Dimension Reduction of Biological Neuron Models by Artificial Neural Networks *

Kenji Doya and Allen I. Selverston
doya@crayfish.ucsd.edu
Department of Biology
University of California, San Diego
La Jolla, CA 92093-0322, USA

December 28, 1992

Abstract

An artificial neural network approach for dimension reduction of dynamical systems is proposed and applied to conductance-based neuron models. Networks with bottleneck layers of continuous-time dynamical units could make a 2-dimensional model from the trajectories of the Hodgkin-Huxley model and a 3-dimensional model from the trajectories of a 6-dimensional bursting neuron model. Nullcline analysis of these reduced models revealed the bifurcations of the neuronal dynamics underlying firing and bursting behaviors.

1 Introduction

The dynamics of neural membrane potential is best described by conductance-based models, for example, the Hodgkin-Huxley (HH) model (Hodgkin and Huxley 1952) and its derivatives, which are usually fourth or higher order differential equation systems. The lower dimensional models that approximate the behaviors of these higher order models have been proposed for faster simulations, rigorous mathematical analyses, and better intuitive understanding of the dynamics. In this paper, we propose a new method for reducing the dimensionality of dynamical systems using artificial neural networks. We applied this method for constructing a 2-dimensional model that approximates the firing behavior of the 4-dimensional HH model and a 3-dimensional model that approximates a 6-dimensional model of bursting neurons recently proposed by Guckenheimer et al. (submitted).

In a standard conductance-based model, the dynamics of the membrane potential v is given by the following form.

$$C_m \dot{v} = I - \sum_j g_j a_j^{p_j} b_j^{q_j} (v - v_j), \tag{1}$$

*submitted to *Neural Computation*

where C_m is the membrane capacitance and I is externally injected current. The subscript j denotes one kind of ionic current and g_j and v_j are the maximal conductance and the reversal potential of the ionic current, respectively. The variables a_j and b_j are the normalized activation and inactivation of the ionic current, which represent the opening probabilities of the gating elements of the ionic channels, and the exponents p_j and q_j represent the number of the gating elements in a channel, which are usually integers between 0 and 4. The dynamics of activation and inactivation variables are typically given by the form

$$\dot{a} = k_a(v)(a_\infty(v) - a), \tag{2}$$

where $k_a(v)$ is a rate constant and $a_\infty(v)$ is a steady state activation or inactivation level, which is a sigmoidal function of the membrane potential v.

The HH model is the first and the best known model of this kind. It incorporates sodium, potassium and leak currents and makes a 4-dimensional differential equation system. As more kinds of ionic currents are discovered and characterized by physiological experiments, the corresponding conductance-based models become quite complex. For example, a recent model of a neuron in the stomatogastric ganglion of the crab is a 14-dimensional differential equation system (Buchholtz et al. 1992). Such complexity of models is inconvenient for mathematical analyses and numerical simulations, especially in multi-cellular or multi-compartmental studies.

2 Dimension Reduction

Trajectories of a dissipative dynamical system usually fall into some submanifold of the state space after some transient time. In such cases, it is convenient to define a lower dimensional coordinate system and a compact form of the vector field on this manifold. The dimension required for such a reduced state space depends on the dimensionality of the attractor solution. Simple limit cycle trajectories, such as the waveforms of tonically firing neurons, can be well approximated by 2-dimensional systems, whereas some of the complex limit cycle trajectories, such as those of bursting neurons, require 3-dimensional state space. Chaotic oscillation requires at least 3-dimensional state space and the dimension must be larger than the Lyapunov dimension of the attractor (Brown and Bryant 1991).

Dimension reduction of conductance-based neuron models is usually performed heuristically by finding constraints between the time courses of the variables (FitzHugh 1961; Rose and Hindmarsh 1989). Most often, the fastest activation variables are regarded as instantaneous functions of the membrane potential. The slower variables are grouped by the similarity of their time-courses and replaced by their linear combination. Recently, (Kepler et al. 1992) proposed a systematic dimension reduction scheme that utilizes a nonlinear transformation called "equivalent potential" and the Lagrange multiplier method, so that the characteristics of equilibrium points are preserved. However, their method cannot be applied to models that include calcium dynamics, which is essential in many models of bursting neurons (Rinzel and Lee 1987; Epstein and Marder 1990; Guckenheimer et al., submitted).

We propose an artificial neural network approach to this problem. Learning in artificial neural networks is a practical method for constructing a model from examples. We designed a special kind of recurrent network to generate a lower dimensional model from examples of the trajectories of a higher dimensional system.

2.1 Dynamical bottleneck network

Multilayer neural networks with a "bottleneck" layer have been used for compression of high dimensional data vectors, or nonlinear principal component analysis (DeMers and Cottrell 1993; Kramer 1991). This architecture is also suitable for applying the "teacher forcing" technique, which has been used to train a recurrent network as an autonomous oscillator (Doya and Yoshizawa 1989; Williams and Zipser 1989). Figure 1 shows the basic architecture of "dynamical bottleneck network" that we designed for dimension reduction of dynamical systems.

When we train the network, the trajectory $X^*(t) = (x_1^*(t), ..., x_n^*(t))$ of the original n-dimensional system is given to the input layer, compressed into an r-dimensional vector $Y(t) = (y_1(t), ..., y_r(t))$ in the bottleneck layer, and then reconstructed as an n-dimensional vector $X(t) = (x_1(t), ..., x_n(t))$ in the output layer (teacher forcing mode; Figure 1(a)). After the actual output $X(t)$ has become very close to the target trajectory $X^*(t)$ by error gradient descent learning, the output $X(t)$ is fed back to the input so that the network evolves as an r-dimensional autonomous system (autonomous mode; Figure 1(b)). Because our purpose is to reduce the dimensionality of state space, only the units in the bottleneck layer are designed to have dynamics and the other units are supposed to operate instantaneously, i.e. the output is determined as an instantaneous function of its input.

The mapping between the input layer and the bottleneck layer is mediated by a hidden layer of sigmoid units so that arbitrary nonlinear vector fields can be implemented (Irie and Miyake 1988). The output $H(t) = (h_1(t), ..., h_m(t))$ of this hidden layer (H-layer) is given as follows.

$$z_i(t) = \begin{cases} x_i^*(t) & \text{teacher forcing mode} \\ x_i(t) & \text{autonomous mode} \end{cases} \qquad (i = 1, ..., n), \tag{3}$$

$$h_i(t) = f\left(\sum_{j=1}^{n} w_{ij}^H z_j(t) + \sum_{j=1}^{p} w_{ij}^{H_S} s_j + b_i^H \right) \qquad (i = 1, ..., m), \tag{4}$$

where $f(x) = 1/(1 + e^{-x})$. The modulatory input $S = (s_1, ..., s_p)$ is provided when we modeled a vector field that changes with some parameter, for example, the level of current injection.

