

5

Realistic Network Models of Distributed Processing in the Leech

Shawn R. Lockery and Terrence J. Sejnowski

The distributed nature of information processing in the brain presents a great challenge to systems neuroscience. Whether one considers the processing of sensory information or the control of motor responses, the output of the nervous system is a function of large populations of simultaneously active neurons. In all but the simplest reflexes, neurons are arranged in bewildering networks of parallel and feedback pathways, making a purely intuitive understanding of the system difficult if not impossible.

Neural models and computer simulations are approaches to this complexity that could help us achieve a deeper understanding of distributed processing. Useful predictions of a model must be directly related to the experiments that are performed. Since much of the current physiological effort in systems neuroscience is devoted to making single-unit recordings, models are needed that predict the responses of individual neurons. However, models sufficiently detailed to predict responses of individual neurons require a very large number of parameters, including the type and distribution of voltage-sensitive channels, the anatomical position of synapses, and the sign and strength of synaptic connections. Since the value of such parameters is generally not known, one approach is to study the effect on the model of systematically varying each parameter over its physiological range. In general, however, this is not an efficient strategy because there are often too many parameters and their physiological range is often quite large.

There are now a variety of network optimization algorithms that adjust parameters such as connection strengths in artificial neural networks (Hinton 1989). The algorithms adjust each parameter in the network to reduce the overall error in the performance of the network. These algorithms differ according to the type of error information available and the way it is used to change the parameters. Optimization techniques are being applied to network models of known biological circuits where the number of parameters is too large to be studied individually. The implications of such models for our understanding of biological networks has been difficult to establish because of the many simplifying assumptions involved in the first generation of neural network models

(Sejnowski et al. 1988). This chapter reviews our application of network training algorithms to a distributed processing system in a well-defined invertebrate preparation, the local bending reflex of the medicinal leech. The chapter starts with an overview of two common modeling strategies: compartmental and connectionist or neural network models. Its central point is that it is now possible to combine the compartmental and connectionist approaches to create realistic models with large numbers of parameters. Much of the chapter is devoted to exploring in detail a specific example of this combined approach using the local bending reflex as a model system. We first review the behavioral physiology of the local bending circuit and the conceptual issues that prompted our modeling efforts. The model is then presented in detail and its performance is evaluated in terms of a distributed processing hypothesis of the local bending reflex. We conclude by suggesting how additional physiological constraints could be added to predict the identity of as yet unidentified interneurons in the leech.

MODELING OVERVIEW

Compartmental Models

In some of the most detailed simulations produced to date, compartmental models are used to represent individual neurons. The morphology of each neuron is divided into a number of isopotential electrical compartments, with neighboring compartments representing nearby regions of the neuron, and currents pass between compartments via axial resistances (figure 5.1A). The voltage V in a passive compartment is given by:

$$\tau \frac{dV}{dt} = -V + RI_m \quad (5.1a)$$

where τ is the time constant of the compartment, R is the input resistance, and I_m is the net transmembrane current (for more details, see Segev et al. 1989). Action potentials are achieved by locating voltage-dependent ion channels in axonal and somatic compartments. Synapses are modeled by putting synaptic conductances in dendritic compartments. The collective behavior of a large network of neurons is then investigated by connecting the model neurons together according to the known circuitry of the brain region of interest (figure 5.1B). This strategy has been used to study the emergent properties of models that incorporate the cell types and connectivity of such systems as the hippocampus (Traub and Wong 1982; Traub et al. 1989), the olfactory bulb and piriform cortex (Bower 1990), and the neocortex (Bush and Douglas 1991). For example, Bush and Douglas (1991) showed that pyramidal cells with different intrinsic bursting frequencies in a model of visual cortex quickly synchronized their bursting as a result of global feedback

Realistic Network Models of Distributed Processing in the Leech

Shawn R. Lockery and Terrence J. Sejnowski

The distributed nature of information processing in the brain presents a great challenge to systems neuroscience. Whether one considers the processing of sensory information or the control of motor responses, the output of the nervous system is a function of large populations of simultaneously active neurons. In all but the simplest reflexes, neurons are arranged in bewildering networks of parallel and feedback pathways, making a purely intuitive understanding of the system difficult if not impossible.

Neural models and computer simulations are approaches to this complexity that could help us achieve a deeper understanding of distributed processing. Useful predictions of a model must be directly related to the experiments that are performed. Since much of the current physiological effort in systems neuroscience is devoted to making single-unit recordings, models are needed that predict the responses of individual neurons. However, models sufficiently detailed to predict responses of individual neurons require a very large number of parameters, including the type and distribution of voltage-sensitive channels, the anatomical position of synapses, and the sign and strength of synaptic connections. Since the value of such parameters is generally not known, one approach is to study the effect on the model of systematically varying each parameter over its physiological range. In general, however, this is not an efficient strategy because there are often too many parameters and their physiological range is often quite large.

There are now a variety of network optimization algorithms that adjust parameters such as connection strengths in artificial neural networks (Hinton 1989). The algorithms adjust each parameter in the network to reduce the overall error in the performance of the network. These algorithms differ according to the type of error information available and the way it is used to change the parameters. Optimization techniques are being applied to network models of known biological circuits where the number of parameters is too large to be studied individually. The implications of such models for our understanding of biological networks has been difficult to establish because of the many simplifying assumptions involved in the first generation of neural network models

(Sejnowski et al. 1988). This chapter reviews our application of network training algorithms to a distributed processing system in a well-defined invertebrate preparation, the local bending reflex of the medicinal leech. The chapter starts with an overview of two common modeling strategies: compartmental and connectionist or neural network models. Its central point is that it is now possible to combine the compartmental and connectionist approaches to create realistic models with large numbers of parameters. Much of the chapter is devoted to exploring in detail a specific example of this combined approach using the local bending reflex as a model system. We first review the behavioral physiology of the local bending circuit and the conceptual issues that prompted our modeling efforts. The model is then presented in detail and its performance is evaluated in terms of a distributed processing hypothesis of the local bending reflex. We conclude by suggesting how additional physiological constraints could be added to predict the identity of as yet unidentified interneurons in the leech.

MODELING OVERVIEW

Compartmental Models

In some of the most detailed simulations produced to date, compartmental models are used to represent individual neurons. The morphology of each neuron is divided into a number of isopotential electrical compartments, with neighboring compartments representing nearby regions of the neuron, and currents pass between compartments via axial resistances (figure 5.1A). The voltage V in a passive compartment is given by:

$$\tau \frac{dV}{dt} = -V + RI_m \quad (5.1a)$$

where τ is the time constant of the compartment, R is the input resistance, and I_m is the net transmembrane current (for more details, see Segev et al. 1989). Action potentials are achieved by locating voltage-dependent ion channels in axonal and somatic compartments. Synapses are modeled by putting synaptic conductances in dendritic compartments. The collective behavior of a large network of neurons is then investigated by connecting the model neurons together according to the known circuitry of the brain region of interest (figure 5.1B). This strategy has been used to study the emergent properties of models that incorporate the cell types and connectivity of such systems as the hippocampus (Traub and Wong 1982; Traub et al. 1989), the olfactory bulb and piriform cortex (Bower 1990), and the neocortex (Bush and Douglas 1991). For example, Bush and Douglas (1991) showed that pyramidal cells with different intrinsic bursting frequencies in a model of visual cortex quickly synchronized their bursting as a result of global feedback

Realistic Network Models of Distributed Processing in the Leech

Shawn R. Lockery and Terrence J. Sejnowski

The distributed nature of information processing in the brain presents a great challenge to systems neuroscience. Whether one considers the processing of sensory information or the control of motor responses, the output of the nervous system is a function of large populations of simultaneously active neurons. In all but the simplest reflexes, neurons are arranged in bewildering networks of parallel and feedback pathways, making a purely intuitive understanding of the system difficult if not impossible.

Neural models and computer simulations are approaches to this complexity that could help us achieve a deeper understanding of distributed processing. Useful predictions of a model must be directly related to the experiments that are performed. Since much of the current physiological effort in systems neuroscience is devoted to making single-unit recordings, models are needed that predict the responses of individual neurons. However, models sufficiently detailed to predict responses of individual neurons require a very large number of parameters, including the type and distribution of voltage-sensitive channels, the anatomical position of synapses, and the sign and strength of synaptic connections. Since the value of such parameters is generally not known, one approach is to study the effect on the model of systematically varying each parameter over its physiological range. In general, however, this is not an efficient strategy because there are often too many parameters and their physiological range is often quite large.

There are now a variety of network optimization algorithms that adjust parameters such as connection strengths in artificial neural networks (Hinton 1989). The algorithms adjust each parameter in the network to reduce the overall error in the performance of the network. These algorithms differ according to the type of error information available and the way it is used to change the parameters. Optimization techniques are being applied to network models of known biological circuits where the number of parameters is too large to be studied individually. The implications of such models for our understanding of biological networks has been difficult to establish because of the many simplifying assumptions involved in the first generation of neural network models

(Sejnowski et al. 1988). This chapter reviews our application of network training algorithms to a distributed processing system in a well-defined invertebrate preparation, the local bending reflex of the medicinal leech. The chapter starts with an overview of two common modeling strategies: compartmental and connectionist or neural network models. Its central point is that it is now possible to combine the compartmental and connectionist approaches to create realistic models with large numbers of parameters. Much of the chapter is devoted to exploring in detail a specific example of this combined approach using the local bending reflex as a model system. We first review the behavioral physiology of the local bending circuit and the conceptual issues that prompted our modeling efforts. The model is then presented in detail and its performance is evaluated in terms of a distributed processing hypothesis of the local bending reflex. We conclude by suggesting how additional physiological constraints could be added to predict the identity of as yet unidentified interneurons in the leech.

