

# Canonical Neural Models<sup>1</sup>

Frank Hoppensteadt<sup>1</sup> and Eugene Izhikevich<sup>2</sup>

## Introduction

Mathematical modeling is a powerful tool in studying fundamental principles of information processing in the brain. Unfortunately, mathematical analysis of a certain neural model could be of limited value since the results might depend on particulars of that model: Various models of the same neural structure could produce different results. For example, if one obtains results studying a Hodgkin-Huxley-type model (see AXONAL MODELING) and then augments the model by adding more parameters and variables to take into account more neurophysiological data, would similar results hold? A reasonable way to circumvent this problem is to derive results that are largely independent of the model and that can be observed in a class or a family of models.

Having understood the importance of considering families of neural models instead of a single model, we carry out this task by reducing an entire family of Hodgkin-Huxley-type models to a canonical model (for precise definitions see Sect. 4.1 in Hoppensteadt and Izhikevich 1997). Briefly, a model is *canonical* for a family if there is a continuous change of variables that transforms any other model from the family into this one, as we illustrate in Figure 1. For example, the entire family of weakly coupled oscillators of the form (1) can be converted into the canonical phase model (6), where  $H_{ij}$  depend on the particulars of the functions  $f_i$  and  $g_{ij}$ . The change of variables does not have to be invertible, so the canonical model is usually lower-dimensional, simple, and tractable. Yet, it retains many important features of the family. For example, if the canonical model has multiple attractors, then each member of the family has multiple attractors.

The major advantage of considering canonical models is that one can study universal neurocomputational properties that are shared by all members of the family since all such members can be put into the canonical form by a continuous change of variables. Moreover, one need not actually present such a change of variables explicitly, so derivation of canonical models is possible even when the family is so broad that most of its members are given implicitly, e.g., in the abstract form (1). For example, the canonical phase model (6) reveals universal computational abilities (e.g., oscillatory associative memory) that are shared by all oscillatory systems regardless

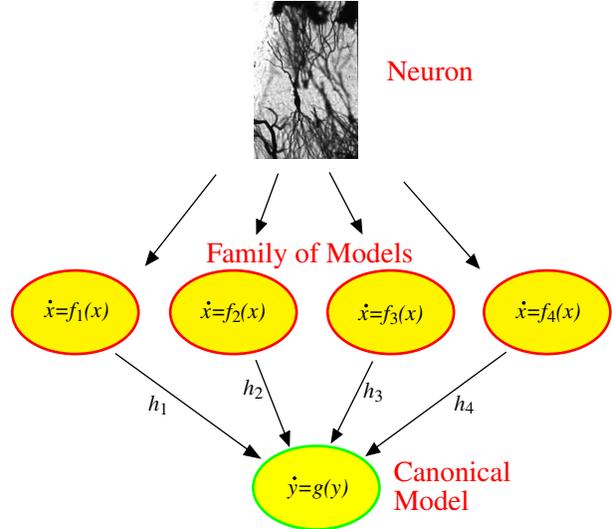


Figure 1: Dynamical system  $\dot{y} = g(y)$ , is a canonical model for the family  $\{f_1, f_2, f_3, f_4\}$  of neural models  $\dot{x} = f(x)$  because each such model can be transformed into the form  $\dot{y} = g(y)$  by the continuous change of variables  $h_i$ .

of the nature of each oscillator or the particulars of the equations that describe it. Thus, the canonical model approach provides a rigorous way to obtain results when only partial information about neuron dynamics is known. Many examples are given below.

The process of deriving canonical neural models is more an art than a science, since a general algorithm for doing this is not known. However, much success has been achieved when we consider weakly connected networks of neurons whose activity is near a bifurcation, which often occurs when the membrane potential is near the threshold value, see DYNAMICS AND BIFURCATION OF NEURAL NETWORKS AND PHASE PLANE ANALYSIS OF NEURAL ACTIVITY. We review such bifurcations and corresponding canonical models. Their rigorous derivation and detailed analysis can be found in the book by Hoppensteadt and Izhikevich (1997).