We employed a continuous-time model for the units in the bottleneck layer. Their dynamics is given by the differential equation

$$\tau_i \dot{y}_i(t) = -y_i(t) + \sum_{j=1}^{m} w_{ij}^Y h_j(t) + b_i^Y(t) \qquad (i = 1, ..., r). \tag{5}$$

The advantage of a continuous-time model is that each unit i can be assigned a distinct time constant τ_i so that events on the time scale near τ_i will be represented primarily by unit i. Furthermore, the time constants can be fine-tuned by the learning algorithm described below. In general, the dynamics of a recurrent network can go through bifurcations (Guckenheimer and Holmes 1983) in the course of learning. In such cases, the output error of the network can be increased abruptly by gradient descent algorithms and the learning equations can be numerically unstable (Doya 1992). Because the recurrence of our network is limited to

the local linear feedback in teacher forcing mode, there can be no problems concerning the bifurcation of the network dynamics as long as the time constants remain positive.

In the reduction of the HH model, a direct mapping between the bottleneck and the output layers was used in order to have a simple correspondence between the reduced state variables and the original biophysical variables. In this case, the output of the network was given as follows.

$$x_i(t) = f\left(\sum_{j=1}^{r} w_{ij}^X y_j(t) + b_i^X\right) \qquad (i = 1, ..., n). \tag{6}$$

In the reduction of the bursting neuron model, a hidden layer (G-layer) was necessary to implement the mapping between the bottleneck and the output layers. Its output $G(t) = (g_1(t), ..., g_l(t))$ was given by

$$g_i(t) = f\left(\sum_{j=1}^{r} w_{ij}^G y_j(t) + b_i^G\right) \qquad (i = 1, ..., l), \tag{7}$$

and the networks output was given as follows.

$$x_i(t) = \sum_{j=1}^{l} w_{ij}^X g_j(t) + b_i^X \qquad (i = 1, ..., n). \tag{8}$$

2.2 Learning algorithms

We used gradient descent learning algorithms to minimize the average error

$$E = \frac{1}{T}\int_0^T e(t)dt, \tag{9}$$

where T is the period of the target trajectory $X^*(t)$ and $e(t)$ is the instantaneous error

$$e(t) = \sum_{i=1}^{n} \frac{1}{2}\left(x_i(t) - x_i^*(t)\right)^2. \tag{10}$$

The error gradients with respect to the weights of the output layer and the G-layer were derived by the standard back-propagation algorithm (Rumelhart et al. 1986). The error gradients with respect to the weights and the time constants of the bottleneck layer and H-layer were derived using a continuous-time version of real-time recurrent learning algorithm (Doya and Yoshizawa 1989; Williams and Zipser 1989) as shown in the Appendix.

We tried two parameter update schemes. In the batch update scheme, the parameters were updated after running the network for period T by

$$w(k+1) - w(k) = -\varepsilon\frac{\partial E}{\partial w} = -\varepsilon\frac{1}{T}\int_0^T \frac{\partial e(t)}{\partial w}dt, \tag{11}$$

where k denotes the number of iterations. In the real-time update scheme, the parameters were updated concurrently by the running average $\Delta w(t)$ of the error gradient as follows.

$$\dot{w} = -\epsilon\Delta w(t), \qquad T\dot{\Delta w}(t) = -\Delta w(t) + \frac{\partial e(t)}{\partial w}. \tag{12}$$

In learning the HH model, both update schemes performed similarly. However, in learning the bursting model, the real-time update was favored because the batch update can be made only after a sequence of many spikes.

All biases b were changed by regarding them as weights from a unitary input. Because the time constants τ_i should always stay positive, it was convenient to use an exponential expression $\tau_i = e^{\sigma_i}$ and to perform gradient descent with respect to the exponent σ_i (Rowat and Selverston 1991). Since $\partial\tau_i/\partial\sigma_i = \tau_i$, this is equivalent to performing gradient descent on τ_i with the learning rate scaled by τ_i^2.

3 Reduction of the Hodgkin-Huxley Model

We tested if our method can reduce the 4-dimensional HH model to a 2-dimensional model. The HH model has the following form.

$$C_m\dot{v} = I - g_L(v - v_L) - g_{Na}m^3h\,(v - v_{Na}) - g_Kn^4(v - e_K), \tag{13}$$

where the subscript L denotes the leak current. The activation m and the inactivation h of the sodium current and the activation n of the delayed rectifier potassium current obey the differential equations of the form (2). This equation system was numerically solved with five levels of depolarizing current I = 0, 15, 30, 45, and 60 ($\mu A/cm^2$). The solution was quiescent at $I = 0$ and oscillatory at higher levels of I. These solutions were used as the target trajectory $X^*(t) = (v(t), m(t), h(t), n(t))$ after rescaling $v(t)$ to the range between 0 and 1.

We provided 8 units in the H-layer and one modulatory input $S = (s_1)$ that represented the level of current injection. One sweep of training consisted of learning of five different waveforms (100ms for the quiescent solution and 100 cycles for oscillatory solutions) associated with the modulatory input level s_1 = 0, 0.15, 0.3, 0.45, and 0.6 ($s_1 = I/100$). The connection weights were initialized randomly (mean=0, standard deviation=0.5). The time constants were initially set as $\tau_1 = 0.5$ms and $\tau_2 = 5$ms.

An example of the output of the network is shown in Figure 2 (a) and (b). After 80 sweeps of learning, the output of the network became very close to the target waveform under teacher forcing (Figure 2(a)). When the teacher forcing was cut off and the output $X(t)$ was fed back to the input, the network kept on oscillating autonomously on its 2-dimensional state space (y_1, y_2) and could simulate the 4-dimensional waveforms (v, m, h, n) fairly well (Figure 2(b)). The time constants of the bottleneck units were $\tau_1 = 0.15$ms and $\tau_1 = 5.0$ms after learning. It was found from the connection weights between the output and the bottleneck layers that the fast variable y_1 was mainly related with v and m and that the slow variable y_2 was related with h and n. Figure 2 (c) is the comparison of the firing frequency curves of the original and the reduced systems. Although the network was trained at only five levels of depolarizing current I (i.e. modulatory input level s_1), its firing frequency changed smoothly with the change of the parameter.

The purpose of model reduction is not only to decrease the amount of computation but also to provide a good insight into the dynamics. In two or three dimensional systems, we can graphically investigate the structure of the state space, for example, using the nullclines of the vector field. In a 2-dimensional state space, we can locate equilibrium points from the

intersections of two nullclines $\dot{y}_1(t) = 0$ and $\dot{y}_2(t) = 0$ and their stabilities can be checked by the slopes of the nullclines at the intersections.