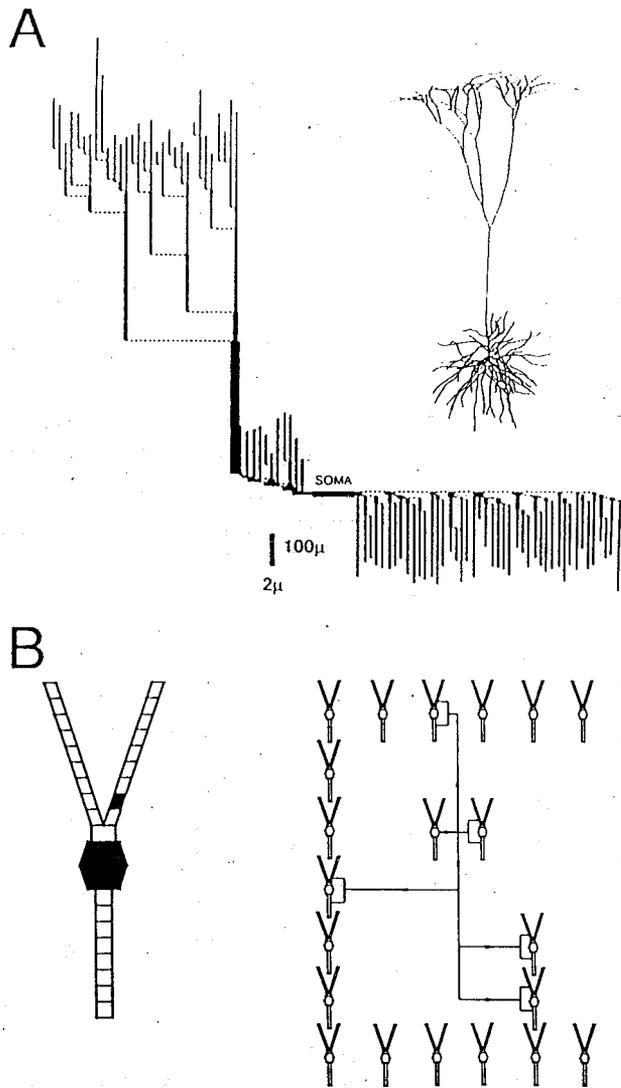
MODELING OVERVIEW

Compartmental Models

In some of the most detailed simulations produced to date, compartmental models are used to represent individual neurons. The morphology of each neuron is divided into a number of isopotential electrical compartments, with neighboring compartments representing nearby regions of the neuron, and currents pass between compartments via axial resistances (figure 5.1A). The voltage V in a passive compartment is given by:

$$\tau \frac{dV}{dt} = -V + RI_m \quad (5.1a)$$

where τ is the time constant of the compartment, R is the input resistance, and I_m is the net transmembrane current (for more details, see Segev et al. 1989). Action potentials are achieved by locating voltage-dependent ion channels in axonal and somatic compartments. Synapses are modeled by putting synaptic conductances in dendritic compartments. The collective behavior of a large network of neurons is then investigated by connecting the model neurons together according to the known circuitry of the brain region of interest (figure 5.1B). This strategy has been used to study the emergent properties of models that incorporate the cell types and connectivity of such systems as the hippocampus (Traub and Wong 1982; Traub et al. 1989), the olfactory bulb and piriform cortex (Bower 1990), and the neocortex (Bush and Douglas 1991). For example, Bush and Douglas (1991) showed that pyramidal cells with different intrinsic bursting frequencies in a model of visual cortex quickly synchronized their bursting as a result of global feedback



Compartmental Models

Figure 5.1 (A) Equivalent compartmental model of a pyramidal cell in visual cortex. Each rectangle is a single compartment whose length and width matches the length and diameter of the corresponding segment of the neuron (inset). Dotted lines indicate branch points in the dendritic tree. Length and width of compartments are shown on different scales, as the scale bar indicates. (Part A reprinted from Lytton and Sejnowski 1991, with permission of the American Physiological Society.) (B) Model of area CA3 of the hippocampus using a network of compartmental neurons. Each neuron in the network is a simplified compartmental neuron comprising a soma and three dendrites represented by a total of 28 compartments. The soma and one dendritic compartment (filled polygons) contain active conductances producing action potentials. Excitatory connections (arrows) in the network were assigned randomly according to the probability of connections among neighboring CA3 neurons. (Part B reprinted from Traub and Wong 1982, © AAAS.)

inhibition and intracortical reexcitation. Thus, synchrony between cortical neurons responding to coherent visual stimuli could be an emergent property of intracortical connectivity.

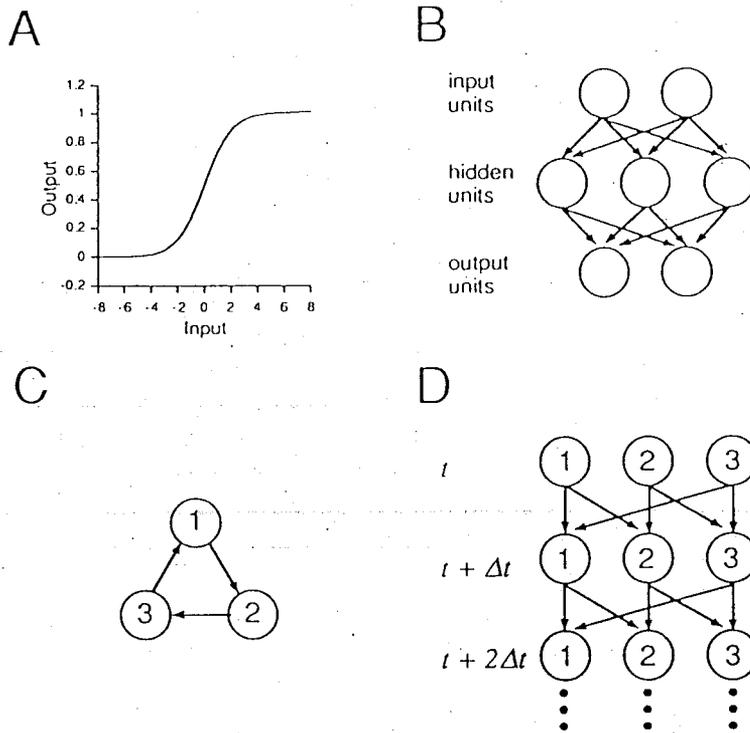
Neural Network Modeling

A neural network model trades the realism of compartmental models for control over the network's input-output function. Network models are composed of simplified neuronlike elements or *units* connected by synapses of various strengths or *weights* (see figure 3.2). Each unit has a state or activity level that is a function of its net input. The net input to a unit is the weighted sum of the activities of the units from which it receives a connection. The output of a unit is generally a nonlinear function of its net input or activity level. Many networks use a sigmoidal function (figure 5.2A). This approximates the relationship in real neurons between net input and firing frequency, since weak inputs produce little or no activity (*cutoff*), and strong inputs drive the unit to its maximum firing rate (*saturation*). Variations in the construction of network models include the use of binary units that are either on or off (Hopfield 1982), and units that multiply their inputs in addition to adding them (Durbin and Rumelhart 1989).

Model networks usually contain three types of units (figure 5.2B; see also figure 3.3): (1) *input units* that receive inputs only from sources external to the network; (2) *output units* that define the output state of the network; and (3) *hidden units* that receive inputs from sensory neurons and provide the inputs to the output units. A particular set of activation values on the input units constitutes an input vector and the set of activations on the output units constitutes the output vector. The input-output function of a network is the mapping or list of associations between input and output vectors, and it is a function of the weights in the network. The power of network models derives from the use of optimization algorithms that can adjust the weights to achieve almost any input-output function, a task that is next to impossible if attempted by hand in large networks.

Backpropagation

Algorithms for optimizing weights fall into three classes: *supervised* procedures in which the algorithm is given a list or *training set* of desired input and output vector pairs (Rumelhart et al. 1986a), *reinforcement* procedures in which each input vector is associated with a scalar output assessment instead of an output vector (Barto and Jordan 1987), and *unsupervised* procedures in which there is no assessment of network output. However, the distinction between supervised and unsupervised procedures is not absolute. Some ostensibly unsupervised algorithms turn out on closer inspection to contain an internal assessment of performance or monitoring. It is thus more useful to distinguish between



Aspects of Network Models

Figure 5.2 (A) The sigmoidal input-output function for the neuronlike units used in many network models. (B) Feedforward network. (C) Recurrent network. (D) The network in C represented as a multilayer, feedforward network in which layers correspond to successive points in time. Units connect between but not within layers. The recurrent network in the feedforward configuration can be trained by backpropagation in much the same way as a feedforward network—by sending the error backward from later to earlier points in time.

monitored and *unmonitored* systems as well as supervised and unsupervised (Churchland and Sejnowski 1992).

Backpropagation is a supervised, unmonitored optimization procedure for networks with one or more layers of hidden units. At the start of optimization, the weights are assigned small values at random. An input vector is presented, and activity flows forward to the output units, determining an output vector. An error measure is then calculated by comparing the output vector to the output specified in the training set. This is repeated for each input-output pair in the training set. The weights in the network are then changed or updated by small amounts based on the total error produced by this pass through the training set. Each pass through the training set and its associated weight update is called a *training cycle*. After many cycles, a set of weights is found that reproduces the desired input-output function.

In mathematical terms, backpropagation is a form of gradient descent.

Weights are updated in the direction of the error gradient, and backpropagation is a way to estimate the derivative of the total error with respect to each weight. The total error produced on any pass through the training set is a function of the weights. The derivative of this function with respect to each weight tells whether increasing the weight will increase or decrease the error. Since our objective is to find weights that minimize the error, we decrease the weights with positive derivatives, and increase weights with negative derivatives. It turns out that the derivative of the error with respect to each weight can be calculated explicitly from the activity of the presynaptic unit, together with the net input and error of the postsynaptic unit (Rumelhart et al. 1986a). The procedure for multilayer networks is called backpropagation because the error of the postsynaptic unit is calculated recursively in terms of the error of the units to which it projects. In practice, this involves sending the error back from the output units to the hidden units. (For additional commentary on backpropagation, see chapter 3.)

