RESTRICTED PHASE-LOCKING DYNAMICS OF PERIODICALLY FORCED NETWORK GAMMA RHYTHMS

JONATHAN CANNON AND NANCY KOPELL

1. Introduction

[[[Ditch theta: use resetting and spiking thresholds, and then let the go to infinity!]]]

Gamma-frequency (30-100Hz) electrical rhythms in the brain are associated with attention, ..., and active sensory processing, and have therefore been subject of many modeling efforts and much theoretical investigation. One major theme in this scholarly dialogue is the Communication Through Coherence (CTC) hypothesis, stating that the brain can dynamically establish a line of communication between two cortical areas by establishing a consistent phase relationship between the local gamma rhythms in those areas. The CTC hypothesis has been investigated in vivo and in simulation, but there has been little progress towards establishing a theoretical groundwork for CTC. In particular, the mechanism by which the brain establishes a specific, optimal interregional phase relationship remains largely unexplored.

Here, we show that the cellular mechanism implicated in gamma rhythms is ideally suited to establishing a predictable interareal phase relationship. We study a simple model of the gamma-rhythmic mechanism underlying interneuronal network gamma (ING) and pyramidal-interneuronal network gamma (PING), subjected to periodic forcing. We draw on previous work in which we showed mathematically that this circuit mechanism is drawn onto an invariant attracting torus in its state space, limiting its behavior to phase locking and precession. In this work, we prove that if this circuit phase locks to periodic forcing (in this case, square pulses of arbitrary strength and duration), it must do so mono-stably: only one characteristic phase relationship between forcing and local network spiking may be stable. We argue that the combination of a persistent attracting invariant torus and monostable phase-locking makes the an ideal candidate for CTC.

2. Model

Here we design a model in order to study the periodically-forced behavior of gamma rhythms generated by two different network mechanisms. In the first, interneuronal network gamma

(ING), a population of fast-spiking interneurons under tonic excitation fire together and then recover and spike again after the inhibition decays over the course of a gamma cycle. In the second, pyramidal-interneuronal network gamma (PING), a population of pyramidal neurons under tonic drive fire together, triggering a rapid synchronous response from an associated population of fast-spiking interneurons, and then recover and spike again after the inhibition decays over the course of a gamma cycle.

The question of when and how populations of cells synchronize into gamma rhythms is interesting and important, but peripheral to our work here, where we are more concerned with their behavior during the generation of gamma; therefore, we propose a simple model in which cells are assumed to fire synchronously. We represent the (synchronized) membrane potential of a population of cells with the voltage V of a QIF neuron [**Latham2000**], and we represent the level of inhibition shared by that population as a scalar s in $[0,1]$. A T_I-periodic forcing current $I(\Phi)$, is delivered to V, where Φ is the phase of the forcing cycle. Thus, the ODE for our model is

$$\begin{cases} \dot{V} = & \frac{1}{\tau}(V^2 + G) \\ \dot{s} = & -\frac{s}{\tau_s} \\ \dot{\Phi} = & 1 \end{cases} \tag{2.1}$$

where

$$G = b - gs + I(\Phi).$$

is the net flux of current; g is the maximal conductance of the inhibitory synapses; τ_s is the decay time constant of inhibition; τ is the membrane time constant of the cells in the population; and $b > 0$ is the baseline level of tonic excitation to the population. The forcing phase $\Phi \in [0, T_I)$ is on $\mathbb{T}^1$, and $I(\cdot)$ is a T_I-periodic piecewise-continuous function representing the periodic drive to the I-population. When V reaches some threshold $V_{spike} \in (0, \infty]$, it "spikes" and resets to $V_{reset} \in [-\infty, 0)$; at each spike, s resets instantly to

$$\rho(s) = 1 + c(s-1). \tag{2.2}$$

We call this synaptic resetting rule ρ "linearly resetting synapses." In [[[ATTRACTING INVARIANT TORI IN PERIODICALLY FORCED NETWORK GAMMA RHYTHMS]]], we show that linearly resetting synapses are a natural simplification of fast synaptic rise dynamics during a spike.

We call this model the "NG (neuronal gamma) oscillator." If we assume that the population of cells described by V is a population of fast-spiking interneurons, the NG oscillator is a simple model of ING where the interneurons are receiving periodic forcing. If we instead assume that V describes a population of pyramidal neurons whose volleys of spikes rapidly trigger an interneuron spike volley and reset the inhibition level, then the NG oscillator is a simple model of PING (where forcing is delivered to the pyramidal cells rather than the interneurons).

Additional Notation:

- When we want to consider a circular variable (e.g., Φ) in its lift to $\mathbb{R}$, we use an overbar (e.g., $\bar{\Phi}$).
- We write V_t to refer to the value of V at time t. Given an ODE and a set of initial conditions V_0, s_0, Φ_0, we write $V_t(V_0, s_0, \Phi_0)$ to refer to the value of V when the system is initialized with these initial conditions and allowed to flow forward by time t. We use s_t and Φ_t similarly.

[[[BE MORE SPECIFIC ABOUT WHAT YOU DO AND DON'T USE]]]

3. Monostable phase locking

We are interested in the conditions under which the only asymptotically stable behavior of this system is a single 1:1 phase-locked orbit, i.e., after a transient the population in question spikes exactly once at the same phase on every forcing cycle. A priori, a two-dimensional oscillatory system subjected to periodic forcing may engage in a wide range of behaviors: the periodically-forced relaxation oscillator, for example, shows period-doubling, chaos, and, most relevantly to our work here, bistability of phase-locked orbits. In [[[ATTRACTING INVARIANT TORI IN PERIODICALLY FORCED NETWORK GAMMA RHYTHMS]]], we show that for sufficiently small $c > 0$ in the definition of the resetting map ρ, the dynamics of the NG oscillator are asymptotically attracted to an invariant torus on which all trajectories are either periodic (phase-locked) or quasi-periodic (precessing regularly relative to the forcing). However, this result does not rule out bi- or multi-stability: any number of stable periodic trajectories may exist on a torus as long as they all wind around the torus the same number of times in each direction, i.e., phase lock with the same number of spikes per forcing cycle.

One tool commonly used in the study of phase locking is the phase response curve. An oscillator's phase response curve is a function that takes as an argument the oscillator's phase and describes the result of an instantaneous perturbation arriving at that phase by returning the resulting advance or delay of the oscillator. The intersections between a phase response curve and a horizontal line at height H represent phases at which the perturbation can shorten the oscillator's period by exactly H, and hence can phase lock the oscillator at a period H-shorter than its natural period. The slope of the phase response curve can be used to evaluate the stability of the corresponding phase-locked orbit: only intersections with sufficiently shallow downward slope can be stable [[[citation]]]. The phase response curve of the Fitzhugh-Nagumo relaxation oscillator shows two distinct peaks, and can therefore be used to show the possibility of two stably phase-locked orbits [[[citation]]].

The phase response curve is best-suited to very short perturbations (though it can in some situations be generalized to describe the effects of extended perturbations [[[citation]]]). It is also limited in that the oscillator is assumed to return to its natural limit cycle between

the arrivals of perturbations (though again, some progress has been made in lifting this restriction [[[citation]]]).

Both of these limitations are easily circumvented by studying phase locking through a different function, the "firing map" [[[citations]]]. Instead of taking advantage of the specific, pulsatile form of the forcing, we take advantage of the spiking structure of the oscillator. The firing map takes as an argument the phase of the periodic forcing signal at one oscillator spike, and returns the phase at the next. This formulation replaces the assumption of pulsatile forcing with an assumption of periodic forcing, and replaces the assumption of return to a natural limit cycle with the assumption that the system state at each spike may be completely characterized by the forcing phase at that spike. In the case of the NG oscillator, when $c = 0$ in (2.2), this is indeed the case: at the right-hand limit of each spike, we have $s = 1$ and $\theta = -\pi$, so the system state is characterized completely by Φ. We let $\mathbf{P}(\Phi)$ denote the firing map from Φ at one spike to Φ at the next. If $\mathbf{P}$ is applied repeatedly, the result is a one-dimensional discrete dynamical system. As such, intersections between $\mathbf{P}$ and the diagonal signify phase locking: when an initial spikes occur at one of these forcing phases, the following spike occurs at the same phase, so an integer number of forcing cycles has elapsed between spikes. The slope of $\mathbf{P}$ at an intersection represents the stability of phase locking at that phase: a slope strictly between 1 and -1 represents a stable phase lock, and a slope with magnitude bounded away from 1 cannot be stable.

We let $\bar{\Phi}$ represent a lift of Φ to $\mathbb{R}$, and we let $\bar{\mathbf{P}}(\bar{\Phi})$ denote the natural lift of the map $\mathbf{P}$ to $\mathbb{R}$ that acts on forcing phase $\bar{\Phi} \in \mathbb{R}$. We use $\bar{\mathbf{P}}$ to define an inter-spike interval (ISI) function $\Psi(\Phi)$:

$$\Psi(\bar{\Phi}) = \bar{\mathbf{P}}(\bar{\Phi}) - \bar{\Phi}. \tag{3.1}$$

$\Psi(\bar{\Phi})$ returns the ISI following a spike at forcing phase $\bar{\Phi}$. It is easy to check that Ψ is T_I-periodic, so we abuse notation and write $\Psi(\Phi)$ to represent the ISI function on $[0, T_I) = \mathbb{T}^1$. Intersections between Ψ and the horizontal at height T_I signify 1:1 phase locks, and only intersections where Ψ crosses from above to below can be stable. Therefore, in order to prove that a system can only phase lock stably at a single phase, it is sufficient to show that Ψ may cross the horizontal of height T_I only twice, and may cross only once from above to below.

In the next section, we use the linearization of the model to study the evolution of variations in V, which allows us to write an expression for $\Psi'(\Phi)$ that we use to show that the number of crossings is limited to two. This expression requires that we assume the forcing to consist of a single periodic square pulse of any height and duration. In principle, the same methods could be used to study the response of the oscillator to any periodic forcing function $I(\Phi)$. Visual inspection of simulations suggests that for pulsatile inputs with a range of pulse shapes, phase locking is also monostable, though our proof does not address these other cases.

[[[Is it worth making some sim results?]]]

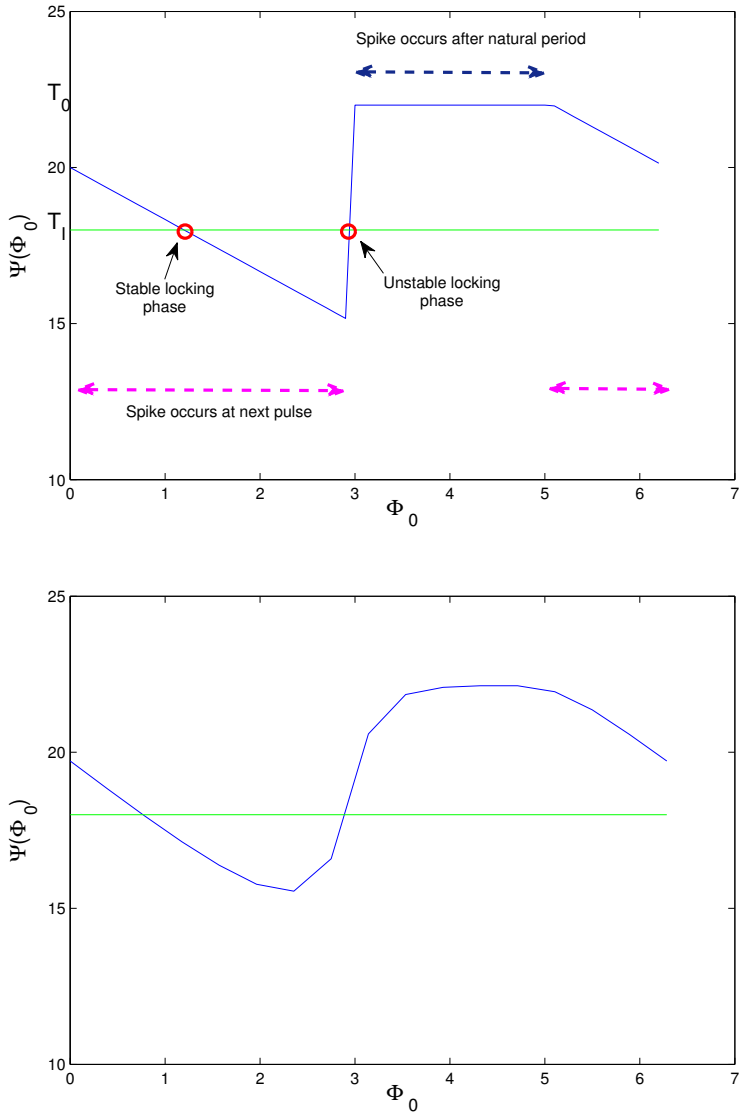

Figure 1: **Top:** The interspike interval function Ψ as a function of initial forcing phase Φ_0 at a spike., for the singular limit case $\tau \to 0$. At the phases where $\Psi(\Phi_0)$ crosses T_I (the green line), phase locking is possible. **Bottom:** Ψ vs. Φ_0 for $0 < \tau \ll 1$. Note the similarity to the singular limit case. We prove that under square-pulse forcing, the direction of Ψ' may only change twice (as it does above), and that therefore Ψ may cross T_I only two times. Only one of the two crossings may be from top to bottom, and this is the requirement for stability, so only one stable phase-locking phase may exist.