## Weakly Connected Neural Networks

The assumption of weak neuronal connections is based on the observation that the typical size of a postsynaptic potential is less than 1 mV, which is small in comparison with the mean size necessary to discharge a cell (around 20 mV) or the averaged size of the action potential (around 100 mV); see detailed review of relevant electrophysiological data in Chapter 1 in Hoppensteadt and Izhikevich (1997). From the mathematical point of view this results in neural

<sup>1</sup>Invited chapter to Arbib MA (Ed) Brain Theory and Neural Networks. Second Edition. The MIT press, Cambridge, MA, 2001

models of “weakly connected” form

$$\dot{x}_i = f(x_i, \lambda_i) + \varepsilon \sum_{j=1}^n g_{ij}(x_i, x_j, \varepsilon), \quad (1)$$

where each vector  $x_i \in \mathbb{R}^m$  describes membrane potential, gating variables, and other electrophysiological variables of the  $i$ -th neuron (see ION CHANNELS: KEYS TO NEURONAL SPECIALIZATION). Each vector  $\lambda_i \in \mathbb{R}^l$  denotes various biophysical parameters of the neuron. The function  $f$  describes the neuron’s dynamics, and the functions  $g_{ij}$  describe connections between the neurons. The dimensionless parameter  $\varepsilon \ll 1$  is small, reflecting the strength of connections between neurons.

## Bistability And Hysteresis

Bistable and hysteretic dynamics are ubiquitous in neural models, and they may play important roles in biological neurons. The cusp bifurcation depicted in

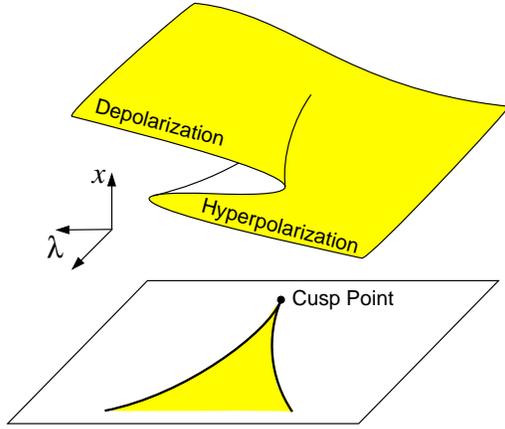


Figure 2: Cusp surface.

Figure 2 is one of the simplest bifurcations leading to such dynamics. For example, the sigmoidal neuron

$$\dot{x} = -x + aS(x), \quad S(x) = 1/(1 + e^{-x}),$$

is at a cusp bifurcation point  $x = 0.5$  when  $a = 4$ . It is bistable when  $a > 4$ . If each neuron in the weakly connected network (1) is near a supercritical cusp bifurcation, then the entire network can be transformed into the canonical form (Hoppensteadt and Izhikevich 1997)

$$y'_i = r_i - y_i^3 + \sum_{j=1}^n s_{ij}y_j \quad (2)$$

where each scalar  $y_i \in \mathbb{R}$  describes re-scaled dynamics of the  $i$ th neuron. Particulars of the functions  $f$

and  $g_{ij}$  and the value of the parameters  $\lambda_i$  do not affect the form of the canonical model, but only particulars of the parameters  $r_i$  and  $s_{ij}$ . Thus, studying the canonical model (2) one can gain some insight into neuro-computational behavior of any neural model near a cusp bifurcation, whether it is a simple sigmoidal neuron or a biophysically detailed conductance-based (Hodgkin-Huxley-type) neuron.

The canonical model (2) is quite simple: Each equation has only one non-linear term, namely,  $y_i^3$ , and two internal parameters,  $r_i$  and  $s_{ij}$ . Still, the Cohen-Grossberg-Hopfield convergence theorem applies, which means that the canonical model has the same neuro-computational properties as the standard Hopfield network (see COMPUTING WITH ATTRACTORS).

**Theorem 1 (Cohen-Grossberg-Hopfield Convergence Theorem)** *If the connection matrix  $S = (s_{ij})$  is symmetric, then the canonical neural network (2) is a gradient system.*

One can easily check that

$$E(y) = - \sum_{i=1}^n (r_i y_i - \frac{1}{4} y_i^4) - \frac{1}{2} \sum_{i,j=1}^n s_{ij} y_i y_j$$

is a potential function for (2) in the sense that  $y'_i = -\partial E / \partial y_i$ , see also ENERGY FUNCTIONS FOR NEURAL NETWORKS.

