Annu. Rev. Neurosci. 1995. 18:255–81

NEUROMORPHIC ANALOGUE VLSI

Rodney Douglas[1,2], *Misha Mahowald*[1], *and Carver Mead*[2]
[1]MRC Anatomical Neuropharmacology Unit, Oxford OX1 3TH, England and
[2]Computation and Neural Systems Program, California Institute of Technology, Pasadena, California 91125

KEYWORDS: CMOS, computation, model, network, neuron

INTRODUCTION

Neuromorphic systems emulate the organization and function of nervous systems. They are usually composed of analogue electronic circuits that are fabricated in the complementary metal-oxide-semiconductor (CMOS) medium using very large-scale integration (VLSI) technology. However, these neuromorphic systems are not another kind of digital computer in which abstract neural networks are simulated symbolically in terms of their mathematical behavior. Instead, they directly embody, in the physics of their CMOS circuits, analogues of the physical processes that underlie the computations of neural systems. The significance of neuromorphic systems is that they offer a method of exploring neural computation in a medium whose physical behavior is analogous to that of biological nervous systems and that operates in real time irrespective of size. The implications of this approach are both scientific and practical. The study of neuromorphic systems provides a bridge between levels of understanding. For example, it provides a link between the physical processes of neurons and their computational significance. In addition, the synthesis of neuromorphic systems transposes our knowledge of neuroscience into practical devices that can interact directly with the real world in the same way that biological nervous systems do.

NEURAL COMPUTATION

The enormous success of general purpose digital computation has led to the application of digital design principles to the explanation of neural computa-

0147-006X/95/0301-0255$05.00

tion. Direct comparisons of operations at the hardware (Koch & Poggio 1987, McCulloch & Pitts 1943, von Neumann 1958) and algorithmic (Marr 1982) levels have encouraged the belief that the behavior of nervous systems will, finally, be adequately described within the paradigm of digital computation. That view is changing. There is growing recognition that the principles of neural computation are fundamentally different from those of general purpose digital computers (Hertz et al 1991, Hopfield 1982, Rumelhart & McClelland 1986). They resemble cooperative phemomena (Haken 1978, Julesz 1971, Marr & Poggio 1976) more closely than theorem proving (McCulloch & Pitts 1943, Newell et al 1958, Winston & Shellard 1990). The differences between the two kinds of computation are most apparent when their performance is compared in the real world, rather than in a purely symbolic one. General purpose digital approaches excel in a symbolic world, where the algorithms can be specified absolutely and the symbols of the computation can be assigned unambiguously. The real world is much less rigid. There the solution must emerge from a complex network of factors all operating simultaneously with various strengths, and the symbols must be extracted from input data that are unreliable in any particular but meaningful overall. Under these fluid conditions neural computations are vastly more effective than general purpose digital methods.

The reasons for the superior performance of nervous systems in the real world are not completely understood. One possibility rests in their different strategies for obtaining precision in the face of the noise inherent in their own components. Digital systems eliminate noise at the lowest level, by fully restoring each bit to a value of zero or one at every step of the computation. This representation is the foundation of Boolean logic. Consistent with this approach, digital architectures rely on each bit in the computation to be correct, and a fault in a single transistor can bring the entire computation to a halt. In contrast, neural systems use unrestored analogue quantities at the base level of their computations, which proceed according to elementary physical laws. Noise is eliminated from these analogue signals by feedback from the next level in the computation, where many signals have had an opportunity to combine and achieve precision at a collective rather than individual level. This collective signal is then used to optimize the performance of individual components by adaptation. The same architectures and adaptive processes that neural systems use to generate coherent action in the presence of imperfect components may also enable them to extract precise information from a noisy and ambiguous environment.