Figure 3 illustrates the nullclines of the network model in autonomous mode. The nullcline of the fast variable $\dot{y}_1(t) = 0$ has an "N-shape", which is a major feature of the FitzHugh-Nagumo (FN) model (FitzHugh 1961; Nagumo et al. 1962). At a low level of depolarizing current (Figure 3 (a)), two nullclines intersect where the slope of the fast nullcline is positive and therefore the equilibrium point is stable. As the level of the current is increased (Figure 3 (b) and (c)), the slope of the fast nullcline at the intersection becomes negative. This makes the equilibrium point unstable and a stable limit cycle trajectory is made around the equilibrium point, that is shown by the thick line.

4 Reduction of a Bursting Neuron Model

The crustacean stomatogastric ganglion (STG) is a well identified neural network that controls the movement of the stomach. Some of the neurons in the STG are conditional bursters that endogenously alternate between "firing" and "resting" phases under the effect of neuromodulators (Selverston and Moulins 1987; Harris-Warrick et al. 1992). Recently, Guckenheimer et al. (submitted) proposed a conductance-based model of a conditionally bursting neuron in the STG. We applied our method to reduce their 6-dimensional model (GGH model) to a 3-dimensional model.

The GGH model incorporates transient potassium current (A current) (Connor et al. 1977), calcium activated potassium current (KCa current), and calcium current in addition to the currents of the HH model. The dynamics of the membrane potential v and the normalized intracellular calcium concentration c are given by the following equations.

$$C_m \dot{v} = I - g_L(v - v_L) - g_{Na} m(v)^3 h\,(v - v_{Na}) - g_K n^4 (v - v_K)$$
$$- g_A a(v)^3 b\,(v - v_K) - g_{KCa} \frac{c}{0.5 + c}(v - v_K) - g_{Ca} \frac{z}{0.5 + c}(v - v_{Ca}), \tag{14}$$

$$\dot{c} = \rho\,(-k_{Ca} \frac{z}{1 + 2c}(v - v_{Ca}) - c). \tag{15}$$

The dynamics of sodium current inactivation h, delayed-rectifier potassium current activation n, A current inactivation b, and calcium current activation z are given by the form (2). Sodium current activation m and A current inactivation a are regarded as instantaneous functions of the membrane potential v. Therefore it is already a slightly reduced conductance-based model.

The input and the output layers had 6 units that represent the state of the GGH model $X = (v, c, n, h, z, b)$. The bottleneck layer had $r = 3$ units and their time constants were set initially as $\tau_1 = 0.5$ms, $\tau_2 = 10$ms, and $\tau_3 = 200$ms. The G-layer and the H-layer had $l = m = 6$ units.

Figure 4 shows an example of the bursting waveforms of the reduced model after training for 1500 cycles (period $T = 984.2$ms). When the network was run with teacher forcing (Figure 4 (a)), the output $X(t)$ (solid lines) were very close to the target waveform $X^*(t)$ (dotted lines). In autonomous mode (Figure 4(b)), although the waveforms were considerably

distorted, the 3-dimensional network model could simulate at least the qualitative aspects of bursting trajectory. The time constants were tuned to $\tau_1 = 0.31$ms, $\tau_2 = 5.1$ms and $\tau_3 = 210$ms by learning. The mapping between the reduced and the original variables was not so simple as in the case of HH model, though it could be seen that the fast component y_1 is mainly correlated with the membrane potential v and the slowest component y_3 roughly represents the calcium concentration c.

Figure 5 (a) shows the three nullclines and the bursting trajectory in the 3-dimensional state space of the network model. The fast nullcline ($\dot{y}_1 = 0$, light grey) forms an arch with a U-shaped cross-section. The medium speed nullcline ($\dot{y}_2 = 0$, medium grey) is approximately a vertical plane. The slow nullcline ($\dot{y}_3 = 0$, dark grey) is almost horizontal plane in the lower left part and then bends rapidly upwards into a roughly vertical plane. The trajectory spirals up the right-hand side of the arch of the fast nullcline in the firing phase, then travels leftward slowly between the narrow space between the fast and the medium nullclines in the resting phase, and eventually comes back to the spiral zone from the lower left part of the arch.

Because y_3 was designed to change very slowly, we can investigate the short term behavior of the system by taking 2-dimensional slices of the state space at some levels of y_3, as shown in Figure 5 (b), (c) and (d). At a low level of y_3 (Figure 5(b)), there is only one equilibrium point (α) in the y_1y_2 subspace. This equilibrium is unstable since it is located in the folded region of the fast nullcline $\dot{y}_1 = 0$ and there is a limit cycle around it, which corresponds to firing activity. As the slow variable y_3 is increased, the fast nullcline $\dot{y}_1 = 0$ makes a new intersection with the medium nullcline $\dot{y}_2 = 0$ and two more equilibrium points (β and γ) emerge (Figure 5 (c)). One of the new equilibria β is stable and the other γ is a saddle point that divides the 2-dimensional state space into two attractor basins. Therefore at this level of y_3, there are two distinct behaviors, one is firing with the limit cycle around α and another is resting at the equilibrium β. These two quasi-stable attractors are located on opposite sides of the nullcline $\dot{y}_3 = 0$ so that the slow variable y_3, which roughly corresponds to the intracellular calcium concentration, increases when the neuron is firing and decreases when it is resting. When y_3 is further increased, the equilibria α and γ collide and disappear (Figure 5 (d)) and the attractor point β dominates the 2-dimensional state space. This means that the firing activity is stopped and the trajectory goes toward the equilibrium γ through a narrow channel between the nullclines $\dot{y}_1 = 0$ and $\dot{y}_2 = 0$. Since γ is on the side where $\dot{y}_3 < 0$, y_3 then decreases until the equilibrium γ disappears at a low level of y_3.

5 Discussion

5.1 What was found by the network models

The low dimensional models that were constructed by the learning of artificial neural networks share several characteristics with the neuron models that were constructed by human experts.

In the 2-dimensional model of firing neuron, the nullcline for the fast component had an "N-shape" or negative resistance curve. As the level of depolarizing current is increased, the intersection of the two nullclines shifts into the negative slope region of the fast nullcline

and the neuron begins firing. This is the basic feature of the FN model (FitzHugh 1961; Nagumo et al. 1962). However, the slower nullcline of our network model was not a straight line as in the FN model but was curved in the hyperpolarized region. A similar curve in the slower nullcline was proposed by Hindmarsh and Rose (1982) for a model of firing with long inter-spike intervals.

In the 3-dimensional model of the bursting neuron, coexistence of two attractors at an intermediate level of the slowest variable, one resting and one firing, was essential. These attractors are located on different sides of the slowest nullcline and vanish at either extremity of the slowest variable. This qualitative mechanism of bursting has been proposed by Hindmarsh and Rose (1984), although their cubic form of vector field is very different from the vector field of our network model.

Our results confirm the plausibility of these former models and also suggest the possibility that artificial neural networks can find novel mechanisms of neuronal function that have not been found by mathematical biologists.