We can gain some intuition for the shape of Ψ by studying the special case in which τ is sufficiently small that spikes occur immediately as soon as $G > 0$. During a pulse, G can only exceed zero if the previous spike took place sufficiently long ago. If the previous spike was too recent, the pulse arrives under slightly too much inhibition to produce a positive net current. When this occurs, the next spike time jumps ahead to the next time the net current is positive, which may be at the next pulse or once the natural period of oscillation has elapsed since the last spike.

Thus, for small τ, $\Psi(\Phi_0)$ a characteristic shape, illustrated in Figure 9, top. When Φ_0 is in the range marked by purple arrows, the next spike is evoked by the next pulse. In this range, the next spike time does not change with Φ_0, so the interspike interval shrinks steadily as Φ_0 increases and $\frac{d}{d\Phi_0}\Psi = -1$. At some initial forcing phase, the next pulse occurs slightly too early to evoke a spike, and here Ψ jumps up sharply. In the range marked by blue arrows, the next spike occurs after the natural oscillator period T_0, before the following pulse. In this range, the ISI is independent of the initial forcing phase Φ_0, so $\frac{d}{d\Phi_0}\Psi = 0$. Beyond this range, the following pulse occurs before T_0 has passed since the last spike, so this pulse evokes the next spike and $\frac{d}{d\Phi_0}\Psi = -1$ again.

For the discussion above, we have assumed that $c = 0$ in (2.2) such that after every spike, $s = 1$ and the system state may be fully characterized by a value of Φ. However, the requirement of a one-dimensional state at each spike may be eased using the natural generalization of the firing map, the return map $\mathbf{R}(\theta, s, \Phi)$ from the full system state at one spike to the system state at the next. If the system state at spikes approaches an attracting invariant graph over forcing phase Φ, then the asymptotically stable dynamics of the system can be completely described by a firing map defined only on the trajectories from this attracting set of points. In [[[ATTRACTING INVARIANT TORI IN PERIODICALLY FORCED NETWORK GAMMA RHYTHMS]]], we show that if c is sufficiently small in (2.2), the NG oscillator does indeed posses such an attracting invariant graph. Therefore, in order to study the set of stably phase-locked orbits for small c, we need only study the dynamics of this firing map and the corresponding ISI function: all other behavior is necessarily transient. In the next section, we use this to generalize our proof of monostable phase locking to sufficiently small $c > 0$.

4. Proof of monostability

Theorem 4.0.1. *When the ING oscillator described in* (4.3) *with* $c = 0$ *or sufficiently small* $c > 0$ *is forced by periodic square pulses, stable* $1:1$ *phase locked spiking can occur at only one forcing phase.*

4.0.1. *Proof Outline.* [[[SWITCH TO V? Remove references to previous appendix. Rethink "transversely"]]]

(1) Working from the ODE with $c = 0$ and its linearization describing the evolution of small variations, we show that $\Psi'(\Phi_0)$ can be written as an integral over the θ trajectory between spikes.

(2) Using the standard change of variables from the theta neuron to the QIF neuron, we write $\Psi'(\Phi_0)$ as an integral over the V trajectory of the corresponding QIF neuron.

(3) We let $I(\cdot)$ be a square pulse of duration σ and positive height I_{step}. A simple argument shows that $\Psi(\Phi_0)$ can only cross T_I on one of two subintervals of $[0, T_I)$: on one, spikes occur between pulses, and on the other, spikes occur during the pulse.

(4) On the first subinterval, we use the integral expression for $\Psi(\Phi_0)$ to show that its sign is opposite the sign of $\int_{u=0}^{\sigma} 2V_{u+t_p}(\Phi_0)du$, where t_p is the time between a spike and the arrival of a pulse; therefore, the sign of $\Psi'(\Phi_0)$ is the same as the sign of $\int_{u=0}^{\sigma} 2V_{u+t_p}(\Phi_0)du$.

(5) We present Lemma 4.0.1, which states that V_t strictly increases for as long after a spike as the input current is flat. Using this lemma, we show that $\int_{u=0}^{\sigma} 2V_{u+t_p}(\Phi_0)du$ increases with t_p, which decreases as Φ_0 increases; therefore $\Psi'(\Phi_0)$ can change signs only once (negative to positive) on the first subinterval of $[0, T_I)$.

(6) An argument paralleling that of the preceding two steps shows that on the second subinterval, $\Psi'(\Phi_0)$ may change signs only once (positive to negative).

(7) A simple argument shows that these conditions on $\Psi'(\Phi_0)$ allow Δ to cross T_I only twice transversely or once tangentially. Only a transverse downward crossing is asymptotically stable, so only one stable 1:1 phase locked trajectory may exist.

(8) The above analysis holds only for $c = 0$. However, for sufficiently small $c > 0$, an attracting invariant torus exists (see previous appendix) that must contain all stable orbits and on which an interspike interval map $\Psi(\cdot)$ is uniquely defined. This map is a differentiable perturbation of the same map for $c = 0$, so for sufficiently small $c > 0$, $\Psi(\cdot)$ may only cross T_I twice and only one stable 1:1 phase locked trajectory may exist.

(9) We prove Lemma 4.0.1.

Remark 4.0.1. *This proof relies heavily on the strictly-decreasing inhibition: without it, the lemma stating that V_t strictly increases while the input current is flat might not hold.*

4.0.2. *Step 1: $\Psi'(\Phi_0)$ expressed as the solution to a variational equation.* Let $t = 0$ denote an initial spike time, and let t_s denote the next spike time. For $c = 0$, the map $\bar{\mathbf{P}}(\bar{\Phi}) = \bar{\Phi} + \Psi(\Phi)$ on the real line (from the forcing phase at $t = 0$ to the forcing phase at t_s) is T_I-periodic. Its derivative is the ratio between a variation $\delta\Phi_0$ in forcing phase at the first spike and the resulting variation $\delta\Phi_{t_s}$ in forcing phase at the second:

$$\mathbf{P}'(\bar{\Phi}) = \frac{\delta\Phi_{t_s}}{\delta\Phi_0}. \tag{4.1}$$

We can write a similar expression for $\Psi'(\Phi)$:

$$\begin{aligned}\Psi(\Phi) =& \mathbf{P}(\bar{\Phi}) - \bar{\Phi} \\ \Psi'(\Phi) =& \mathbf{P}'(\bar{\Phi}) - 1 = \frac{\delta\Phi_{t_s}}{\delta\Phi_0} - 1\end{aligned} \tag{4.2}$$

The evolution of a small variation in state $(\delta V_t, \delta s_t, \delta\Phi_t)$ between two spikes is described by the variational equation produced by linearizing (4.3) about a trajectory (V_t, s_t, Φ_t):

$$\begin{cases}\delta\dot{V}_t = & \frac{1}{\tau}(2V_t\delta V_t - g\delta s_t + I'(\Phi_t)\delta\Phi_t) \\ \delta\dot{s}_t = & -\frac{\delta s_t}{\tau_s} \\ \delta\dot{\Phi}_t = & 0\end{cases} \tag{4.3}$$

We can immediately solve the uncoupled equations for $\delta\dot{s}_t$ and $\delta\dot{\Phi}_t$ in terms of the initial variation in state $(\delta V_0, \delta s_0, \delta\Phi_0)$:

$$\begin{cases}\delta\Phi_t = & \delta\Phi_0 \\ \delta s_t = & \delta s_0 e^{-\frac{t}{\tau_s}}\end{cases} \tag{4.4}$$

The third equation can be solved with the help of an integrating factor:

$$\delta\dot{V}_t - \frac{2}{\tau}V_t\delta V_t = \frac{1}{\tau}(-g\delta s_t + I'(\Phi_t)\delta\Phi_t) \tag{4.5}$$

$$\delta\dot{V}_t e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr} - \frac{2}{\tau}V_t\delta V_t e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr} = \frac{1}{\tau}(-g\delta s_t + I'(\Phi_t)\delta\Phi_t)e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr} \tag{4.6}$$

$$\frac{d}{dt}\left(\delta V_t e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr}\right) = \frac{1}{\tau}(-g\delta s_t + I'(\Phi_t)\delta\Phi_t)e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr} \tag{4.7}$$

Integrating from $t = 0$ to t_s^-, where t_s^- is the left-hand limit of spike time t_s,

$$\delta V_{t_s^-} e^{-\frac{2}{\tau}\int_0^{t_s^-} V_r\delta V_r dr} - \delta V_0 = \frac{1}{\tau}\int_0^{t_s^-}(-g\delta s_t + I'(\Phi_t)\delta\Phi_t)e^{-\frac{2}{\tau}\int_0^t V_r\delta V_r dr}dt \tag{4.8}$$

$$\delta V_{t_s^-} = \delta V_0 e^{\frac{2}{\tau}\int_0^{t_s^-} V_r\delta V_r dr} + \frac{1}{\tau}\int_0^{t_s^-}(-g\delta s_t + I'(\Phi_t)\delta\Phi_t)e^{\frac{2}{\tau}\int_t^{t_s^-} V_r\delta V_r dr}dt \tag{4.9}$$

Substituting from (4.4),

$$\delta V_{t_s^-} = \delta V_0 e^{\frac{2}{\tau}\int_0^{t_s^-} V_r \delta V_r dr} + \frac{1}{\tau}\int_0^{t_s^-} (-g\delta s_0 e^{-\frac{t}{\tau_s}} + I'(\Phi_t)\delta\Phi_0) e^{\frac{2}{\tau}\int_t^{t_s^-} V_r \delta V_r dr} dt \tag{4.10}$$

In the previous appendix, we used a saltation matrix to project a variation $\hat{\zeta}_0$ at initial spike time 0 onto the plane $\Phi = \Phi_0^*$; we followed it to the next spike at time t_s using the linearized ODE; we projected it onto the plane $\theta = \pi$; and we followed it through a synaptic resetting event. In this appendix, we will skip the first step and use the linearized ODE to track a variation that is initialized at the plane $\theta = -\pi$. Following the form of equation (**??**) from the previous appendix but leaving off the initial projection by the saltation matrix $\hat{\mathbf{M}}_0$, we write

$$\hat{\zeta}_{t_s} = D\rho \mathbf{M}_{t_{s-}} \mathbf{B} \hat{\zeta}_0$$

where $\mathbf{B}$ is defined in (**??**), $\mathbf{M}_{t_{s-}}$ is defined in (**??**), and $D\rho$ is defined in (**??**).

