22

Nonlinear dynamics and neuronal networks : proceedings of the 63rd W.-E.-Heraeus-Seminar, Friedrichsdorf, 1990 / ed. by H. G. Schuster. – Weinheim ; New York ; Basel ; Cambridge : VCH, 1991
(Nonlinear systems ; Vol. 2)
ISBN 3-527-28342-0 (Weinheim ...) Gb.
ISBN 1-56081-167-6 (New York) Gb.

Small Nervous Systems and Neural Network Models

by *D. Kleinfeld, H. J. Chiel,* and *H. Sompolinsky*[1]

*** AT&T Bell Laboratories, Murray Hill, NJ 07974.**
† Department of Biology, Case Western Reserve University, Cleveland, OH 44106.

Abstract

The relation between experimental observation and the results of phenomenological theories is examined for three small, invertebrate nervous systems: (1) Recurrent circuits, constructed in vitro, that exhibit bistable output behavior. These circuits act as neuronal memory circuits. (2) The recurrent network that controls escape swimming in the mollusc Tritonia. (3) The feedforward network that meditates the local bending reflex in the leech. We suggest that the dynamic behavior of these networks can be understood by simplified models that treat neurons as threshold or sigmoid-like units and synapses as linear elements. We discuss simple rules that may link the underlying synaptic connectivity to the pattern of output activity in these circuits.

Introduction

Nervous systems are capable of producing spatiotemporal patterns of electrical activity. These patterns may be controlled by inputs from higher nervous centers or external stimuli. Alternatively, the patterns may be sustained as a consequence of internal interactions. Neurons and the connections between neurons have complicated physiological properties. It is a matter of continued controversy as to how much of this complication must be taken into account in understanding the basis for patterned output. We approach this issue by direct comparison between phenomenological models of nervous activity and the results of experiments on small nervous systems.

---

1 On leave from the Racah Institute of Physics, Hebrew University, Jerusalem, 91904 Israel.

The models we consider treat neurons as threshold devices that integrate over individual action potentials and treat the connections between neurons as linear elements. This description allows one to define rules that relate the form of the output pattern to the underlying synaptic connectivity. In the present work, we examine the application of such rules to a number of small nervous system.

In vitro circuits with bistable outputs

The first topic we consider is the theory and experimental characterization of circuits that exhibit bistable output behavior (Kleinfeld *et al.*, 1990a). These circuits have two possible persistent output states and encompass the old idea of 'reverberating loops'. Introduction of a brief stimulus pulse, *i.e.*, just a few action potentials, can trigger sustained firing in the output of one or neurons in the circuit (Fig. 1). Bistable circuits are the most rudimentary example of networks that have multiple, persistent output states. Such systems have been proposed as models of associative memory (Hopfield, 1982; 1984). It would be useful to understand the dynamics of a minimal circuit that can perform this fundamental neurobiological task.

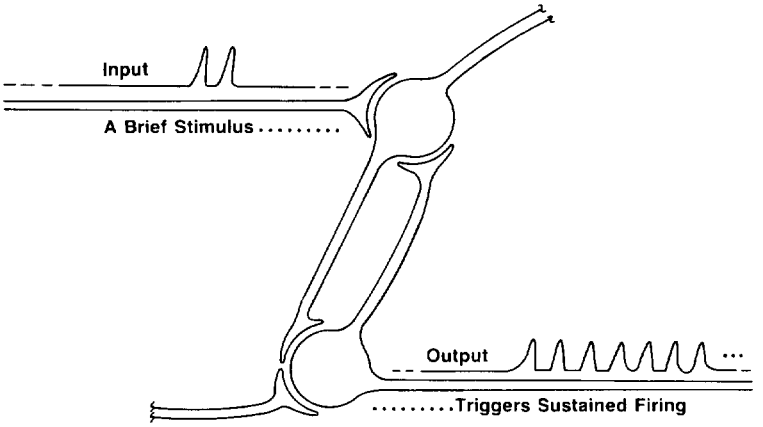

Figure 1. Schematic of a circuit that functions as a short-term memory. An input of only a few action potentials is capable of changing the output from quiescence to spiking for a relatively long period of time.

Theory. Bistable output activity in populations of neurons has been studied by Harth *et al.* (1970), Wilson & Cowan (1972) and Hopfield (1984). Bistability can, in principle, be realized in a circuit with only two neurons. These and other studies of simple circuits suggest that the neurons and the interactions between these neurons, should approximately fulfill the following criteria in order for circuits of two neurons to exhibit bistable output activity (Fig. 2):

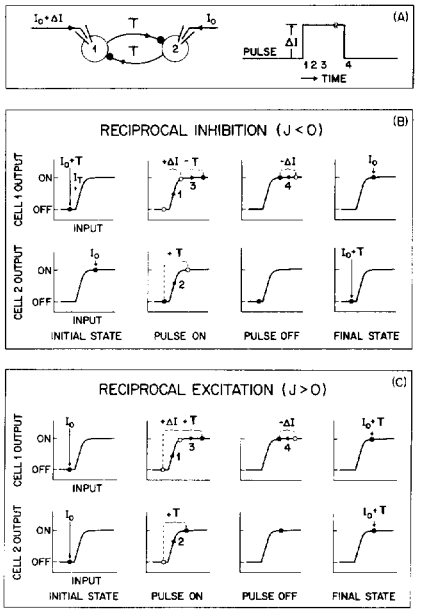

Figure 2. Idealized bistable behavior for circuits of two neurons connected by slow, reciprocal connections. **(A)** Schematic of the circuit. The connections have strength T and a decay-time τ_L. The current injected into each cell consists of two components, a constant bias current I_0 and a transient component ΔI. **(B)** The dynamic properties of a circuit with reciprocal, inhibitory connections. The input-output relation for each neuron is taken to be a saturating non-linear function characterized by a threshold I_T. The width was ignored in our analysis. A transition between states ON/OFF and OFF/ON is elicited by a current-pulse ΔI. **(C)** The dynamic properties of a circuit with reciprocal, excitatory connections. A transition from the quiescent state to the active state is elicited by a current-pulse. Adapted from Kleinfeld *et al.* (1990a).

1. To clearly identify the output states, the activity of the neurons must have discernible quiescent and active levels; we refer to these levels as 'OFF' and 'ON', respectively. Thus the rate of firing of the neurons must change in a non-linear manner as a function of the input current (Fig. 2B).

2. To provide a feedback pathway to stabilize the two output states, the neurons are connected by reciprocal connections of the *same* sign (Fig. 2A). Thus there are two possible circuits, one has reciprocal inhibitory connections and the other has reciprocal excitatory connections.

3. To provide the temporal integration that allows each neuron to maintain a constant level of activity, the duration of the postsynaptic response is long compared with the period between action potentials in the presynaptic cell. Note that temporal integration in our small circuits plays the role of population averaging in large circuits.

The input to each neuron contains contributions from both its presynaptic partner and external sources. A dynamical system that realizes the above features is described by

$$V_1(t+\tau_S) = \Theta\left(T_{12}\int_{t-\tau_L}^{t}\frac{dt'}{\tau_L}V_2(t') + I_{01} - I_{T1}\right)$$

$$V_2(t+\tau_S) = \Theta\left(T_{21}\int_{t-\tau_L}^{t}\frac{dt'}{\tau_L}V_1(t') + I_{02} - I_{T2}\right) \qquad [1]$$

where the outputs V_1 and V_2 vary between 0 (quiescent) and 1 (maximally firing), T_{12} and T_{21} are the values of the synaptic strength, taken as the value of the current that enters the postsynaptic cell when the presynaptic cell is firing at its steady-state rate, I_{01} and I_{02} are external bias currents and I_{T1} and I_{T2} are the intrinsic threshold levels of the neuron. The step function, $\Theta(x)$, is defined by

$$\Theta(x) = \begin{cases} +1; & x > 0 \\ 0; & x \leq 0 \end{cases} \qquad [2]$$

The time τ_S corresponds to the period between action potentials and the τ_L corresponds to the time-constant of the synaptic integration. In the present study $\tau_L >> \tau_S$.

The steady-state solutions of the above system (Eqs. **1,2**), in which neurons are treated as threshold elements, correspond to the stable states of the system. They are given by

$$V_1 = \Theta\left(T_{12}V_2 + I_{01} - I_{T1}\right)$$

$$V_2 = \Theta\left(T_{21}V_1 + I_{02} - I_{T2}\right) \qquad [3]$$

We consider first the case of a circuit with reciprocal inhibitory connections. There are four possible output states: OFF/OFF, ON/OFF, OFF/ON and ON/ON. The output of a circuit may be stable in only one state, *i.e.*, monostable, or it may be stable in both the states ON/OFF and OFF/ON, *i.e.*, bistable (Fig. 3). The ranges of bias currents for which bistable behavior takes is given by $I_{T1} \leq I_{01} \leq I_{T1} - T_{12}$ and $I_{T2} \leq I_{02} \leq I_{T2} - T_{21}$ (Fig. 2B). A transition between the stable states is initiated by injecting a pulse of current, $I_0 \rightarrow I_0 + \Delta I$, into the quiescent neuron (cell 1 in Fig. 2B). Similarly, bistable output can occur in the circuit with reciprocal excitation when the bias currents satisfy $I_{T1} - T_{12} \leq I_{01} \leq I_{T1}$ and $I_{T2} - T_{21} \leq I_{02} \leq I_{T2}$ (Fig. 2C). Bistable output in both circuits results from the non-linear

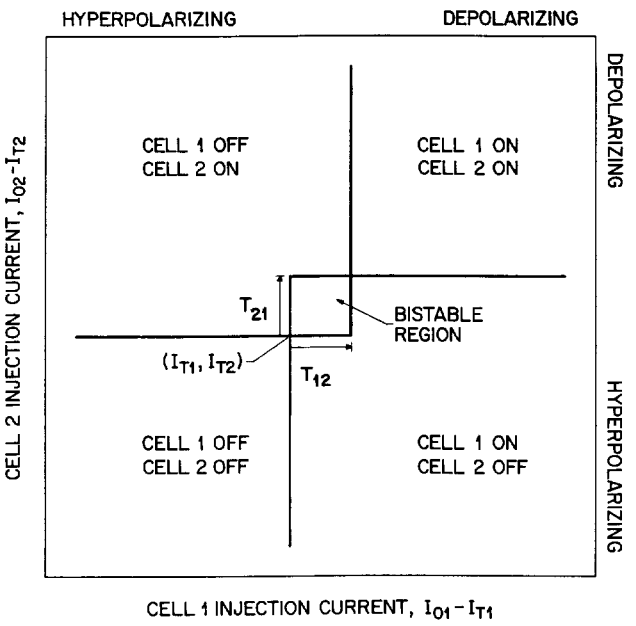

Figure 3. Analysis of an idealized circuit of two neurons that are connected by reciprocal inhibitory connections. The stable output states are shown as a function of the relative bias current into each cell, $I_{01} - I_{T1}$ and $I_{02} - I_{T2}$ respectively. The synaptic input from cell 2 to cell 1 is T_{12} and the input from cell 1 to cell 2 is T_{21}. When the bias currents lie within the bistable region the output of the circuit is stable in both the state ON/OFF and the state OFF/ON. A transient, external perturbation to the circuit can cause a transition between these states. Adapted from Kleinfeld *et al.* (1990a).

firing characteristics of each neuron and the feedback between the two neurons. The simplicity of our circuits allows aspects of the expected behavior to be tested experimentally.