5.2 Technical issues

Although teacher forcing avoids several problems while learning, it can cause a problem when we run the network autonomously. Even if the error $X(t) - X^*(t)$ is very small under teacher forcing, the existence and the stability of the autonomous solution is not guaranteed. This is not usually a problem in simple tasks such as learning sinusoidal oscillations (Doya and Yoshizawa 1989) and modeling the HH model. In learning bursting trajectories, the network often remained resting or tonically firing in autonomous mode, even if the error was very small ($E \simeq 10^{-4}$) under teacher forcing. We are experimenting with several schemes that enforce the maintenance of the teacher-forced trajectory in autonomous mode.

In the reduction of GGH model, we used a single bursting pattern for training. However, one of the most interesting issues in bursting neuron models is the change of the oscillatory modes with the change of parameters, such as the amplitudes of ionic conductances (Epstein and Marder 1990; Guckenheimer et al. submitted). Another important point we have yet to consider is the effects of synaptic input on the neuronal dynamics. Whether the network model can find a low dimensional model of modulatory effects and synaptic inputs is the subject of future studies.

A feed-forward network with tapped delay-lines in the input layer has been used for modeling dynamical systems (Widrow 1985). This approach was not suitable for our purpose of dimension reduction because a very long array of delay units is required to model a dynamical system with multiple time scales. For example, in order to model a bursting neuron, the network must capture the fast spikes in the time scale less than 1ms and also have to respond to the very slow bursting rhythm in the time scale of seconds. Another common approach is the use of a fully connected recurrent neural network with random initial connection weights and uniform time constants (Doya and Yoshizawa 1989; Williams and Zipser 1989). . However, error gradient descent is not as reliable a strategy for recurrent networks as it is for feed-forward networks (Kuan et al. 1990; Doya 1992). Our successful results depend on the special architecture: the network is nearly feed-forward during training and the continuous-time units are provided with different time constants.

6 Conclusions

The dynamical bottleneck network approach proposed in this paper provides a general method for constructing a low dimensional dynamical system from specific trajectories of a high dimensional dynamical system. Analysis of the phase-portraits of the lower dimensional system gives deeper insight into the dynamics of the higher dimensional system. In particular, the dynamics of a complex, 6-dimensional model of a bursting neuron could be approximated by a 3-dimensional network model and it enabled us to investigate the dynamical mechanisms underlying the alternation between resting and firing phases.

Acknowledgments

We thank Peter Rowat for his detailed comments on the manuscript. We are grateful to Gary Cottrell, Dave DeMers, Fu-Sheng Tsung, and Mike Casey for helpful discussions. This research was supported by the grant from Office of Naval Research N00014-91-J-1720.

References

Brown, R. and Bryant, P. 1991. Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Physical Review A* **43**, 2787–2806.

Buchholtz, F., Golowasch, J., Epstein, I. R., and Marder, E. 1992. Mathematical model of an identified stomatogastric ganglion neuron. *Journal of Neurophysiology* **67**, 332–340.

Connor, J. A., Walter, D., and McKown, R. 1977. Neural repetitive firing, modifications of the Hodgkin-Huxley axon suggested by experimental results from crustacean axons. *Biophysical Journal* **18**, 81–102.

DeMers, D. and Cottrell, G. 1993. Non-linear dimensionality reduction. In Giles, C. L., Hanson, S. J., and Cowan, J. D., editors, *Advances in Neural Information Processing Systems 5*. Morgan Kaufmann, San Mateo, CA.

Doya, K. 1992. Bifurcations in the learning of recurrent neural networks. *Proceedings of 1992 IEEE International Symposium on Circuits and Systems* **6**, 2777–2780.

Doya, K. and Yoshizawa, S. 1989. Adaptive neural oscillator using continuous-time back-propagation learning. *Neural Networks* **2**, 375–386.

Epstein, I. R. and Marder, E. 1990. Multiple modes of a conditional neural oscillator. *Biological Cybernetics* **63**, 25–34.

FitzHugh, R. 1961. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal* **1**, 445–466.

Guckenheimer, J., Gueron, S., and Harris-Warrick, R. M. The dynamics of a conditionally bursting neuron. (submitted).

Guckenheimer, J. and Holmes, P. 1983. *Nonlinear Oscillation, Dynamical Dystems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, NY.

Harris-Warrick, R. M., Marder, E., Selverston, A. I., and Moulins, M. 1992. *Dynamic Biological Networks — The Stomatogastric Nervous System*. MIT Press, Cambridge, MA.

Hindmarsh, J. L. and Rose, R. M. 1982. A model of the nerve impulse using two first-order differential equations. *Nature* **296**, 162–164.

Hindmarsh, J. L. and Rose, R. M. 1984. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London. Series B* **211**, 87–102.

Hodgkin, A. L. and Huxley, A. F. 1952. A quantitative description of membrane currents and its application to conduction and excitation in nerve. *Journal of Physiology* **117**, 500–544.

Irie, B. and Miyake, S. 1988. Capabilities of three-layered perceptrons. *Proceedings of International Conference on Neural Networks* **1**, 641–648.

Kepler, T. B., Abbott, L. F., and Marder, E. 1992. Reduction of conductance-based neuron models. *Biological Cybernetics* **66**, 381–387.

Kramer, M. A. 1991. Nonlinear principal component analysis using autoassociative neural networks. *American Institute of Chemical Engineers Journal* **37**, 233–243.

Kuan, C., Hornik, K., and White, H. 1990. Some convergence results for learning in recurrent neural networks. Discussion Paper 90-42, UCSD Department of Economics.

Nagumo, J., Arimoto, S., and Yoshizawa, S. 1962. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE* **50**, 2061–2070.

Rinzel, J. and Lee, Y. S. 1987. Dissection of a model for neuronal parabolic bursting. *Journal of Mathematical Biology* **25**, 653–675.

Rose, R. M. and Hindmarsh, J. L. 1989. The assembly of ionic currents in a thalamic neuron, I-III. *Proceedings of the Royal Society of London. Series B* **237**, 267–334.

Rowat, P. F. and Selverston, A. I. 1991. Learning algorithms for oscillatory networks with gap junctions and membrane currents. *Network* **2**, 17–41.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature* **323**, 533–536.

Selverston, A. I. and Moulins, M. 1987. *The Crustacean Stomatogastric System.* Springer-Verlag, New York, NY.

Widrow, B. and Stearns, S. D. 1985. *Adaptive Signal Processing.* Prentice Hall, Englewood Cliffs, NJ.

Williams, R. J. and Zipser, D. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation* **1**, 270–280.