The fundamental difference in these computational primitives is reflected in the way that general purpose digital computers simulate neural models (Figure 1). The binary representation requires that logical functions form the foundation of the symbols and operations needed to model the biological

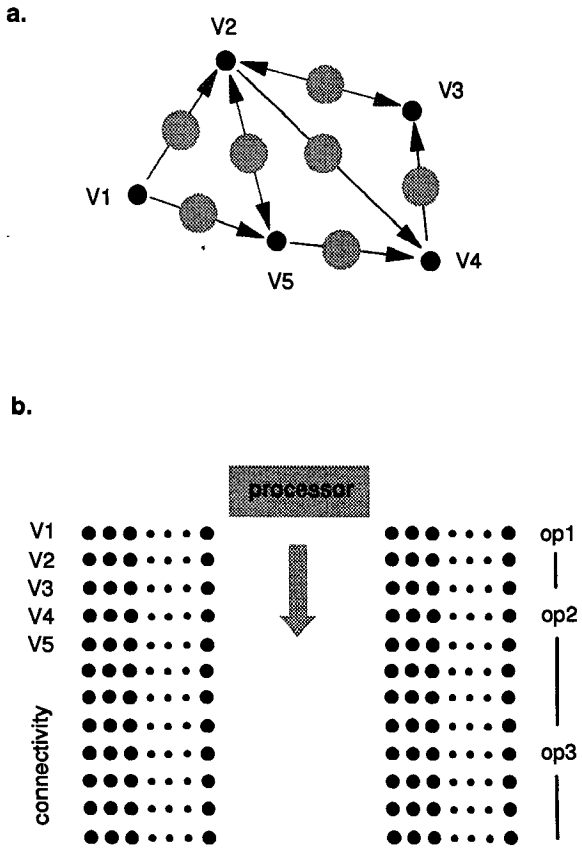


Figure 1 Comparison of the organization of a model on an analogue and on a general purpose digital simulator. (*a*) The analogue in this example represents five state variable nodes (*black filled circles*) of the modeled system. The variables at these nodes are continuous and can encode many bits of information. The variables interact along the paths indicated by arrows. The operations (*gray filled circles*) that occur along the paths are physical processes analogous to those of the modeled system (see Figure 5). The computations of all the state variables proceed in parallel, synchronized by their interactions and mutual dependence on time. (*b*) The digital system comprises many isolated nodes (*black filled circles*), each of which has a binary state, which represents only one bit. The multibit state variables of the modeled system are encoded on group nodes shown to the right of each variable (V1–V5). The connectivity of the modeled system is similarly encoded, as are numerical abstractions of the operations (op1–op*n*) performed along each path. This binary encoding scheme requires many more nodes than does the analogue implementation. In the digital system the interaction between nodes is not continuous but sampled. At each simulated time step, the digital processor evaluates the operations along each of the paths and updates the variables at each node. In the digital case, time is not a physical characteristic of the computation. Instead, it is an abstract variable similar to V1–V5 and must be updated by the processor after the completion of each time step.

system. It follows that the modeled system must be translated explicitly into a mathematical form. This representation is not efficient in terms of the time and number of devices required to perform the computation. First, the state variables and parameters of the model are encoded as abstract binary numbers, which occupy a large area on the surface of a chip relative to a single analogue variable. Even time must be abstracted so that the natural temporal relationships between processes are not guaranteed. Because the primitives are not natural physical processes, every relationship between the system variables must be explicitly defined. Even the most natural properties, such as statistical variation and noise, must be specified.

The efficiency of neuromorphic analogue VLSI (aVLSI) rests in the power of analogy, the isomorphism between physical processes occurring in different media (Figure 1). In general, neural computational primitives such as conservation of charge, amplification, exponentiation, thresholding, compression, and integration arise naturally out of the physical processes of aVLSI circuits. Calculation of these functions using digital methods would require many amplifiers for storage and many digital clock cycles to execute the algorithms that generate the functions (Hopfield 1990). Thus, the computational advantage of aVLSI and neural systems is due to the spatially dense, essentially parallel and real-time nature of analogue circuits (Hopfield 1990, Mead 1989). Of course, only a small number of functions can be generated directly by the physics of the analogue circuits. The analogue approach scores where those functions are the ones required, and where only modest precision (1:10–1:10^3) is necessary. Fortunately, the functions and precision of aVLSI circuits have proven effective for emulating neural phenomena.