In vitro connectivity. Our goal was to test the above ideas in a biological circuit. We adopted a new strategy and attempted to construct the desired circuits *in vitro*. The pioneering work of Nicholls and Ready (1979) showed that identified neurons could be isolated from invertebrate ganglia, leech ganglia in their case, and that these neurons reestablished connections *in vitro*. We used identified neurons from the mollusc *Aplysia* to form our circuits, following the techniques of Schacher and Proshansky (1983). A typical co-culture is shown in Fig. 4.

An important finding of our work (Kleinfeld *et al.*, 1990b) was that identified *Aplysia* neurons can form strong, stereotyped specific connections *in vitro* that are different from those in the intact ganglion. An example involving the connections formed by a particularly well characterized interneuron, L10, to a select group of *in vivo* followers and non-followers is

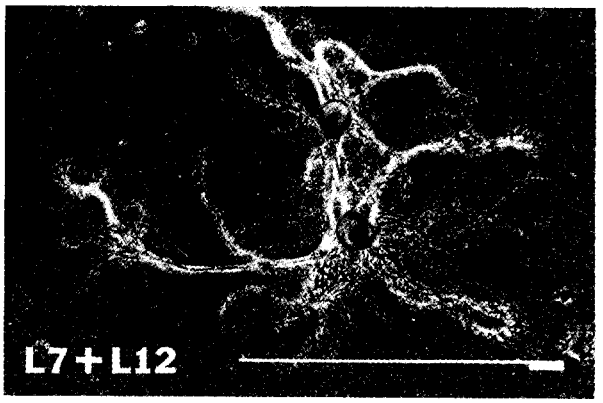

Figure 4. A co-culture of L7 and L12 neurons. The extensive array of fine neuronal processes corresponds to regenerated outgrowth. Photographed under dark field illumination 3 d after the initial plating. The scale bar corresponds to 1 mm.

shown in Fig. 5. Thus, while reciprocal connections of the same sign are rarely found *in vivo* (Koester and Alevizos, 1989), our studies allowed us to identify pairs of neurons that formed reciprocal connections *in vitro*.

Co-cultures of L10 and left upper quadrant (LUQ) neurons formed relatively robust reciprocal, inhibitory connections. Following a burst of action potentials induced in L10, we observed an inhibitory post-synaptic potential in LUQ that decayed with a time-constant of $\tau_L \sim 10$ s (Fig. 6A). An IPSP with similar response properties was observed in L10 following a burst of action potentials in L10 (Fig. 6A). A second co-culture consisted of L7 and L12 neurons and formed reciprocal, excitatory connections (Fig. 6B). These two co-cultures were selected as candidates for circuits with possible bistable outputs. We first describe the firing properties of the individual neurons and the steady-state synaptic response between pairs of neurons. We then discuss the dynamic properties of these circuits.

Neuronal properties. The input-output properties of the neurons were measured to assess conditions under which they behaved like threshold elements. Our results will be illustrated for the case of L10. The response of L10 to a succession of increasingly strong steps

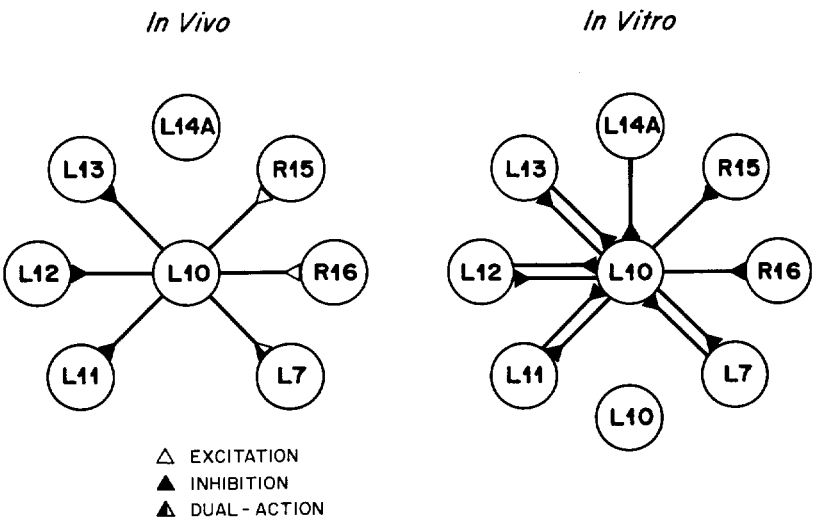

Figure 5. Schematic of the connections formed by L10 to selected *in vivo* followers and non-followers (Kandel *et al.*, 1967; Wachtel & Kandel, 1971). Reproduced from Kleinfeld *et al.* (1990b).

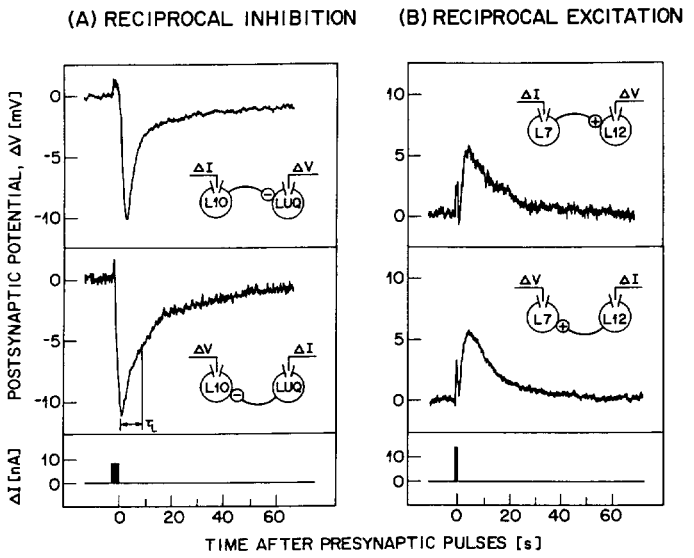

Figure 6. The reciprocal interactions between the pairs of neurons used to construct each circuit. **(A)** The interactions between L10 and LUQ. These cells formed reciprocal inhibitory connections. The hyperpolarization was measured in response to a train of action potentials generated in the presynaptic cell. **(B)** The interactions between neurons L7 and L12. These cells formed reciprocal excitatory connections. The depolarization was measured in response to a train of action potentials generated in the presynaptic cell. The initial spike in the records results from electrotonic coupling between the cells. Reproduced from Kleinfeld *et al.* (1990a).

of intracellularly injected depolarizing current is shown in Fig. 7A. The crucial feature is that an increase in the amplitude of the current by 0.01 nA could change the output of L10 from quiescence to one of continuous firing. The firing rate increased relatively little when the amplitude of the current steps was further increased; compare the traces for 0.10 nA and 0.30 nA in Fig. 7A. The complete input-output relation for L10, measured under quasi-static conditions, is shown in Fig. 7B. Although the firing of L10 showed adaptation, there was an abrupt change from quiescence to firing at all times (Fig. 7B).

The input-output relations for LUQ, L7 and L12 were roughly similar to that for L10. The neurons exhibited a sharp transition from the 'OFF' to an 'ON' level for input currents as small as 0.01 nA and thus behaved as threshold elements on this scale. We note that the input-output relation for L12 only was significantly altered by continuous activity, a point we will return to later.

The stability of the output activity will depend on there being a sufficiently strong synaptic currents under steady-state conditions. These currents must be capable of driving the postsynaptic cell between quiescence and firing in order to supply the feedback necessary to stabilize the two output states. This implies that the input must be large compared to the minimum current required to drive a neuron between its ON and OFF levels. The current that flows in the postsynaptic neuron as a results of activity in the presynaptic neuron is defined as the synaptic strength. Thus the magnitudes of the strengths T_{12} and T_{21} must satisfy $|T_{12}| >> 0.01$ nA and $|T_{21}| >> 0.01$ nA, respectively.

We consider first the connections between LUQ and L10. The L10 fires at $\sim$ 2 spikes/s when it is biased just above its threshold values. A measure of the postsynaptic response in an LUQ biased near its threshold value, with L10 firing at approximately 2 spikes/s, sets a lower bound on the magnitude of the connection strength. A similar argument holds for the postsynaptic response in L10 with LUQ firing near its minimum rate of $\sim$ 0.5 spikes/s.

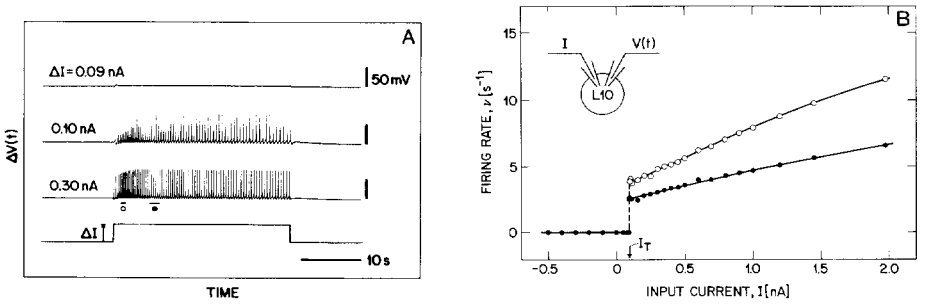

Figure 7. The input-output relation observed for L10. **(A)** The firing behavior observed with current-pulses, ΔI, of increasing amplitude. Note the change from quiescence to firing that occurs when the current is increased by only 0.01 nA, *i.e.*, from 0.09 nA to 0.10 nA. **(B)** The instantaneous firing rate as a function of the input current. The threshold current, I_T, corresponded to the minimum current for which the neuron fired continuously. Open circles (○) correspond to an average over the $2nd$ through $6th$ interspike interval, *i.e.*, the initial rate, and the filled circles (•) correspond to an average over the $15th$ through $20th$ interspike interval, *i.e.*, the steady-state rate; see part (A). Reproduced from Kleinfeld *et al.* (1990a).

The hyperpolarization observed in LUQ in response to a sustained train of action potentials in L10 is shown in Fig. 8A. For rates in the range $\nu = 2$ to 3 spikes/s, the average steady-state connection strength for the L10 to LUQ connection was $T_{21} \simeq -0.1$ nA, found by taking the ratio of the IPSP to the measured resistance of the cell. Here, as in general, the magnitude of the hyperpolarization remained essentially constant over the time-course of the measurements. However, the specific value of the connection strength from L10 to LUQ depended roughly linearly on the rate of firing of the presynaptic L10. In contrast, the strength of the connection from LUQ to L10 was essentially independent of the rate of firing of the presynaptic LUQ, with $T_{12} \simeq$ - 0.1 nA (Fig. 8B). The requirement that the magnitude of synaptic strength had a value greater than 0.01 nA, thus allowing transitions to occur, was fulfilled for both inhibitory connections.

A similar result was found for the L7 to L12 and L12 to L7 connections in the excitatory circuit, for which $T_{12} \sim T_{21} \sim +0.05$ nA. However, after sustained activity in these cells the strength of the L12 to L7 connection decreased.