Appendix: Error gradient calculation

The error gradients of the weights in the output layer and the G-layer and those of the output of the bottleneck layer were derived by the standard back-propagation algorithm (Rumelhart et al. 1986) . In the following, g' and h' denote the derivatives of the sigmoid function at the output values $g = f(\ )$ and $h = f(\ )$, i.e. $g' = g(1-g)$ and $h' = h(1-h)$.

$$\frac{\partial e(t)}{\partial w_{ij}^X} = (x_i(t) - x_i^*(t))\, g_j(t),$$

$$\frac{\partial e(t)}{\partial g_i(t)} = \sum_{k=1}^{n} (x_k(t) - x_k^*(t))\, w_{ki}^X,$$

$$\frac{\partial e(t)}{\partial w_{ij}^G} = \frac{\partial e(t)}{\partial g_i(t)}\, g_i'(t) y_j(t),$$

$$\frac{\partial e(t)}{\partial y_i(t)} = \sum_{k=1}^{l} \frac{\partial e(t)}{\partial g_k(t)}\, g_k'(t) w_{ki}^G.$$

The effect of a small change in a parameter of the bottleneck layer and the H-layer was estimated by the solution of the following variational equations (Doya and Yoshizawa 1989; Williams and Zipser 1989) .

$$\tau_i \frac{\dot{\partial y_i}}{\partial w_{ij}^Y}(t) = -\frac{\partial y_i}{\partial w_{ij}^Y}(t) + h_j(t), \qquad \frac{\partial y_i}{\partial w_{ij}^Y}(0) = 0,$$

$$\tau_i \frac{\dot{\partial y_i}}{\partial \tau_i}(t) = -\frac{\partial y_i}{\partial \tau_i}(t) - \dot{y}_i(t), \qquad \frac{\partial y_i}{\partial \tau_i}(0) = 0,$$

$$\tau_k \frac{\dot{\partial y_k}}{\partial w_{ij}^H}(t) = -\frac{\partial y_k}{\partial w_{ij}^H}(t) + w_{ki}^Y h_i'(t) x_j(t), \qquad \frac{\partial y_k}{\partial w_{ij}^H}(0) = 0.$$

From their solutions, we have the error gradients as follows.

$$\frac{\partial e(t)}{\partial w_{ij}^Y} = \frac{\partial e(t)}{\partial y_i(t)} \frac{\partial y_i}{\partial w_{ij}^Y}(t),$$

$$\frac{\partial e(t)}{\partial \tau_i} = \frac{\partial e(t)}{\partial y_i(t)} \frac{\partial y_i}{\partial \tau_i}(t),$$

$$\frac{\partial e(t)}{\partial w_{ij}^H} = \sum_{k=1}^{r} \frac{\partial e(t)}{\partial y_k(t)} \frac{\partial y_k}{\partial w_{ij}^H}(t).$$

Figure Legends

Figure 1

The architecture of dynamical bottleneck network. (a) Teacher forcing mode. (b) Autonomous mode.

Figure 2

Comparison of the solutions of the HH model and the reduced model. In (a) and (b), the dotted lines show the target trajectory $X^*(t) = (v, m, h, n)$ and the solid lines show the reconstructed output $X(t)$ and the state of the bottleneck layer $Y(t)$. The ordinate is in an arbitrary unit. The level of depolarizing current was $I = 60(\mu A/cm^2)$ and the corresponding modulatory input to the network was $s_1 = 0.6$. (a) After learning for 80 sweeps with teacher forcing. (b) Oscillation in autonomous mode. (c) The curren-frequency curves of the HH model (dotted line) and the 2-dimensional network model (solid line). The modulatory parameter s_1 of the network was set as $s_1 = I/100$.

Figure 3

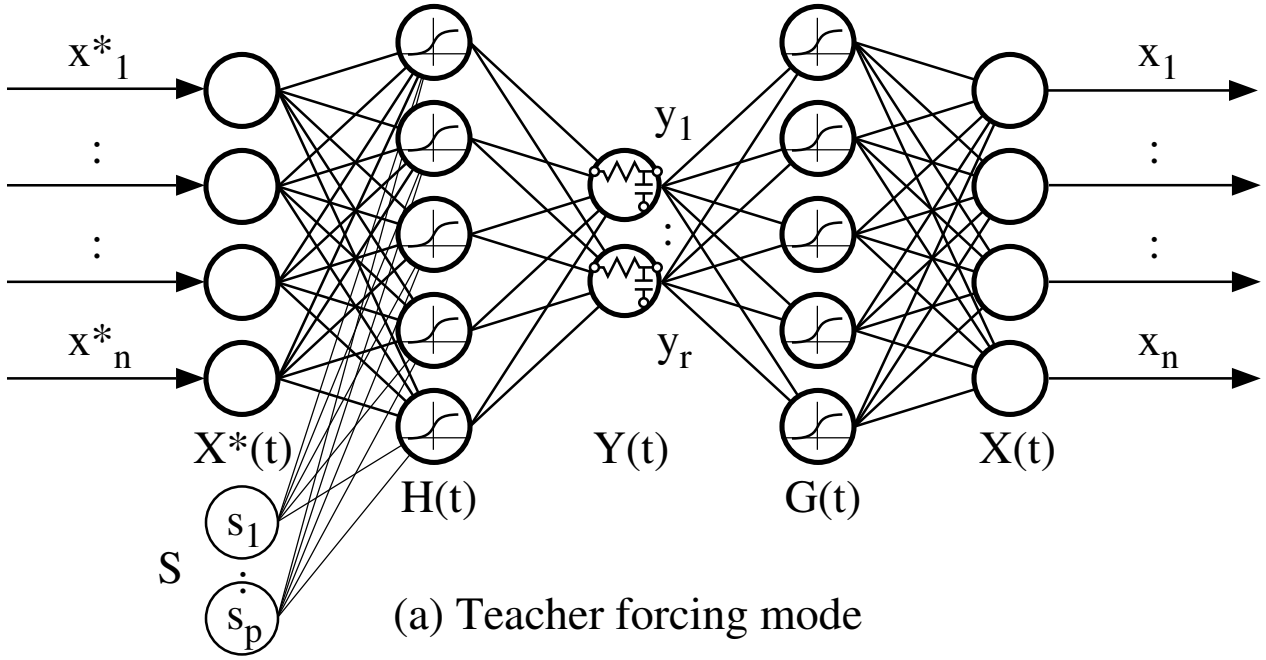
The nullclines of the 2-D model of the HH equations with different levels of the modulatory input s_1 that represent the level of depolarizing current I. The nullclines of the fast variable $\dot{y_1} = 0$ are shown by the solid lines and those of the slow variable $\dot{y_2} = 0$ are shown by the dotted lines. (a) $s_1 = 0$. The slope of the fast nullcline $\dot{y_1} = 0$ at the intersection with the slow nullcline $\dot{y_2} = 0$ is positive so that the equilibrium is stable. (b) $s_1 = 0.3$. The equilibrium becomes unstable and a limit cycle emerges around it. (c) $s_1 = 0.6$.