## Small Amplitude Oscillations

Many biophysically detailed neural models can exhibit small-amplitude (damped) oscillations of the membrane potential, especially when the system is near transition from rest state to periodic activity. In the simplest case this corresponds to the supercritical Andronov-Hopf bifurcation (Figure 3). Many weakly connected networks (1) of such neurons can be transformed into the canonical model

$$z'_i = (r_i + i\omega_i)z_i - z_i|z_i|^2 + \sum_{j=1}^n c_{ij}z_j \quad (3)$$

by a continuous change of variables (Aronson et al. 1990). Here  $i = \sqrt{-1}$ , and each complex variable  $z_i \in \mathbb{C}$  describes oscillatory activity of the  $i$ th neuron. Again, particulars of the form of the functions  $f$  and  $g_{ij}$  in (1) affect only the values of the parameters  $r_i$ , and  $\omega_i$  and the complex-valued synaptic coefficients  $c_{ij} \in \mathbb{C}$ .

Even though the canonical model (3) exhibits oscillatory dynamics, one can still prove the following

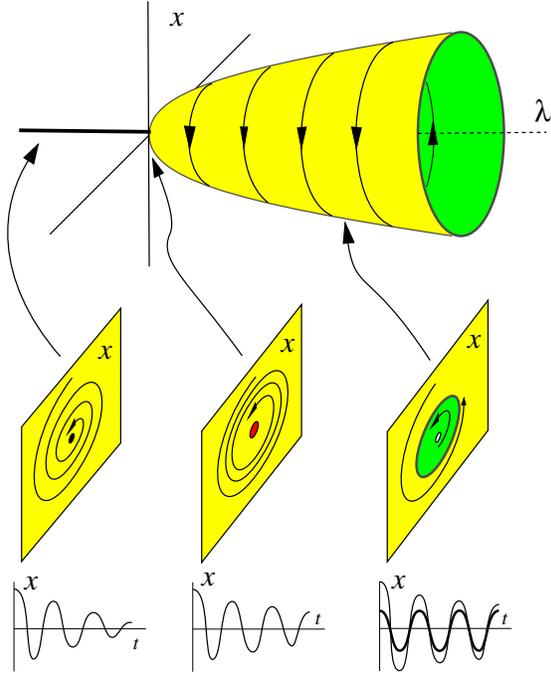


Figure 3: Supercritical Andronov-Hopf bifurcation in  $\dot{x} = f(x, \lambda)$ . *Left*: The rest state is stable. *Middle*: The rest state is losing stability giving birth to a stable limit cycle corresponding to periodic activity. *Right*: The system exhibits periodic activity.

analogue of the Cohen-Grossberg convergence theorem, which implies that the canonical model (3) has oscillatory associative memory; that is, it can memorize and retrieve complex oscillatory patterns (Hoppensteadt and Izhikevich 1996); see Fig. 4.

**Theorem 2 (Synchronization Theorem for Oscillatory Neural Networks)** *If in the canonical neural network (3) all neurons have equal frequencies  $\omega_1 = \dots = \omega_n$  and the connection matrix  $C = (c_{ij})$  is self-adjoint, i.e.,*

$$c_{ij} = \bar{c}_{ji} \quad \text{for all } i \text{ and } j, \quad (4)$$

*then the network always converges to an oscillatory phase-locked pattern; that is, the neurons oscillate with equal frequencies and constant, but not necessarily identical, phases. There could be many such phase-locked patterns corresponding to many memorized images.*

The proof follows from the existence of an orbital energy function

$$E(z) = - \sum_{i=1}^n (r_i |z_i|^2 - \frac{1}{2} |z_i|^4) - \sum_{i,j=1}^n c_{ij} \bar{z}_i z_j$$

for (3), see ENERGY FUNCTIONS FOR NEURAL NETWORKS.