Recurrent Backpropagation

Training networks with recurrent or feedback connections provides a special challenge because the effect of a weight on the error is different at different times (figure 5.2C). Several closely related algorithms exist for adjusting connections in feedback networks (Pearlmutter 1989; Pineda 1987, 1989; Rumelhart et al. 1986a; Williams and Zipser 1990). The most intuitively clear version is called *backpropagation through time* (Williams and Zipser 1990). The idea is to convert the feedback network into an equivalent feedforward network by breaking its activity into discrete time steps (figure 5.2D). The activity of each neuron at each time step is represented as a single neuron in an equivalent feedforward network. The resulting network can then be optimized as if it were a feedforward network. This procedure is referred to as *recurrent backpropagation* to distinguish it from the original form of backpropagation (Rumelhart et al. 1986a) in which consecutive layers do not represent time steps. The original form is sometimes called *static backpropagation*.

Models of Real Circuits Optimized by Static Backpropagation

The first application of network models for real-life neurophysiological modeling arose from the insight that optimization algorithms originally developed for the study of cognitive processes could be used to train networks to reproduce physiologically relevant input-output functions (Lehky and Sejnowski 1988; Zipser and Andersen 1988a). Analysis of the response properties of hidden units in such models often gave new insight into the possible function of neurons in the biological network.

Zipser and Andersen (1988a) used a three-layered feedforward network model as a means of investigating the possible spatial processing performed by neurons in area 7a of parietal cortex. The network was

optimized by static backpropagation to compute the location of a visual stimulus in head-centered coordinates from the input units that signaled either the location of the stimulus on the retina or the angle of gaze of the eyes with respect to the head. After optimizing, the hidden units in the model network had complicated, nonlinear spatial receptive fields that were qualitatively indistinguishable from the receptive fields of area 7a neurons. The correspondence between units in the model and area 7a neurons provided a missing conceptual link between area 7a and spatial localization. In addition, it provided a new hypothesis concerning how the spatial receptive fields of area 7a neurons could participate in the representation of head-centered spatial information.

In a related study, Lehky and Sejnowski (1988) optimized a network to compute the three-dimensional shape of an object from the two-dimensional pattern of light and dark on the retina. The model showed that receptive fields that seem designed for edge detection arise in networks whose sole purpose is to determine three-dimensional shape. Their results pointed out that receptive field information in isolation is an unreliable predictor of the function of neurons in distributed processing systems. In later investigations (Lehky et al. 1992), the activity of complex cells in monkey striate cortex was recorded in response to hundreds of visual images. For each cell, a network model was optimized to reproduce the recorded responses when presented with the same stimuli. After optimization, the model gave accurate responses to novel stimuli, indicating it captured more of the input-output transfer function of the complex cell than previous analytical methods.

Anastasio and Robinson (1990) optimized a network model of the vestibulo-ocular reflex (VOR) to produce eye velocity signals appropriate to head velocity and stimulus-pursuit commands reaching the vestibular nucleus. After optimization, hidden units in the model received both head velocity and pursuit inputs, but to varying degrees, as has been observed experimentally. Their model showed that mixed velocity and pursuit inputs could be an integral part of a distributed vestibulo-oculomotor system.

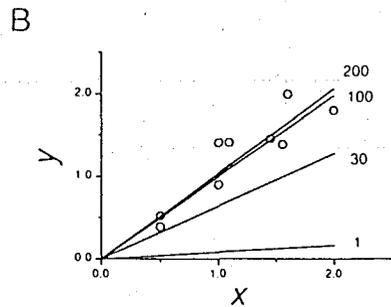
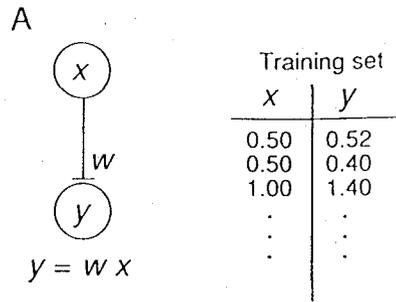
Exploiting the advantages of an invertebrate system, Lockery et al. (1989) were able to measure directly the input-output function of the local bending reflex in the leech, a simple withdrawal response, by stimulating sensory neurons and recording the resulting synaptic potentials in the motor neurons controlling the reflex. Their model incorporated biological constraints such as lateral inhibition between motor neurons, left-right symmetry among pairs of homologous interneurons, and solely excitatory sensory-to-interneuron connections. After optimization to the known input-output function, the receptive and projective fields of model interneurons closely matched those of identified local bending interneurons. The model showed that local bends in response to dorsal, ventral, or lateral stimuli could be produced by a single population of multifunctional interneurons. More recently Wittenberg

and Kristan (1992) have extended the model to include a second withdrawal behavior—shortening. Their results suggest that at least two new types of interneuron involved in leech withdrawal behaviors remain to be found.

Models of Real Circuits Optimized by Recurrent Backpropagation

Static backpropagation is sufficient for problems in which the biological network is, for all intents and purposes, a feedforward network. However, the great majority of neural networks possess feedback connections. Feedback connections require the use of recurrent backpropagation because such connections act over time. One place where feedback connections play a prominent role is in central pattern generators. Pearlmutter (1989) and Doya and Yoshizawa (1989) showed how recurrent backpropagation can be applied to idealized central pattern generators in which each neuron is a passive summing node. Rowat and Selverston (1991) extended this approach for specific cases like the gastric mill central pattern generator of the lobster, where the neurons have active conductances and electrical synapses. Lockery et al. (1990) applied recurrent backpropagation to a model of the leech local bending reflex that included two important features omitted from the previous static model (Lockery et al. 1989): feedback connections and the time course of physiologically recorded synaptic potentials. The recurrent backpropagation model of local bending is described in detail below. The numerous feedback connections in the vestibulo-ocular reflex have prompted a number of models using recurrent backpropagation. Anastasio (1991a) showed that inhibitory feedback connections among vestibular neurons in the brain stem could be the site of integration of the head-velocity signal in the VOR. This model also reproduces the recovery of the response properties of vestibular neurons following ablation of afferent inputs (Anastasio 1992), and suggests sites of plasticity underlying the recovery. Recurrent backpropagation has also been used to construct plausible models of the complex firing patterns exhibited by cortical neurons during the performance of various memory (Zipser 1991) and motor tasks (see chapter 7). A further model considers the possible role of basal ganglia in switching between motor tasks (Brotchie et al. 1991).

Two other groups have constructed dynamic models of the VOR, but used optimization strategies other than recurrent backpropagation. Arnold and Robinson (1990) used numerical differentiation. In this technique, the derivative of the error with respect to a given weight is estimated numerically by increasing the weight slightly and measuring the amount by which error increases or decreases. After doing this for every weight, the weights are updated by decreasing weights with positive derivatives and increasing weights with negative derivatives. Quinn et al. (1992) used a global optimization procedure in which minima of the error function are found by finding the roots of polynomial fits to the error function (Bremermann 1970). Both of these procedures are computationally



The Analogy Between Backpropagation and Curve Fitting

Figure 5.3 (A) On the left is a network with one input unit, one output unit, and one weight. The activity of the output unit y is the product of the weight w and the activity of the input unit x . In backpropagation, the task is to find the value of w that best reproduces the training data shown on the right. Note that x and y are used differently in chapter 3. (B) Optimization of w by backpropagation. The circles are the data points in the training set. The number of iterations of the training set is shown to the right of each w shown. A random initial value of w gives a poor fit to the data (1 iteration). The fit improves with each cycle of optimization (30, 100, and 200 iterations shown).

more expensive than backpropagation and so work best for relatively small networks.

Backpropagation and Biological Plausibility

The historical tendency to speak in terms of training when describing the process of adjusting connection strengths via backpropagation or other optimization algorithms has led some to question the relevance of network models to biology. In fact, none of the examples above used backpropagation as a model for the process by which connection strengths are established in the animal during development (ontogenesis) or during evolution (phylogenesis). The interest is strictly in the end point of the optimization process. In mathematical terms, the task of optimizing a model network to reproduce a desired input-output function is identical to a curve-fitting procedure in which the weights are the coefficients or parameters to be adjusted, and the input-output relations in the training set are the data to be fit. This point is illustrated in figure 5.3. Suppose

we wished to fit a line starting at the origin through the data points in figure 5.3B. In neural network terms, this amounts to supposing that the x coordinate of each point is the activation of an input unit and the y coordinate is the desired activation of an output unit (figure 5.3A). If the input-output function of the output unit is linear, then $y = wx$, and we wish to find the w that minimizes the error in predicting for each x the corresponding y in the data set. A random initial value of w gives a poor fit, but after 200 cycles through data, a satisfactory fit is achieved. Backpropagation is best viewed as one of many techniques for finding an optimum set of parameters when fitting a function. The issue of biological plausibility is moot as far as optimization is concerned. For example, linear regression is not biologically plausible, yet it is a perfectly legitimate means of fitting functions to biological data. This having been said, approximations to backpropagation that have greater biological plausibility have been studied (Mazzoni et al. 1991), and some undesigned-for similarities between backpropagation and neuronal networks are discussed in chapter 3.

Combinations of Compartmental and Neural Network Models

With the introduction of recurrent backpropagation it is possible to combine the strengths of compartmental and neural network models. In a recurrent network model, the state of each unit is its activity level. However, if the state is taken instead to represent a voltage, then each unit in the network model is formally equivalent to an electrical compartment in a compartmental model. In such a model, units communicate by passing currents rather than weighted activations, and the sigmoidal input-output function represents the relationship between presynaptic voltage and postsynaptic current. It thus becomes possible to use neural network training algorithms such as backpropagation to adjust connections between compartments in realistic models (Rowat and Selverston 1991).