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2}\frac{s_{t_{s-}}}{\mu} & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \mathbf{B}\hat{\zeta}_0$$

For $c = 0$:

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \mathbf{B}\hat{\zeta}_0$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} \kappa & -2g\Sigma & \Omega \\ 0 & e^{-\frac{t_s}{\tau_s}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{\zeta}_0$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2}\kappa & 0 & -\frac{1}{2}\Omega + 1 \end{pmatrix} \hat{\zeta}_0$$

$\hat{\zeta}_0$ is a variation at the plane $\theta = -\pi$, so $\Delta\hat{\theta}_0 = 0$:

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{2}\kappa & 0 & -\frac{1}{2}\Omega + 1 \end{pmatrix} \begin{pmatrix} 0 \\ \Delta\hat{s}_0 \\ \Delta\hat{\Phi}_0 \end{pmatrix}$$

$$\begin{pmatrix} \Delta\hat{\theta}_{t_s} \\ \Delta\hat{s}_{t_s} \\ \Delta\hat{\Phi}_{t_s} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ (-\frac{1}{2}\Omega + 1)\Delta\hat{\Phi}_0 \end{pmatrix}$$

$$\Delta\hat{\theta}_{t_s} = (-\frac{1}{2}\Omega + 1)\Delta\hat{\Phi}_0 \tag{4.11}$$

Substituting into (**??**),

$$\Psi'(\Phi_0) = \frac{(-\frac{1}{2}\Omega + 1)\Delta\hat{\Phi}_0}{\Delta\hat{\Phi}_0} - 1 = -\frac{1}{2}\Omega + 1 - 1 = -\frac{1}{2}\Omega \tag{4.12}$$

or, substituting from (**??**) and (**??**) and using $\Phi_t^* = \Phi_0^* + t$,

$$\Psi'(\Phi_0) = -\frac{1}{2\tau}\int_0^{t_s} (1+\cos(\theta_t^*))I'(\Phi_0+t)e^{\frac{1}{\tau}\int_t^{t_s}(1-G_r^*)\sin(\theta_r^*)dr}dt. \tag{4.13}$$

$$\begin{aligned}
\Psi'(\Phi_0) &= -\frac{1}{2}\int_0^{t_s} (1+\cos(\theta_t^*))I'(\Phi_0+t)e^{\int_t^{t_s}(1-G_r^*)\sin(\theta_r^*)dr}dt \\
&= -\frac{1}{2}\int_0^{t_s} e^{\ln(1+\cos(\theta_t^*))}I'(\Phi_0+t)e^{\int_t^{t_s}(1-G_r^*)\sin(\theta_r^*)dr}dt \\
&= -\frac{1}{2}\int_0^{t_s} e^{\int_{\theta_t^*}^{\pi}\frac{\sin(\theta)}{1+\cos(\theta)}d\theta}I'(\Phi_0+t)e^{\int_t^{t_s}(1-G_r^*)\sin(\theta_r^*)dr}dt
\end{aligned}$$

We change coordinates in the integral to integrate along the path θ_t^*, using $d\theta = \dot{\theta}_t^* dt$ and $\theta_{t_s}^* = \pi$:

$$\begin{aligned}
&= -\frac{1}{2}\int_0^{t_s} e^{\int_t^{t_s}\frac{\sin(\theta_r^*)}{1+\cos(\theta_r^*)}\dot{\theta}_r^* dr}I'(\Phi_0+t)e^{\int_t^{t_s}(1-G_r^*)\sin(\theta_r^*)dr}dt \\
&= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s}\frac{\sin(\theta_r^*)}{1+\cos(\theta_r^*)}\dot{\theta}_r^* + (1-G_r^*)\sin(\theta_r^*)dr}dt
\end{aligned}$$

Substituting for $\dot{\theta}_t^*$ from (**??**),

$$\begin{aligned}
&= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s}\frac{\sin(\theta_r^*)}{1+\cos(\theta_r^*)}\left[1-\cos(\dot{\theta}_r^*)+(1+\cos(\dot{\theta}_r^*))G_r\right]+(1-G_r^*)\sin(\theta_r^*)dr}dt \\
&= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s}\frac{\sin(\theta_r^*)(1-\cos(\dot{\theta}_r^*))}{1+\cos(\theta_r^*)}+\sin(\theta_r^*)dr}dt \\
&= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s}\frac{2\sin(\theta_r^*)}{1+\cos(\theta_r^*)}dr}dt
\end{aligned}$$

Applying the trigonometric identity $\tan\left(\frac{\theta}{2}\right) = \frac{\sin(\theta)}{1+\cos(\theta)}$,

$$= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s} 2\tan\left(\frac{\theta_r^*}{2}\right)dr}dt$$

And changing to the V variable of the QIF neuron using (5.1),

$$= -\frac{1}{2}\int_0^{t_s} I'(\Phi_0+t)e^{\int_t^{t_s} 2V_r^* dr}dt$$

V_r^* depends on the initial forcing phase Φ_0. We reintroduce this dependency into our notation:

$$\Psi'(\Phi_0) = -\frac{1}{2}\int_0^{t_s} I'(\Phi_0 + t) e^{\int_t^{t_s} 2V_r^*(\Phi_0)dr} dt \tag{4.14}$$

4.0.3. *Step 3: Phase locking to a square pulse.* Let us consider a T_I-periodic input consisting of a periodic square pulse of height I_{step} and duration $\sigma < T_I$. The input as a function of forcing phase Φ can be written as

$$I(\Phi) = \begin{cases} I_{step} & \text{when } \Phi \in [0, \sigma) \\ 0 & \text{when } \Phi \in [\sigma, T_I) \end{cases} \tag{4.15}$$

and we can write its distributional derivative as

$$I'(\Phi) = I_{step}\delta(\Phi) - I_{step}\delta(\Phi - \sigma) \tag{4.16}$$

where δ is the Dirac delta function on the circle $\mathbb{T}^1 = [0, T_I)$.

We showed in section 5.1.1 that the map $\mathbf{P}(\bar{\Phi})$ on the real line from the forcing phase at $t = 0$ to the forcing phase at t_s is a T_I-periodic orientation-preserving homeomorphism on $\mathbb{R}$. The image of $[0, \sigma)$ under $\mathbf{P}$ is therefore an interval on $\mathbb{R}$. 1:1 phase locking occurs at any phase Φ where $\mathbf{P}(\bar{\Phi}) = \bar{\Phi} + T_I$. By the T_I-periodicity of $\mathbf{P}$, a 1:1 locking phase $\bar{\Phi} \in \mathbb{R}$ exists if and only if there exists a 1:1 locking phase $\bar{\Phi} \in [0, T_I)$. If this point is on $[\sigma, T_I)$, then it must fall in the subinterval $A = [\sigma, T_I) \cap [\mathbf{P}(\sigma) - T_I, \mathbf{P}(T_I) - T_I)$; if it is in $[0, \sigma)$, then it must fall in the subinterval $A' = [0, \sigma) \cap [\mathbf{P}(0) - T_I, \mathbf{P}(\sigma) - T_I)$. In the first case, phase locked spiking occurs between pulses; in the second, phase locked spiking occurs during the pulse. (See Figure 4.0.3.)

In the following two steps, we shall assume the first case ($\Phi_0 \in A$), such that the upward step of current arrives before the downward step. The upward step arrives at time t_p, where

$$t_p = -\Phi_0 \bmod T_I \tag{4.17}$$

and the downward step arrives at time $t_p + \sigma$. The same argument will apply to the second case ($\Phi_0 \in A'$), in which the downward step arrives before the upward step; we simply need to replace I_{step} with $-I_{step}$ and $t_p + \sigma$, the time of the second step, with $t_p + (T_I - \sigma)$.

4.0.4. *Step 4: $sgn(\Psi'(\Phi_0)$ for a square pulse.* We substitute the derivative (4.16) into (4.14):

$$\Psi'(\Phi_0) = -\frac{1}{2\tau}\int_0^{t_s} (1 + \cos(\theta_t^*)) \left(I_{step}\delta(\Phi_0 + t) - I_{step}\delta(\Phi_0 + t - \sigma)\right) e^{\frac{1}{\tau}\int_t^{t_s}(1 - G_r^*)\sin(\theta_r^*)dr} dt$$

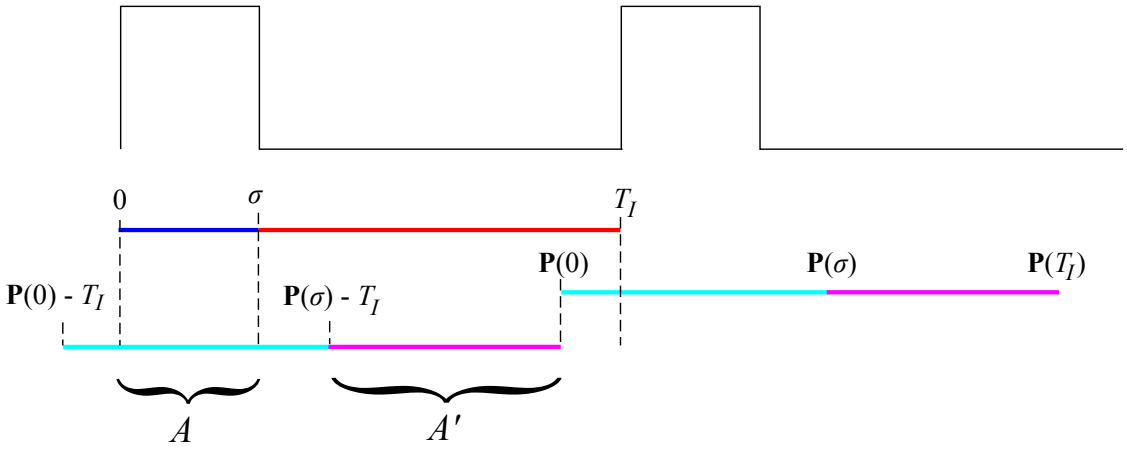

Figure 2: Illustration of the intervals in which 1:1 phase locking can occur. If phase locked spikes occur at phase $\bar{\Phi}$, then $\mathbf{P}(\bar{\Phi}) - T_I = \bar{\Phi}$. Therefore, if $\bar{\Phi} \in [\sigma, T_I)$, i.e., spikes occur between pulses, then $\bar{\Phi} \in A$, where $A = [\sigma, T_I) \cap [\mathbf{P}(\sigma) - T_I, \mathbf{P}(T_I) - T_I]$. Similarly, if $\bar{\Phi} \in [0, \sigma)$, i.e., spikes occur during pulses, then $\bar{\Phi} \in A'$, where $A' = [0, \sigma) \cap [\mathbf{P}(0) - T_I, \mathbf{P}(\sigma) - T_I]$. Phase locking cannot occur for $\bar{\Phi}$ outside these intervals.

From (4.17), we see that these delta functions will pick out $t = t_p$ and $t = t_p + \sigma$, respectively, so

$$\Psi'(\Phi_0) = -\frac{I_{step}}{2\tau}\left((1+\cos(\theta^*_{t_p}))e^{\frac{1}{\tau}\int_{t_p}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} - (1+\cos(\theta^*_{t_p+\sigma}))e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}\right) \tag{4.18}$$

Factoring out $(1+\cos(\theta^*_{t_p+\sigma}))e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}$ from both exponentials:

$$= -\frac{I_{step}}{2\tau}e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}\left(\frac{1+\cos(\theta^*_{t_p})}{1+\cos(\theta^*_{t_p+\sigma})}e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}(1-G^*_r)\sin(\theta^*_r)dr} - 1\right) \tag{4.19}$$

$$= -\frac{I_{step}}{2\tau}e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}\left(e^{\ln(1+\cos(\theta^*_{t_p}))-\ln(1+\cos(\theta^*_{t_p+\sigma}))}e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}(1-G^*_r)\sin(\theta^*_r)dr} - 1\right) \tag{4.20}$$

$$= -\frac{I_{step}}{2\tau}e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}\left(e^{\int_{\theta^*_{t_p}}^{\theta^*_{t_p}+\sigma}\frac{\sin(\theta)}{1+\cos(\theta)}d\theta}e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}(1-G^*_r)\sin(\theta^*_r)dr} - 1\right) \tag{4.21}$$

We take our integral from $\theta^*_{t_p}$ to $\theta^*_{t_p+\sigma}$ along the path θ^*_r, and change variables from θ to the time variable r using $d\theta = \dot{\theta}^*_r dr$.