The self-adjoint synaptic matrix arises naturally when one considers complex Hebbian learning rules (Hoppensteadt and Izhikevich 1996)

$$c_{ij} = \frac{1}{n} \sum_{s=1}^k \xi_i^s \bar{\xi}_j^s \quad (5)$$

where each vector  $\xi^s = (\xi_1^s, \dots, \xi_n^s) \in \mathbb{C}^n$  denotes a pattern of phase relations between neurons to be memorized, see also HEBBIAN SYNAPTIC PLASTICITY. Notice that the problem of negative (mirror) images does not arise in oscillatory neural networks, since both  $\xi^k$  and  $-\xi^k$  result in the same phase relations.

The key difference between the Hopfield-Grossberg network and the oscillatory network (3) is that memorized images correspond to equilibrium (point) attractors in the former and to limit cycle attractors in the latter. Pattern recognition by an oscillatory neural network involves convergence to the corresponding limit cycle attractor, which results in synchronization of the network activity with an appropriate phase relation between neurons, as in Fig. 4; see also COMPUTING WITH ATTRACTORS.

## Large Amplitude Oscillations

Suppose that neurons in the weakly connected network (1) exhibit periodic spiking; see Figure 5 and CHAINS OF COUPLED OSCILLATORS; COLLECTIVE BEHAVIOR OF COUPLED OSCILLATORS and PHASE-PLANE ANALYSIS OF NEURAL ACTIVITY. If they have nearly equal frequencies, then the network can be transformed into the phase canonical model

$$\varphi'_i = \omega_i + \sum_{j=1}^n H_{ij}(\varphi_j - \varphi_i) \quad (6)$$

where each  $\varphi_i \in \mathbb{S}^1$  is a one-dimensional (angle) variable that describes the phase of the  $i$ th oscillator along the limit cycle attractor corresponding to its periodic spiking (see Figure 5), and each  $H_{ij}$  is a function that depends on  $f$  and  $g_{ij}$  that can be explicitly computed using Malkin's Theorem (Theorem 9.2 in Hoppensteadt and Izhikevich 1997).

The phase canonical model (6) describes frequency locking, phase locking, and synchronization properties of the original system (1). Therefore, to understand these and other non-linear phenomena that might take place in oscillating neural networks, it usually suffices to consider the phase model. In particular, one can glimpse *the universal computational*

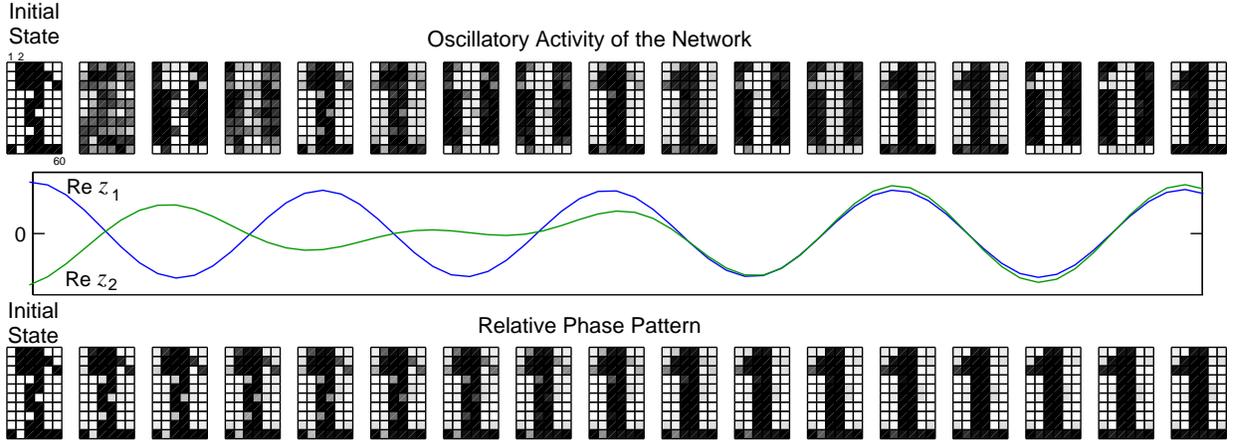


Figure 4: Pattern recognition via phase locking by the oscillatory canonical model (3). Complex Hebbian learning rule (5) was used to memorize patterns “1”, “2”, and “3”. When the distorted pattern “1” is presented as an initial state, the neurons synchronize with the phase relations corresponding to the memorized pattern “1”.