A CASE STUDY: THE LEECH LOCAL BENDING REFLEX

Background

The local bending reflex of the leech proved to be an ideal preparation in which to apply the combined compartmental and network modeling approach to a real biological system. First, the reflex is subserved by a comparatively small number of sensory neurons, interneurons, and motor neurons, each of which is re-identifiable from animal to animal. Second, the input-output function of the reflex could be measured with unusual precision by pairwise intracellular recordings from sensory neuron and motor neurons, providing a more realistic training set than is normally available. Third, interneurons in the reflex receive multiple sensory inputs and form a distributed representation of stimulus location, mak-

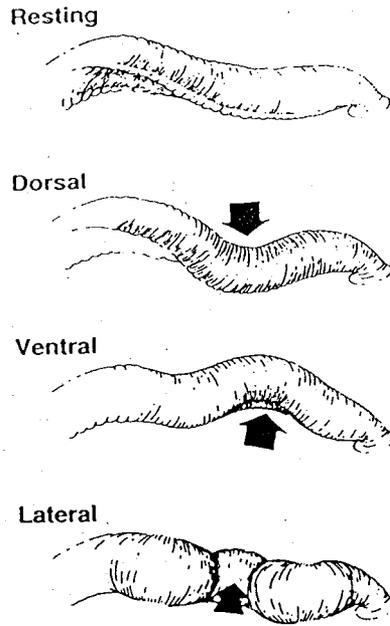
ing it difficult to construct a model by hand. Finally, physiological estimates of the connection strengths of interneuron inputs and outputs could be made. This allowed explicit comparisons between the connectivity of model and actual interneurons.

Local bending is a directed escape reflex (figure 5.4A). In response to a moderate mechanical stimulus, the leech withdraws from the site of contact. This is accomplished by contracting longitudinal muscles beneath the stimulus, and relaxing longitudinal muscles on the opposite side of the body, resulting in a U-shaped local bend. Thus, a dorsal stimulus causes a dorsal bend, a ventral stimulus a ventral bend, and a lateral stimulus a lateral bend. The local bending reflex can be elicited in a semi-intact preparation consisting of a single midbody ganglion and the associated segment of body wall. It has thus been possible to identify the sensory neurons, motor neurons, and a subpopulation of the interneurons contributing to the reflex (Kristan 1982; Lockery and Kristan 1990b; Nicholls and Baylor 1968). Each ganglion contains sensory neurons responding to touch and pressure (Nicholls and Baylor 1968). Three pairs of T cells respond selectively to touch on the dorsal (TD), ventral (TV), and lateral (TL) body surface. In addition, there are two pairs of P cells, which respond to pressure on the dorsal (PD) or ventral (PV) body surface, respectively. Several types of motor neurons innervate the longitudinal muscles that control the reflex (Ort et al. 1974; Sawada et al. 1976; Stuart 1970). These include excitors and inhibitors of the dorsal longitudinal muscles (DEs and DIs), excitors and inhibitors of the ventral longitudinal muscles (VEs and VIs), an excitor of lateral longitudinal muscle, and the L cell, which has a combined dorsal and ventral field. The innervation of each motor neuron type is restricted to a single side of the body (left or right).

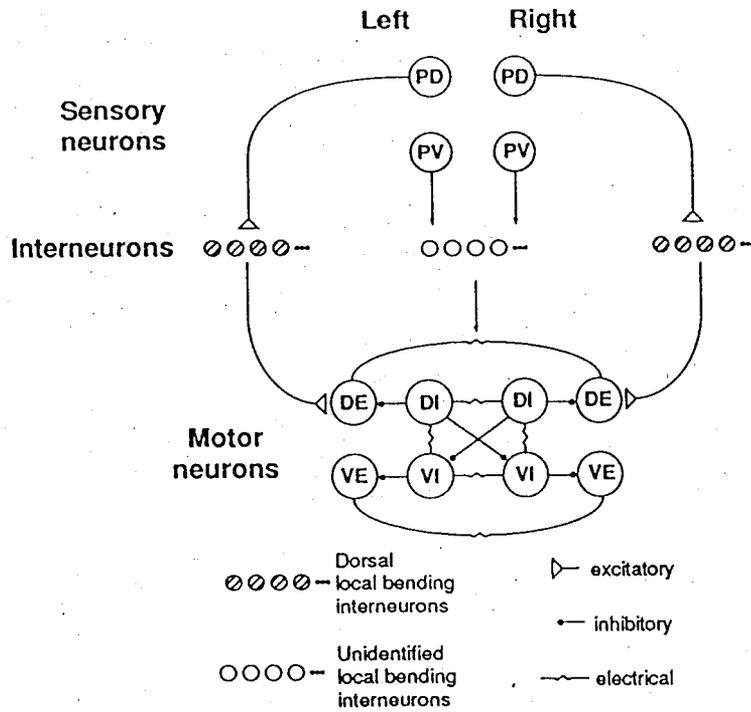
As a first step in understanding how local bending is produced, we focused on a representative subset of the sensory and motor neurons contributing to local bending (figure 5.4B). We used activation of P cells as sensory input to the reflex since dorsal, ventral, and lateral stimuli each activate at least one P cell, and the effect of P cells is greater than that of the T cells (Kristan 1982). Motor outputs were monitored by recording from DEs, DIs, VEs, and VIs. The lateral motor neuron was not monitored in these initial investigations because, for technical reasons, it is somewhat more difficult to locate and identify. The L cell was not monitored in this initial study because its response is much briefer than the response of the other motor neurons (Kristan 1982).

The simplified local bending network retains essential features of the complete network. The spatial location of sensory input is encoded on an array of sensory neurons, and motor output on an array of motor neurons, which are highly interconnected by both chemical and electrical synapses. The connection between sensory neurons and motor neurons is mediated by a layer of interneurons. In computational terms, the task of the interneurons is to compute the input-output function of the

A



B



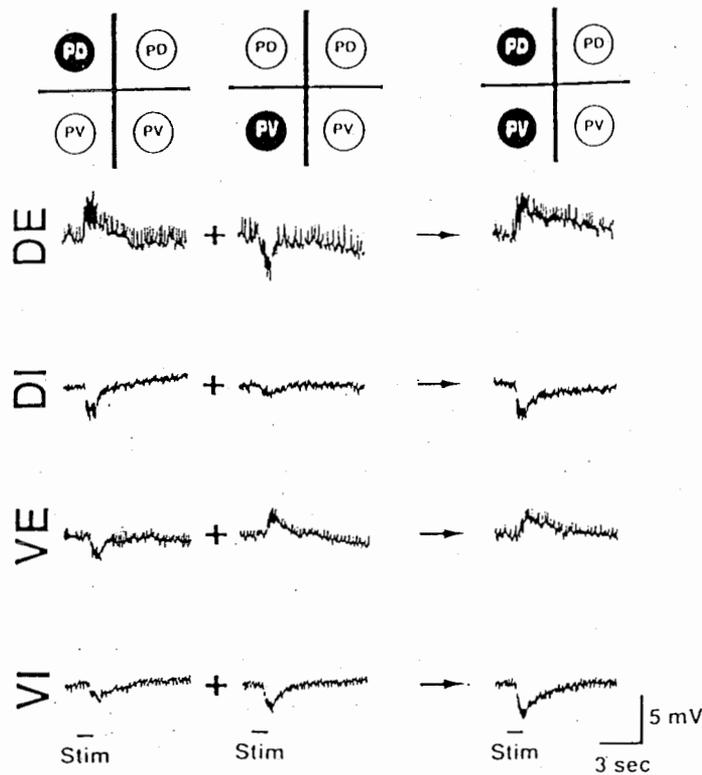
reflex: the mapping between each pattern of P cell stimulation and the pattern of motor neuron excitation and inhibition sufficient to withdraw from the stimulus. Synaptic transmission from sensory neurons to interneurons involves action potentials in the presynaptic neuron (Lockery and Kristan 1990b), but transmission from interneurons to motor neurons (and between motor neurons) does not; as in the vertebrate retina (Roberts and Bush 1981), synaptic activation is a graded function of presynaptic voltage.

Measuring the Input-Output Function

We began our analysis with the input-output function of the local bending reflex (figure 5.5; Lockery and Kristan 1990a). Using a standard stimulus sufficient to elicit local bending responses, P cells were activated, in pairs or individually, while intracellular recordings were made from all eight motor neurons in the simplified network. We examined in detail the motor neuron synaptic potentials produced by eight patterns of sensory input. These included four pairs of P cell stimulation (the left and right PD, the left and right PV, left PD and PV, right PD and PV), as well as each P cell stimulated individually. In general, the response of the network to stimulation of pairs of P cells was well predicted by the sum of the responses to the same two P cells stimulated individually. For example, the motor neuron VI, in response to ipsilateral stimulation of PV and PD, receives an IPSP that is approximately the linear sum of the IPSP produced by individual stimulation of PV and PD. However, there were exceptions to this, as in the case of the ipsilateral VE, which receives an EPSP in response to ipsilateral stimulation of PV and PD, despite the fact that the individual effects of PV and PD tend to cancel each other out. In this case, the nonlinearity of the response to the paired P cell pattern is in the direction that is behaviorally efficacious, since increased excitation of VE promotes withdrawal from the site of stimulation. Similar results were found for other patterns of paired P cell stimulation. Thus, the basic patterns of motor output underlying local bending behavior are produced by the linear combination of individual

Local Bending in the Leech: Behavior and Circuitry

Figure 5.4 (A) Local bending behavior. The leech withdraws from the site of contact by contracting longitudinal muscles beneath the stimulus, and relaxing longitudinal muscles on the opposite side of the body. (B) Simplified local bending circuit. The strongest input to the reflex is provided by pressure sensitive mechanoreceptors, the P cells, which have dorsal (PD) or ventral (PV) receptive fields. Behavior is produced, in part, by dorsal and ventral excitatory (DE, VE) and inhibitory (DI, VI) motor neurons. The contributions of longitudinal muscle motor neurons with lateral projective fields, as well as the motor neurons controlling circular and oblique muscles, have not been studied in detail. Input from sensory to motor neurons is mediated by a layer of interneurons, a subpopulation of which have been identified.