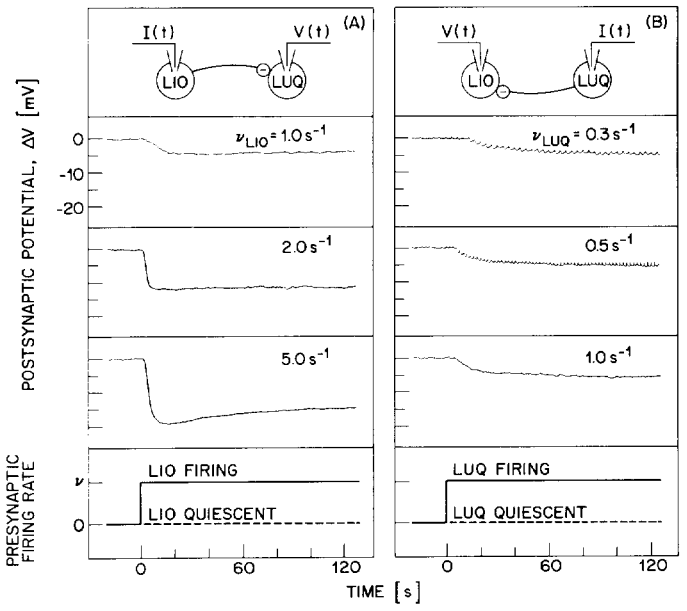

Figure 8. The steady-state interaction between L10 and LUQ neurons biased near their threshold values. **(A)** The inhibitory response measured in an LUQ for different rates of firing in L10. Action potentials were elicited in L10 by current-pulses. The minimum firing rate of L10 in our circuits was typically $\nu = 2$ spikes/s, for which the hyperpolarization in this cell corresponded to a value $T_{12} \simeq -0.2$ nA. **(B)** The inhibitory response measured in L10 in response to different rates of firing in LUQ. The firing rate of the LUQs in our circuits were typically $\nu = 0.3$ spikes/s to $\nu = 0.5$ spikes/s, for which the hyperpolarization in L10 corresponded to a value $T_{21} \simeq -0.1$ nA. The spikes in the data record were caused by weak electrotonic coupling. Reproduced from Kleinfeld *et al.* (1990a).

The stability of the output states also depends upon temporal integration of the synaptic input. This implies that the duration of the postsynaptic response must be long compared to the time between action potentials in the presynaptic cell. The decay-time of the synaptic input for all connections was $\tau_L \sim 10$ s (Figs. 6A and 6B). This time was greater than the longest period between action potentials, $\nu^{-1} < 2$ s, fulfilling this condition. Variations in firing rates on the time-scale of τ_L should not affect the stability of the circuits.

To summarize, we utilized three criteria as guides for the construction of circuits that should exhibit bistable output. These were (1) neurons with a non-linear input-output relation, (2) reciprocal connections of the same sign, and (3) integration over past activity by the relatively long decay-time of the interactions. All three criteria could be satisfied by co-cultures of L10 and LUQ neurons. Co-cultures of L7 and L12 neurons satisfied these criteria only for the first $\sim$ 100 s following the onset of activity.

Circuit dynamics: Reciprocal inhibition. The dynamic behavior of the circuits with reciprocal inhibition was probed in co-cultures of L10 and LUQ. The dynamic response of one such circuit is shown in Fig. 9. We first set the proper bias currents by: (1) Adjusting the currents to maintain the neurons at their quiescent level. (2) Increasing the current to L10 to a value $\lesssim 0.1$ nA above its threshold current. (3) Increasing the current to the LUQ to values $\lesssim 0.1$ nA below the value at which the neurons began to fire. The circuit remained in the state with L10 active and the LUQ quiescent until a current-pulse was simultaneously injected into the LUQ (Fig. 9A). The circuit remained in the new state until a pulse was injected into L10 (Fig. 9A). This pulse caused the output to return to the original state.

The data in Fig. 9 illustrates many of the features of these circuits.

1. Both the state ON/OFF, with LUQ active and L10 quiescent, and the companion state OFF/ON were stable for times that were very long compared to the synaptic integration time (Fig. 9A and 9B).

2. The alternation between states could be repeated without obvious fatigue of the output (Fig. 9C).

3. The output states were stable when one of the cells fired erratically (Fig. 9A). This shows that the temporal integration of the synaptic input made the output of the circuits relatively insensitive to fluctuations in the firing rate that are fast compared to τ_L (Fig. 6).

4. Relatively few action potentials were required to induce transitions between the two output states (Fig. 9D). This shows that transitions can be induced by relatively low firing rates.

In the above discussion of bistable output, the bias currents were adjusted so that the output of each neuron was near its optimal sensitivity to changes in its synaptic inputs. We

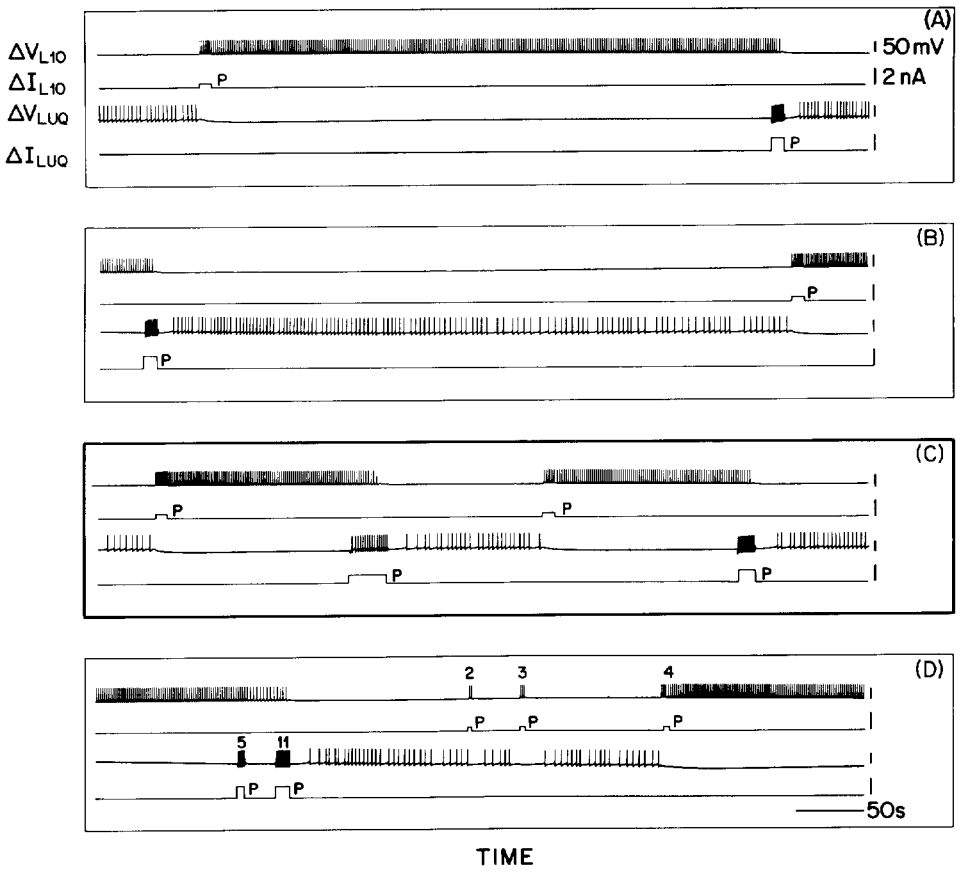

Figure 9. Bistable behavior in an inhibitory circuit consisting of co-cultured LUQ and L10 neurons. Injection of a current pulse is indicated by a 'P'. **(A)** Demonstration of the long-term stability, $\sim$ 500 s, of the state OFF/ON. Note the relatively erratic firing of LUQ. The fluctuations are faster than the integration period provided by the decay time, τ_L, of the connection. **(B)** Demonstration of the long-term stability, $\sim$ 500 s, of the state ON/OFF. **(C)** The basic bistable response of the circuit. Note that relatively weak current-pulses were used to cause transitions between the states ON/OFF and OFF/ON. **(D)** The effectiveness of the duration of the current-pulse in initiating a transition between states. A pulse to LUQ that elicited 11 action potentials, but not one that elicited 5 action potentials, caused a transition. A pulse to L10 that elicited 4 action potentials, but not pulses that elicited either 2 or 3 action potentials, caused a transition. Reproduced from Kleinfeld *et al.* (1990a).

next discuss the stability of the output as a function of the bias levels of each cell.

An example of data taken at different levels of bias current is shown in Fig. 10A. Initially the bias currents were adjusted so that LUQ was active and L10 was quiescent. We probed the stability by injecting a strong current-pulse into L10. The LUQ was briefly inhibited but resumed firing. The output was judged monostable in the state ON/OFF for this set of bias currents. We next increased the level of the bias current to L10 only and again

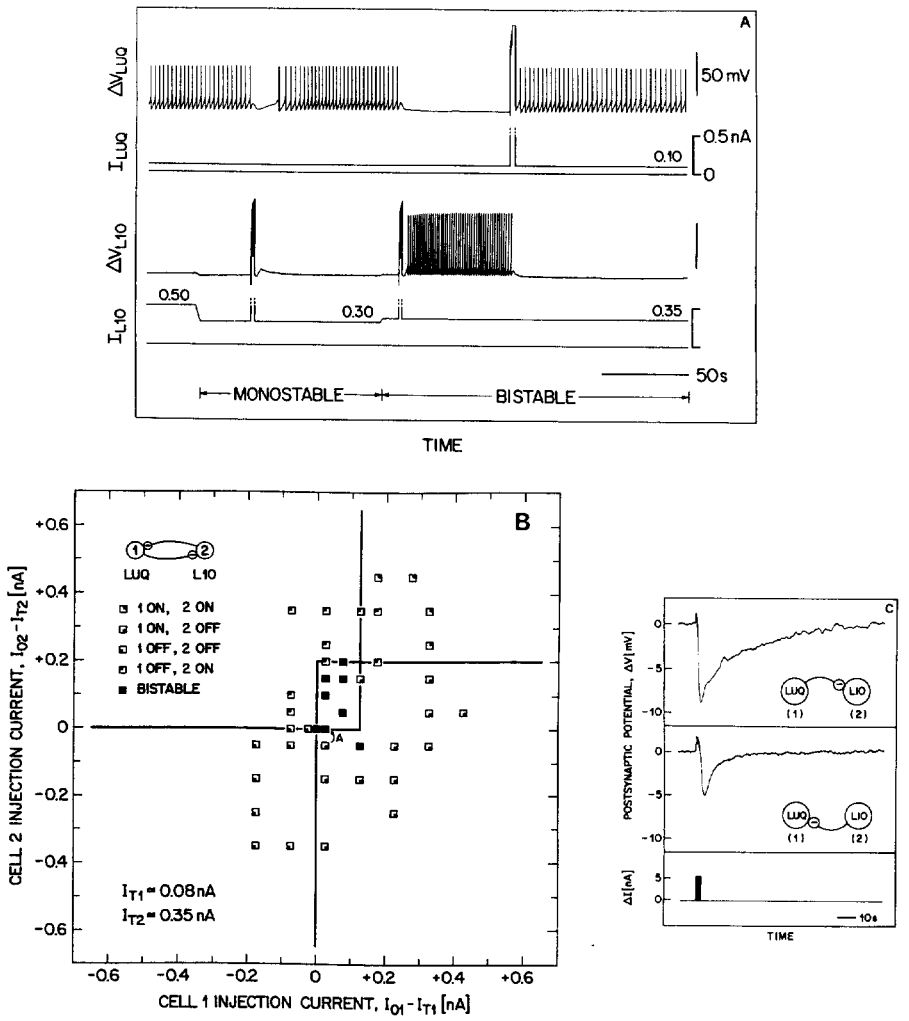

Figure 10. Analysis on the LUQ/L10 inhibitory circuit as a function of the bias currents. **(A)** Example of monostable output and bistable outputs. The bias currents were first adjusted to I_{01} = 0.10 nA and I_{02} = 0.30 nA, for which LUQ was active and L10 quiescent. Pulsing L10 momentarily inhibited LUQ, but the original activity soon recovered and thus the output was monostable. The bias current to L10 only was changed to I_{02} = 0.35 nA. Pulsing L10 now elicited a transition to the state OFF/ON. Subsequently pulsing LUQ caused a transition back to the previous state. The output was thus bistable. **(B)** Compilation of the results of the analysis. The data labeled 'A' corresponds to that in part (A). Note that there was a small range of bias currents for which the output was bistable in the states ON/OFF and OFF/ON and that the relative bias currents that separated the bistable region from monostable regions, and separated different monostable regions, agreed with theory (Fig. 9). **(C)** The relative connection strengths between the L10 and LUQ neurons. The inhibitory responses were elicited by a train of action potentials in the presynaptic cell. The connection from L10 to LUQ is roughly twice the strength of the connection from LUQ to L10. This is consistent with a bistable region that was nearly twice as large for currents injected into the LUQ as for currents injected into L10, as observed in part (B). Reproduced from Kleinfeld *et al.* (1990a).

probed the output. Injection of a pulse into L10 now caused a transition to the state OFF/ON. We observed that the output returned to its original state when we subsequently injected a current-pulse into LUQ. The output was judged bistable for this set of bias currents. This data shows how a change in the bias current to a single cell changed the output behavior of the circuit from monostability to bistability.