At present, both digital and analogue technologies are orders of magnitude less powerful than is necessary to embody the complete nervous system of any animal, even if we knew how. The problem at this time is not one of complete understanding but of how best to embody the principles of simpler neural subsystems. The choice between digital simulation and analogue emulation depends on a number of factors, such as ease of implementation and the goal of the experiment. Digital simulation is precise, the results are easily stored and displayed, and the configuration of the simulation is as easy as recompiling a program. Digital simulation is effective in many applications where the size of the machine, its power consumption, and its time of response are not important. Analogue emulation is an attractive alternative to simulation if the performance of the system in simulation would be so slow that it is impractical, and as in the case of prosthetic devices, the size and power-consumption of the device are critical. Analogue VLSI emulations at present require specialized skills on the part of the designer. This requirement may be less of an obstacle in the future as the potential of aVLSI systems is developed.

EARLY SENSORY PROCESSING

The extraction of behaviorally relevant signals from sensory data is a natural starting point for neuromorphic aVLSI design. Sensory systems have many similar elements laid out in regular spatial arrays of one or two dimensions, with a high proportion of local interactions between elements. They must provide noise-resistant, alias-free operation over a wide range of stimulus intensities by using physical elements of finite dynamic range. The aVLSI medium is well suited to the synthesis of such systems. By creating neuromorphic sensory systems, we build a concrete ground that can be used to explore the interaction of medium and computation in both biological and silicon systems. We also create processors that can provide input to behaviorally functional neuromorphic systems and that could be developed as prosthetic input devices for humans.

Sensors are the critical link to the world. If the source data about the world are not properly captured and transmitted to later processing stages, no amount of computation can recover them. One approach to this problem is to collect and transmit as much raw data as possible. This strategy, which is adopted in traditional artificial sensors such as charge-coupled device (CCD) cameras, makes strong demands on the dynamic range of the sensors and the communication bandwidth of the transmission channel to the later processing stages. For example, many natural visual scenes have dynamic ranges of three orders of magnitude during steady illumination, and the dynamic range of the same scene viewed from sunny afternoon to dusk may vary by seven orders of magnitude. Not only do these ranges exceed the limits of current digital imaging technology, but the torrent of low-level data generated by simple gray-level imagers cannot be analyzed in real-time even by multi-processor general purpose computers (Scribner et al 1993).

Biological systems have evolved an alternative approach to sensing. Rather than transmitting all possible information to the brain, they extract only that which is salient to the later processing stages. One of the advantages of computing high-order invariants close to the sensors is that the bandwith of communication to subsequent processors is reduced to a minimum. Only salient data are transmitted to the next stage of processing. The degree of invariance achieved by early sensory processing is a compromise that is made differently by various modalities and species. For example, the space available for additional circuitry restricts the sophistication of signal processing that can be accomplished near to an array of sensors such as the retina. Also, invariance of response implies that sensory information has been discarded. The question of what to discard cannot be context sensitive if invariants are hard wired into early sensory processing. Understanding the trade-offs made by biological

sensors requires a knowledge of a broad range of constraining factors, from the bandwidth of the individual computational elements to the behavior of the organism as a whole.

Vision

Early visual processing is a popular topic for research, and a number of algorithms have been described for detecting various features of the visual scene. The huge flux of information in algorithmic visual processing has, until recently, restricted these computations to high-performance digital computers. But mass-market applications are now driving the search for implementations of these algorithms on compact, low-power chips (Koch & Li 1994). In general, these VLSI circuits aim to capture the mathematics of vision, rather than emulate the biological methods of visual processing.

One of the earliest attempts to use aVLSI technology to emulate the principles of biological sensory processing outlined above was an electronic retina that emulates the outer plexiform layer of the mud puppy (Mahowald & Mead 1989, Mead & Mahowald 1988). This silicon retina includes components that represent the photoreceptor, horizontal cell, and bipolar cell layers of the retina. The photoreceptor transduces light into a voltage that is logarithmic in the intensity of the stimulus. The synaptic interactions between the cell types are implemented in analogue circuits. The photoreceptors drive the horizontal cells (a noninverting synapse) via a transconductance (Figure 2*a*). A resistive network emulates the gap junction connections between horizontal cells (Figure 2*b*). Thus, the voltage at each node of the horizontal cell network represents a spatially weighted average of the photoreceptor inputs to the network. As in the biological retina, the electrotonic properties give rise to an exponentially decreasing spatial receptive field in the horizontal cell network. The antagonistic center-surround receptive field of the bipolar cell is implemented with a differential amplifier that is driven positively by the photoreceptors but inhibited by the horizontal cell output.