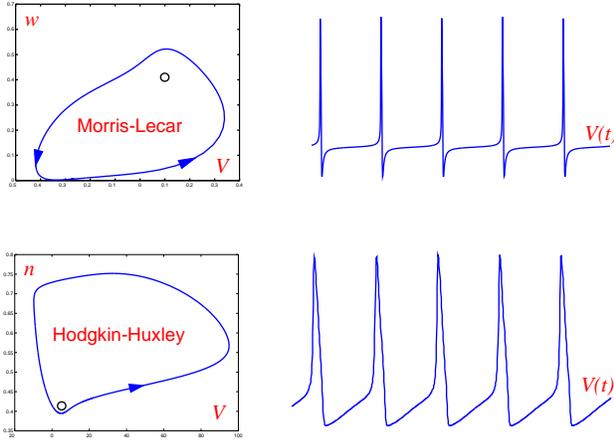


Figure 5: Examples of large amplitude limit cycle attractors corresponding to periodic spiking in two biophysically detailed neural models (Morris-Lecar (1981) and Hodgkin-Huxley (1952)).

abilities that are shared by all oscillatory systems regardless of the nature of each oscillator or the particulars of the equations that describe it. Indeed, one can prove the following analogue of Theorem 2

**Theorem 3 (Synchronization Theorem for Oscillatory Neural Networks)** *If all oscillators in (6) have equal frequencies; i.e.,  $\omega_1 = \dots = \omega_n$ , and the connection functions  $H_{ij}$  have pairwise odd form; i.e.,*

$$H_{ij}(-\psi) = -H_{ji}(\psi) \quad (7)$$

for all  $i$  and  $j$ , then the canonical phase model (6) converges to a phase-locked pattern  $\varphi_i(t) \rightarrow \omega_1 t + \phi_i$  for all  $i$ , so the neurons oscillate with equal frequencies ( $\omega_1$ ) and constant phase relations ( $\varphi_i(t) - \varphi_j(t) = \phi_i - \phi_j$ ). In this sense the network dynamic is synchronized. There could be many stable synchronized patterns corresponding to many memorized images.

The proof is based on the observation that the phase canonical model (6) has the energy function

$$E(\varphi) = \frac{1}{2} \sum_{i,j=1}^n R_{ij}(\varphi_j - \varphi_i)$$

where  $R_{ij}$  is the antiderivative of  $H_{ij}$ ; that is,  $R'_{ij} = H_{ij}$ , see Theorem 9.15 in Hoppensteadt and Izhikevich 1997, and ENERGY FUNCTIONS FOR NEURAL NETWORKS.

For example, Kuramoto’s model (1984)

$$\varphi'_i = \omega_i + \sum_{j=1}^n s_{ij} \sin(\varphi_j + \psi_{ij} - \varphi_i) \quad (8)$$

has such an oscillatory associative memory when  $\omega_1 = \dots = \omega_n$

$$s_{ij} = s_{ji} \quad \text{and} \quad \psi_{ij} = -\psi_{ji},$$

for all  $i$  and  $j$ . If we denote  $c_{ij} = s_{ij} e^{i\psi_{ij}}$ , then these conditions have the form (4). The energy function for Kuramoto’s model is

$$E(\varphi) = -\frac{1}{2} \sum_{i,j=1}^n s_{ij} \cos(\varphi_j + \psi_{ij} - \varphi_i).$$

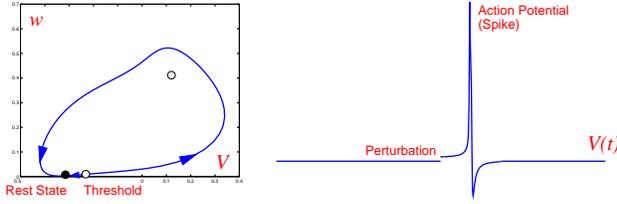


Figure 6: Neural excitability in Morris-Lecar (1981) neuron having fast  $\text{Ca}^{2+}$  and slow  $\text{K}^{+}$  voltage-gated currents. The rest state (black circle) is stable, but small perturbations can push the voltage beyond the threshold (white circle) thereby causing a large amplitude excursion – action potential. The voltage variable changes slowly near the rest states, but fast during the generation of action potentials.