The compiled results for all bias levels with a particular L10/LUQ co-culture are shown in Fig. 10B. We used sets of currents that were separated in value by 0.05 nA in acquiring this data. There are three noteworthy features:

1. There was a range of bias currents for which the output was stable in both the state ON/OFF and the state OFF/ON, *i.e.*, bistable (■; Fig. 10B). The order of magnitude of this range was ~ 0.1 nA, which corresponds to the magnitude of the steady-state value of the synaptic inputs.

2. The range of bias currents for which bistable output occurred was about twice as large for currents injected into L10 as for currents injected into LUQ. From theory (Fig. 3), this ratio is related to the synaptic strengths by $(I_{01}-I_{T1})/(I_{02}-I_{T2}) = T_{12}/T_{21} \sim 2$. The observed connection strengths for these cells are consistent with the expected value (Fig. 10C).

3. The relative bias currents that separated the region of bistable output from the monostable regions, and separated different monostable regions, were in good agreement with the predictions of a simple model (*cf.* Figs. 3 and 10B).

In summary, the properties of this circuits were consistent with the behavior predicted from theoretical arguments for bistability.

Circuit dynamics: Reciprocal excitation. We now consider the output properties of co-cultures of L7 and L12. These neurons form slow, reciprocal excitatory interactions (Fig. 6B) and are expected to produce a distinct, bistable output (Fig. 2C). Thus the behavior of the excitatory circuits provides information that is complementary to that obtained with the inhibitory circuits.

The basic response of the excitatory system is shown in Fig. 11A. We first set the bias currents just below the threshold values for the cells. The neuronal output remained quiescent. The subsequent injection of a current-pulse into L7 caused L7 to fire immediately and L12 to fire 1½ s later. The latter delay coincides with the rise-time of the depolarization (Fig. 6B). Sustained activity in both cells was observed following cessation of the pulse. The brief inactivity of L7 following termination of the pulse was caused by a hyperpolarizing afterpotential. The change in the output of the circuit from the quiescent to the active state is in agreement with the model (Fig. 2C). A similar transition was observed when the pulse was injected into L12, rather than L7, in accord with the approximate symmetry of the connections (see below; trial 6 in Fig. 11B).

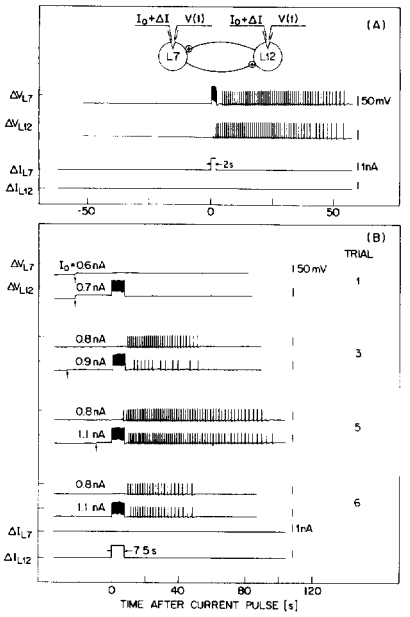

Figure 11. The behavior of an excitatory circuit consisting of co-cultured L7 and L12 neurons. **(A)** The basic response. The bias currents were set $\sim$ 0.05 nA below their threshold values for both neurons. The transition from the quiescent to the active state was induced by injecting a brief current-pulse into L7. **(B)** The behavior as a function of the bias current to each neuron. The four data records are taken from a set of consecutive trials. The value of the bias current was increased in 0.1 nA steps between trials; the arrows indicate the time of the increase. Current-pulses were injected into L12 at 120 s intervals. The partial synchrony in the firing of the neurons results from the electrotonic coupling. Reproduced from Kleinfeld *et al.* (1990a).

The duration of the active state was limited by at least two mechanisms. First, the strength of the L12 to L7 connection appreciably decreased over the time-course of the active output, as mentioned above. Secondly, the value of the threshold current increased over the same period. These effects can cause a circuit in its active state to relax to its quiescent state.

The limited duration of the state ON/ON in a bistable L7/L12 circuit preventing us from measuring the output behavior for an extensive range of bias currents. However we could observe the output over a restricted range, in which the response of the circuit was probed in a series of measurements using increasing values of bias current (Fig. 11B). A relatively long current-pulse was injected into L12 to insure that the postsynaptic response in L7 consistently reached its maximum amplitude. At low values of bias current the circuit remained in the quiescent state following a pulse (trial 1, Fig. 11B). As the bias currents were increased, in steps of 0.1 nA, we observed the beginnings of sustained activity following a pulse. For the response in trial 3 (Fig. 11B) the model suggests that the bias current of L7 is properly set but that of L12 is too low. We thus increased the bias level of L12. The circuit

exhibited sustained firing in response to a pulse (trial 5; Fig. 11B), in agreement with the model. The duration of the active output was long compared with the decay-time of the connections but decreased with successive trials (*c.f.* trials 5 and 6, Fig. 11B).

Interlude

Our analysis of the output behavior and stability of the circuits with reciprocal connections used an extremely simple model. Neurons were taken as threshold elements and details of the synaptic interactions, such as the non-linear behavior of the LUQ to L10 connection, were ignored. The model provided an accurate account of the basic bistable output behavior and details of the stability of the output for the circuit with reciprocal inhibitory connections. The circuit with reciprocal excitatory contained time-dependent neuronal input-output relations and synaptic interactions, yet the model accounted for its behavior for many synaptic time-constants.

Our results with the *in vitro* circuits suggests that simplified models may be appropriate for understanding the dynamics of *in vivo* circuits. We next discuss the analysis of the circuit that controls the muscles involved in escape swimming in the mollusc *Tritonia*.

Escape Swimming in Tritonia.

Tritonia diomedea is a mollusc that exhibits a pattern of undulatory flexions as an escape mechanism from predatory starfish (Willows, 1967). The *Tritonia* first retracts from the site of contact with the starfish and then swims by alternate dorsal and ventral flexions of its body (Fig. 12). The swim flexions last for two to twenty cycles.

Studies on the neural control of the escape swim response began with the work of Willows (1967) and Willows and Hoyle (1969). The neural elements that form the circuit that controls the swim rhythm, referred to as a central pattern generator (CPG), have been elucidated by the work of Getting (1981; 1983a; 1983b; Getting & Dekin, 1985).

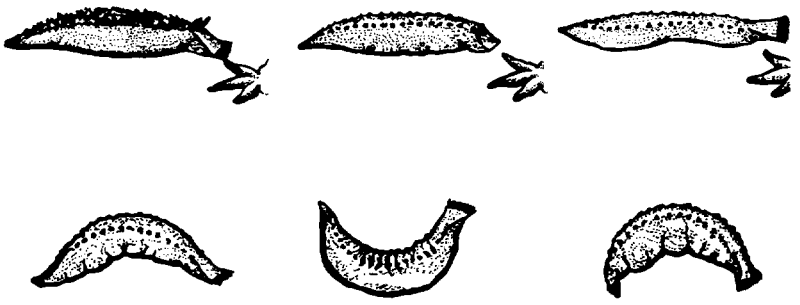

Figure 12. Cartoon of the escape swimming response in *Tritonia* (clockwise from upper left hand corner). The *Tritonia* withdraws from an attacking starfish, extends its body, and exercises ventral-dorsal undulatory motions to gain hydrodynamic lift. Adapted from Willows (1967).

The CPG in *Tritonia* consists of four neural groups, denoted by *VSI−A*, *VSI−B*, *C2* and *DSI*. The *VSI* neurons are the ventral swim interneurons, *C2* is a cerebral neuron and *DSI* represents the dorsal swim interneurons. The observed output pattern consists of bursting output from *VSI−A* and *VSI−B* neurons alternating with bursts from the *C2* and *DSI* neurons; Fig. 13. The time interval between consecutive action potentials within a bursting state is ~0.01 s to 0.1 s and the duration of each state is, on average, approximately 5 s.

Three observations (Getting 1981; 1983b) were of primary importance in motivating our theoretical study of this circuit.

1. The CPG in *Tritonia* functions without a pacemaker cell, *i.e.*, a single neuron whose firing properties determines the output period of the network. This implies that the rhythmic output is a collective property of the network.

2. The output appears an an alternation between two well defined output states, one with C2 and the DSI firing and VSI-A and VSI-B quiescent, and vice versa. This suggests that the behavior of this circuit can be understood in terms of a model whose output can alternate between two states.

3. The synaptic connections have components that act on two different time-scales. For example, the synaptic input to *C2* from *DSI* shows a rapid excitatory response followed

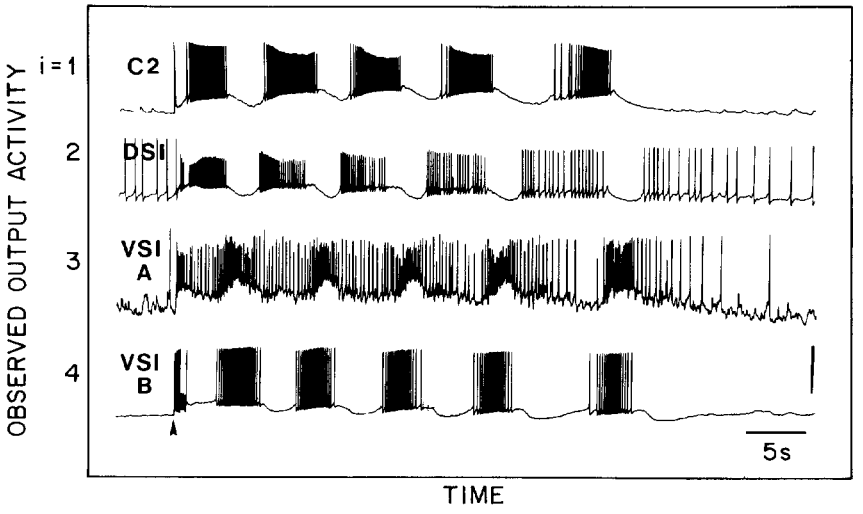

Figure 13. The output activity simultaneously measured from a *C2*, *DSI*, *VSI−A* and *VSI−B* neuron in an isolated brain preparation from *Tritonia*. These neurons comprise the CPG that controls the escape swim sequence. Their output corresponds to $V_1(t)$, $V_2(t)$, $V_3(t)$ and $V_4(t)$, respectively, in the analysis presented in the text. The arrow indicates the initiation of the sequence. Note that in the present work we are concerned only with the oscillatory behavior of the CPG, and not with the gradual dephasing that leads to its turning off. Vertical bar: 50 mV for *C2*, *DSI* and *VSI−B* and 25 mV for *VSI−A*. Adapted from Getting (1983b).

by a much slower inhibitory response; Fig. 14.