The silicon retina performs like a video camera in that the image is transduced directly on the chip and the output is displayed on a monitor screen continuously in time. However, unlike the CCD camera, the silicon retina reports contrast rather than absolute brightness and so it is able to see comparable detail in shaded and bright areas of the same scene. This contrast encoding emulates an essential property of retinae and is due to the center-surround receptive field of the bipolar cell, which computes the difference of the logarithmic photoreceptor and horizontal cell outputs. By using the local average of the horizontal cells as a reference signal, the redundant features of the image are suppressed while novel features are enhanced. This occurs because large areas of uniform luminance produce only weak visual signals. In these regions the output from any single photoreceptor is canceled by the spatial average

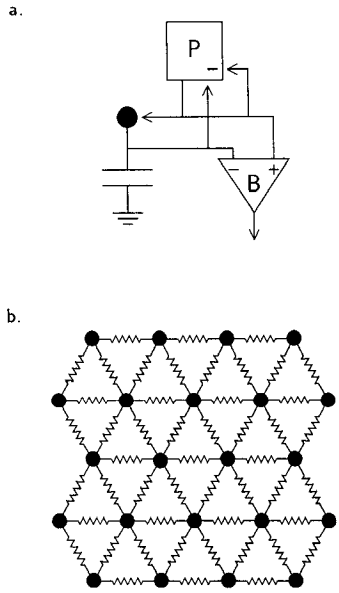


Figure 2 The silicon retina is a two-dimensional hexagonal array of pixels on a chip. The retina transduces a light image focussed directly on its surface, processing the stimulus by principles found in the outer plexiform layer of the vertebrate retina. Various versions of silicon retinae contain from 2304 (Mahowald & Mead 1989) to 8250 pixels (Boahen & Andreou 1992) on a 6.8 mm × 6.9 mm square of silicon using a 2-μm fabrication process. (*a*) A pixel from the silicon retina produced by Mahowald (1991) contains an analogue of a photoreceptor, P, a bipolar cell, B, and a node of the horizontal cell network (*filled circle*). (*b*) A hexagonal resistive network emulates the horizontal cells of the retina. The horizontal cells extend parallel to the surface of the retina and are resistively coupled to each other by gap junctions. The values of these resistors is controlled by applying an analogue voltage to the chip, so the degree of connectivity between horizontal cells can be varied after the chip is fabricated. In addition to the retinal array, the retina chips also have instrument circuits that monitor the outputs of the pixels. The pixels can be monitored individually or scanned out sequentially to a video monitor.

signal from the horizontal cell network. Novel luminance features, such as edges, evoke strong retinal signals because the receptors on either side of the edge receive significantly different luminance. A similar principle applies in the time domain, where the relatively slow temporal response of the horizontal cell network enhances the visual system's response to moving images. In fact, as in the physiology of biological retinal cells, the spatial and temporal characteristics of the retinal processing are simultaneously expressed, as can be seen in the response of the silicon retina to flashed lights of varying spatial extent illustrated in Figure 3.

The silicon retina encourages a functional interpretation of retinal processing. From an engineering point of view, the purpose of the outer-plexiform