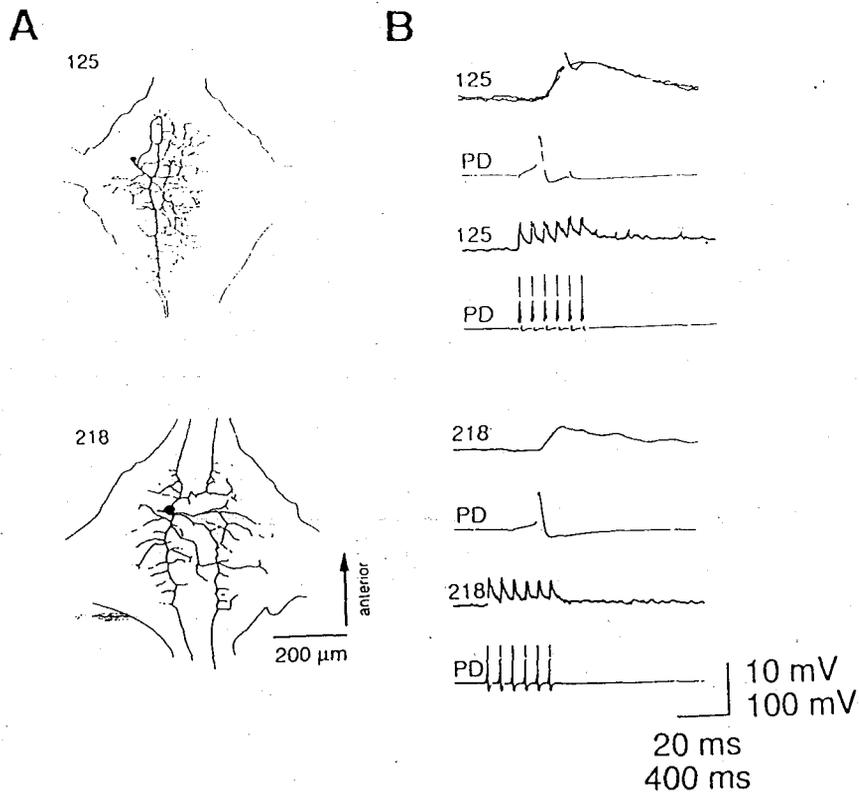
Local Bending Input-Output Function

Figure 5.5 Synaptic potentials recorded intracellularly from four motor neurons in response to single and paired P cell stimulation. The filled circles at the top of each column indicate which P cells were stimulated. Each row shows the synaptic potential recorded in the indicated motor neuron. Motor neurons have projective fields ipsilateral to the sensory neurons; recordings from contralateral motor neurons were also made but are not shown. The duration of the stimulus is shown by the bars below. All recordings are from the same preparation. Similar experiments were made for all possible patterns of paired P cell stimulation. (Reprinted from Lockery and Kristan 1990a by permission of the *Journal of Neuroscience*.)

P cell effects, with minor enhancements due to a nonlinearity in the system (Lockery 1989).

Identification of Interneurons

As a working hypothesis of sensorimotor integration in the reflex, we adopted a *dedicated interneuron* model in which each form of local bending is under the control of a single interneuron, or type of interneuron. We imagined that dorsal local bending is produced by one or more specific, dorsal bending interneurons that are excited by dorsal P cells and receive no input from (or are inhibited by) ventral P cells. The unique pattern of excitation and inhibition of longitudinal muscle motor neurons is determined by the pattern of excitatory and inhibitory con-



Identifying Properties of Local Bending Interneurons

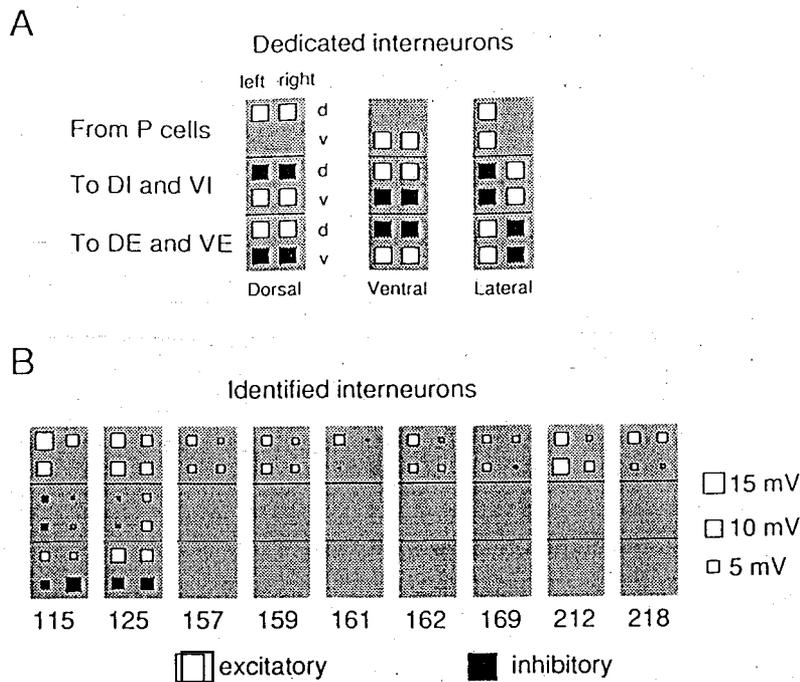
Figure 5.6 (A) Morphology. Camera lucida drawings were made from whole mounts of cells injected ionophoretically with Lucifer yellow. Identifying numbers were assigned according to the most common location of their cell body on a standard map of the ganglion (Muller et al. 1981). (B) Physiological properties. Within each panel, the top trace shows the response of the interneuron to the single PD action potential shown in the second trace. The third trace shows the response to the train of PD impulses shown in the fourth trace. In some traces, two or three sweeps (several seconds apart) are superimposed to show the extent to which latencies were consistent. Vertical calibration: first and third traces, 10 mV; second and fourth traces, 100 mV. Horizontal calibration: first and third traces, 20 ms; second and fourth traces, 400 ms. (Reprinted from Lockery and Kristan 1990b by permission of the *Journal of Neuroscience*.)

nections from dorsal bending interneurons to motor neurons. Similar types of interneurons are specific for ventral and lateral bends. Because one or the other of the P cells stimulated in lateral local bending is also stimulated in dorsal or ventral bending, the lateral bending interneurons must be capable of overcoming the effects of the dorsal and ventral bending interneurons. This could be accomplished by inhibitory connections from the lateral local bending interneurons onto other interneurons.

To determine whether local bending is produced by dedicated interneurons of the type we hypothesized, we sought to identify interneu-

rons in the leech contributing to local bending and to determine their sensory and motor neuron connections. As a first step, we searched for interneurons contributing to dorsal local bending; that is, interneurons that received excitatory input from at least one dorsal P cell, and in turn excited at least one DE motor neuron. Interneurons were impaled and a dorsal P cell was stimulated to fire a train of action potentials sufficient to elicit local bending (10 Hz, 0.5 s). Any neuron that received an excitatory synaptic potential from the dorsal P cell was then depolarized by current injection while action potentials of the DE motor neuron were recorded extracellularly. Neurons that were excited by the dorsal P cell and excited DE were considered dorsal bending interneurons. Each was filled with a fluorescent dye and its morphology was recorded. A survey of more than 70 ganglia yielded a sample of 235 dorsal bending interneurons. On the basis of morphological and physiological criteria (figure 5.6), nine types of dorsal bending interneuron were repeatably identifiable from preparation to preparation. Eight of these were bilaterally paired neurons and one was an unpaired cell; thus, each ganglion contained 17 dorsal bending interneurons. Each interneuron sent at least one axon into the interganglionic connective (figure 5.6A), very likely projecting to the adjacent ganglion where it may make both input and output connections (Wittenberg and Kristan 1992). As a group, the 17 interneurons send 16 axons into the anterior connective and 12 axons to the posterior connective. Thus, each ganglion potentially contains functional processes of at least 28 additional local bending interneurons, bringing the number of local bending interneurons to 45. If the axons of local bending interneurons extend more than one ganglion, like other interneurons in the leech (Brodfuehrer and Friesen 1986), the number rises to 73 or more. Since it seems likely that a search for interneurons contributing principally to ventral or lateral bends would identify other interneurons, this number is expected to go even higher.

Each morphologically defined type of local bending interneuron was associated with a unique intracellular recording in response to P cell stimulation (figure 5.6B). Thus, interneurons differed systematically in the size and shape of the synaptic potential produced by single P cell action potentials, the presence or absence of P cell evoked action potentials, as well as in the presence of a fast or slow component in the response to trains of P cell action potentials sufficient to evoke the behavior. The presence or absence of action potentials is interesting because it suggests that the effects of some interneurons are mostly local, given the P cell stimulus intensities used, while others might have both local and distal effects. Similarly, the difference in fast and slow responses to trains of P cell action potentials suggests that longer sensory stimuli might select a different subset of interneurons than are selected by short stimuli, since the response of interneurons with substantial slow components will be selectively enhanced. In principle, therefore, the heterogeneity among interneurons in their response to



Connectivity of Hypothesized and Identified Local Bending Interneurons

Figure 5.7 (A) Hypothesized dedicated interneurons. Gray dominoes show the input and output connections of a single interneuron. The upper third of each shows connections from sensory neurons to the interneuron; the middle third shows the connections from the interneuron to the inhibitory motor neurons; the bottom third shows the connection from the interneuron to the excitatory motor neurons. White boxes indicate excitatory connections; black boxes indicate inhibitory connections. Interneurons specific for dorsal, ventral, and lateral are shown. (B) Identified dorsal local bending interneurons. Symbols are as in A, except that box area is proportional to the amplitude of the connection determined from intracellular recordings of interneurons or motor neurons. Blank spaces denote connections whose strength has not been determined.

sensory input could be the basis for differences in the spatial and temporal patterns of motor output as a function of stimulus intensity and duration, thus increasing the computational repertoire of the network.

The dedicated interneuron model predicts that dorsal bending interneurons will have input and output connections like those in figure 5.7A. To determine if this was the case, we measured the strength of the connections from each P cell to a given interneuron by activating each P cell individually with a standard intracellular stimulus and noting the peak amplitude of the synaptic potential in the interneuron. The strength of output connections to motor neurons was measured in a similar fashion. However, we were hampered by the fact that the motor neurons lie on the dorsal surface of the ganglion, while all but two of the interneurons lie on the ventral surface. At present, therefore, output connections are known only for cells 115 and 125. The connectivity of

the dorsal bending interneurons was consistent with the dedicated interneuron model in that most interneurons received substantial excitatory input from both dorsal P cells, and cells 115 and 125 excited both DEs and inhibited both VEs (figure 5.7B). However, the results were inconsistent with the dedicated interneuron model in two respects. First, with one exception, each interneuron received substantial input from at least one ventral P cell. This means that these interneurons are active during ventral local bends, when their excitatory effects on DEs are inappropriate. This also means these interneurons are active during lateral local bending, when the effect on the contralateral DE is inappropriate. Second, the effect of cells 115 and 125 on the inhibitory interneurons was not always the opposite of the effect on the excitatory motor neuron with the same projective field; for example, cell 125 excites both DE and DI on the right side.