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} \frac{\sin(\theta^*_r)}{1+\cos(\theta^*_r)}\dot{\theta}^*_r dr} e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}(1-G^*_r)\sin(\theta^*_r)dr} - 1 \right) \tag{4.22}$$

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} \frac{\sin(\theta^*_r)}{1+\cos(\theta^*_r)}(1-\cos(\theta^*_r)+(1+\cos(\theta^*_r))G^*_r dr} e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}(1-G^*_r)\sin(\theta^*_r)dr} - 1 \right) \tag{4.23}$$

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} \sin(\theta^*_r)G^*_r + \frac{\sin(\theta^*_r)(1-\cos(\theta^*_r))}{1+\cos(\theta^*_r)} dr} e^{\frac{1}{\tau}\int_{t_p}^{t_p+\sigma}\sin(\theta^*_r)-\sin(\theta^*_r)G^*_r dr} - 1 \right) \tag{4.24}$$

Combining the integrals,

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} \frac{\sin(\theta^*_r)(1-\cos(\theta^*_r))}{1+\cos(\theta^*_r)} + \sin(\theta^*_r)dr} - 1 \right) \tag{4.25}$$

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} \frac{2\sin(\theta^*_r)}{1+\cos(\theta^*_r)} dr} - 1 \right) \tag{4.26}$$

Applying the trigonometric identity $\tan\left(\frac{\theta}{2}\right) = \frac{\sin(\theta)}{1+\cos(\theta)}$,

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} 2\tan\left(\frac{\theta^*_r}{2}\right) dr} - 1 \right) \tag{4.27}$$

And changing to the V variable of the QIF neuron using (5.1),

$$= -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} 2V^*_r dr} - 1 \right) \tag{4.28}$$

V^*_r depends on the initial forcing phase Φ_0. We reintroduce this dependency into our notation:

$$\Psi'(\Phi_0) = -\frac{I_{step}}{2\tau} e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr} \left(e^{\int_{t_p}^{t_p+\sigma} 2V^*_r(\Phi_0)dr} - 1 \right) \tag{4.29}$$

In order to show that $\Psi'(\Phi_0)$ changes sign only once on interval A, it is sufficient to note the $\pm$ sign of $\Psi'(\Phi_0)$:

$$sgn\left(\Psi'(\Phi_0)\right) = sgn\left(-\frac{I_{step}}{2\tau}\right) sgn\left(e^{\frac{1}{\tau}\int_{t_p+\sigma}^{t_s}(1-G^*_r)\sin(\theta^*_r)dr}\right) sgn\left(e^{\int_{t_p}^{t_p+\sigma} 2V^*_r(\Phi_0)dr} - 1\right) \tag{4.30}$$

$$= -\, sgn\left(I_{step}\right) sgn\left(e^{\int_{t_p}^{t_p+\sigma} 2V^*_r(\Phi_0)dr} - 1\right) \tag{4.31}$$

$$= -\, sgn\left(I_{step}\right) sgn\left(\int_{t_p}^{t_p+\sigma} 2V^*_r(\Phi_0)dr\right) \tag{4.32}$$

To eliminate the dependence of the bounds of integration on t_p, we change variables in the integral to $q = r - t_p$:

$$= - \, sgn\,(I_{step})\, sgn\left(\int_0^{\sigma} 2V^*_{q+t_p}(\Phi_0)dq\right) \tag{4.33}$$

$$\tag{4.34}$$

4.0.5. *Step 5: $\Psi'(\Phi_0)$ may change signs only once.* Before the arrival of the pulse, $V_t^*(\Phi_0)$ is independent of Φ_0 because the cell receives the same flat current before the first step no matter the initial forcing phase, so we write V_t^*. We prove in Lemma 4.0.1 below that during this time, $\frac{d}{dt}V_t^* > 0$. Therefore the value of V when the pulse arrives increases with t_p, the time between the initial spike and the arrival of the pulse.

For $q \in [0, \sigma)$, $V^*_{q+t_p}$ is an autonomous 1-D dynamical system with respect to q:

$$\frac{\partial}{\partial q}V^*_{q+t_p} = {V^*_{q+t_p}}^2 - ge^{-\frac{q+t_p}{\tau_s}} + I_{step} \tag{4.35}$$

with initial condition $V^*_{t_p}$ at $q = 0$. Trajectories cannot cross, so as the initial condition increases with t_p, so does $V^*_{q+t_p}$ for any $q \in [0, \sigma)$, at which point the pulse ends.

Since $\frac{\partial}{\partial t_p}V^*_{q+t_p} > 0$, we also have

$$\frac{\partial}{\partial t_p}\int_0^{\sigma} 2V^*_{q+t_p}dr = \int_0^{\sigma} 2\frac{\partial}{\partial t_p}V^*_{q+t_p}dq > 0. \tag{4.36}$$

From (4.17), we see that t_p decreases with Φ_0, so $\int_{q=0}^{\sigma} 2V^*_{q+t_p}dq$ decreases with Φ_0. This integral can change signs (from positive to negative) only once as Φ_0 increases, and must cross zero transversely. So by (4.33), $\Psi'(\Phi_0)$ may change signs once (from negative to positive) with the increase of Φ_0, and must also cross zero transversely.

4.0.6. *Step 6: Steps 3 and 4 for $\Phi_0 \in [0, \sigma)$.* The preceding two steps assumed that Φ_0 was on the arc $A = [\sigma, T_I) \cap [\mathbf{P}(\sigma)T_I, \mathbf{P}(T_I)T_I)$. As previously explained, Δ may also cross T_I on the arc $A' = [0, \sigma) \cap [\mathbf{P}(0) - T_I, \mathbf{P}(\sigma) - T_I)$. In this case, we can repeat the same derivation substituting $-I_{step}$ for I_{step} and $T_I - \sigma$ for σ in (4.32), and conclude that

$$sgn(\Psi'(\Phi_0)) = -sgn(-I_{step})sgn\left(\int_0^{T_I-\sigma} 2V^*_{q+t_p}(\Phi_0)dq\right) = sgn\left(\int_0^{T_I-\sigma} 2V^*_{q+t_p}(\Phi_0)dq\right) \tag{4.37}$$

where $\int_0^{T_I-\sigma} 2V_{q+t_p}(\Phi_0)dq$ decreases with Φ_0. Therefore, on A', $\Psi'(\Phi_0)$ can change signs only once (from positive to negative) as Φ_0 increases, and must cross zero transversely.

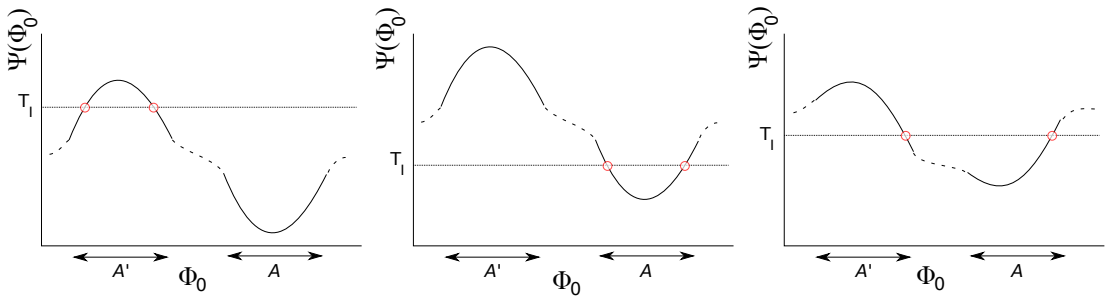

Figure 3: If $\Psi(\Phi_0)$ crosses T_I, it may do so only on A (on which it must be concave) and A' (on which it must be convex), and only in three ways. **Left:** If two crossings occur on A', none can occur on A. **Center:** If two crossings occur on A, none can occur on A'. **Right:** If one crossing occurs on A, only one can occur on A'.

4.0.7. *Step 7: Ψ may only cross T_I twice.* On A, Ψ may switch from decreasing to increasing, and may therefore cross T_I once downwards and then once upwards. In this case, $\Psi(\Phi) > T_I$ at both ends of A. Any other crossings must occur on A'; therefore, $\Psi(\Phi) > T_I$ at both ends of A'. For additional crossings to occur on A', Ψ would have to decrease and then increase on this interval; but we have shown that the sign of the derivative can switch signs from positive to negative on this interval, so no additional crossings are possible.

If only one crossing occurs on A, then $\Psi(\Phi) > T_I$ on one side of A' and $\Psi(\Phi) < T_I$ on the other. Therefore an odd number of crossings must occur on A'; but for three crossings to occur on A', the sign of the derivative would have to switch twice, which is not permitted. Therefore, in this case only one crossing may occur on A'.

A parallel argument to the first shows that if there are no crossings on A, a maximum of two may occur on A'.

In all of these cases, when two crossings occur one is from above to below and the other from below to above. An asymptotically stable 1:1 phase lock is only possible when the map $\mathbf{P} - T_I$ has a fixed point Φ at which $-1 \leq \mathbf{P}'(\Phi) \leq 1$. We recall that $\mathbf{P}(\Phi) = \Psi(\Phi) + \Phi$, so at such a fixed point we must have $-1 \leq \Psi'(\Phi) + 1 \leq 1$, or $-2 \leq \Psi'(\Phi) \leq 0$. In other words, only when Ψ crosses T_I downwards can the associated fixed point be asymptotically stable. Therefore, only one asymptotically stable 1:1 phase locked trajectory may exist.

See Figure 3 for illustration.

4.0.8. *Step 8: for $c > 0$.* It is not immediately clear how to define Ψ for $c > 0$: at any initial forcing phase Φ, the system may have a range of initial values of s, so the subsequent interspike interval is not uniquely determined. However, if the return map $\mathbf{R}$ from (s, Φ) just after one spike to (s, Φ) at the next possesses an asymptotically stable invariant circle (see previous appendix) *and* that circle is a graph of s over Φ, then after a sufficient transient,

the initial value of s after a spike at any forcing phase Φ *is* uniquely determined by the graph, and $\Psi(\Phi)$, the interspike interval following this initial condition, is uniquely defined. For $c = 0$, the graph $s = 1$ serves this purpose: immediately after any spike, $s = 1$, so this circle of initial conditions forms an attracting invariant circle for the return map.

The only part of the evolution of the system that depends on c is the resetting map, which depends smoothly on c; therefore, by Fenichel's theorem for maps [**Fenichel**], for a sufficiently small c the invariant circle persists and depends smoothly on c. Therefore, for sufficiently small c, the invariant circle is still a graph over Φ, and the interspike intervals following the initial condition parametrized by Φ depend smoothly on c. Thus, $\Psi(\Phi)$ depends smoothly on c. For $c = 0$, Δ only crosses T_I transversely, so for a sufficiently small smooth perturbation no additional crossings can be created. Therefore, for sufficiently small $c > 0$, Δ can cross T_I only twice, and only one crossing can correspond to an asymptotically stable 1:1 phase lock; in other words, 1:1 phase locking is monostable.

4.0.9. *Step 9: Proof of Lemma 4.0.1.*

Lemma 4.0.1. *If an ING oscillator receives a flat current I_0 after a spike, its voltage strictly increases.*

Proof. In the V vs. t plane, the V-nullcline is the set $V = \sqrt{ge^{-\frac{t}{\tau_s}} - b - I_0}$. This nullcline vanishes at a saddle-node bifurcation when $b - ge^{-\frac{t}{\tau_s}} + I_0 < 0$, which occurs when $t > \tau_s \ln\left(\frac{b+I_0}{g}\right)$. After a spike at $t = 0$, the voltage begins below the V-nullcline. The branch of the nullcline accessible from below slants up and right, and on the nullcline all vectors point horizontally to the right, so the nullcline cannot be crossed from below. Therefore V remains below this nullcline, and $\dot{V} > 0$ until $t > \tau_s \ln\left(\frac{b+I_0}{g}\right)$. And for all $t > \tau_s \ln\left(\frac{b+I_0}{g}\right)$, we also have $\dot{V} = V^2 + b - ge^{-\frac{t}{\tau_s}} + I_0 > V^2 + 0 > 0$. See Figure 4 for illustration. □

Invariant tori are particularly useful in spiking systems. In a neural model with a Poincare section corresponding to a "spike," we can define a Poincare map from the state at one spike to the state at the next. If the full state space is only two-dimensional, this is a one-dimensional map, so the value of a single variable at a spike determines the system state at the next spike. In the following sections and in a number of other publications, this variable is time or the phase of an ongoing forcing oscillation, and the map is called a "spike map" [**Brette2003**] (though other publications refer to it as a "firing map" [**Carrillo2001**], "firing time map" [**Goel2002**] or "spike-time map" [**Tiesinga2002a**]). In higher dimensional systems, no single variable at a spike can determine the state at the next. But if the Poincare map possesses an attracting invariant circle (i.e. the full system possesses an attracting invariant torus), a one-dimensional map may be defined on the invariant circle; after the system becomes close to the torus, this map begins to act as a spike map and can be used to study the asymptotic dynamics of the system.