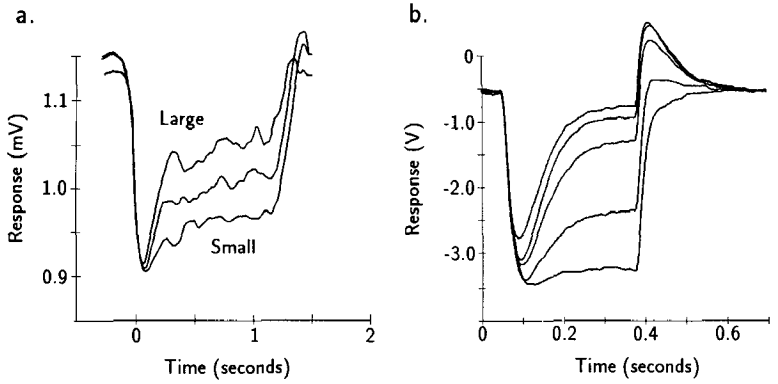


Figure 3 Comparison of the response of bipolar cells in a vertebrate and a silicon retina to a flashing light stimulus of varying spatial extent. (*a*) Response of a bipolar cell in the mud puppy, *Necturus maculosus* (data from Werblin 1974). (*b*) Output of a pixel in the silicon retina. Test flashes of the same intensity but of different diameters are centered on the receptive field of the unit. Larger flashes increase the excitation of the surround. The surround response takes time to build up to the capacitance of the resistive network. Because the surround level is subtracted from the center response, the output shows a decrease for long times. This decrease is larger for larger flashes. The overshoot at stimulus offset decays as the surround returns to its resting level. The peak response diminishes as the test flash becomes larger spatially. The larger test flash charges the network to a higher value. The charging of the network works against the finite response time of the receptor to produce the observed decrease in peak response at the bipolar cell level.

layer is to keep the output of retina from saturating over several orders of magnitude change in background illumination while allowing it to report the small contrast differences in a typical scene reliably. The encoded retinal output still contains sufficient information to fully support general visual processing in the brain, but it can be transmitted at a much lower bandwidth than can a gray-scale image. The efficient encoding of visual information is necessitated by a simple physical constraint, the finite range of retinal cell outputs. The hypothesis underlying the early silicon retina architecture was that lateral inhibition was sufficient to deal with the problem of dynamic range in retinal processing (Barlow 1981). Lateral inhibition removes redundant information caused by correlation of luminance across the image. However, lateral inhibition exacerbates another form of image redundancy, noise. Noise is a form of redundancy because it does not supply additional information about the world, so bandwidth is wasted by transmitting it (Atick & Redlich 1990, 1992). In the retina, noise arises from photon fluctuations at low light levels. But, it can also arise from the photoreceptors themselves, as dynamic thermal noise and as static miscalibration of photoreceptors' gains and operating points. The early silicon retinae were plagued by static mismatches between transistors, which

caused each photoreceptor to give a different response to identical inputs. The solution to this problem was to incorporate more of the features of outer plexiform processing that had been observed in biological systems.

New generations of silicon retinae remove the intrinsic sources of noise by using adaptive elements (Mead 1989, Mahowald 1994) and resistive coupling between the receptors (Boahen & Andreou 1992). These retinae include feedback from the horizontal cells to the photoreceptors. This feedback allows the photoreceptor to respond over a large input range with higher gain than the photoreceptors in the earlier feedforward retinae. The feedback between the photoreceptors and the horizontal cells affects the time constants of the cells and the electrotonic spread of signals in the restive networks.

The inner-plexiform layer of the retina is more poorly understood than the outer one. It is believed to play a role in motion processing, and in lower vertebrates, such as the frog, it has complex synaptic interactions that give rise to invariant feature detectors. This complex processing is desirable when the output of the retina is used directly to mediate a specialized behavior (Horiuchi et al 1991, 1992). The cost of incorporating more complex processing in the retinal array is that the photodetectors are spread farther apart to make room for additional circuitry. An example of a complex silicon retina is the velocity-tuned retina of Delbrück (1993). This retina contains analogue delay elements that correlate the spatio-temporally distributed signals that arise from moving targets. At each pixel location, the retina generates three independent outputs that are tuned to a particular speed and direction of motion, irrespective of stimulus contrast. This retina contains 676 pixels on a 6.8 mm × 6.9 mm die fabricated in a 2 μm process. The computation is energy efficient. The analogue processing in the retina (not including video display) consumes 1.5 mW in the dark to 8 mW in the light, which is less than 5 μW per pixel, compared with 10 μW per pixel for a typical CCD pixel that does no computation.