There are various estimates of the storage capacity of the network, as discussed by Vicente et al. (1996). In particular, those authors found a time scale during which oscillatory networks can have better performance than Cohen-Grossberg-Hopfield-type attractor neural networks.

Since neither the form of the functions  $f$  and  $g_{ij}$ , nor the dimension of each oscillator in (1) were specified, one could take the above result to the extreme and claim that *anything that can oscillate can also be used for computing*, as for associative pattern recognition, etc. The only problem is how to couple the oscillators so that (7) is satisfied.

## Neural Excitability

An interesting intermediate case between rest and periodic spiking behavior is when a neural system is *excitable*; that is, it is at rest, but can generate a large-amplitude spike in response to a small perturbation; see Figure 6 and PHASE-PLANE ANALYSIS OF NEURAL ACTIVITY and OSCILLATORY AND BURSTING PROPERTIES OF NEURONS. A simple but useful criterion for classifying neural excitability was suggested by Hodgkin (1948), who stimulated cells by applying currents of various strengths. When the current is weak the cell is quiet. When the current is strong enough, the cell starts to fire repeatedly. He suggested the following classification according to the emerging frequency of firing (see Figure 7)

- **Class 1 neural excitability.** Action potentials can be generated with arbitrarily low frequency. The frequency increases with increasing the applied current.

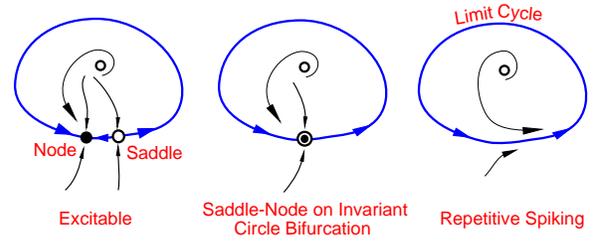


Figure 8: Class 1 neural excitability via saddle-node on invariant circle bifurcation: The threshold state (saddle) approaches the rest state (node), they coalesce and annihilate each other leaving only limit cycle attractor. The oscillation on the attractor has two time scales: slow transition through the “ghost” of the saddle-node bifurcation and fast rotation along the rest of the limit cycle.

- **Class 2 neural excitability.** Action potentials are generated in a certain frequency band that is relatively insensitive to changes in the strength of the applied current.

Their class of excitability influences neuro-computational properties of cells (see review by Izhikevich 2000). For example, Class 1 neural systems have a well-defined threshold manifold for their state variables, beyond which they generate a large amplitude spike. They generate an all-or-none response, and they act as *integrators*, meaning that the higher the frequency of the incoming pulses, the sooner they fire. In contrast, Class 2 neural systems do not have a threshold manifold. They can generate spikes of arbitrary intermediate amplitude, and they act as *resonators*. That is, they respond to certain resonant frequencies of the incoming pulses. Increasing the incoming frequency may delay or even terminate their response.

A canonical model for Class 1 excitable systems is described below, while the canonical model for Class 2 systems has yet to be found.

## Class 1 Excitable Systems

Class 1 excitable systems are understood relatively well (Rinzel and Ermentrout 1989, Ermentrout 1996, Hoppensteadt and Izhikevich 1997, Izhikevich 2000). The transition from rest to periodic spiking in such systems occurs via a saddle-node on invariant circle bifurcation, as we depict in Figure 8, see also DYNAMICS AND BIFURCATION OF NEURAL NETWORKS and OSCILLATORY AND BURSTING PROPERTIES OF NEURONS. A weakly connected network of such neurons can be transformed