The observed synaptic response in *Tritonia* appears to be analogous to the form of the interaction in a Hopfield-like neural network for the generation of temporal patterns (Sompolinsky & Kanter, 1986; Kleinfeld, 1986). We will analyze the CPG in *Tritonia* within the framework of this model. We first review the model, focusing on two issues: (1) The rules for forming synaptic connections in terms of the desired output states. (2) Simplified dynamics of the networks. A full description of the correspondence between the model and the CPG in *Tritonia* is given elsewhere (Kleinfeld & Sompolinsky, 1988; 1989).

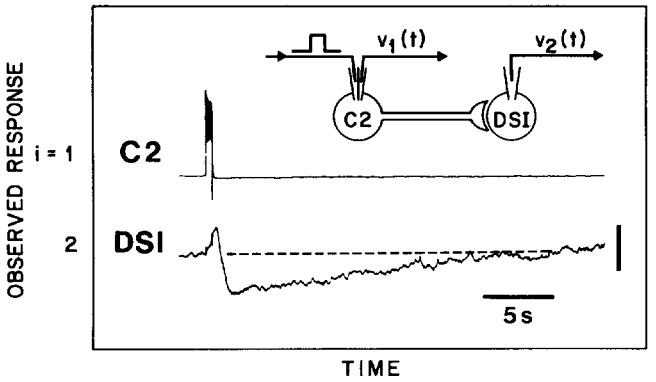

Figure 14. An example of the synaptic interaction between two neurons in the CPG in *Tritonia*. Shown is the pre-synaptic activity measured in the $C2$ neuron, $v_1(t)$, and the post-synaptic response measured in a *DSI* neuron, $v_2(t)$, as the result of a short pulse of current injected into $C2$. The measurement was performed under conditions which insured that only mono-synaptic connections contributed to the observation. The observed response applies to two out of the three DSI neurons (DSI_B and DSI_C); the other DSI neuron (DSI_A) exhibits only a slow response. The area under the initial, positive going response corresponds roughly to T_{21}^S; that under the slowly decaying response corresponds to T_{21}^L. The time dependence of the slow decay corresponds to the time dependence of the slow synaptic response function, $w(t)$. Vertical bar: 40 mV for $C2$ and 2 mV for *DSI*. Adapted from Getting (1981).

Network model: Connectivity. Our network consists of highly interconnected model neurons whose essential feature is a non-linear relation between their inputs and their firing rate. The output patterns are encoded in the strength of the synaptic connections between pairs of neurons. Rhythmic output emerges as a collective property of the network.

The output of each neuron, $V_i(t)$, varies between zero (quiescent) and unity (maximum firing rate). The state of the network is specified by the output activity of all of its neurons and represented by $V(t) = \{V_i(t)\}_{i=1}^N$ where N is the number of neurons. A pattern is defined as a temporal sequence of meta-stable states, denoted by $V^\mu = \{V_i^\mu\}_{i=1}^N$. A cyclic pattern of length r consists of the sequence

$$\cdots \longrightarrow V^1 \longrightarrow V^2 \longrightarrow V^3 \longrightarrow \cdots \longrightarrow V^{r-1} \longrightarrow V^r \longrightarrow V^1 \longrightarrow \cdots .$$

We define the synaptic connection between the j-th pre-synaptic neuron and the i-th post-synaptic neuron as T_{ij}. A central feature of the model is that each connection T_{ij} is *functionally* separated into two components, denoted T_{ij}^S and T_{ij}^L. The two components are hypothesised to have different characteristic response times. The synaptic connections T_{ij}^S act on the shorter of the two time-scales. This time-scale, τ_S, determines the time required for the network to settle in each of the stable states. The synaptic connections denoted T_{ij}^L act on the longer of the two times. This time-scale, τ_L ($\tau_L >> \tau_S$), sets the time for the onset of the transitions between consecutive states in the pattern. Thus the duration of an individual state in a pattern will be $\sim\tau_L$, while the transitions between states occurs on the faster time-scale of τ_S.

The role of the connection strengths T_{ij}^S is to stabilize the network until a transition to the next meta-state occurs. This is achieved by defining the T_{ij} in terms of a Hebb-like learning rule *i.e.*,

$$T_{ij}^S = \sum_{\mu=1}^{r} (2V_i^\mu - 1)(2V_j^\mu - 1), \qquad i \neq j \qquad [4]$$

where r is the length of the pattern and $T_{ii}^S = 0$. The variable $(2V_i^\mu - 1)$ has a value of either -1 (quiescent) or +1 (maximally firing) so that inhibitory as well as excitatory synapses are formed.

The role of the connection strengths T_{ij}^L is to induce transitions from the μ-th meta-stable state to the $(\mu + 1)$-th state. We define

$$T_{ij}^L = \lambda \sum_{\mu=1}^{r-1} (2V_i^{\mu+1} - 1)(2V_j^\mu - 1), \qquad i \neq j, \ \lambda > 0 \qquad [5]$$

$T_{ii}^L = 0$, $V^r = V^1$ and λ is a scaling parameter for the transition strength that determines the period of the rhythmic output.

The rules we described for specifying the synaptic connections assume the existence of connections between all pairs of neurons. Biological networks may intrinsically contain a much smaller set of connections. The performance of the network model is only marginally affected when a substantial fraction of the fast components of the synaptic connections (T_{ij}^S) are eliminated at random. The main effect of eliminating a fraction of these components is to decrease the maximum number of states that can be used to construct an output pattern. Eliminating the slow components of the synaptic connections (T_{ij}^L) has a negligible effect on this number. However, random elimination of a fraction of the T_{ij}^L synapses will decrease the ability of the network to make a transition between the stable states. This decrease can be offset by a compensating increase in the value of λ. Thus the value of λ is bounded by

$$\lambda \gtrsim \frac{<\text{Strength of the } T_{ij}^{L} \text{ synapses}>}{<\text{Strength of the } T_{ij}^{S} \text{ synapses}>} . \qquad [6]$$

The value of λ will determine the output period of the network. This relation will allow us to link the observed synaptic strengths in *Tritonia* with the measured output period.

Network model: Dynamics. The integrated synaptic input to each model neuron is assumed to be a *linear* summation of the outputs of the pre-synaptic neurons. An example of the time-course of a post-synaptic response to a short pre-synaptic stimulus is illustrated in Fig. 15A. The dynamic evolution of the network is described by a set of circuit equations

$$\tau_S \frac{du_i(t)}{dt} + u_i(t) = \frac{1}{N} \sum_{j=1}^{N} \left(T_{ij}^{S} V_j(t) + T_{ij}^{L} \int_0^{\infty} dt' \, V_i(t - t') \, w(t') \right) + I_i \qquad [7]$$

where $u_i(t)$ is the net input to the i-th neuron, I_i represents an external input (Fig. 15B), the synaptic response function $w(t)$ for the slow components is a non-negative function that is normalized to unity and characterized by a mean time-constant τ_L. The output of a model neuron, $V_i(t)$, is related to its net input, $u_i(t)$, by a non-linear gain function

$$V_i(t) = g\left(u_i(t) - \tfrac{1}{2}\right) \qquad [8]$$

where as the threshold level of the neuron has, for simplicity in the present discussion, been taken as ½. The dynamic features of the network do not depend on the details of the gain function; Fig. 15C illustrates an appropriate form.

The analysis of the dynamic properties of the network is greatly simplified when the neurons are approximated by threshold elements, *i.e.*, the neurons are either quiescent or fully active and the slow synaptic response function is approximated as a pure time delay, *i.e.*, $w(t) = (1/\tau_L)\delta(t - \tau_L)$. In this limit the dynamics are governed by a set of update rules

$$V_i(t + \tau_S) = \Theta\left(\sum_{j=1}^{N} \left(T_{ij}^{S} (2V_j(t) - 1) + T_{ij}^{L} (2V_j(t - \tau_L) - 1) \right) \right) \qquad [9]$$

where $\Theta(x)$ is defined by Eq. **2**.

We illustrate the temporal output expected from the model defined by Eq. **9** by simulating a toy network. The network contains 7 neurons and 3 stable states, with connections formed using the Hebb-like rules (Eqs. **4,5**). Starting from a random initial

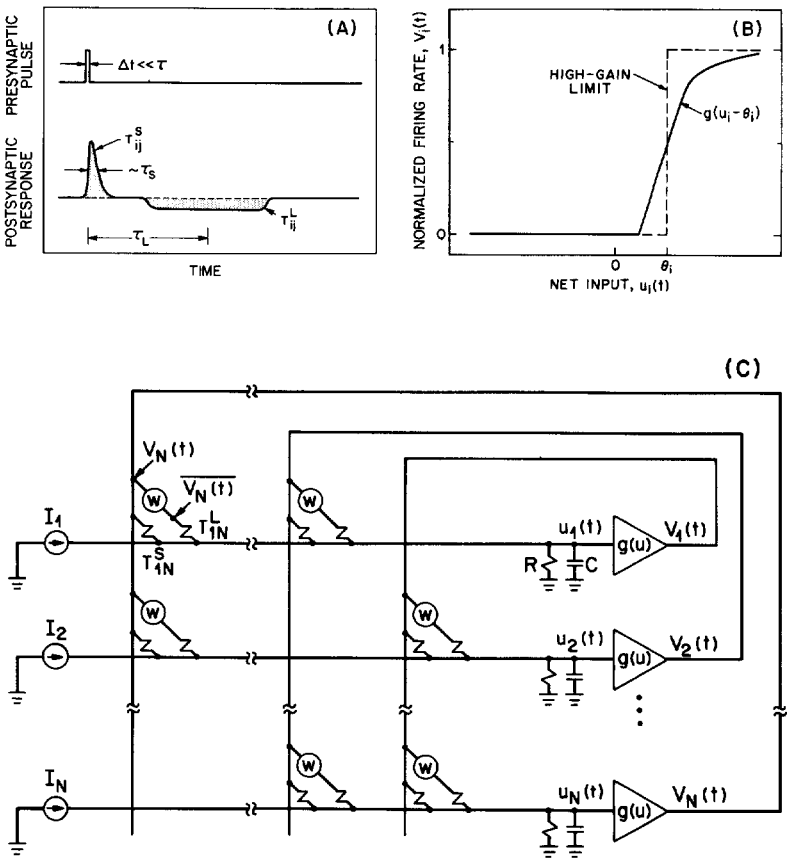

Figure 15. An associative network model for the generation of rhythmic patterns. **(A)** Illustration of a two-component synaptic connection from the j-th to the i-th neuron. The components are resolved following a short pulse of activity in the pre-synaptic neuron. The area under the fast synaptic response is equal to T_{ij}^S, taken to be excitatory in this example. The area under the slow synaptic response is equal to T_{ij}^L, taken to be inhibitory in this example. The ratio of these areas, averaged over all pairs of synapses, equals the transition strength λ. The time-course of the slow-synaptic response corresponds to the response function w(t); it has a time-constant of τ_L. **(B)** Schematic representation of a saturating gain function for a neuron. This function relates the output, or firing frequency of a neuron, $V_i(t)$, to the value of the net input, $u_i(t)$, and the mean operating level, θ_i. We took $\theta_i = \frac{1}{2}$ (Eq. 8) for simplicity in our analysis. **(C)** Schematic representation of the circuit diagram for the model network. Neurons are represented by saturating amplifiers with a charging time of RC, where R represents the *net* input resistance of the neuron. Synaptic connections between each pair of neurons are represented by conductances proportional to T_{ij}^S (fast synaptic components) or T_{ij}^L (slow synaptic components) Reproduced from Kleinfeld and Sompolinsky (1988).

condition, the network converged toward a stable state and subsequently made near-periodic transitions between all three states. The resulting firing pattern is illustrated in Fig. 16.

We are now ready to draw a connection between our model and Gettings (1981; 1983b) measurements on the central pattern generator controlling the swim rhythm in *Tritonia*.