Motivation for the Model

That most local bending interneurons are active in dorsal, ventral, and lateral local bends suggested that the simple, dedicated interneuron model is incorrect. Rather, the interneurons appear to form a distributed representation of sensory input. In the distributed model, each form of the local bending behavior is produced by the simultaneous action of many interneurons, whose combined outputs produce the associated motor output pattern. The distributed model thus has the advantage that it can account for ventral inputs to dorsal bending interneurons (figure 5.7B) and the conflicts between interneuron outputs to motor neurons. In this view of the network, connections of an interneuron that are inconsistent with one motor output pattern are offset by behaviorally consistent connections from other interneurons.

Although the distributed processing hypothesis was attractive, it remained to be shown whether it was consistent with the physiological details of the biological network. Specifically, was the distributed model consistent with the many lateral connections among the motor neurons, with the amplitude and time course of motor neuron synaptic potentials in response of all different patterns of sensory input, and with the type of interneurons identified?

Architecture of the Model

To test the distributed processing hypothesis, we constructed a model of the reflex and used backpropagation to adjust connection strengths between neurons (Lockery et al. 1990). The model comprised four sensory neurons (P cells), up to 40 interneurons, and eight motor neurons (figure 5.8A). Sensory neurons responding to touch (T cells; Nicholls and Baylor 1968) were excluded from this preliminary model, since their effects on behavior are small in comparison to the effects of the P cells (Kristan 1982). The number of interneurons represented, in round num-

bers, an upper estimate of total number of local bending interneurons, based on the number of local bending interneurons identified to date within a single ganglion, and the number of axons from local bending interneurons projecting into the ganglion from adjacent ganglia (Lockery and Kristan 1990b). Each of the eight motor neurons in the model represented one of eight types of longitudinal muscle motor neurons, in which there are two to four neurons per type (Ort et al. 1974; Stuart 1970). These comprise the excitors and inhibitors of dorsal longitudinal muscle, DE and DI, respectively, and the excitors and inhibitors of ventral longitudinal muscle, VE and VI, respectively.

All known chemical and electrical synaptic connections between motor neurons were included and connection strengths were determined from previous physiological recordings (Granzow et al. 1985; see below). Weights of feedforward connections from sensory neurons to interneurons, and from interneurons to motor neurons were adjusted by the backpropagation algorithm (Rumelhart et al. 1986a) so that the model reproduced the amplitude and time course of motor neuron synaptic potentials recorded in response to single and paired P cell stimulation in eight different patterns (Lockery and Kristan 1990a). The training procedure was not allowed to insert connections between interneurons, since functional connections of this type have not been found; nor was the training procedure allowed to insert feedback connections from motor neurons to interneurons, since only one such connection had been identified (Friesen 1989b).

The present model is more realistic than the previous model (Lockery et al. 1989), which excluded electrical and chemical synapses that would introduce feedback among the motor neurons. Feedback leads to networks whose activity evolves in time. Thus recurrent backpropagation (Pearlmutter 1989) was used instead of static backpropagation (Rumelhart et al. 1986a) to train output units (motor neurons) in the network to follow a predetermined time course. The present model is also more realistic because the response of model motor neurons matches the real motor neurons not only in amplitude, but also in time course. By incorporating temporal dynamics, the model can predict the time course of synaptic potentials in the interneurons. (An additional discussion of dynamic recurrent networks is provided in chapter 7.)

Model Neurons

Neurons were modeled as passive, single electrical compartments having in parallel an input resistance R and capacitance C to ground. In later models, multicompartmental neurons could be used, but one compartment was sufficient for the theoretical issues raised by the physiological data currently available. The rate of change of membrane potential (dV/dt) was determined by the present voltage and the sum of the

current introduced by synaptic inputs using equation 5.1a, indexed by each neuron i :

$$\tau_i \frac{dV_i}{dt} = -V_i + RI_i \quad (5.1b)$$

where τ_i is the neuronal time constant, and I_i is the sum of the current introduced by chemical and electrical synapses. The use of single rather than multicompartmental model neurons is a simplifying assumption based on the assumption that the interneurons are nearly isopotential during the 500 ms long stimulus trains used in physiological experiments (Lockery and Kristan 1990a).

Trains of action potentials in sensory neurons were modeled as stepwise increases in presynaptic voltage. To reduce computation, action potentials within the trains were not represented individually, since the duration of the action potential is short with respect to the time constants for chemical synaptic transmission. The amplitude of the step was scaled to keep the range of synaptic potentials in the interneurons within the range observed in intracellular recordings (Lockery and Kristan 1990b).

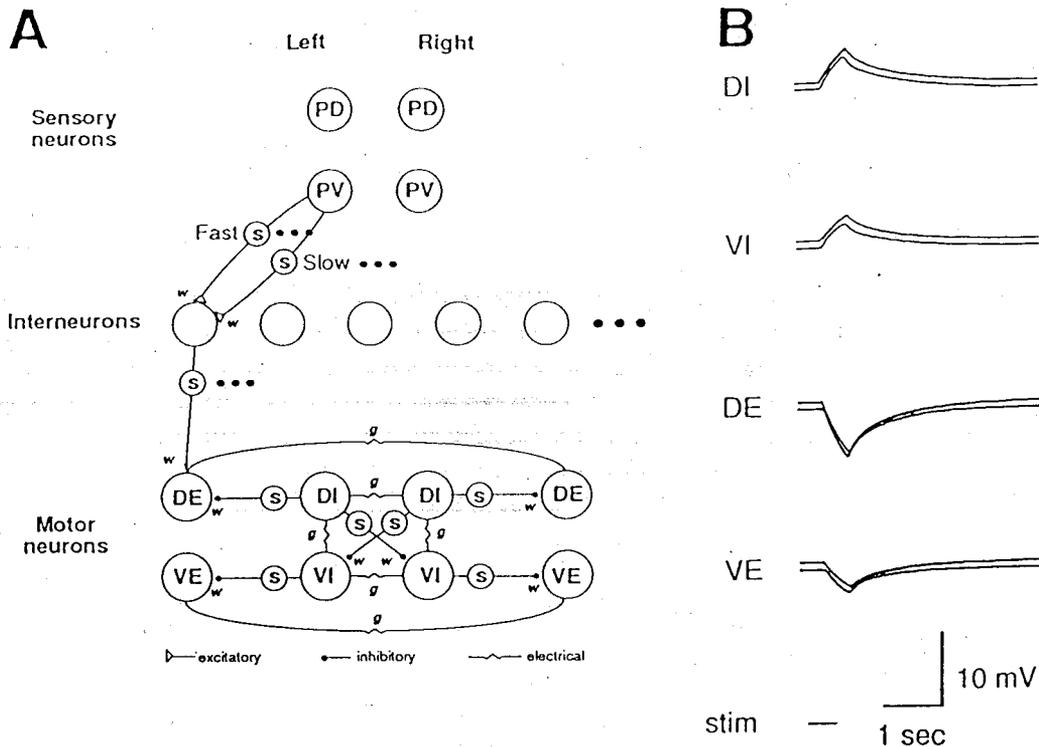
Synapses

Electrical synapses were modeled as ohmic conductances. Current introduced by electrical synapses was given by Ohm's law, and summed linearly according to the equation:

$$I_j^{elec} = \sum_i g_{ji} (V_i - V_j) \quad (5.2)$$

where g_{ji} is the conductance of the electrical synapse between neurons i and j . Within the ganglion, chemical synaptic transmission between inhibitory and excitatory motor neurons (Granzow et al. 1985), as well as from interneurons to other interneurons in the leech (Angstadt and Calabrese 1991; Friesen 1985; Thompson and Stent 1976), is a graded function of presynaptic voltage; it does not require the occurrence of action potentials, nor is it substantially affected by them. The synapse between inhibitory and excitatory motor neurons was used as a model for all chemical synapses in the network, since this connection has been most intensively studied. In physiological experiments (Granzow et al. 1985) a stepwise increase in presynaptic voltage in DI produced a slowly rising postsynaptic potential in DE. To account for the long synaptic rise time, synapse units (s-units, figure 5.8A) were inserted between pairs of neurons connected by chemical synapses. These can be viewed as additional compartments in which the state variable is synaptic activation rather than voltage. The activation of each s-unit S_{ji} was given by:

$$\tau_{ji} \frac{dS_{ji}}{dt} = -S_{ji} + f(V_i) \quad (5.3)$$



Network Model of the Local Bending Circuit

Figure 5.8 (A) Four sensory neurons were connected to eight motor neurons via a layer of 10 interneurons. Neurons were represented as single electrical compartments whose voltage varied as a function of time. Known electrical and chemical connections among motor neurons were assigned fixed connection strengths (g and w in the motor layer) determined from intracellular recordings. Interneuron input and output connections were adjusted by recurrent backpropagation. Chemical synaptic delays were implemented by inserting s -units between chemically connected pairs of neurons. S -units with different time constants were inserted between sensory and interneurons to account for fast and slow components of synaptic potentials recorded in interneurons. (B) Response of the model network to stimulation of the left ventral P cell after synaptic connections had been optimized using backpropagation. Each tracing shows the simulated intracellular recording from a single motor neuron, together with the target response from the data set against which connections in the model were optimized. To facilitate comparison of model and target synaptic potentials, the latter have been shifted upward by 1 mV. (Reprinted from Lockery and Sejnowski 1992 by permission of the *Journal of Neuroscience*.)

where τ_{ji} is the synaptic time constant and $f(V_i)$ is a sigmoidal function ($0 \leq f(V_i) \leq 1$) relating presynaptic and postsynaptic membrane potential.