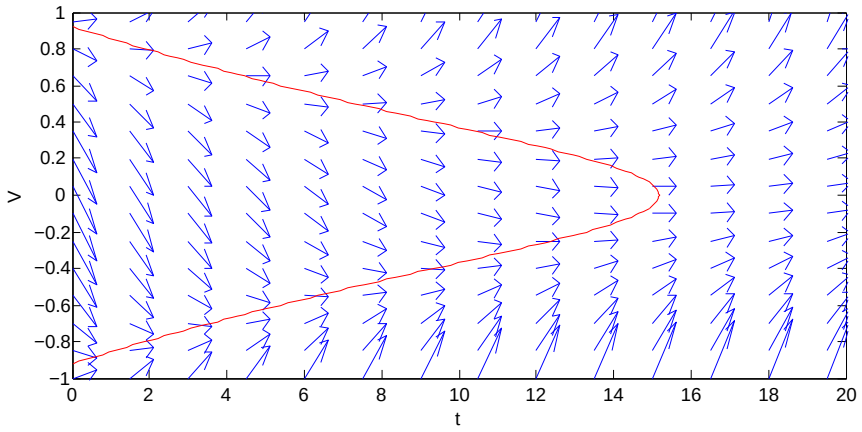

Figure 4: System is receiving flat current I_0 after a spike at $t = 0$. In the V vs. t plane, the V nullcline (red) cannot be crossed from underneath. All arrows to the right of the nullcline point upwards, so V is always increasing.

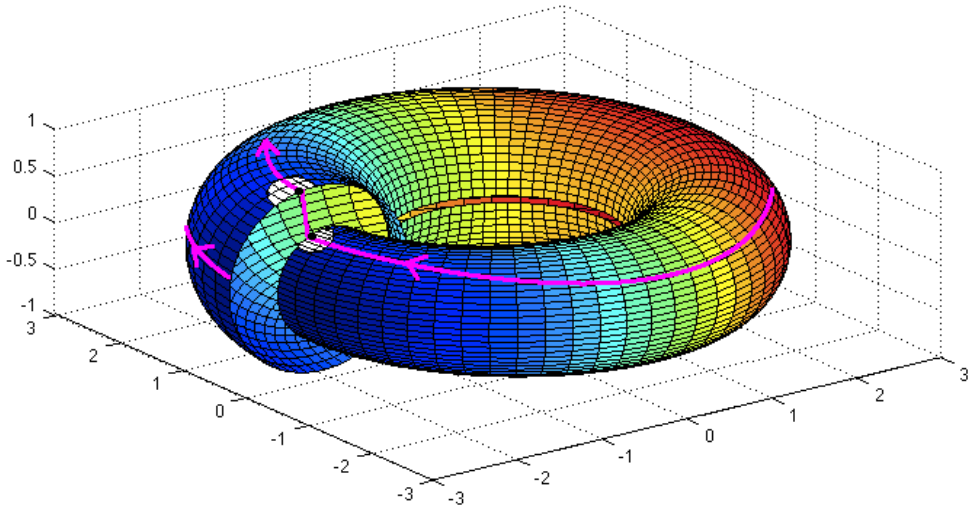

Figure 5: A flow on a torus with a discontinuous jump. Points are identified across the jump (e.g., the black points on the magenta trajectory), and in the induced topology the white set is open. The jump is a homeomorphism on the circle, so this topology glues the two circular faces together one-to-one and continuously to create an unbroken torus on which flows are continuous.

4.1. **The Relaxation Oscillator and Phase-Locking.** Here we study the phase-locking properties of a periodically-forced relaxation oscillator. As a generic exemplar of such an oscillator, we study the periodically-forced Fitzhugh-Nagumo (FN) oscillator, defined by the equations

$$\begin{cases} \tau \dot{v} = & v - v^3 - w + \epsilon I(\Phi) \\ \dot{w} = & v - c - rw \\ \dot{\Phi} = & 1 \end{cases} \tag{4.38}$$

In these equations, $\Phi \in \mathbb{T}^1 = [0, T_I)$, $I(\cdot)$ is a piecewise-continuous function from $\mathbb{T}^1$ to $\mathbb{R}^+$ and hence a periodic current with period T_I, and $0 < \tau \ll 1$. This model has been studied in [**Alexander1990**](in which only the w variable is forced), [**Guckenheimer2006**], [**Croisier2009**], [**Izhikevich2000**], and [**Coombes2000**], to name a few; our analysis is intended only to compare these results to the properties of ING oscillators.

4.1.1. *The Breakup of the Invariant Torus.* Though any system with a globally attracting limit cycle is globally attracted to an invariant torus for sufficiently weak periodic forcing, the strength of forcing required to break up the torus varies by system. One factor that can

reduce the requisite forcing strength necessary to break the torus is a separation of time scales. This effect is apparent in the case of the forced relaxation oscillator.

Croiser et. al [**Croisier2009**] conduct a numerical study of the FN oscillator under forcing. They find that interesting dynamics emerge at a forcing strength ϵ that decreases precipitously as $\tau \to 0$. These dynamics include the period-doubling of stable 1:1 phase-locked orbits, and bistability between 1:1 and 2:1 phase-locked solutions, both for forcing periods close to the system's natural period. As discussed above, neither of these behaviors is possible if there exists a globally attracting invariant torus; therefore, their observations prove that when τ is small, the attracting invariant torus breaks up at very small ϵ.

4.1.2. *Bistable Phase-Locking.* Even when the forced FN oscillator possesses an invariant torus, it still tends to support stable 1:1 phase-locking at two different phases. Croiser et. al [**Croisier2009**] demonstrate by numerical simulation that when a FN oscillator (with small, nonzero ϵ) is given pulsatile periodic forcing, the stable phase-locking tongue has bistable regions for forcing periods immediately to the right and left of the natural period.

Croisier et al. point out that the bistability is directly related to the fact that the phase-response curve (PRC) makes two downward excursions from the zero line (see Figure 6). Here, the phase response curve (PRC) is defined as a function from an oscillator's phase at the arrival of a temporally localized pulse to the resulting advance (or delay) in phase [**Winifree**] [**Ermentrout2002**] [**Canavier2012**]. It can be used to produce a "phase-resetting map" from the oscillator's phase at one stimulus to its phase at the next:

$$\phi_{i+1} = \phi^i + f(\phi^i) + T_I \bmod T_0 \tag{4.39}$$

where $f(\cdot)$ is the PRC, $\phi \in \mathbb{T}^1 = [0, T_0)$ (see [**Croisier2009**]), and T_I and T_0 are the forcing period and natural period as before. When the PRC crosses $T_0 - T_I$, the phase-resetting map crosses the diagonal and phase-locking is possible; when this crossing is from above to below, the phase-resetting map crosses from above the diagonal to below, so phase-locking may be stable (as long as the slope of the phase-resetting map at the crossing is greater than -1). The PRC for the FN oscillator is reproduced from [**Croisier2009**] in Figure 6. Note its two downward excursions. If the forcing period is just above the natural period, then $T_0 - T_I$ is just below zero, and these downward excursions may cause the PRC to cross $T_0 - T_I$ downwards twice with shallow slopes, giving rise to two stably phase-locked trajectories.

5. ING: Robust, Monostable Phase-Locking on an Invariant Torus

Here we investigate the phase-locking properties of the forced ING mechanism.

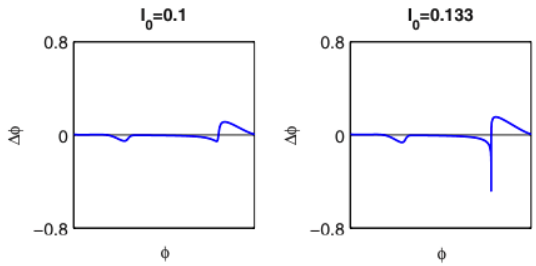

Figure 6: The phase response curve (PRC) of the FN oscillator with small τ in response to two different pulse strengths (both very weak), reproduced from Croisier et al. 2009 [**Croisier2009**].

5.1. **Existence of a Globally Attracting Invariant Torus for Non-Small ϵ.** *Here we discuss the broad conditions under which a periodically-forced ING system as described in (4.3) possesses a globally attracting invariant torus.*

In order to avoid the inconvenience of resetting membrane potentials at a spike, we switch from a QIF neuron with voltage V to the equivalent theta neuron with phase θ, using the change of variables presented in [**Ermentrout1986**]:

$$V = \tan\left(\frac{\theta}{2}\right) \tag{5.1}$$

where $\theta \in \mathbb{T}^1 = [0, 2\pi)$. We can replace the first equation in (4.3) with

$$\dot{\theta} = \frac{1}{\tau}\left[1 - \cos(\theta) + (1 + \cos(\theta))G\right]. \tag{5.2}$$

In θ coordinates, the invariant manifold (when it exists) is a torus with one jump discontinuity at $\theta = \pi$ corresponding to the instantaneous synaptic resetting event at each spike. If each trajectory is connected across the jump by identifying the point (π, s, Φ) at the left-hand limit of a spike time with $(\pi, \rho(s), \Phi)$ at the right-hand limit of the spike, then the invariant set is a topological torus in phase space, as discussed in Section **??** and illustrated in Figure 7. Importantly, though the vector field on this torus is not continuous across $\theta = \pi$, it points in the positive direction on either side of it, so trajectories have unique continuations across the boundary and it is accurate to describe the invariant set as a flow constrained to a torus [**Alexander1990**]. As discussed in Section **??**, when an attracting invariant torus exists, the long-term dynamics of the ING oscillator cannot include multiple phase-locked trajectories with different locking ratios, and cannot period-double as parameters change.

As with any limit-cycle oscillator, the invariant torus which exists for $\epsilon = 0$ (no forcing) persists for sufficiently small $\epsilon > 0$. However, due to its distinctive structure, the ING

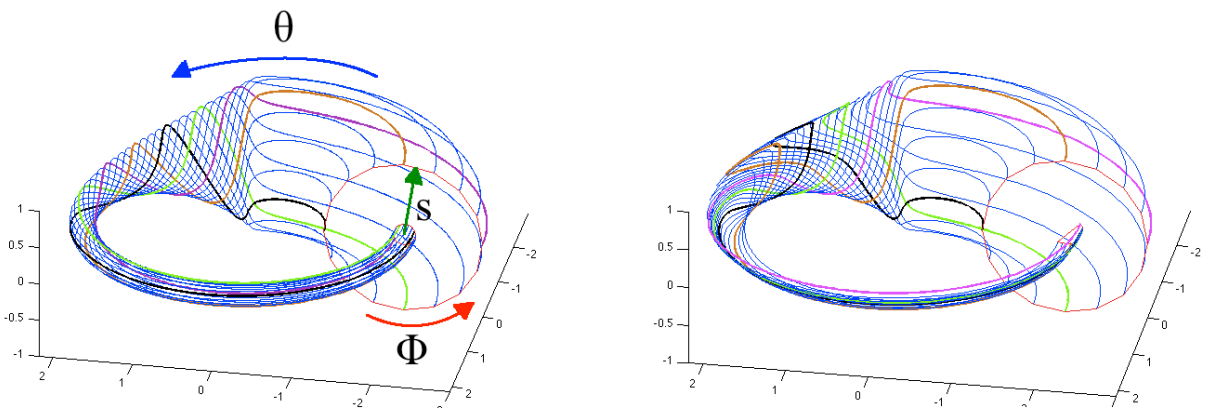

Figure 7: **Left:** Trajectories of ING go to an invariant torus with a jump discontinuity at $\theta = \pi$. θ is represented by the angle around a ring in the x/y plane; s is represented by the length of a vector extending perpendicularly out from this central ring; and Φ is represented by the angle of this vector. around the circular cross-section of the torus. The large red circle is the set $\theta = \pi$, $s = 1$. Sixteen trajectories are initialized from this set, each at a different forcing phase. Four are colored for visibility. **Right:** Under periodic forcing, some trajectories increase and then decrease in the θ direction before spiking; ultimately, they cluster together towards a single trajectory on the cyclinder, giving rise to stable phase-locking.

oscillator has a much broader regime throughout which its invariant torus provably persists. First we show that

As discussed in Section **??**, the existence of the attracting invariant torus is equivalent to the existence of an attracting invariant circle for the map from the system state at one spike to the next. We let $\mathbf{R}$ denote the map from the (s, Φ) state of the system at the right-hand limit of a spike time at $t = 0$ to its (s, Φ) state at the right-hand limit of the next spike time t_s:

$$\begin{aligned} \mathbf{R} : [0,1] \times \mathbb{T}^1 &\to [0,1] \times \mathbb{T}^1 \\ (s_0, \Phi_0) &\to (s_{t_s}, \Phi_{t_s}) \end{aligned}$$

In the following, we assume $\epsilon = 1$ because an assumption of weak forcing is not required. We show that an attracting invariant circle exists for $\mathbf{R}$ (and hence an attracting broken invariant torus exists for the dynamical system) when $c = 0$, when c is sufficiently small, and when g is sufficiently large.