Correspondence between the model and Tritonia. The focus of our analysis is to determine if the properties of the CPG in *Tritonia* support the mechanism we propose for generating patterns. We ask:

1. Are the observed synaptic strengths consistent with those calculated from the form of the observed output states?
2. Are the simple update rules [Eq. 9] sufficient to demonstrate the emergence of an oscillatory output that qualitatively resembles the observed pattern?
3. Is the period of the observed output pattern accounted for in terms of the magnitude and form of the observed slow synaptic response?

The observed output sequences (Fig. 13) will be approximated by an oscillation between a state V^+ and its antiphase $V^- \equiv (1 - V^+)$, where

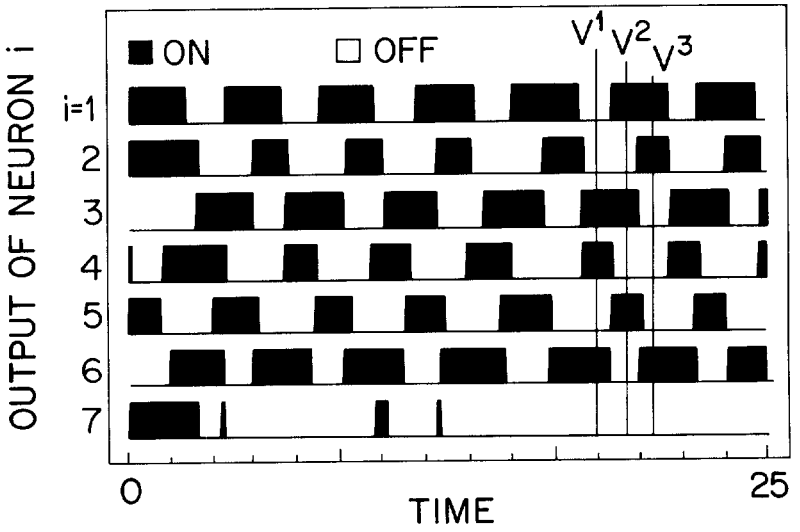

Figure 16. Simulation of a network containing 7 neurons and 3 output states that generates cyclic sequences. The output states are $V^1 = (\,0011010\,)^T$, $V^2 = (\,1010100\,)^T$ and $V^3 = (\,1100010\,)^T$. Time is denoted in units of $\tau_L = 6{\cdot}\tau_S$. Adapted from Kleinfeld (1986).

$$V^+ = \begin{pmatrix} C2 & \text{active} \\ DSI & \text{active} \\ VSI-A & \text{quiescent} \\ VSI-B & \text{quiescent} \end{pmatrix} = \begin{pmatrix} +1 \\ +1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad V^- = \begin{pmatrix} 0 \\ 0 \\ +1 \\ +1 \end{pmatrix}.$$

These states are used as the stable in our model. The short-term connection strengths T_{ij}^S, and the long-term connection strengths, T_{ij}^L, were deduced from the outputs V^+ and V^- using the Hebb-like rules (Eqs. **4,5**). For example, the value of the predicted fast, stabilizing connection from DSI to C2 is

$$T_{21}^S = (2V_2^+ - 1)(2V_1^+ - 1) = +1 .$$

The value of the predicted slow, transition causing connection from DSI to C2 is

$$T_{21}^L = \lambda\,(2V_2^- - 1)(2V_1^+ - 1) = -\lambda .$$

Note that the rules allow one to predict all possible connections. Only a subset of these connections need be present for the model to function, as discussed above.

How do the predicted synaptic strengths compare with the observed values? The strength of a synaptic connection is proportional to the integral, with respect to time, of the conductance changes ($\sigma(t)$) induced in the post-synaptic neuron by a short pulse of activity in the pre-synaptic neuron. These integrals were estimated from measurements (Fig. 14) of the potentials ($V_{obs}(t)$) induced in the post-synaptic neuron by a short burst of action potentials in the pre-synaptic neuron.

We made a crude classification of the observed synaptic strengths in *Tritonia* based on the pair-wise measurements of Getting (1981; 1983b) and Getting's detailed analysis (1989). The observed response was classified as either a fast component, T_{ij}^S, or a slow component, T_{ij}^L, according to the time-scale of the response.

In our simple analysis we need only consider the average magnitude of the fast synaptic components relative to that of the slow components, *i.e.*, λ, in addition to the sign of the measured response. For example, the sign of the fast components of the observed synaptic connection from *C2* to *DSI* (Fig. 14) was estimated by

$$T_{21}^S = \text{sign}\left[\int_0^\infty dt\ \sigma^S(t)\right] = \text{sign}\left[\int_0^\infty dt\ V_{obs}^S(t)\right] = +1 .$$

Similarly, the slow component of this connection is

$$T_{21}^{L} = \lambda \cdot \text{sign}\left[\int_0^\infty dt\ \sigma^{L}(t)\right] = \lambda \cdot \text{sign}\left[\int_0^\infty dt\ V_{obs}^{L}(t)\right] = -\lambda\ .$$

The determination of λ contains a large uncertainty, in part because of the difficulty in separating the fast and slow contributions to the measured synaptic response. We estimate

$$\lambda \simeq 5 \text{ to } 10.$$

The signs of the experimentally observed synaptic strengths match those of the theoretically predicted strengths (Fig. 17). Three of the possible twelve synaptic connections show both a short-term and a long-term response. Connections $(i,j) = (3,1)$ and $(i,j) = (3,2)$ both show short-term inhibition followed by a long-term excitation, while connection $(i,j) = (2,1)$ shows short-term excitation followed by long-term inhibition.

The observed connections in *Tritonia* suggest that Hebb-like rules are a useful prescription for specifying synaptic connectivity. The form of these connections further illustrates how the sign of the net synaptic input to a neuron can change over time. Lastly, the relatively larger number of T_{ij}^{L} connections that are absent in *Tritonia* in comparison with the T_{ij}^{S} components is consistent with the constraint on λ (Eq. **6**)

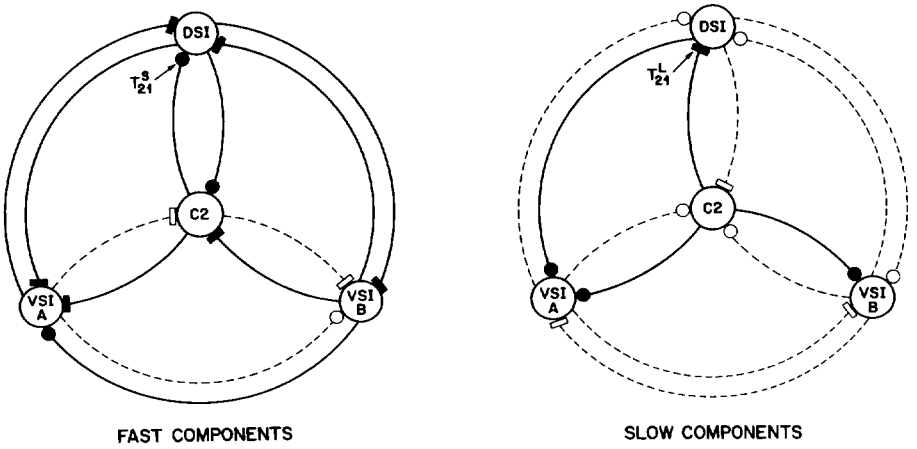

Figure 17. The synaptic strengths for *Tritonia*. Circles represent excitatory connections and rectangles represent inhibitory connections. All theoretically possible connections, calculated for the observed output states in *Tritonia* using the Hebb-like rules in the model, are shown. Solid lines and symbols indicate those connections observed in *Tritonia*. Fast components refer to the T_{ij}^{S} while slow components refer to the T_{ij}^{L}.

We now examine whether the observed network parameters in *Tritonia* indeed give rise to rhythmic output in the network model. We begin our analysis with the simplified model that accents the role of the observed synaptic connections in generating stable oscillations (Eq. **8**). Immediately after the network has stabilized in a new state V^+, the output of the i-th neuron is $V_i(t) = V_i^+$, but the delayed output is $V_i(t-\tau_L) = V_i^-$. The output of the i-th neuron after the next update is

$$V_i(t+\tau_S) = \Theta\left[\sum_{j=1}^{4} T_{ij}^S (2V_j^+ - 1) + T_{ij}^L (2V_j^- - 1)\right]$$

$$= \Theta\left[\begin{pmatrix} 0 & +1 & 0 & -1 \\ +1 & 0 & -1 & -1 \\ -1 & -1 & 0 & +1 \\ 0 & -1 & 0 & 0 \end{pmatrix}\begin{pmatrix} +1 \\ +1 \\ -1 \\ -1 \end{pmatrix} + \lambda \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ +1 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \\ +1 \\ +1 \end{pmatrix}\right]$$

$$= \Theta\left[\begin{pmatrix} 2 \\ 3+\lambda \\ -3-2\lambda \\ -1-\lambda \end{pmatrix}\right] = \begin{pmatrix} +1 \\ +1 \\ 0 \\ 0 \end{pmatrix} = V_i^+ \qquad \text{for} \quad \lambda > 0 .$$

Thus the output of the network is stable on the time-scale of τ_S. After the network has remained in the state V^+ for a time τ_L, the delayed output changes to $V_i(t-\tau_L) = V_i(t) = V_i^+$. The output of the i-th neuron after the next update is

$$V_i(t+\tau_L+\tau_S) = \Theta\left[\sum_{j=1}^{4} T_{ij}^S (2V_j^+ - 1) + T_{ij}^L (2V_j^+ - 1)\right]$$

$$= \Theta\left[\begin{pmatrix} 2 \\ 3-\lambda \\ -3+2\lambda \\ -1+\lambda \end{pmatrix}\right] = \begin{pmatrix} +1 \\ 0 \\ +1 \\ +1 \end{pmatrix} \qquad \text{for} \quad \lambda > 3 .$$

The network is now in a mixed, unstable state. Using this new value for the current state in the update procedure gives

$$V_i(t+\tau_L+2\tau_S) = \Theta\left[\sum_{j=1}^{4} T_{ij}^S (2V_j(t+\tau_L+\tau_S) - 1) + T_{ij}^L (2V_j^+ - 1)\right]$$

$$= \Theta\left[\begin{pmatrix} -2 \\ -1-\lambda \\ 1+2\lambda \\ 1+\lambda \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \\ +1 \\ +1 \end{pmatrix} = V_i^- .$$

The network has now completed a transition from the state V^+ to the state V^-. It will remain in this state for a time $t_0 \simeq \tau_L$, after which the cycle will repeat itself. The output of the network will oscillate only if the transition strength is $\lambda > 3$. This value is consistent with the observed value of $\lambda = 5$ to 10. Our analysis demonstrates that, within the framework of our model, even a small network can function with the elimination of many of its theoretically possible connections.

The simplified analysis presented above suggests that the observed connection strengths can give rise to rhythmic output in the model network. We examined the steady-state behavior of the network model for *Tritonia* using analog dynamics (Eq. 6) and an appropriate synaptic response function that is a smooth function of time. The equivalent electrical circuit is shown in Fig. 18. Stable oscillations of the form described by the previous simplified analysis were observed. The output activity for the transition strength $\lambda = 10$ is shown in Fig. 19.