The s-unit time constants combined the temporal dynamics of all the steps in transmitter release. Single s-units were inserted between interneurons and motor neurons, and between pairs of motor neurons. Preliminary simulations showed that the time course of motor neuron synaptic potentials, which exhibit a fast and slow exponential decay, could not be produced by a local bending model having only one s-unit with a single time constant at each connection. Therefore, two s-units—one with a fast and one with a slow time constant—were inserted between sensory neurons and interneurons. This was consistent with intracellular recordings of synaptic potentials from interneurons in response to P-cell stimulation, which showed both fast, slow, and mixed fast and slow synaptic potentials (Lockery and Kristan 1990b).

In equation 5.3, the sigmoidal function $f(V_i)$ represented the many processes which limit synaptic current, including limits on presynaptic calcium accumulation and number of vesicles released, the saturation of binding of transmitter to receptors, and the reversal potential of the synaptic conductance. The form of the sigmoidal function was determined by extrapolating from existing data on steady-state postsynaptic voltage as a function of presynaptic voltage at the synapses between two different inhibitory-excitatory motor neuron pairs (Granzow et al. 1985). Threshold for synaptic transmission was placed at the resting potential ($V = 0$), consistent with the fact that the inhibitory motor neurons do not release transmitter at rest (Lockery and Kristan, unpublished). Total current due to chemical synapses I_j^{chem} was given by:

$$I_j^{chem} = \sum_i w_{ji} S_{ji} \quad (5.4)$$

where w_{ji} is the strength of the chemical synapse to neuron j from neuron i , and is functionally similar to the product of the maximum synaptic conductance and the driving force ($E_{rev} - V_m$). Thus, in the model, currents from chemical synapses added linearly. Combining equations 5.1b through 5.4 yields the final equation:

$$\tau_j \frac{dV_j}{dt} = -V_j + R_j \left[\sum_i g_{ji} (V_i - V_j) + \sum_i w_{ji} S_{ji} \right] \quad (5.5)$$

Input Resistance and Time Constants

Motor neuron input resistance (20 M Ω) and time constant (10 ms) were estimated from the time course and amplitude of the postsynaptic potential in DE in response to depolarizing current step (1.5 nA) injected into DI (Granzow et al. 1985). Interneuron input resistance was assumed to be

higher (40 M Ω) because the somata of most local bending interneurons are smaller than the somata of the motor neurons. For simplicity, the time constants for the interneurons were the same as for the motor neurons. The time constants for the s-units between motor neurons were chosen to reproduce the time course of postsynaptic voltage in DE upon injection of a square step of depolarizing current in the presynaptic motor neuron cell DI. For simplicity, the same value was chosen for the s-units for the connection from interneurons to motor neurons. This connection was chosen as the standard synapse because it is an established monosynaptic connection (Granzow et al. 1985); the same time constant was used for s-units between interneurons and motor neurons. Time constants for the fast (10 ms) and slow (1500 ms) s-units between sensory and interneurons were fit by hand, using a simplified model network comprising one sensory neuron, one interneuron, and one motor neuron. The values selected provided the best fit to the time course of the intracellularly recorded motor neuron synaptic potential in response to P cell stimulation.

Fixed Weights

The coupling resistance between pairs of electrically coupled motor neurons was calculated to achieve an effective coupling ratio of 0.1, in accordance with pairwise intracellular recordings from motor neurons (Friesen, unpublished; Lockery and Kristan, unpublished). Weights for chemical synapses between motor neurons were adjusted by hand to reproduce the steady-state response of the postsynaptic motor neuron to maximum depolarization of the presynaptic neuron (Granzow et al. 1985).

Optimized Weights

Connection strengths to and from interneurons were optimized by backpropagation to match the time course and amplitude of motor neuron synaptic potentials in the model to those in physiological recordings. The synaptic potential time course of a representative motor neuron recording was adopted as a template (figure 5.8B, upper traces). To construct the training set for backpropagation, the amplitude of the template was scaled to the average peak amplitude for each motor neuron in response to each of the eight patterns of single and paired P cell stimulation (Lockery and Kristan 1990a; figures 5.4 and 5.5). In accordance with the backpropagation algorithm, connections from sensory neurons to interneurons and interneurons to motor neurons were initially small and randomly assigned. For each iteration of the optimization procedure, an input pattern from the data set was presented and the network's behavior was simulated for 100 time steps (10 ms/step). The time course and amplitude of the simulated synaptic potential in each motor neuron was then compared to the desired time course for

the relevant motor neuron in response to the same pattern of P cell stimulation; this gave a measure of the instantaneous error of the network at each time step. This procedure was repeated for each pattern in the data set. Each connection in the network was then adjusted by a small amount as specified by the algorithm and the process was repeated until the average instantaneous error was less than 0.19 mV.

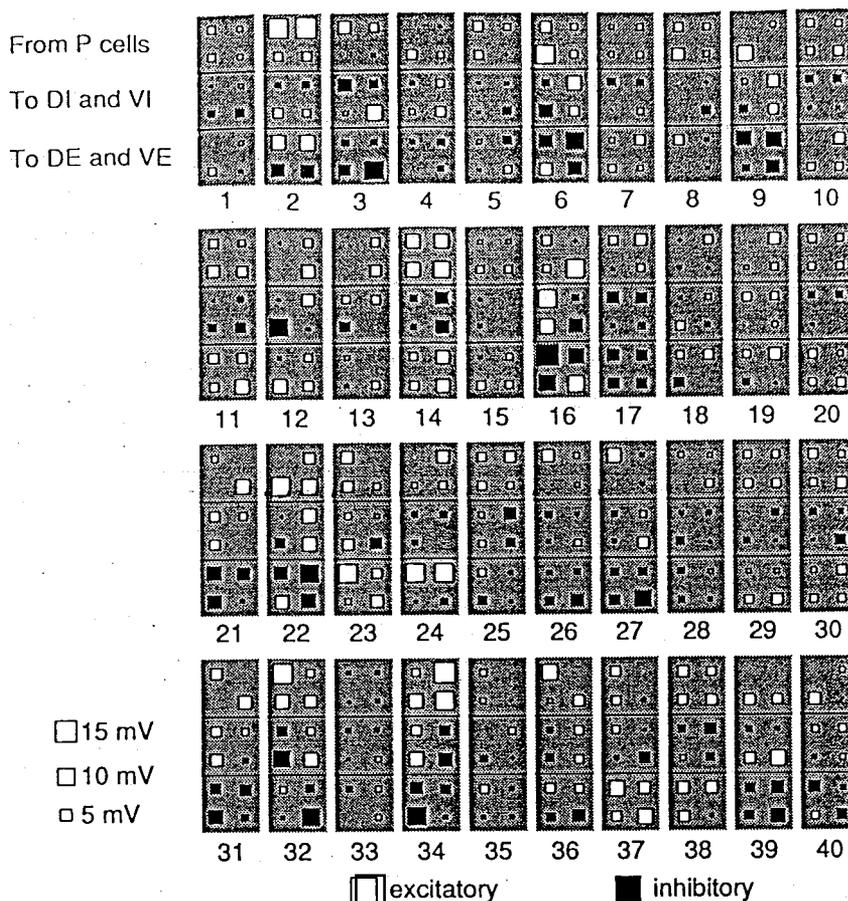
Weights from sensory neurons to interneurons were constrained to be positive, since no inhibitory connections have been observed between such pairs. No constraints were placed on connections from interneurons to motor neurons. After 10 to 20 thousand iterations, a good match was obtained between simulated and desired motor neuron synaptic potentials (figure 5.8B, lower traces) for each input-output relation in the data set.

Performance of the Model

To compare interneurons in the model network to actual interneurons, simulated physiological experiments were performed. Interneuron input connections were determined by recording the amplitude of the postsynaptic potential in a model interneuron while each of the P cells was stimulated with a standard current pulse. Output connections were determined by recording the amplitude of the postsynaptic potential in each motor neuron when an interneuron was stimulated with a standard current pulse. This analysis yielded strong similarities between the model and biological network. First, model interneurons (figure 5.9), like those in the real network (figure 5.7B), generally received three or four substantial connections from P cells and thus received both dorsal and ventral P cell inputs. This means that interneurons in the model contributed to dorsal, ventral, and lateral local bends, as the distributed hypothesis holds. Second, individual interneurons had significant effects on seven or eight of the motor neurons. As in the biological network, the effect on an inhibitory motor neuron was not always the opposite of the effect on the excitatory motor neuron with the same projective field. For example, interneuron 12 (figure 5.9) excites both the right-hand DE and the right-hand DI. Finally, the range of fast, slow, and mixed synaptic potentials recorded in the interneurons in response to P cell stimulation was also reproduced by the model (figure 5.10). An interneuron in the model responded with a fast, slow, or mixed synaptic potential depending on the relative weights assigned to the fast and slow synaptic processes represented by the two different connections between each sensory neuron and interneuron in the model (figure 5.8A).