Figure 4

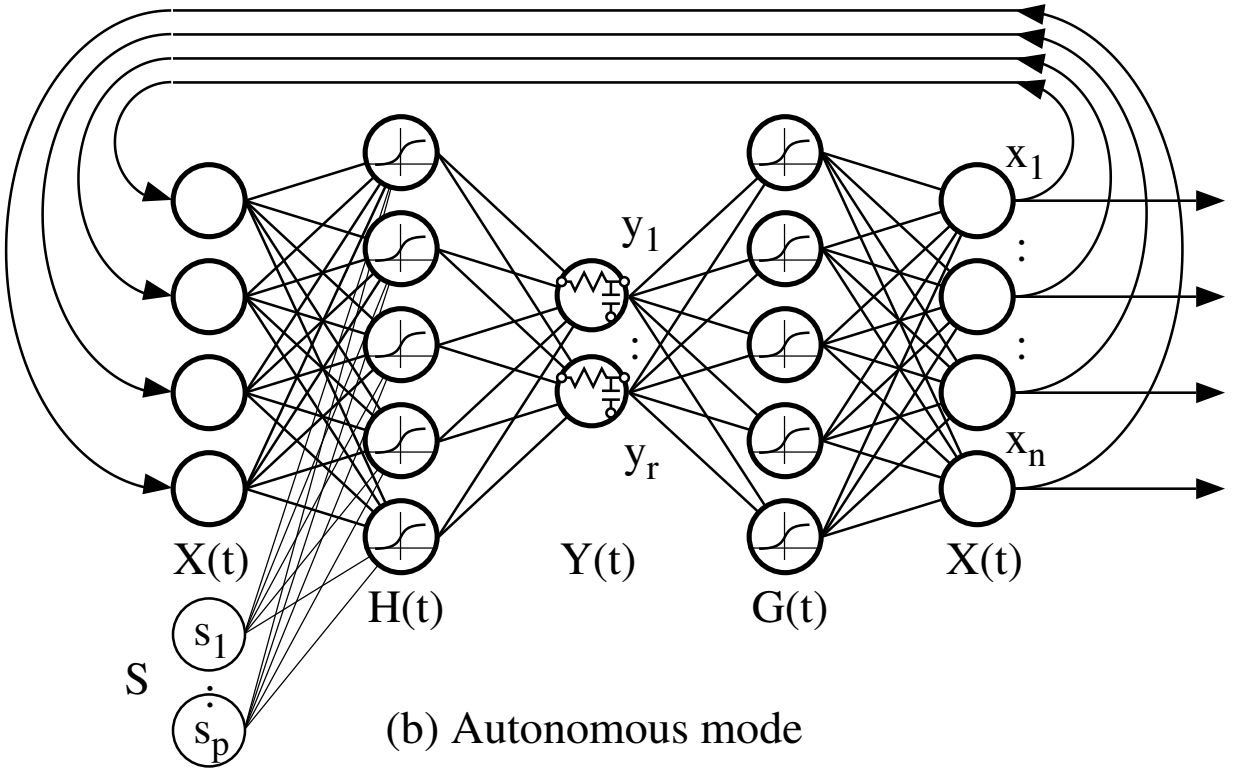
Comparison of the waveforms of the bursting neuron models. The dotted lines show the trajectory of the GGH model $X = (v, c, n, h, z, b)$ and the solid lines show the output X and the reduced state Y of the 3-dimensional network model. (a) After 1500 cycles of learning with teacher forcing. The output $X(t)$ is very close to the target $X^*(t)$. (b) Bursting trajectory of 3-dimensional system in autonomous mode.

Figure 5

The nullclines and the trajectory of the 3-dimensional bursting model. (a) The nullcline of the fast variable $\dot{y_1} = 0$ is shown in light grey, the medium variable $\dot{y_2} = 0$ in medium grey, and the slowest variable $\dot{y_3} = 0$ in dark grey. The thick line shows the bursting trajectory. (b) (c) and (d) Two-dimensional sections ($y_1 y_2$ plane) of the nullclines at different levels of the slow variable y_3. The nullclines $\dot{y_1} = 0$, $\dot{y_2} = 0$, and $\dot{y_3} = 0$ are shown by thick, thin, and dotted lines respectively. (b) $y_3 = -0.8$. (c) $y_3 = -0.4$. (d) $y_3 = 0$. See text for details.