The period observed for the output of the CPG in *Tritonia* is $2t_0 = 7$ to 10 s (Fig. 13), while the time-constant for the slow synaptic response lies in the range $\tau_L = 3$ to 6 s ($\tau_L \simeq 6$ s for the data shown in Fig. 14). Is this value for the period accounted for by the model? As we discussed earlier, the predicted value for t_0 depends on the values of τ_L and λ and on the form of the response function $w(t)$. We used the analog equations for the network (see above) and calculated the dependence of $2t_0$ on λ (Fig. 20). The lower estimate, which is in accord with

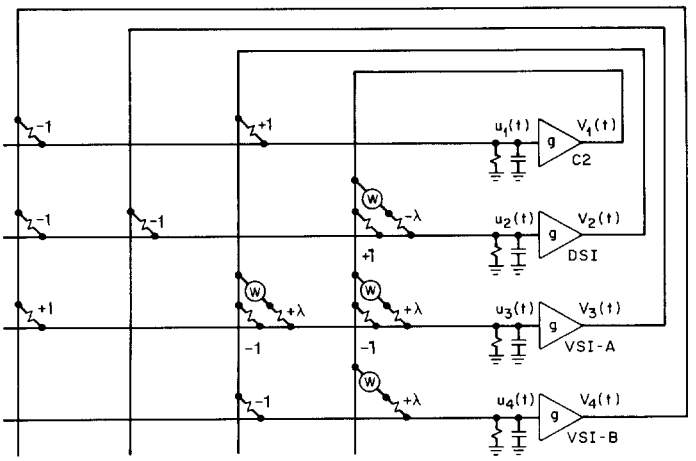

Figure 18. Schematic representation of the equivalent circuit for the analog network model describing the CPG in *Tritonia*. The synaptic strengths contained in this circuit correspond to the observed connections T_{ij}^S and T_{ij}^L. Adapted from Kleinfeld and Sompolinsky (1988).

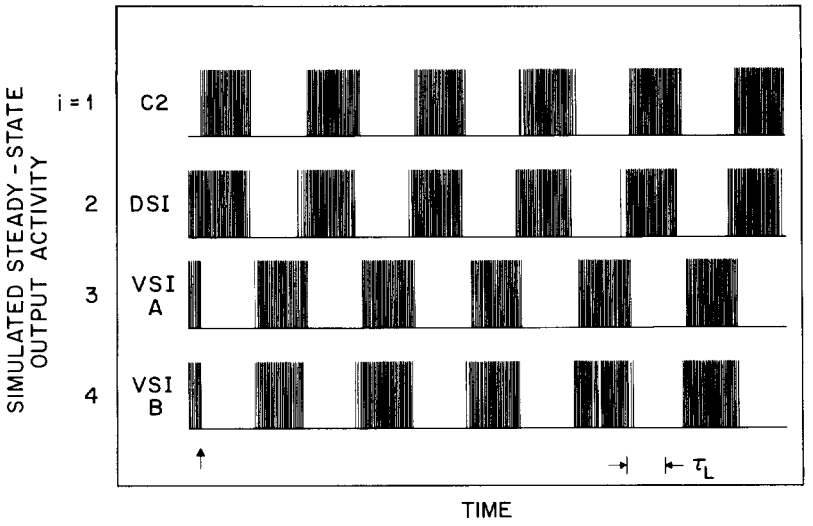

Figure 19. Simulated output activity from the analog network model describing the CPG in *Tritonia*. The arrows indicates the start of the simulated output from the initial states $V(t<0) = \overline{V(t<0)} = (0\,1\,1\,1)^T$. The network equations were simulated using the observed values of T_{ij}^S and T_{ij}^L. Reproduced from Kleinfeld and Sompolinsky (1988).

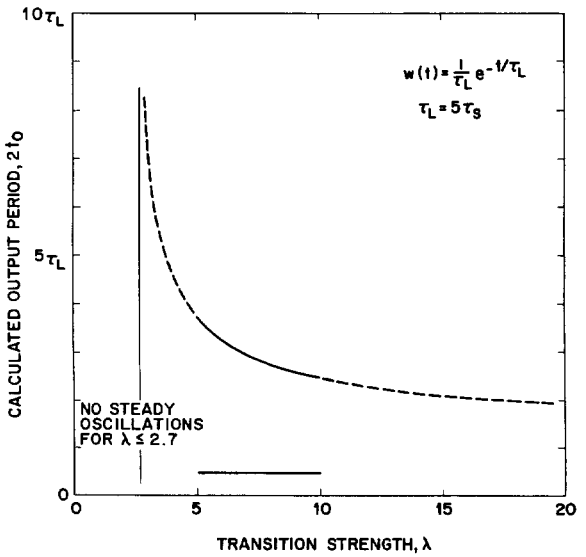

Figure 20. The output period, $2t_0$, of the analog network model for *Tritonia* as a function of the average transition strength λ. A period of $2t_0 \simeq$ 7s to 10s corresponds to the period observed in *Tritonia*. The solid line delimits the range of values for λ estimated from the measured connection strengths. For the value $\lambda = 10$, the period deduced from the model is in accord with the observed period. Reproduced from Kleinfeld and Sompolinsky (1989).

the experimental value, corresponds to $\lambda \simeq 10$.

How is the output of the network in *Tritonia* affected by a change in bias currents? We performed an analysis similar to that for the two-cell circuits with static, bistable outputs (Eq. **3**). We consider the behavior of the circuit for changes in the bias currents to C2 and DSI only (Fig. 21). There is a range of currents for which the circuit will produce oscillatory output. Outside of this range the output will remain constant in time. Two interesting points of this analysis are

1. The bias level to the DSI neurons appears to be too low in the absence of external inputs. This is consistent with the experimental observation that activation of the rhythmic output in *Tritonia* requires an external excitatory input to the DSI neurons (Getting & Dekin, 1985)

2. In principle, the CPG in *Tritonia* should be able to operate with a smaller number of neurons. The consequence of this decrease in number is a reduction in the range of bias levels for which oscillatory output can occur. The results calculated for functional removal of VSI-B are shown in Fig. 21.

It would be interesting to test these ideas. Neurons can be functionally removed from a circuit by hyperpolarization or, when they are electrically coupled to other neurons, by photostimulation techniques (Miller & Selverston, 1979).

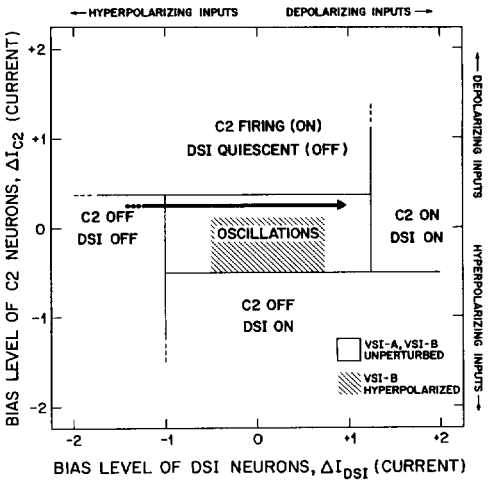

Figure 21. Analysis for the output from the model network for *Tritonia* as a function of the bias currents to neurons C2 and DSI. In the absence of external input, the bias level of the DSI neuron is hyperpolarized and the CPG is inactive. Depolarizing DSI (the 'ramp depolarization' in *Tritonia*; Getting and Dekin (1985)) brings its bias level into the range for which the CPG can oscillate (arrow). The model predicts that when VSI-B is functionally removed from the CPG, *e.g.*, by strong hyperpolarization, there is a range of operating levels for which the CPG can still oscillate (shaded region). The magnitude of the currents used in calculating the results in this figure are similar to the synaptic currents in *Tritonia*.

Interlude

We used our associative network model to analyze the CPG controlling the swim rhythm in the mollusc *Tritonia*. This is a small network, yet it contains many of the basic features inherent in our model. The sign and time course of the observed synaptic strengths were in accord with the values predicted by Hebb-like rules. This suggest the utility of such rules for predicting the strength of the underlying synaptic connections from the observed output states. The rhythmic output in *Tritonia* could be understood by a simplified analysis that employed threshold units as neurons and that replaced the response function of the slow synapses by a simple time-delay.

Our discussion of *Tritonia* showed that a simple prescription of neuronal connectivity, *i.e.*, Hebb rules, was applicable to a small network with recurrent connections. This raises the question as to whether simple principles that relate neuronal connectivity to the output activity of the network exist for more complex systems. In particular, the output of many systems cannot be approximated by a set of few persistent states. Hence, Hebb rules may not be appropriate. Several investigators have examined the applicability of unsupervised (Barlow, 1989; Linsker, 1990) as well as supervised (Zipser & Andersen, 1988; Lehky & Sejnowski, 1988; Lockery *et al.*, 1989) learning rules to a variety of biological systems. We discuss next the application of supervised learning rules to the relatively small network that controls local bending in the leech.

Local bending in the leech

The leech exhibits a variety of behaviors whose neurophysiological correlates have been studied in much detail (Muller *et al.*, 1981). One behavior, elucidated by Kristan (1982), concerns a localized withdrawal response known as the local bending response. The body-wall of the leech will flex, forming a u-shaped bend, in response to localized mechanical stimulation. A dorsal stimulus causes a dorsal bend and a ventral stimulus causes a ventral bend (Fig. 22).

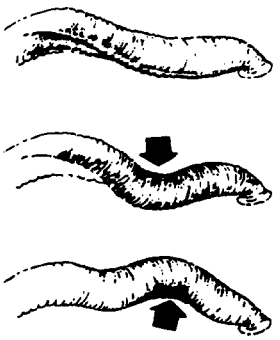

Figure 22. The local bending reflex in the leech. The posterior half of the animal is shown. **(Top)** Resting position. **(Middle)** Flexure following dorsal stimulation. **(Bottom)** Flexure following ventral stimulation. Adapted from Kristan (1982).

The circuitry that governs this behavior has been studied by Lockery and Kristan (1990a; 1990b), who worked out many of the details for a functionally dominant subset of this circuit. Within each of multiple ganglia, input to the circuit occurs via the activation of sensory neurons known as pain, or P, cells. These neurons make connections onto a set of interneurons that in turn make connections onto a set of excitatory and inhibitory motor neurons. There are 4 sensory input cells (right dorsal (PD), ... , left ventral (PV)), 16-17 interneurons and 8 motor neurons (Fig. 23A). The majority of connections between the sensory neurons and the interneurons, but not between the interneurons and the motor neurons, are known (Lockery & Kristan, 1990b).

The input from any P cell results in either a depolarizing postsynaptic potential (PSP) or a hyperpolarizing PSP in every motor neuron in the ganglion (Fig. 23B). Interestingly, when pairs of sensory cells were jointly stimulated the response in the motor neurons did not necessarily correspond to the linear sum of the responses to individual sensory cells. For example, stimulating both the PD and the PV sensory cells leads to a strong depolarization in the DE motor neuron, even though the individual responses sum to essentially zero (Fig. 23B). This suggests that interactions from sensory cells to motor neurons are mediated in a non-linear fashion by the interneurons.