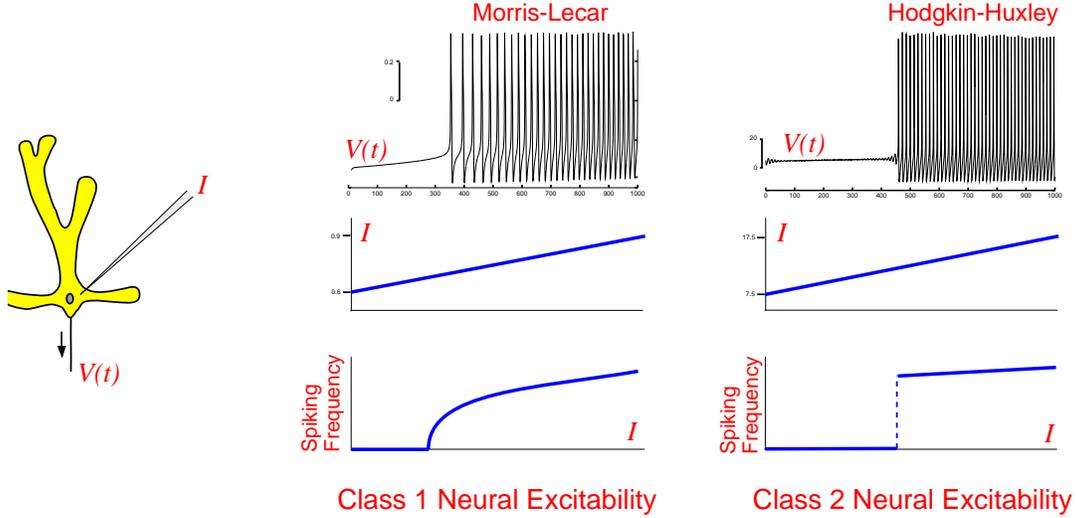


Figure 7: Transition from rest to repetitive spiking in two biophysical models when the strength of applied current,  $I$ , increases. The neural excitability is classified according to the frequency of emerging spiking.

into a canonical model, which can be approximated by

$$\vartheta'_i = 1 - \cos \vartheta_i + (1 + \cos \vartheta_i) \left( r_i + \sum_{j=1}^n s_{ij} \delta(\vartheta_j - \pi) \right) \quad (9)$$

where  $\vartheta_i \in \mathbb{S}^1$  is the phase of the  $i$ th neuron along the limit cycle corresponding to the spiking solution. Again, particulars of the functions  $f$  and  $g_{ij}$  in (1) do not affect the form of the canonical model above, but only affect the values of the parameters  $r_i$  and  $s_{ij}$ , which can be computed explicitly (Hoppensteadt and Izhikevich 1997, Chapter 8). Notice that the canonical model (9) is pulse-coupled, whereas the original weakly coupled network (1) is not. The qualitative reason for pulse coupling is that the voltage changes are extremely slow most of the time because of the proximity to the rest state, but they are relatively instantaneous during the generation of an action potential. Hence the duration of coupling looks infinitesimal on the slow time scale.

The neuron is quiescent when  $r_i < 0$  (Fig. 8, left) and fires periodically when  $r_i > 0$  (Fig. 8, right). It fires a spike exactly when  $\vartheta_i$  crosses the value  $\pi$ , which results in a step-like increase in the phases of other neurons. Hence, the canonical model (9) is a *pulse coupled neural network* (Izhikevich 1999). It has many important physiological features including absolute and relative refractory periods; see Figure 9. Indeed, the effect of every incoming pulse depends on the internal state of the neuron since it is multiplied by the function  $(1 + \cos \vartheta_i)$ . The effect is maximal

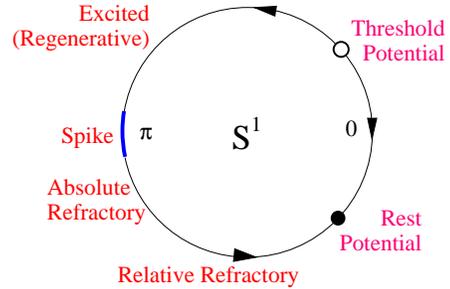


Figure 9: Diagram of the canonical model (9) for class 1 neural excitability (From Hoppensteadt and Izhikevich 1997).

when the neuron is near rest since  $(1 + \cos \vartheta_i) \approx 2$  when  $\vartheta_i \approx 0$ . It is minimal when the neuron is generating a spike since  $(1 + \cos \vartheta_i) \approx 0$  when  $\vartheta_i \approx \pi$ .