5.1.1. *Invariant torus when $c = 0$.* If $c = 0$ (synapses saturate fully at each spike), we can immediately see that since $\rho(s) = 1$ for all s, trajectories reaching $\theta = \pi$ jump onto the circle $s = 1$ in (s, Φ) space. This circle is clearly invariant and attracting under $\mathbf{R}$, and the flow from this circle forms an invariant surface in the full phase space. All we need

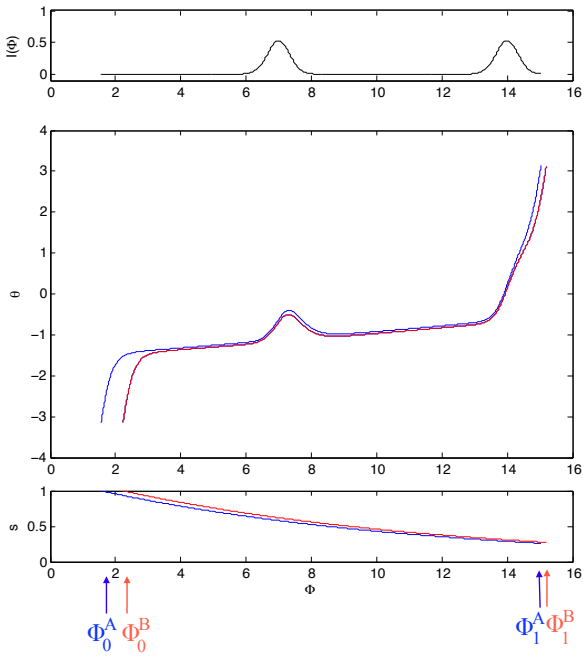

Figure 8: Two trajectories are initialized with $s = 0$, at phases $\bar{\Phi}_0^B > \bar{\Phi}_0^A$. The population phase θ of trajectory B (red) is below that of trajectory A (blue), and cannot cross it because it is under more inhibition. Therefore, trajectory A reaches a spike at an earlier forcing phase: $\bar{\Phi}_1^B > \bar{\Phi}_1^A$.

to show is that this surface is a torus when trajectories are identified across the jump; for this, it is sufficient to show that when the circle $s = 1$ at $\theta = -\pi$ is followed along the flow to the next spike and reset by ρ (i.e., subjected to the return map $\mathbf{R}$), the result is an orientation=preserving homeomorphism on the circle.

We let $\mathbf{P}_{\mathbb{T}^1} : \mathbb{T}^1 \to \mathbb{T}^1$ denote the restriction of $\mathbf{R}$ to the circle $s = 1$, which takes as an argument an initial forcing phase Φ_0 and returns the value of Φ at which a trajectory initialized at $(-\pi, 1, \Phi_0)$ reaches its next spike. If we can show that $\mathbf{P}_{\mathbb{T}^1}$ is an orientation-preserving homeomorphism on the circle, then after points are identified across the jump, the forward flow from $(-\pi, 1, \Phi)$ back to itself traces out a torus.

Theorem 5.1.1. $\mathbf{P}$ *is an orientation-preserving homeomorphism on the circle* $(-\pi, 1, \Phi)$.

The proof below is illustrated in Figure 8.

Proof. Let $\bar{\theta}$ and $\bar{\Phi}$ denote the lifts of θ and Φ to $\mathbb{R}$. Let $\mathbf{P} : \mathbb{R} \to \mathbb{R}$ denote the map $\mathbf{P}_{\mathbb{T}^1}$ for forcing phase lifted to $\mathbb{R}$. Since $\mathbf{P}$ is a lift of the map $\mathbf{P}_{\mathbb{T}^1}$ to $\mathbb{R}$, $\mathbf{P}_{\mathbb{T}^1}$ is an

orientation-preserving homeomorphism on $\mathbb{T}^1$ if and only if $\mathbf{P}$ is an orientation-preserving homeomorphism on $\mathbb{R}$.

Consider two trajectories, A and B, which spike at initial time $t = 0$ at forcing phases $\bar{\Phi}_0^A$ and $\bar{\Phi}_0^B$, with $\bar{\Phi}_0^B > \bar{\Phi}_0^A$. We let the state at the right-hand limit of the A spike, $\begin{pmatrix} -\pi \\ 1 \\ \bar{\Phi}_0^A \end{pmatrix}$, flow forward to forcing phase $\bar{\Phi}_0^B$, where we will have $\bar{\theta}^A > -\pi$ and $s^A < 1$. From this point forward, when the trajectories are at the same forcing phase, the inhibition on trajectory A will be lower. From (5.2), we see that for fixed θ, $\dot{\theta}$ strictly increases as inhibition decreases, so a trajectory under more inhibition can not cross from below to above a trajectory with higher inhibition. From this point forward, $\bar{\theta}^B$ cannot cross $\bar{\theta}^A$ from below, and will stay beneath it until trajectory A reaches another spike at some forcing phase $\bar{\Phi}_1^A$. Hence trajectory b must reach a spike at some $\bar{\Phi}_1^B > \bar{\Phi}_1^A$.

By the preceding argument, $\mathbf{P}$ (the map from $\bar{\Phi}_0^A$ to $\bar{\Phi}_1^A$ described above) preserves ordering. It is continuous because it is a Poincare map on a continuous flow; it is therefore a homeomorphism on $\mathbb{R}$.

The map $\mathbf{P}$ is the lift of the map $\mathbf{P}_{\mathbb{T}^1}$ acting on $\mathbb{T}^1$. Since $\mathbf{P}$ is a periodic order-preserving homeomorphism on $\mathbb{R}$, $\mathbf{P}_{\mathbb{T}^1}$ from Φ_0^A to Φ_1^A must be an orientation-preserving homeomorphism on the circle.

□

As discussed above, this condition is sufficient to prove the existence of a broken invariant torus that can be repaired by identifying points across the jump, and this set absorbs all initial conditions in finite time. $\mathbf{P}_{\mathbb{T}^1}$ is the spike map discussed in Section **??**, which takes the forcing phase at the right-hand limit of one spike to the forcing phase at the right-hand limit of the next. Once the system has reached a spike, it becomes constrained to the invariant torus, and this spike map fully determines its dynamics.

This proof works because when all other things are equal, a system state with lower s reaches a spike before a system state with higher s. Trajectory B, which spiked more recently than trajectory A, must have a higher value of s at any given time, and therefore its next spike must be later than the next spike of trajectory A. By contrast, a trajectory on the right branch of the FN oscillator will reach a spike sooner if w is higher, whereas a trajectory on the left branch will spike sooner if w is lower, so this proof does not apply.

5.1.2. *Invariant torus for small $c > 0$.* An attracting torus also exists in the case $c > 0$ as long as trajectories contract together sufficiently strongly with time. One possible cause of this contraction is the convergence of trajectories due to the resetting map ρ. When $c = 0$, ρ (and hence the return map $\mathbf{R}$) forces all spiking trajectories onto the circle $s = 1$; when c is small, ρ (and hence $\mathbf{R}$) pushes trajectories close to the circle $s = 1$, and as a result the phase space contracts significantly at each application of the resetting map. Since the

net excitation G exceeds zero in bounded time, the time between spikes is bounded, and this contraction occurs regularly; therefore, we expect it to lead to a steady contraction of the whole phase space. In the tradition of contraction-mapping theorems, we expect a strong contraction to lead to the existence of an invariant set, which in this case will be an invariant circle under $\mathbf{R}$ (like the circle $s = 1$ in the $c = 0$ case) and a corresponding broken invariant torus in the full phase space.

Intuitively, it would make sense to apply Niel Fenichel's result on the persistence of invariant manifolds to show that the torus persists when c is perturbed away from zero; however, Fenichel gives his result only for diffeomorphisms and continuous flows, whereas our flow experiences discontinuities and our return map $\mathbf{R}$ is not a diffeomorphism for $c = 0$ (due to the degeneracy of the map ρ in this case). In Appendix **??**, we use a contraction-mapping result taken from [**Kolesov2003**] and [**shilnikov98**] called the Annulus Principle to show that for any set of system parameters, $c > 0$ may be chosen sufficiently small that there still exists an attracting invariant circle for $\mathbf{R}$ and hence a broken attracting invariant torus for the full system.

5.1.3. *Invariant torus for large g.* A second source of contraction in phase space of the ING system is the convergence of trajectories under sustained inhibition. This second factor in the creation of an invariant torus has been referred to in [**diener85**] and [**dorea2009**] as "rivering" because many voltage trajectories converge tightly under sustained inhibition to form a "river" of trajectories.

In the ING system, s and Φ trajectories evolve independently of θ: $(s_t, \Phi_t) = (s_0 e^{-\frac{t}{\tau_s}}, \Phi_0 + t)$. When the parameters of the ING system force θ to remain sufficiently negative for a sufficiently long time, rivering causes the θ coordinates of sets of trajectories on the same (s, Φ) trajectory to converge into tight rivers. As a result, the time for any system state (θ, s, Φ) to reach a spike comes to depend exclusively on s and Φ and becomes largely independent of θ; equivalently, the next spike time comes to depend only on what (s, Φ) trajectory the system joins at a spike, and becomes largely independent of the specific point along that trajectory that the spike occurs. When $c = 0$, fully-resetting synapses create an attracting invariant torus by causing the set of trajectories reaching a spike at the same forcing phase Φ to converge onto the same trajectory; when strong rivering of θ trajectories occurs, it creates an attracting invariant torus by causing the set of trajectories proceeding from a spike along the same (s, Φ) trajectory to converge on the same trajectory.

For any one-dimensional ODE, we can define a quantitative measure of rivering. Consider the (possibly nonautonomous) ODE $\dot{v} = F(v, t)$. For two nearby initial conditions v_0^* and $v_0^* + \Delta v_0$, we define Δv_t as the difference between the trajectories initialized at these two points after time t. Δv_t evolves according the the linearization of the ODE about the "base trajectory" v_t^*: $\Delta \dot{v}_t = F_v(v_t^*, t)\Delta v_t$. This linear ODE can be solved by ordinary methods:

$$\Delta v_t = \Delta v_0 e^{\int_0^t F_v(v_r^*, r)dr}.$$

We set $\kappa = e^{\int_0^{t_s} F_v(v_t^*, t)dt}$, where t_s is the next spike time after $t = 0$. κ is a measure of the strength of rivering between $t = 0$ and $t = t_s$. If it is close to zero, nearby trajectories converge almost completely; if it is near 1, nearby trajectories stay approximately the same distance apart; if it is large, nearby trajectories diverge. κ is implicitly a function of the initial condition v_0^*.