Audition

Auditory processing is another interesting example of nonlinear sensory transduction. Here too, insights have been gained by the development of neuromorphic auditory processors. Early efforts focussed on emulating the cochlea (Lyon & Mead 1988, Mead 1989), the fluid-mechanical traveling wave structure that encodes sound into spatio-temporal patterns of discharge in the auditory neurons.

The sound pressure in the outer ear is coupled mechanically into vibrations of the cochlea's basilar membrane by the ossicles of the middle ear. The membrane behaves like a spatially distributed low-pass filter with a resonant peak. The resonant frequency of the filter decreases exponentially along the basilar membrane from the oval window toward the apex of the cochlea, so

the optimal response of the membrane is spread in space according to frequency. This response can be modified by the outer hair cells, which are motile and able to apply forces to the basilar membrane at frequencies throughout the auditory spectrum. Hair cells provide mechanical feedback that modifies the damping of the basilar membrane, so changing its resonance. The cochlea can use this mechanism to preferentially amplify low-amplitude traveling waves. Thus, the cochlea acts as an automatic gain control mechanism for sound, analogous to the gain control for luminance in the outer-plexiform layer of the retina.

The silicon cochlea of Lyon & Mead (1988) uses a unidirectional cascade of second-order sections with exponentially scaled time constants to approximate the traveling wave behavior of the basilar membrane. Because the sections can be tuned to generate a gain greater than one over a range of frequencies, they can approximate the resonance behavior caused by the outer hair cells. The behavior of this silicon cochlea is remarkably similar to its biological counterpart. However, the degree of resonance of each stage of the filter cascade is quite sensitive to device mismatches. To investigate this problem, a more detailed aVLSI cochlea has been made that uses a resistive network to model the cochlea fluid, and a special purpose circuit to model regions of the basilar membrane (Watts 1993). The collective nature of resistive grid makes this cochlea less sensitive to device mismatch than the filter cascade design. The price of this robustness is an increase in circuit complexity and size.

This chip highlights an interesting methodological point—because neuromorphic models are constrained by what can be implemented in a physical medium, they can provide insight into biological design. The response of the basilar membrane scales—that is, the spatial pattern of the response is invariant with frequency except for a displacement along the membrane. There are two physical models that give rise to scaling. In the first—constant mass scaling—the mass of the membrane and the density of the fluid are constant, but their stiffness changes exponentially along its length. In the second model—increasing mass scaling—all three change exponentially along the length of the membrane. Although the behavior of these two models is indistinguishable, their implications for physical implementation are radically different. The constant mass model requires that membrane stiffness change by a factor of about one million. Such a range of variation can be simulated on a digital computer, but it cannot be implemented easily in a physical device. This suggests that the increasing mass model, in which the range of variation can be absorbed by three parameters rather than one, should be adopted. In this case no single parameter need change by a factor of more than 100–1000, which is feasible in real devices such as neurons and chips. Moreover, each of the three parameters could be controlled by just one independent parameter,

the width of the basilar membrane, which increases linearly with displacement along its length (Watts 1993). Watts' (1993) resistive network silicon cochlea follows this idea. The exponentially varying membrane stiffness, membrane mass, and participatory fluid mass are all controlled by a single control line whose voltage increases linearly with membrane length, as does width. This chip emulates well the basilar membrane performance predicted by numerical and analytical methods.

The resistive network cochlea does not yet contain active elements that could emulate the outer hair cells. But, a simple circuit that models the sensory transduction and motor feedback required to do this has been designed and tested separately (Watts 1993). Basically, the outer hair cell circuit is a feedback system containing a saturating nonlinear element in its feedback loop (Watts 1993, Yates 1990). The circuit senses the local basilar membrane displacement and applies motor feedback there. These circuits preferentially amplify low-amplitude waves, because under these conditions the feedback gain is high. As the wave becomes amplified, the feedback element is driven into saturation, and the gain becomes smaller.