The local bending circuit was modeled by a feedforward network by Kristan *et al.* (1989) and Lockery *et al.* (1989). Symmetries in the underlying connectivity allow us to consider a reduced circuit that contains all 4 sensory cells, but only half the number of interneurons and motor neurons that are in the original circuit (S. R. Lockery, personal communication). The

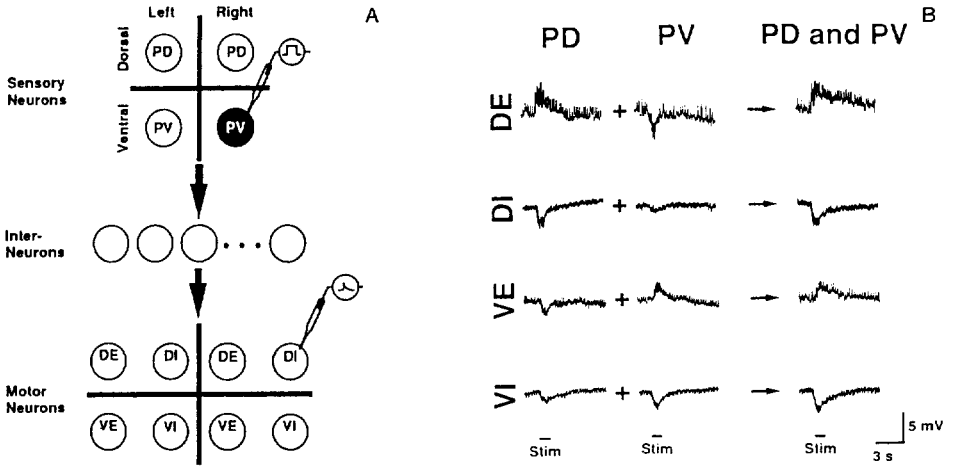

Figure 23. Anatomy and physiology of local bending in the leech. **(A)** Schematic of the neuronal circuit. There are four sensory neurons, the left and right dorsal pain cells (PDs) and ventral pain cells (PVs), at least 17 interneurons and 8 motor neurons, left and right dorsal excitors (DEs), dorsal inhibitors (DIs), ventral excitors (VEs) and ventral inhibitors (VIs). **(B)** Intracellular records from the soma of four motor neurons following stimulation of one or both sensory neurons that are ipsilateral to the motor neurons. Adapted from Lockery and Kristan (1990).

resultant circuit is shown in Fig. 24. The underlying connection strengths were estimated using supervised learning algorithm, *i.e.*, back-propagation. A non-linear input/output relation was assumed for the interneurons (Fig. 14B). The algorithm attempted to find a set of connections so that the sensory-input to motor-output of the model network matched that of an experimentally derived training set of 9 to 12 pairs of sensory inputs and motor neuron PSPs.

Lockery *et al.* (1989) found that the algorithm converged to a number of solutions that reproduced the observed sensory-input to motor-output relations of the circuit. Further, the distributed nature of the connectivity in these networks is consistent with the available physiological data.

The above result raises the question as to whether the observed relation between sensory input and motor neuron PSPs could be accounted for by a circuit containing a smaller number of interneurons. This question was considered by Kristan *et al.* (1989), who examined the difference between the observed values of the PSPs and the output of the model network as a function of the number interneurons. The results are shown in Fig. 25. There was a substantial improvement in the performance of the network when the number of interneurons was increased from one to three. Only a marginal improvement was realized by an additional increase in the number of interneurons.

These intriguing results indicate that the learning algorithm used for this circuit provides insufficient constraints on the underlying connectivity. One possible additional constraint that may account for the larger network in the leech is stability against variations in the strength of the connections. A second possibility in that the same interneurons are used by other behaviors (Kristan *et al.*, 1989). Indeed, one interneuron is known to play a role in the leech swim circuit (Stent *et al.*, 1978).

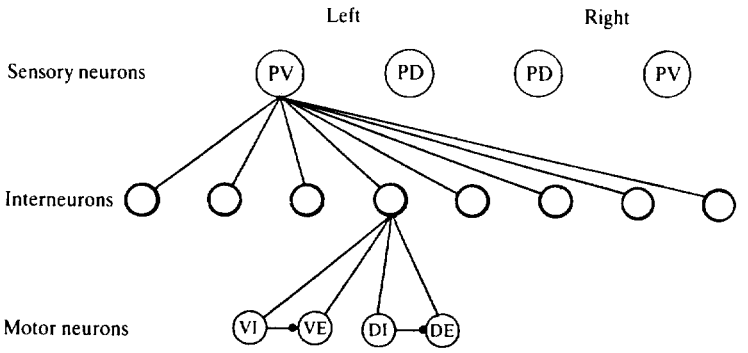

Figure 24. Layered network model for the circuit controlling the bending reflex in the leech. Only one group of connections from each layer is shown. A symmetry in the relation between the sensory-inputs and the motor-outputs is accounted for by including only 8 interneurons and 4 motor neurons. The inhibitor motor neurons make fixed inhibitory connections to their respective excitor motor neurons. Adapted from Kristan *et al.* (1989) following discussions with Dr. S. R. Lockery.

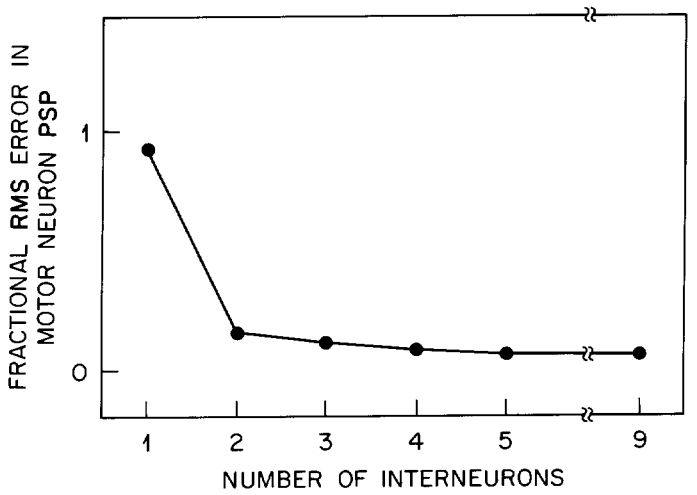

Figure 25. The root-mean-square (RMS) error in the output from the model network as a function of the number of interneurons in the reduced leech circuit (Fig. 24). Adapted from Kristan *et al.* (1989) following discussions with Dr. S. R. Lockery.

Summary

We have discussed the relation between the pattern of output activity and the underlying synaptic connectivity in small nervous systems. In three circuits, two that produce switchable, bistable output states and the circuit that controls swimming in *Tritonia*, the output pattern could be understood in terms of neurons that behaved as threshold devices and connections between the neurons that behaved as linear elements. In these circuits there was a clear correspondence between the output pattern and the sign of the underlying synaptic connections. The analysis of a fourth circuit, mediating the local bending response in the leech, demonstrated some of the issues that arise when applying simple design principles to the connectivity in larger networks.

We thank S. R. Lockery for valuable discussions and for sharing unpublished data on the leech local bending circuit with us. We thank our colleagues A. L. Obaid, T. D. Parsons, F. Raccuia-Behling and B. M. Salzberg for their contributions to the studies on *in vitro* circuits. This work was supported, in part, by the Fund for Basic Research administered by the Israeli Academy of Arts and Sciences and by the US-Israeli Binational Science Foundation.

References

Barlow, H. B. (1989) Unsupervised learning. *Neural Comput.* **1** 295-311.

Getting, P. A. (1981) Mechanism of pattern generation underlying swimming in *Tritonia*. I. Neuronal network formed by monosynaptic connections. *J. Neurophys.* **46**, 65-79.

Getting, P. A. (1983a) Mechanism of pattern generation underlying swimming in *Tritonia*. II. Network reconstruction. *J. Neurophys.* **49**, 1017-1035.

Getting, P. A. (1983b) Mechanism of pattern generation underlying swimming in *Tritonia*. III. Intrinsic and synaptic mechanisms for delayed excitation. *J. Neurophys.* **49**, 1036-1050.

Getting, P. A. & Dekin, M. S. (1985) Mechanism of pattern generation underlying swimming in *Tritonia*. IV. Gating of central pattern generator. *J. Neurophys.* **53**, 466-480.

Getting, P. A. (1989) Reconstruction of Small Neural Networks. *In* Methods in Neuronal Modeling: From Synapses to Networks. Koch, C., & Segev, I., eds., pp. 171-194, MIT Press, MA.

Harth, E. M., Csermely, T. J., Beek, B. & Lindsay, R. D. (1970) Brain functions and neural dynamics. *J. Theoret. Biol.* **26**, 93-120.

Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational properties. *Proc. Nat'l. Acad. Sci. USA* **79**, 2554-2558.

Hopfield, J. J. (1984) Neural with graded response have collective computational properties like those of two-state neurons. *Proc. Nat'l. Acad. Sci. USA* **81**, 3088-3092.

Kleinfeld, D. (1986) Sequential state generation by model neural networks. *Proc. Nat'l. Acad. Sci. USA* **83**, 9469-9473.

Kleinfeld, D. & Sompolinsky, H. (1988) Associative neural network model for the generation of temporal patterns: Theory and application to central pattern generators. *Biophy. J.* **54**, 1039-1051.

Kleinfeld, D. & Sompolinsky, H. (1989) Associative neural network models for central pattern generators. *In* Methods in Neuronal Modeling: From Synapses to Networks. Koch, C. & Segev, I., eds., pp. 195-246, MIT Press, MA.

Kleinfeld, D., Raccuia-Behling, F. & Chiel, H. J. (1990a) Circuits constructed from identified *Aplysia* neurons exhibit multiple patterns of persistent activity. *Biophy. J.* **57**, 697-715.

Kleinfeld, D., Parsons, T. D., Raccuia-Behling, F., Salzberg, B. M. & Obaid, A. L. (1990b) Foreign connections are formed *in vitro* by *Aplysia* interneuron L10 and its *in vivo* followers and non-followers. *J. Exp. Biol.* in press.

Koester, J. & Alevizos, A. (1989) Innervation of the kidney by L10, the LUQ cells, and an identified peripheral motoneuron. *J. Neurosci.* **9**, 4078-4088.

Kristan, W. B. (1982) Sensory and motor neurones responsible for the local bending reflex in leeches. *J. Exp. Biol.* **96**, 161-180.

Kristan, W. B., Lockery, S. R., Wittenberg, G. & Cottrel, G. W. (1989) Behavioural choice: In theory and in practice. *In* The Computing Neuron. Durbin, R., Miall, C. & Mitchison, G., eds., pp. 180-204, Addison-Wesley, NY.

Lehky, S. R. & Sejnowski (1988) Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature* **333**, 452-454.

Linsker, R. (1990) Perceptual neural organization: Some approaches based on network models and information theory. *Ann. Rev. Neurosci.* **13**, 257-281.

Lockery, S. R. & Kristan, W. B. (1990a) Distributed processing of sensory information in the leech. I. Input-output relations of the local bending reflex. *J. Neurosci.* in press.

Lockery, S. R. & Kristan, W. B. (1990b) Distributed processing of sensory information in the leech. II. Identification of interneurons contributing to the local bending reflex. *J. Neurosci.* in press.

Lockery, S. R., Wittenberg, G., Kristan, W. B. & Cottrell, G. W. (1989) Function of identified interneurons in the leech elucidated using neural networks trained by backpropagation. *Nature* **340**, 468-471.

Miller, J. P. & Selverston, A. I. (1979) Rapid killing of single cells by irradiation of intracellularly injected dye. *Science* **206**, 702-704.

Muller, K. J., Nicholls, J. G. & Stent, G. S. (1981) *Neurobiology of the Leech*, Cold Spring Harbor Laboratory, NY.

Schacher, S. & Proshansky, E. (1983) Neurite regeneration by *Aplysia* neurons in dissociated cell culture: Modulation by *Aplysia* hemolymph and the presence of the initial axonal segment. *J. Neurosci.* **3**, 2403-2413.

Sompolinsky, H. & Kanter, I. (1986) Temporal association in asymmetric neural networks. *Phys. Rev. Lett.* **57**, 2861-2864.

Stent, G. S., Kristan, W. B., Friesen, W. O., Ort, C. A., Poon, M. & Calabrese, R. L. (1978) Neuronal generation of the leech swimming movement *Science* **200**, 1348-1357.

Willows, A. D. O. (1967) Behavioral acts elicited by stimulation of single, identifiable brain cells. *Science* **157**, 570-574.

Willows, A. D. O. & Hoyle, G. (1969) Neuronal network triggering a fixed action pattern *Science* **166** 1549-1551.

Wilson, H. R. & Cowan, J. D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophy. J.* **12**, 1-23.

Zipser, D. & Andersen, R. A. (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679-484.