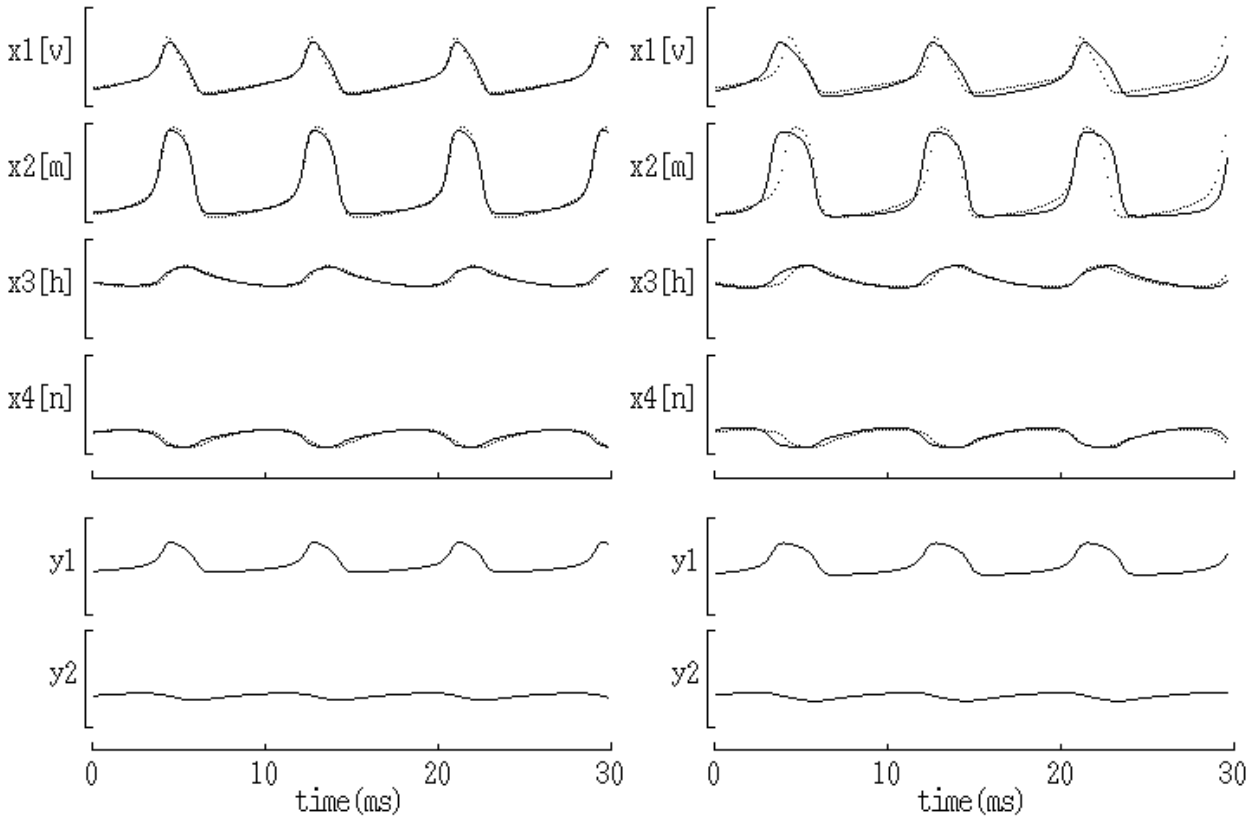


(a) Teacher forcing mode

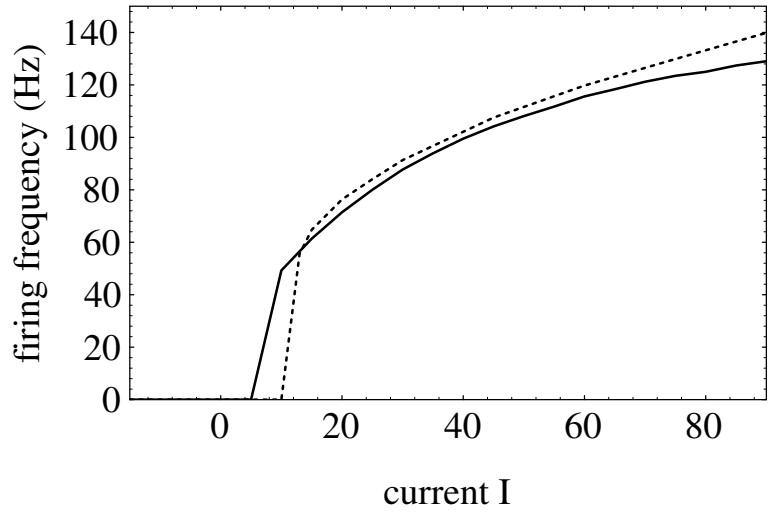


(b) Autonomous mode

Figure 1

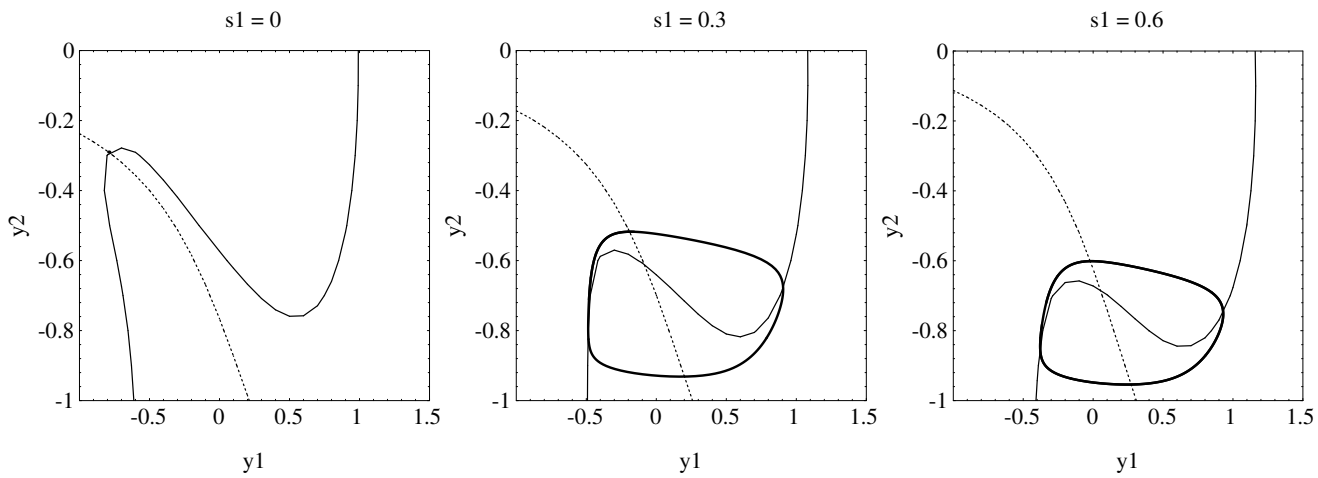


(a) teacher forcing mode (b) autonomous mode



(c)

Figure 2



(a) (b) (c)