For the ING system described by (4.3), we can try to define κ for the membrane potential variable V:

$$\Delta V_{t_s} = \kappa \Delta V_0, \text{ where } \kappa = e^{\int_0^{t_s} 2V_t^* dt}.$$

Unfortunately, by this definition the integral in κ explodes at both ends of the interval, where $V = \pm\infty$. Fortunately, we can circumvent this problem by using the change of variables from V to θ as described by (5.2) to define κ for the theta neuron phase θ:

$$\Delta\theta_{t_s} = \kappa \Delta\theta_0, \text{ where } \kappa = e^{\int_0^{t_s} \sin(\theta_t^*)(1-G_t^*)dt}$$

where G_t^* is the net current at time t along the base trajectory $(\theta_t^*, s_t^*, \Phi_t^*)$:

$$G_t^* = b - g s_t^* + I(\Phi_t^*).$$

We can use a change of variables to fully exploit the contraction resulting from the rivering of θ trajectories. We replace the coordinate Φ with $\phi = \Phi + \tau_s \ln(s)$, and then define a map $\tilde{\mathbf{R}}$ analogous to $\mathbf{R}$ that takes the system state (s_0, ϕ_0) at the right-hand limit of a spike at time 0 and returns the system state (s_{t_s}, ϕ_{t_s}) at the right-hand limit of the next spike:

$$\begin{aligned} \tilde{\mathbf{R}} : [0,1] \times \mathbb{T}^1 &\to [0,1] \times \mathbb{T}^1 \\ (s_0, \phi_0) &\to (s_{t_s}, \phi_{t_s}) \end{aligned} \tag{5.3}$$

This new coordinate system is useful because all spikes occurring at the same value of ϕ_0 experience the same timecourse of s and Φ until the next spike, and therefore tend to river together. Thus, though $\mathbf{R}(s_0, \Phi_0)$ may depend strongly on both s_0 and Φ_0, $\tilde{\mathbf{R}}(s_0, \phi_0)$ depends mainly on ϕ_0 and therefore contracts trajectories together strongly in the s direction.

In Appendix **??**, we show that by choosing the inhibitory conductance parameter g sufficiently large and bounding the allowable magnitude of the forcing current $I(\cdot)$, we can guarantee that κ is small (while other important quantities remain bounded from below). Using the contraction-mapping result described above, we show that if κ is sufficiently small

(relative to these other bounded quantities), then $\tilde{\mathbf{R}}$ possesses an attracting invariant torus, and hence a broken attracting invariant torus exists in the full space.

Remark 5.1.1. *In the proof in Appendix* **??***, a second type of convergence of trajectories also occurs for large g: since large g prevents spiking for a long time, all trajectories reach spikes with s close to zero. A measure of convergence of s trajectories due to decay of inhibition is* $e^{-\frac{t_s}{\tau_s}}$*. In our proof, we show that like* κ*, this quantity also becomes arbitrarily small with large g, and also contributes to the contraction of phase space that causes the system to meet the conditions of the Annulus Principle and converge on an attracting invariant manifold.*

5.2. **Monostability.** *Here we use the inter-spike interval function* Ψ *to study the stability of ING phase-locking, and prove that phase-locking to square pulses is monostable.*

The bistability of the FN oscillator discussed in 4.1.2 is related to its constraint to a plane: excitatory forcing pushes the oscillation phase forward when v is rising but retards the phase when v is falling. In our ING system, excitatory forcing can only cause trajectories to reach spikes earlier. This is possible because the voltage variable lives on a circle, and therefore can continue to move in the same direction throughout the ING period. The uniformity of the ING oscillator's response to excitatory pulses suggests that ING might not lend itself to bistability in the same way as the FN oscillator.

We first consider the case of the forced ING system modeled by the equations (4.3) with $c = 0$, as discussed in section 5.1.1. We may study the number of stably phase-locked states using the spike map $\mathbf{P}$ (from forcing phase at one spike to forcing phase at the next, as defined in section 5.1.1) and its action on $\bar{\Phi}_0 \in \mathbb{R}$: where $\mathbf{P}$ intersects the diagonal line $\mathbf{P}(\bar{\Phi}_0) = \bar{\Phi}_0 + T_I$, phase-locking occurs. Equivalently, we may study the inter-spike interval (ISI) function, defined as

$$\Psi(\Phi_0) = \mathbf{P}(\bar{\Phi}_0) - \bar{\Phi}_0. \tag{5.4}$$

Phases Φ_0 where Ψ crosses the line $\Psi(\Phi_0) = T_I$ are locking phases. By the basic theory of iterated maps (explored and explained more thoroughly in Appendix **??**), only phases where the crossing is from top to bottom may be stable. In order to show that the ING oscillator may only have one stably phase-locked orbit in response to a particular forcing input, it is sufficient to show that Ψ can only cross T_I in the downward direction once, or equivalently that it can only cross T_I twice.

If we assume that τ is sufficiently small that spikes occur immediately as soon as $G > 0$, we can analytically describe $\Psi(\Phi_0)$ (and illustrate it; see Figure 9). During a pulse, G can only exceed zero if the previous spike took place sufficiently long ago. If the previous spike was too recent, the pulse arrives under slightly too much inhibition to produce a positive net current. When this occurs, the next spike time jumps ahead to the next time the net current is positive, which may be at the next pulse or to the time a spike would naturally occur unforced (after the natural period of oscillation).

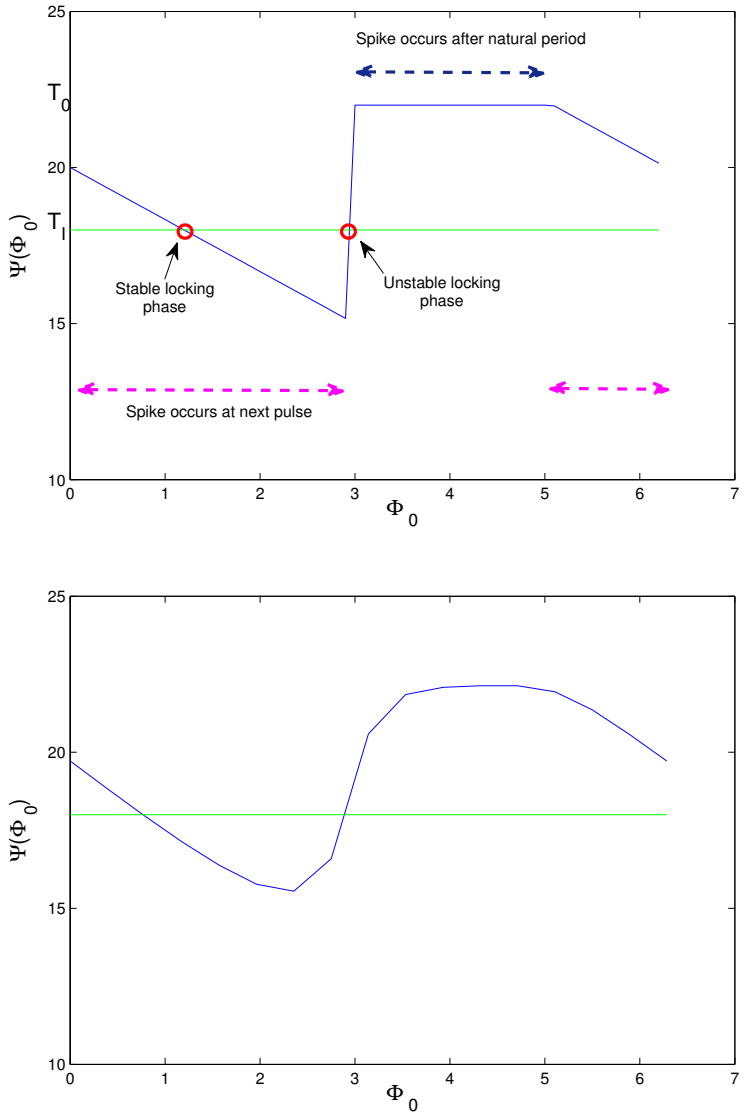

Figure 9: **Top:** The interspike interval function Ψ as a function of initial forcing phase Φ_0 at a spike., for the singular limit case $\tau \to 0$. At the phases where $\Psi(\Phi_0)$ crosses T_I (the green line), phase locking is possible. **Bottom:** Ψ vs. Φ_0 for $0 < \tau \ll 1$. Note the similarity to the singular limit case. We prove that under square-pulse forcing, the direction of Ψ' may only change twice (as it does above), and that therefore Ψ may cross T_I only two times. Only one of the two crossings may be from top to bottom, and this is the requirement for stability, so only one stable phase-locking phase may exist.

Thus, for small τ, $\Psi(\Phi_0)$ a characteristic shape, represented in Figure 9, top. When Φ_0 is in the range marked by purple arrows, the next spike is evoked by the next pulse. In this range, the next spike time does not change with Φ_0, so the interspike interval shrinks steadily as Φ_0 increases and $\frac{d}{d\Phi_0}\Psi = -1$. At some initial forcing phase, the next pulse occurs slightly too early to evoke a spike, and here Ψ jumps up sharply. In the range marked by blue arrows, the next spike occurs after the natural oscillator period T_0, before the following pulse. In this range, the ISI is independent of the initial forcing phase Φ_0, so $\frac{d}{d\Phi_0}\Psi = 0$. Beyond this range, the following pulse occurs before T_0 has passed since the last spike, so this pulse evokes the next spike and $\frac{d}{d\Phi_0}\Psi = -1$ again.

When Ψ takes this shape, it can cross the horizontal line $\Psi(\Phi_0) = T_I$ in the downwards direction no more than once in each forcing period. Only at this crossing point can stable phase-locking occur, so phase-locking (if it occurs at all) must be monostable.

Remark 5.2.1. *We make a much more thorough study of the limiting case of small τ in [[[COMPANION PAPER]]].*

When τ is not vanishingly small, the shape of $\Psi(\Phi_0)$ is a smoother approximation of this shape that still has only one peak and one trough (see Figure 9, bottom), so we still expect it to cross $\Psi(\Phi_0) = T_I$ in the downward direction only once.

However, we seek a more rigorous proof of monostability. It is difficult or impossible to write an explicit expression for Ψ, but solving a linear ODE allows us to write an expression for $\Psi'(\Phi)$. We do so in Appendix **??**, and use it to study the number of stably locked states. We prove that in the case of square pulses, the ISI function on the circle can cross T_I from above only once, proving the monostability of $1:1$ phase-locking in response to square-wave forcing. We also extend this result from the $c = 0$ case to the case of sufficiently small $c > 0$.

Empirical observation of simulations suggests that phase locking to non-square periodic pulses is also generally monostable. A proof for non-square pulses would be more difficult, but might follow along similar lines.

6. Discussion

6.1. **Major Results.** The ING network's relative simplicity and its fundamental differences from other studied oscillators make it interesting from a purely theoretical perspective: it exemplifies an oscillatory mechanism that is fundamentally different from other archetypical oscillators. This work investigates the dynamics of the simple gamma-rhythmic ING circuit under periodic forcing and contrasts them against those of relaxation oscillators. It identifies two key properties of the ING circuit:

- If synapses saturate sufficiently at each spike volley and/or parameters allow for sufficient rivering of voltage trajectories as inhibition decays, then no matter the

forcing strength, the periodically forced ING circuit possesses an attracting invariant torus. As a result, it may achieve only periodic and quasiperiodic dynamics, to the exclusion of period-doubling, coexistence of 1:1 and 2:1 orbits, and any other behaviors impossible on the two-torus. We find that this property is shared by the PING oscillator if the rise time of the inhibitory population is small.

- In the restricted case of saturating synapses and square forcing pulses, the ING circuit can possess only one stable, phase-locked orbit at a time. Phase-locking is also monostable for any pulse shape when the membrane time constant is sufficiently small to make the ISI function take on a shape similar to its shape in the limit as the membrane time constant goes to zero (see Figure 9).

Both of these properties are not shared by the forced relaxation oscillator, which has regimes of period doubling and bistability. The differences can be traced back to the ING oscillator's rapidly-resetting, slowly-decaying feedback inhibition, and the unidirectional influence of decreasing inhibition and excitatory forcing made possible by the ING oscillator's circular (rather than linear) fast variable.

In our models, we have modeled fast-spiking interneurons with theta neurons. This simplification is not completely justified: fast-spiking neurons have been observed to show type-2 rather than type-1 excitability, i.e. they do not initiate spiking by a saddle-node bifurcation like the theta neuron but instead by a Hopf bifurcation. Consistent with this result is the observation that fast-spiking neurons resonate about their resting voltages at slow-gamma-like frequencies. Resonance may compromise the validity of our ING results. In particular, it may not always be the case that the spike map is monotonic: earlier *or* later spike timing may determine whether input pulses align or fail to align with periods of increased excitability, significantly advancing or delaying the next spike. However, preliminary results show that resonance at or near the spiking frequency does not lead to nonmonotonicity of the spike map. Furthermore, in our PING model, our results depend on the E-population being type-1 excitable but do not depend strongly on the excitability type of the I-population as long as they can respond quickly to strong excitatory input.