The transduction of cochlea signals into the early stages of neural processing has also been investigated by using neuromorphic methods. In the biological cochlea the spatial patterns of vibrations of the basilar membrane are sensed by the inner hair cells, which transduce the mechanical displacement into the release of neurotransmitter and thereby affect the rate of action potential generation by their associated auditory neurons. Consequently, each neuron has a bandpass tuning for frequency whose optimal response depends on the position of the inner hair cell on the basilar membrane. Thus, the combined pattern of mean rate responses in auditory neurons could represent the spectral shape of the input signal. But this explanation is not sufficient to explain auditory perception, because the bandpass of the neurons is not amplitude invariant even at the moderate amplitudes occurring during normal speech.

The temporal structure of action potential discharge offers an alternative representation of frequencies below about ~5 kHz. Auditory neurons operating in this range tend to discharge on one polarity of an input waveform, so their probability density function for spike generation is a half-wave rectified version of the analogue signal of their hair cell. This phase encoding persists even during fully saturated discharge (Evans 1982). Since individual auditory neurons rarely fire above 300 Hz, the phase-dependent spikes cannot occur on every cycle. Instead, the outputs of several neurons driven by the same hair cell must be combined to provide a reliable signal. For example, the outputs could be combined by a postsynaptic neuron that acts as a matched filter for spike repetition rate (Carr 1993, Lazzaro 1991, Suga 1988).

The matched filter is an autocorrelator that receives the unmodified auditory fiber action potentials as input and the same signal delayed by the match

interval. In principle, the output of this correlator could have strong peaks not only at the match frequency but also at its harmonics. This problem is resolved by the basilar membrane filter, which attenuates the harmonics leaving a single peak at the match frequency. Thus, the time-delay matched-filter hypothesis is attractive but leaves open the questions of how the various time delays could arise in biological circuits and how precise these delays need be.

By building a silicon auditory processor that uses the time-delay correlation, Lazzaro & Mead (Lazzaro 1991, Lazzaro & Mead 1989, Lyon & Mead 1988) have shown that this approach can work robustly despite component variability (Figure 4). Their chip is composed of the filter cascade cochlea and circuitry that emulates the auditory neurons and their postsynaptic matched filters. The filter sections of the cochlea have exponentially scaled time constants and span a frequency range of 0.4 Hz to 4 kHz. Each section drives a circuit that models inner hair cell transduction. Each silicon hair cell, in turn, excites a group of simple neurons. The action potentials of all the groups are led both directly and via time delays to an array of correlators. The time delays of the correlators in the array increase exponentially and are roughly matched to the time constants of the silicon basilar membrane.

The auditory processor correctly identifies the frequency of a test signal in real time, which suggests that time-delay correlation is a feasible explanation for the operation of the early stages of auditory perception. The chip also offers interesting prospects for prostheses and auditory processing for computers (Lazzaro et al 1993).

RECONFIGURABLE NETWORKS

The sensory aVLSI systems described above have specialized neurons with hard-wired connections. Fortunately, sensory subsystems have a very regular structure that is dominated by local connectivity. In these cases it is possible to satisfy the connections of many thousands of neural nodes on a single chip. But beyond the sensory processing stage, the connectivity quickly becomes more complex, as do the characteristics of the neurons, which receive thousands of inputs along their branching dendritic trees. Such highly connected networks can only be constructed by distributing the network across multiple chips. There are many strategies for dividing the network among chips. These strategies typically recognize three different elements: the synapses onto the neurons, the input-output relation of the neurons, and the communication pathways of the axons.

Whatever the strategy employed, it is desirable to reuse the same hardware for exploring various neural architectures. This task requires the development of general silicon neurons, whose biophysical properties and connectivity can be reconfigured to match a range of neuronal types, and a communication