Figure 3

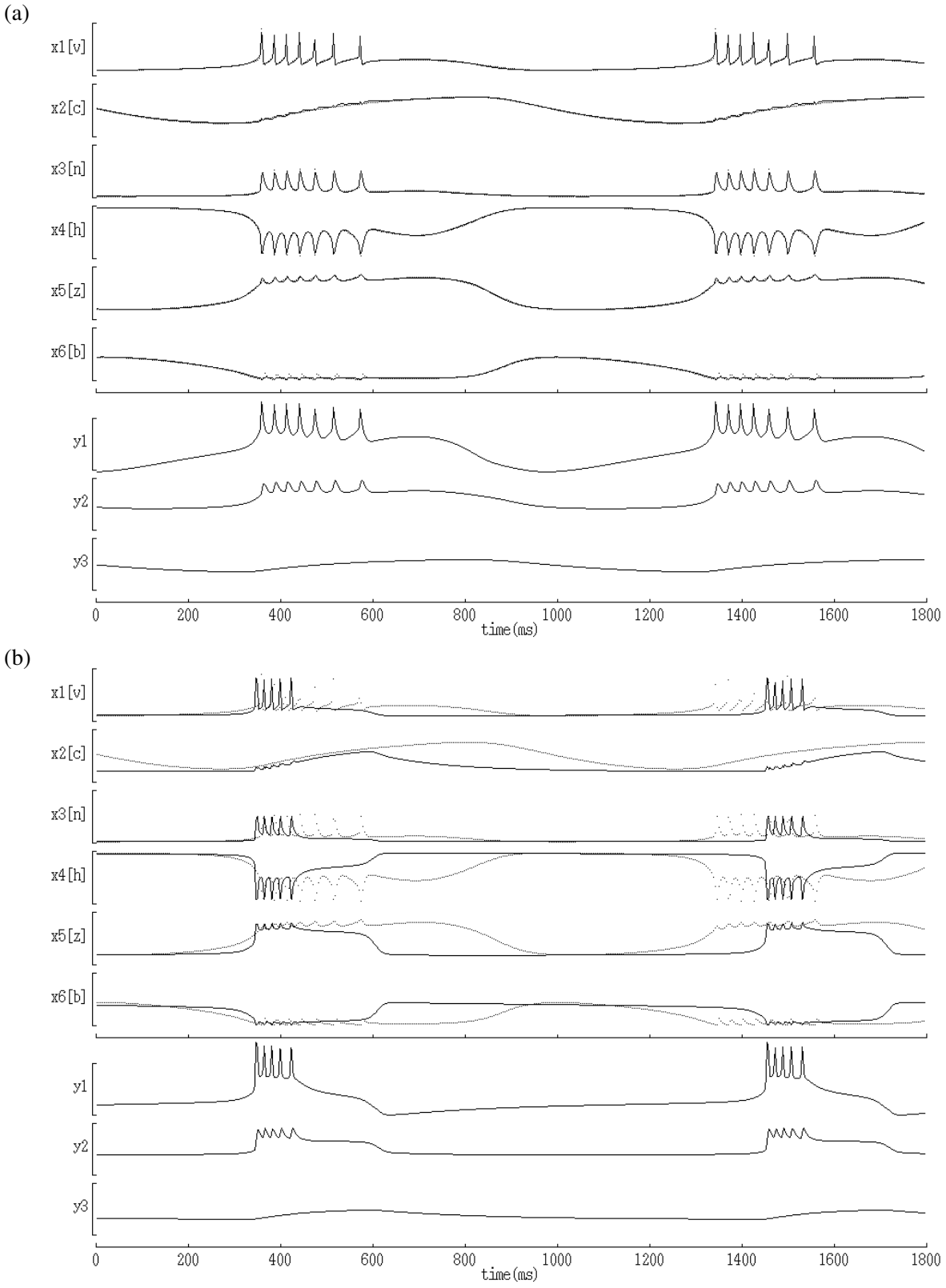
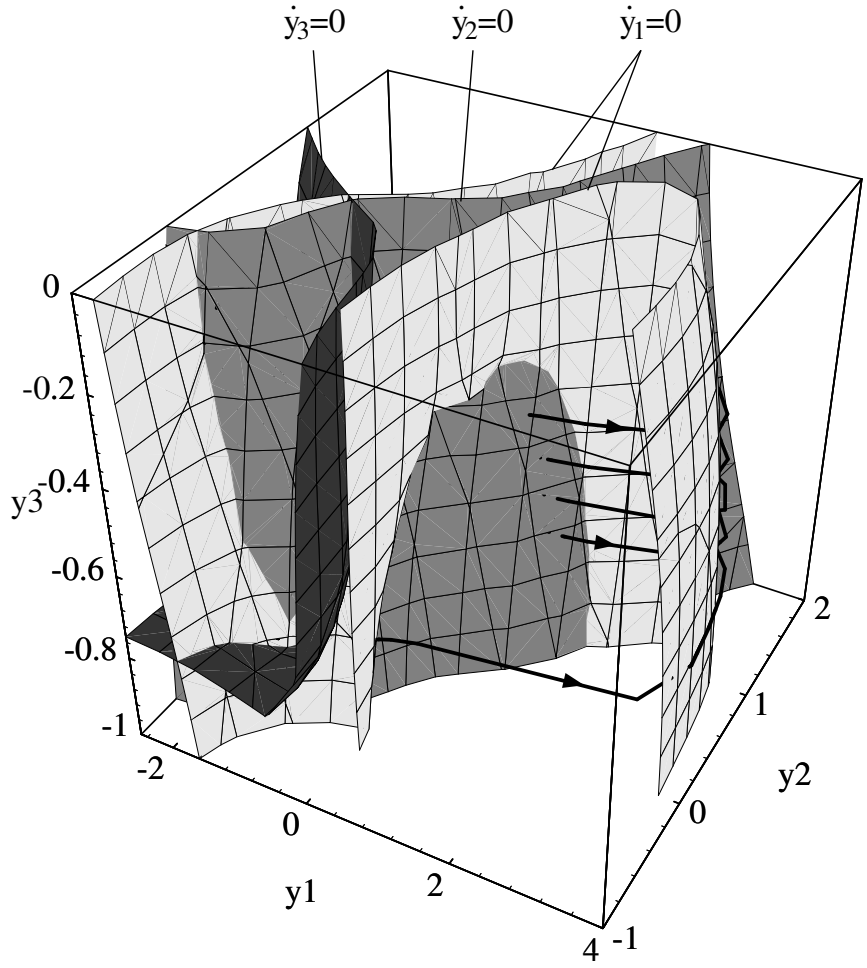
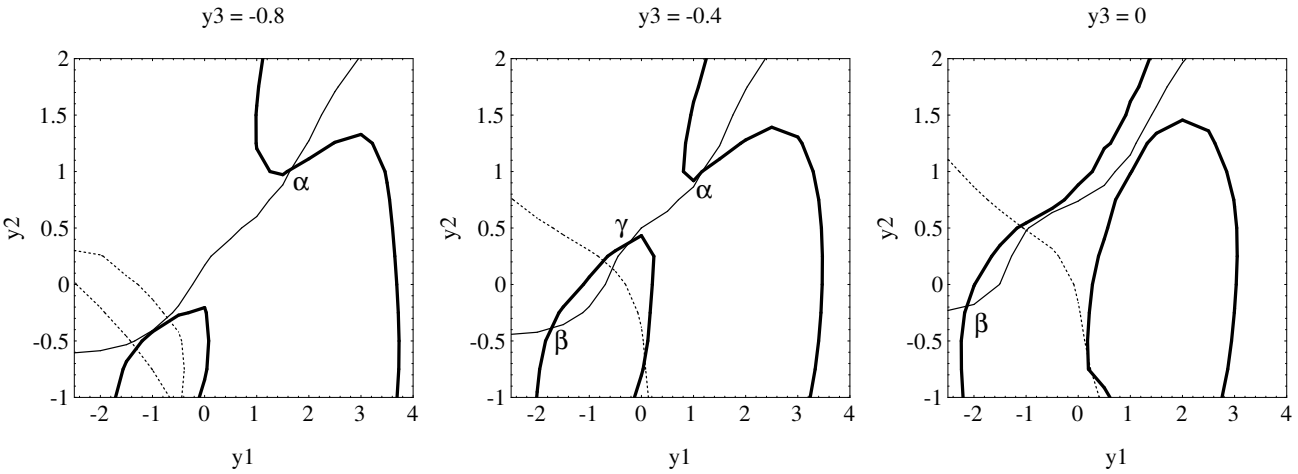

(a)
x1[v]
x2[c]
x3[n]
x4[h]
x5[z]
x6[b]
y1
y2
y3
0
200
400
600
800
1000
1200
1400
1600
1800
time(ms)
(b)
x1[v]
x2[c]
x3[n]
x4[h]
x5[z]
x6[b]
y1
y2
y3
0
200
400
600
800
1000
1200
1400
1600
1800
time(ms)

Figure 4



(a)



(b) (c) (d)

Figure 5