One major assumption in our models was that each population of cells fires synchronously or not at all. It has repeatedly been observed that during some episodes of PING (dubbed "sparse" or "weak" PING), only a fraction of the E-cells fire on each cycle [**Buhl1998**] [**Burchell1998**] [**Fisahn1998**]. However, our work does not rest heavily on the assumption of synchronous E-cells as long as the I-volley is triggered all at once, creating a sudden onset followed by a slow decay of inhibition. We believe that with a reasonable set of assumptions, sparse PING could be shown to obey the same dynamic restrictions as strong PING.

It has also been observed that some I-cells may fire on only a fraction of cycles, and that different amounts of inhibition may be recruited on each gamma cycle [**Atallah2009**]. The effects of variable inhibitory recruitment on phase-locking has been studied in [**Serenevy**]. Though the authors find that this effect makes phase-locking more robust, their results also show that a volley occurring at a later phase of the periodic drive may be followed by an

earlier second volley due to less inhibitory recruitment, creating a non-monotonic spike map and qualitatively different behavior than the dynamics described here. We hypothesize that in some parameter regimes, this effect leads to chaos.

Our brief discussion of PING makes two additional major assumptions: the rise time of the inhibitory population following an E-spike must be small, and the forcing must be delivered only to pyramidal cells. The first assumption has strong basis in experimental findings: Atallah and Scanziani observe that this lag time is about 2ms in vivo. This may be due partially to low membrane time constants. Indeed, fast-spiking interneurons do have lower membrane time constants than most other cortical cells [**Pike2000**]. It also may be due to the strong E-to-I connectivity in these networks.

The second assumption is a caricature: there is reason to believe that fast-spiking interneurons involved in PING do receive input from other cortical areas, as well as evidence that forcing these interneurons can entrain a gamma rhythm [**Cardin2009**]. If we continue to assume that the E- and I-populations are each firing synchronously, we do not expect this complication to significantly impact our conclusions. If excitatory forcing is delivered to the I-population in addition to the E-population, inhibitory spike volleys may be initiated with or without excitatory participation; but in either case, these volleys are created by forcing that pushes them above decaying inhibition, and cause the inhibition to reset quickly, wiping out most history dependence. These dynamics are not significantly different than the dynamics when periodic forcing is delivered exclusively to pyramids.

6.2. **Relationship to Other Work.** To our knowledge, our work is the first to directly compare the network gamma mechanism to the relaxation oscillator, and the first to identify dynamic constraint inherent to the network gamma mechanism under forcing.

In an effort to answer questions about schizophrenia, Vierling-Claassen and Kopell [**dorea2009**] study periodically forced PING circuits using a model very similar to ours. They create a one-dimensional map similar to our spike map using the assumption that the various possible trajectories of both cells river together completely under inhibition and that synapses saturate at every spike. In this manuscript we show that in the case of ING, either one of those assumptions or the combination of partial rivering and partial saturation is sufficient to create an invariant 2-torus, allowing us to define a one-dimensional map describing the asymptotic dynamics. We also show that the same is true of PING if the time between an E-spike and the subsequent I-spike is small. Their work is aimed at explaining a specific observation in schizophrenic patients, and assume a specific profile of periodic forcing; ours is aimed at deepening our understanding of any instance of forced network gamma with analytical results valid for any periodic input.

The work of Serenevy and Kopell [**Serenevy**] is closely related to ours. It studies forced ING using maps from one spike volley to the next, and demonstrates that variable I-cell participation adds robustness to phase locking. Their work assumes complete rivering of

trajectories in order to define a one-dimensional map description of the interesting phase-locking dynamics that may result from variable participation. Our work applies the language of invariant manifolds and assumes a fixed population of participating cells, allowing us to generalize and relax the assumption of complete rivering as discussed above.

Various authors, including Ermentrout et al. [**Ermentrout2001**] and Kilpatrick and Ermentrout [**Kilpatrick2011**] have studied QIF neurons with adaptation in coupled and forced settings. The QIF-with-adaptation models they use are very similar to our model of ING: instead of decaying and resetting inhibition, they have a decaying and resetting adaptive current. Our work differs from theirs in at least two important ways. First, other work generally uses PRC's that assume weak coupling/forcing or pulsatile forcing; here, we prove the existence of invariant tori given no assumptions about the forcing signal except Lipschitz continuity, and prove monostability assuming forcing by square pulses of finite length. Second, we generalize our result to PING, an interaction of two cell populations. To our knowledge, we are the first to draw a connection between the QIF-with-adaptation model (our ING model) and the PING circuit.

Engelbrecht and Mirollo [**Engelbrecht2012**] also study the existence of attracting invariant manifolds for neuronal systems. In their study of the interaction between periodic forcing and noise in individual neurons, they find that even high-dimensional neuron models asymptote to two-dimensional surfaces (like our ING and PING models), allowing them to describe a forced Hodgkin-Huxley neuron with a one-dimensional spike map. They explain this effect as the persistence of an invariant torus from the unforced oscillatory state, but they do not further examine the factors that allow this torus to persist.

Burden et al. [**Burden2011**] study the existence of attracting invariant manifolds for hybrid dynamical systems like the ones we study here. They prove that if a resetting map in a hybrid system with a periodic orbit is of rank r, then the periodic orbit lies in a $r+1$-dimensional manifold that attracts all initial conditions in finite time. When $c = 0$ in our ING model, the resetting map is rank 1, and we prove that this implies the existence of a two-dimensional invariant manifold. In later unpublished work, they also discuss exponentially attracting manifolds in hyperbolic systems. However, all of the work done by Burden et al. assumes the existence of a periodic orbit, whereas our work proves the existence of invariant tori even if no periodic orbit exists.

Rajan et al. [**Rajan2010**] provide analytical results indicating that strong ongoing input to a chaotic neural network suppresses chaotic behavior. Though their results are for firing rate models of random networks and do not relate to rhythmic behavior, their results tell a story very similar to ours: a driven network displays a limited, non-chaotic repertoire of dynamics, and can therefore respond reliably and reproducibly to an input. Any network receiving ongoing temporally-varying input must strike a balance between its intrinsic dynamics and its response to the input. Rhythms may prove to play an important role in adjusting and maintaining this balance.

6.3. **Implications for Communication Through Coherence.** The properties of gamma rhythms under forcing are particularly interesting in the context of the Communication-Through-Coherence (CTC) hypothesis [**Womelsdorf2007b**]. Briefly, this hypothesis states that oscillations (and in particular gamma rhythms) in neuronal populations create periodic windows during which they are susceptible to input, alternating with windows of insensitivity to input. For two oscillating populations to effectively communicate, the output from one oscillation must be phase-aligned with the susceptible windows of the other.

For CTC to operate, it is necessary (though not sufficient) that a mechanism exist for reliably establishing a consistent phase relationship between the rhythms of the "sending" and "receiving" populations. Several studies have demonstrated that gamma rhythms in the brain cannot be relied upon to sustain a metronome-like periodicity [**Burns2011**] [**Xing2012**], so any phase alignment between gamma rhythms must actively maintained [**Nikolic2013**]. One straightforward method of creating and maintaining such a phase relationship is to allow the rhythmicity of the sending population's output to entrain the receiving population directly. But entrainment is not sufficient: the phase relationship between the two populations must be stable and predictable. As is demonstrated by the case of the FN system, an oscillator given excitatory forcing of moderate strength at a period close to its natural period may align differently with alternate forcing cycles due to period-doubling; or its alignment with the forcing may depend strongly on initial conditions due to bi-stability of phase-locked orbits. The properties of ING identified above guarantee that neither of these behaviors may interfere with the phase alignment of the entrained receiving population, facilitating CTC.

We have proven that phase-locking occurs with a specific, reliable phase relationship to the forcing; but we have not specified *what* phase relationship. This topic is explored in depth in [[[COMPANION PAPER]]], where it is shown that volleys in the gamma circuit generally follow input pulses with short delay. This result places gamma rhythms in a very specific position among the set of all possible oscillators, which may entrain to any phase relationship with periodic forcing. The specific phase relationship created by network gamma rhythms under forcing is optimal for CTC with these networks: when forcing pulses (which presumably also encode information) reach the receiving population, it has been nearly a full gamma period since the last spike volley, so inhibition is low and cells can readily fire in response. In [[[COMPANION PAPER]]], it is also shown that when two gamma circuits are mutually coupled and phase-locking occurs, a phase relationship is created that is optimal for unidirectional transmission of information from the more driven population to the less driven one.

The techniques used here all assume that $\Phi \in \mathbb{T}^1$, i.e., forcing is periodic. However, with almost exactly the same methods, we could study forcing by N periodic signals of different periods by considering $\vec{\Phi} \in \mathbb{T}^N = [0, T_{I,1}) \times ... \times [0, T_{I,N})$ (and setting $\dot{\vec{\Phi}}$ to a vector of ones). A return map could be defined for this system that took the state $(s, \vec{\Phi})$ at one spike to the state at the next; an attracting invariant torus for this map that could be written as a graph of s over $\vec{\Phi}$ would allow us to study the asymptotically stable dynamics of the system as the

dynamics of a highly constrained map on $\mathbb{T}^N$. This method could answer questions about when the system could phase-lock to one periodic signal without interference by others, as discussed and studied through simulation in [**Christoph2008**], and potentially reveal the importance of the network gamma mechanism in selective communication through phase locking.

6.4. **Broader Implications.** The ING/PING mechanism is interesting outside the context of gamma rhythms due to its resemblance to other rhythmic processes across the nervous system. Some bursting mechanisms may be described by a process of overcoming a threshold that decays slowly and resets quickly. For instance, the high-threshold bursting of thalamic relay neurons implicated in alpha rhythms is characterized by the slow activation of h-type cation channel while the cell rests at low voltage, followed by its rapid inactivation during a burst of spikes. As the h-current activates, it requires less excitation to evoke a burst. The ING circuit as defined here might serve as a first approximation of this process, with s replaced by the activation of the h-current and spiking events replaced by bursting events. To the extent that such a model accurately describes the generation of a rhythm, this rhythm may be expected to share the distinctive properties of ING under forcing.

Though the systems treated here are low-dimensional for analytical tractability, the ideas and techniques presented apply equally to higher dimensional systems, e.g. a gamma rhythm generated by Hodgkin-Huxley neurons or by populations that are not completely synchronous. In particular, if all of the variables in a system approach a fixed point that varies slowly with the decay of a single slow variable and persists for sufficient time, the rivering effect will create a strong contraction in phase space. As discussed briefly in Section 5.1.3, this effect may occur when fast variables describing the cell state converge during the slow decay of inhibition, or when synapses rise rapidly towards saturation during a spike. We show here that variational equations may be used to find conditions under which rivering induces sufficient contraction in state space to force the existence of an invariant manifold. If this rivering occurs under inhibition that strictly decays, spikes occurring later relative to the forcing signal remain under greater inhibition until the next spike, and are therefore followed by a later spike (as discussed in Section **??**). We expect that such a monotonic relationship combined with strong rivering will force higher dimensional systems onto an invariant torus as well.

In many neural processes, individual spikes or spike volleys form natural landmarks in the dynamics, so maps from one spiking event to the next are a convenient framework in which to study neural dynamics. The framework of invariant manifolds significantly extends the applicability of spike maps. A map from the time or forcing phase of one spiking event to the time or forcing phase at the next is only well-defined if the system's state at a spike can be completely specified by a single variable, so such a map can only be defined for a system that is effectively two-dimensional. However, even a high-dimensional system may become effectively two-dimensional due to contraction onto an invariant 2-torus. Similarly, the map from the state at one spike to the state at the next may be reduced to a second- or third-order spike map (a map from two or three variables at one spike to the same

variables at the next) if the system converges onto a three- or four-dimensional manifold, respectively.

Conversely, maps from one spike to the next may be used to prove the existence of attracting manifolds of the full system, as we have done here. The search for low-dimensional activity patterns and their causes in the brain is ultimately a search for low-dimensional attracting invariant manifolds in large state-spaces; the techniques we have presented here may prove valuable to that search.

7. Appendix

Here we prove the theorem:

8. References