. Note that y_{\pm} are only defined for real parts at which

$$\left(1 - \frac{x}{\beta}\right) < |\mu| e^{x\Delta}. \quad (\text{A3})$$

For $|\mu| < 1$, the possible x values are thus restricted to positive values and the perturbations are stable. If the eigenvalue μ is real and positive, $\mu = 1$ yields the solution $z = 0$ and thus

marks the verge of stability, separating unstable ($\mu > 1$) and stable ($\mu < 1$) perturbations.

The phase part of Eq. (A1) yields the implicit equations

$$\Delta y_{\pm}(x) + \arctan\left(\frac{y_{\pm}(x)}{\beta - x}\right) = -\arg(\mu). \quad (\text{A4})$$

More specifically, as complex eigenvalues of real matrices M always come in pairs of complex conjugates, for any solution with imaginary part y there is also one with imaginary part $-y$ (corresponding to the complex conjugate eigenvalue). The phase equation (A4) also reveals that, since $\Delta > 0$, the imaginary part y_{\pm} of the solution vanishes only for $\arg(\mu) = 0$ and then the real part is uniquely determined by $1 - x/\beta = \mu e^{\Delta x}$. Nonzero imaginary parts of μ always also result in imaginary parts of z .

-
- [1] J. O'Keefe and M. Recce, *Hippocampus* **3**, 317 (1993).
 [2] E. Pastalkova, V. Itskov, A. Amarasingham, and G. Buzsaki, *Science* **321**, 1322 (2008).
 [3] G. Dragoi and S. Tonegawa, *Nature (London)* **469**, 397 (2010).
 [4] C. D. Brody, R. Romo, and A. Kepecs, *Curr. Opin. Neurobiol.* **13**, 204 (2003).
 [5] M. N. Shadlen and W. T. Newsome, *Curr. Opin. Neurobiol.* **4**, 569 (1994).
 [6] E. Franssen, E. Tahvildari, A. V. Egorov, M. E. Hasselmo, and A. A. Alonso, *Neuron* **49**, 735 (2006).
 [7] G. Major and D. Tank, *Curr. Opin. Neurobiol.* **14**, 675 (2004).
 [8] H. R. Wilson and J. D. Cowan, *Biophys. J.* **12**, 1 (1972).
 [9] H. R. Wilson and J. D. Cowan, *Kybernetik* **13**, 55 (1973).
 [10] P. C. Bressloff, *Phys. Rev. E* **82**, 051903 (2010).
 [11] C. K. Machens and C. D. Brody, *Neural Comput.* **20**, 452 (2008).
 [12] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster, *Neuron* **62**, 578 (2009).
 [13] W. Gerstner, *Phys. Rev. E* **51**, 738 (1995).
 [14] W. Gerstner, *Neural Comput.* **12**, 43 (2000).
 [15] J. Eggert and J. L. van Hemmen, *Neural Comput.* **13**, 1923 (2001).
 [16] C. Leibold, *Phys. Rev. Lett.* **93**, 208104 (2004).
 [17] W. Gerstner, J. L. van Hemmen, and J. D. Cowan, *Neural Comput.* **8**, 1653 (1996).
 [18] R. D. Traub, M. A. Whittington, S. B. Colling, G. Buzsaki, and J. G. Jefferys, *J. Physiol.* **493**, 471 (1996).
 [19] N. Brunel, *J. Comput. Neurosci.* **8**, 183 (2000).
 [20] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M. B. Moser, *Nat. Rev. Neurosci.* **7**, 663 (2006).
 [21] J. Buhmann and K. Schulten, *Europhys. Lett.* **4**, 1205 (1987).
 [22] S. Dehaene, J. P. Changeux, and J. P. Nadal, *Proc. Natl. Acad. Sci. USA* **84**, 2727 (1987).
 [23] M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. I. Abarbanel, and G. Laurent, *Phys. Rev. Lett.* **87**, 068102 (2001).
 [24] G. B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience* (Springer, New York, 2010).
 [25] T. Prager, H.-P. Lerch, L. Schimansky-Geier, and E. Schöll, *J. Phys. A* **40**, 11045 (2007).
 [26] C. K. Machens, R. Romo, and C. D. Brody, *Science* **18**, 1121 (2005).
 [27] C. Leibold, A. Gundlfinger, R. Schmidt, K. Thurley, D. Schmitz, and R. Kempter, *Proc. Natl. Acad. Sci. USA* **105**, 4417 (2008).
 [28] G. Mongillo, O. Barak, and M. Tsodyks, *Science* **319**, 1543 (2008).