The qualitative similarity between the connectivity and temporal response properties of model and actual interneurons established the central result that a population of interneurons resembling the identified dorsal local bending interneurons could mediate local bending in a distributed

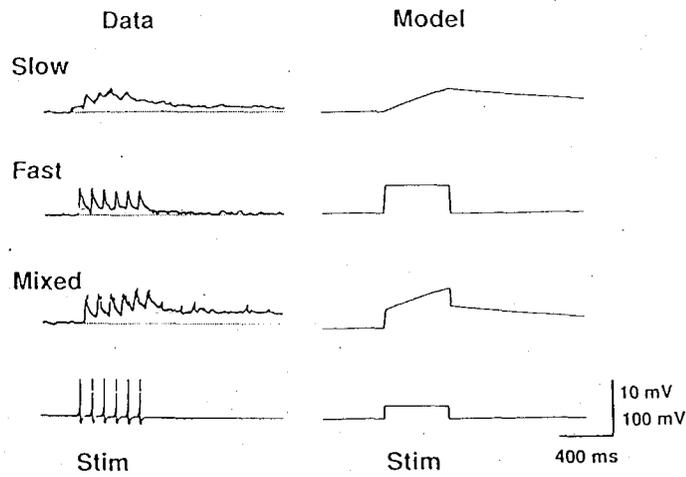
Model interneurons



Input and Output Connections of Model Local Bending Interneurons

Figure 5.9 Model interneurons, like the actual interneurons, received substantial inputs from three or four sensory neurons and had significant effects on most of the motor neurons. Symbols as in figure 5.7.

processing system without additional interneurons specific for different forms of local bending. This constitutes a plausibility claim or an existence proof (Crick 1989) that the actual local bending circuitry could operate as we suppose. This was an important result because beforehand we thought that dedicated interneurons might be necessary to achieve accurate amplitudes and time courses for each motor neuron in response to each pattern of sensory input. Note that a negative result—failure of the optimization procedure to find a solution to the local bending input-output function—would have told us nothing. This is because the true absence of a solution cannot be distinguished from the case in which the solution is simply missed by the optimization procedure. This is a general problem with optimization via gradient descent, since it can become trapped in a local minimum, i.e., stuck with a set



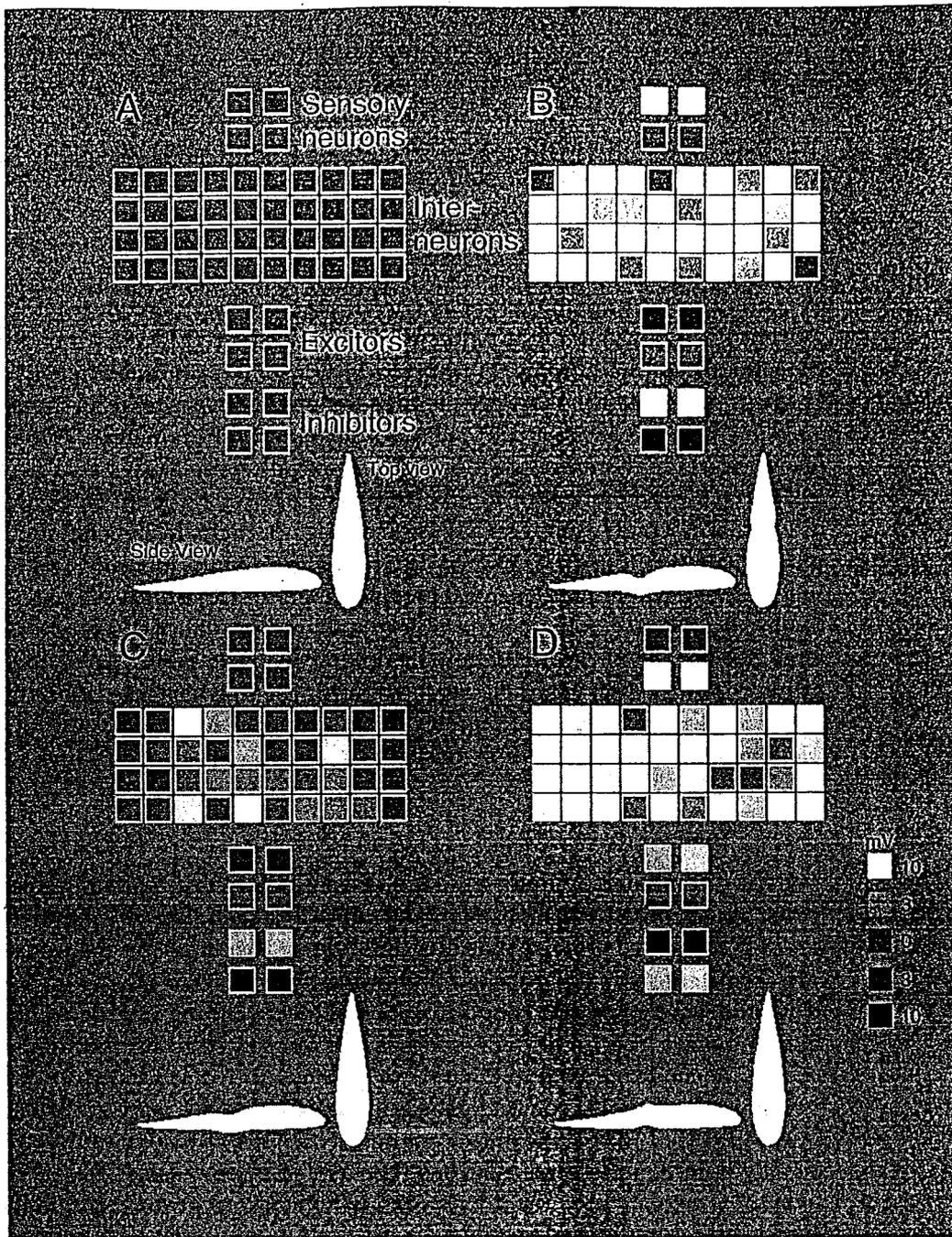
Recorded and Simulated Synaptic Potentials

Figure 5.10 Comparison of recorded and simulated synaptic potentials in interneurons in response to stimulation (Stim) of a single P cell. (Data) Physiological recordings are shown on the left. (Model) Simulated synaptic potentials are shown on the right. Interneurons with fast, slow, and mixed responses were seen in physiological recordings; similar types were seen in the model after training. In the simulations, the trains of P cell impulses were modeled as a step increase in average membrane potential in the sensory neuron. (Reprinted from Lockery and Sejnowski 1990 with permission.)

of connection strengths that are only a partial solution to the problem. The danger of local minima can be mitigated to some degree by training many networks from different random initial weight values, and by using different algorithms (Kirkpatrick et al. 1983).

The construction of a complete and realistic model of the local bending reflex makes it possible for the first time to visualize the response of the entire network as it evolves in time (figure 5.11). This analysis showed that in the model almost every interneuron contributed to the rapid rising phase of the synaptic potentials in the motor neurons (figure 5.11B). By contrast, the slow falling phase was under the control of many fewer interneurons (figure 5.11C). This observation suggests that, in the model, a single pattern of sensory input is represented by different patterns of interneuron activation at different times relative to stimulus onset and offset. Similarly, a single pattern of motor neuron excitation and inhibition is represented by different interneuron activity patterns at different times relative to the stimulus. The concept of time-dependent sensory and motor representations has received relatively little attention because models of sensorimotor integration have tended to focus on static representations. The idea of time-dependent representations can now be tested by physiological experiments to determine whether the biological network exhibits a similar time dependence.

Comparing the response to different patterns of P cell stimulation showed that most interneurons contributed to more than one form of



The Response of the Complete Local Bending Network as a Function of Time

Figure 5.11 (A) Resting. (B) Rising phase of the response to dorsal stimulation. (C) Falling phase of the response to dorsal stimulation. (D) Rising phase of the response to lateral stimulation. In each panel, the shading of the squares shows the voltage of the indicated sensory, motor, or interneuron, according to the gray scale at the left. The leech's profile is shown in side view, with anterior to the left, and in top view, in which anterior is up.

local bending, but each pattern of sensory input was associated with a unique spatial pattern of interneuron activation (figure 5.11D). The spatial pattern of interneuron activation is decoded by the pattern of synaptic connections to the motor neuron such that each interneuron activation pattern generates a unique—and behaviorally correct—pattern of motor neuron excitation and inhibition. An interesting theoretical question is whether distributed representations across as many as 40 interneurons are necessary. Two possibilities arise. First, the large number of interneurons may be required to accommodate additional complexities in the input-output function of the reflex. So far, we have measured the motor output in response to eight patterns of single or paired P cell stimulation, and these measurements were made at a single stimulus intensity and duration. Clearly, there are many more input patterns to investigate as one varies the number of P cells and the stimulus parameters. These experiments may reveal additional input-output complexities that explain the large numbers of interneurons that appear to be involved in the reflex. Second, each interneuron may also participate in behaviors other than local bending. Indeed, cell 115 was initially identified as a member of the swimming central pattern generator (Friesen 1989a), and both cell 115 and 125 have detectable effects on shortening (Wittenberg and Kristan 1992). Participation in other behaviors could force the connections of each interneuron to compromise between its different behavioral roles, necessitating additional interneurons to make up for the attendant functional shortcomings.

CONCLUSION

We have combined compartmental and neural network modeling approaches to construct a realistic model of the local bending reflex in the leech. The model showed that identified dorsal bending interneurons are consistent with a distributed processing hypothesis of the network. The model suggests that subpopulations of local bending interneurons that remain to be identified—such as the ventral bending interneurons—are similar in kind to the interneurons identified so far. Addition of more physiological constraints would make the model more realistic, and could lead to specific and accurate predictions of the connectivity of as yet unidentified interneurons (Lockery and Sejnowski 1992). For example, we can now train networks in which one group of model interneurons is constrained to have all the known input and output connections of the dorsal bending interneurons. A second, undifferentiated group of model interneurons would then be free during training to develop connections that predict the complementary input and outputs of the actual interneurons that act in concert with the identified dorsal bending interneurons. These predictions could then be tested by identifying the remaining local bending interneurons in the ganglion.

Backpropagation and other network training algorithms enable us to

construct working models of distributed processing networks that were not possible a few years ago. By using neural network units to represent electrical and synaptic compartments, it is possible to make these models as realistic as compartmental models themselves. In the future, we can expect this approach to be applied to other issues in systems neuroscience, including mechanisms of central pattern generation in invertebrates and vertebrates, the processing of temporal information, and the prediction of sites of synaptic-plasticity. This chapter has focused on networks of neurons. However, the network modeling approach has potential biological applications far wider than the nervous system, since the units in the network can represent other biological variables, including environmental (Lockery and Sejnowski 1991) and developmental factors (Mjolsness et al. 1991), as well as aspects of the immune response (Vertosick and Kelly 1991).

ACKNOWLEDGMENT

Supported by a Postdoctoral Fellowship from the National Institutes of Health, The Howard Hughes Medical Institute, The Bank of America-Giannini Foundation for Medical Research, and The Mathers Foundation.