A canonical model for *slowly* connected Class 1 excitable neurons with spike frequency adaptation has the form (Izhikevich 2000)

$$\begin{aligned} \vartheta'_i &= 1 - \cos \vartheta_i + (1 + \cos \vartheta_i) \left( r_i + \sum_{j=1}^n s_{ij} w_j \right) \\ w'_i &= \delta(\vartheta_i - \pi) - \eta w_i \end{aligned}$$

where  $w_i$  describes slow synaptic processes. The term  $s_{ii} w_i$  denotes not a self-synapse, but a slow spike frequency adaptation ( $s_{ii} < 0$ ) or facilitation ( $s_{ii} > 0$ ) process.

## Discussion

The canonical model approach to computational neuroscience provides a rigorous way to derive simple yet accurate models that describe single cell or network dynamics (see also SINGLE-CELL MODELS). Such a derivation is possible even when no assumptions are made regarding the detailed form of equations describing neural activity. Indeed, we specify neither  $f$  nor  $g_{ij}$  in (1). The only assumptions we make are those concerning the dynamics of each neuron – whether it is quiescent, excitable, periodic spiking, etc. Nevertheless, any such neural system can be transformed into a canonical model by a piece-wise continuous change of variables.

Derivation of canonical models can be a daunting mathematical task. However, once found, the canonical models provide invaluable information about universal neuro-computational properties shared by a large family of neural systems. For example, studying the canonical model (9) sheds light on behavior of *all* Class 1 excitable systems and their networks regardless of the details of equations describing their dynamics.

**Road Map:** BIOLOGICAL NEURONS AND NETWORKS; COOPERATIVE PHENOMENA; DYNAMIC SYSTEMS AND OPTIMIZATION

**Background:** DYNAMICS AND BIFURCATION OF NEURAL NETWORKS; PHASE PLANE ANALYSIS OF NEURAL ACTIVITY

**Related Reading:** CHAINS OF COUPLED OSCILLATORS; COLLECTIVE BEHAVIOR OF COUPLED OSCILLATORS; COMPUTING WITH ATTRACTORS; COOPERATIVE PHENOMENA; ENERGY FUNCTIONS FOR NEURAL NETWORKS; PATTERN FORMATION, BIOLOGICAL

## References

- Aronson D.G., Ermentrout G.B., and Kopell N., 1990, Amplitude response of coupled oscillators. *Physica D*, 41:403–449.
- Ermentrout G.B., 1996, Type I Membranes, Phase Resetting Curves, and Synchrony. *Neural Computation*, 8:979–1001.
- Hodgkin A.L., 1948, The local electric changes associated with repetitive action in a non-medulated axon. *Journal of Physiology*, 107:165–181.
- Hodgkin A.L. and Huxley A.F., 1952, A quantitative description of membrane current and application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544.
- Hoppensteadt F.C. and Izhikevich E.M., 1997, *Weakly Connected Neural Networks*, Springer-Verlag, NY
- Hoppensteadt F.C. and Izhikevich E.M., 1996, Synaptic organizations and dynamical properties of weakly connected neural oscillators. II. Learning of phase information. *Biological Cybernetics*, 75:129–135.
- Izhikevich E.M., 1999, Class 1 Neural Excitability, Conventional Synapses, Weakly Connected Networks, and Mathematical Foundations of Pulse-Coupled Models, *IEEE transactions on neural networks*, 10:499–507
- Izhikevich E.M., 2000, Neural Excitability, Spiking, and Bursting, *International Journal of Bifurcation and Chaos* vol.10, n.3.
- Kuramoto Y., 1984, *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, New York.
- Morris C. and Lecar H., 1981, Voltage Oscillations in the Barnacle Giant Muscle Fiber, *Biophysical Journal*, 35:193–213.
- Rinzel J. and Ermentrout G.B., 1989, Analysis of Neural Excitability and Oscillations. In Koch C., Segev I. (eds) *Methods in Neuronal Modeling*, The MIT Press, Cambridge, Mass.
- Vicente C.J., Arenas A., and Bonilla L. L. (1996) On the short-term dynamics of networks of Hebbian coupled oscillators, *J. Phys. A*, L9-L16.