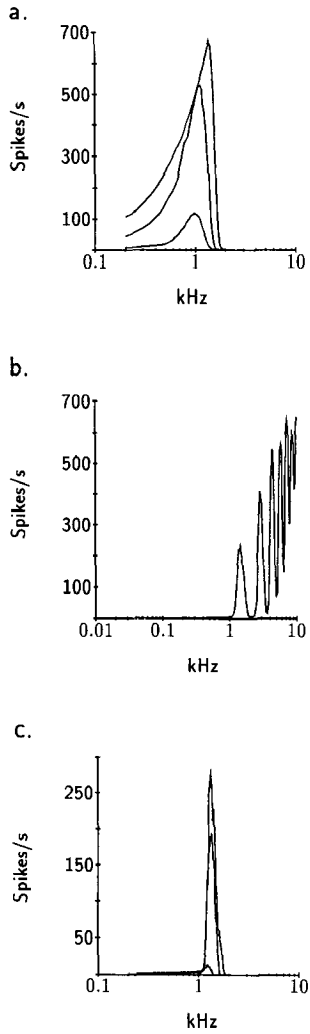


Figure 4 Response of a silicon auditory processor that emulates a cochlea, auditory neurons, and proposed postsynaptic neurons that act as correlators. (*a*) Average discharge rate of the auditory neurons in response to sinusoidal stimuli (2, 5, and 20 mV peak amplitude, bottom to top) applied to the cochlea. Notice that the bandwidth of the response is sensitive to input amplitude. (*b*) Output from a correlator postsynaptic to the auditory fibers shown in *a*, with cochlea filtering disabled. The correlator responds at its optimal frequency (about 1.5 kHz) with narrow bandwidth, but it also responds at higher harmonics. (*c*) Output of the correlator to the stimuli described in *a*, with cochlea filtering enabled. The response is logarithmic over the stimulus range 2–20 mV. There are no harmonics. The bandwidth is insensitive to stimulus amplitude. The auditory processor responds to frequencies over a range of approximately 0.3–3 kHz (data not shown). Modified from Lazzaro 1991.

strategy that can handle the huge convergence and divergence of general neuronal systems.

Neurons

There have been a number of reports of electronic neurons built of discrete components that emulate real neurons with various degrees of accuracy (Hoppensteadt 1986, Keener 1983) and that could be wired together to form small networks. In recent years the development of VLSI technology has made it possible to envisage very large networks of realistic CMOS neurons that could operate in real time and interact directly with the world via sensors.

VLSI designers have taken a wide range of approaches to the design of neurons for artificial neural networks. Examples of recent work in the field can be found in Murray & Tarassenko (1994), Ramacher & Ruckert (1991), Zaghloul et al (1994), *IEEE Transactions on Neural Networks*, and the *Advances in Neural Information Processing Systems* series. Many of these designs use pulse-based encoding, similar to the action potential output encoding of real neurons (Murray et al 1989, 1991; Murray & Tarrasenko 1994). However, in most of these cases, the connectivity between neurons is the focus of effort. The responses of the individual neurons are simplified to fixed input-output relations that conform to the models used in artificial neural network research. Less effort is devoted to the study of the complex biophysical mechanisms that underly neuronal discharge, and what their significance might be.[1]

Neurons act as elaborate spatio-temporal filters that map the state of their approximately 10^4 input synapses into an appropriate pattern of output action potentials. This mapping is dynamic and arises out of the interaction of the many species of active and passive conductances that populate the dendritic membrane of the neuron and that serve to sense, amplify, sink, and pipe analogue synaptic currents en route to the spike generation mechanism in the soma. The members of this society of conductances are each effective for a specific ion, or profile of ions, and they interact via their mutual dependence on either membrane potential or intracellular free calcium or both. The macroscopic conductance for a particular ion depends on the states of individual channels embedded in the neuronal membrane. These channels are charged, thus their performance is sensitive to voltage gradients across the membrane. The fraction of the channels that are conductive (open) at any membrane potential can be described by the Boltzmann distribution (Hille 1992). The fraction of channels that are open is given by

$$O\,(O + C) = 1/(1 + e^{-zV/kT}),$$

[1]However, DeYong et al (1992) have designed a neural processing element that is more clearly neuromorphic.

where O is the number of open channels; C, the number of closed channels; V, the transmembrane voltage; z, the channel charge; and kT, the thermal energy per charge carrier. This activation function has a sigmoidal form, saturating when all of the channels are open. Its dynamics are controlled by a time constant that may itself be voltage dependent. The activation is often opposed by an inactivation process that attenuates the conductance.

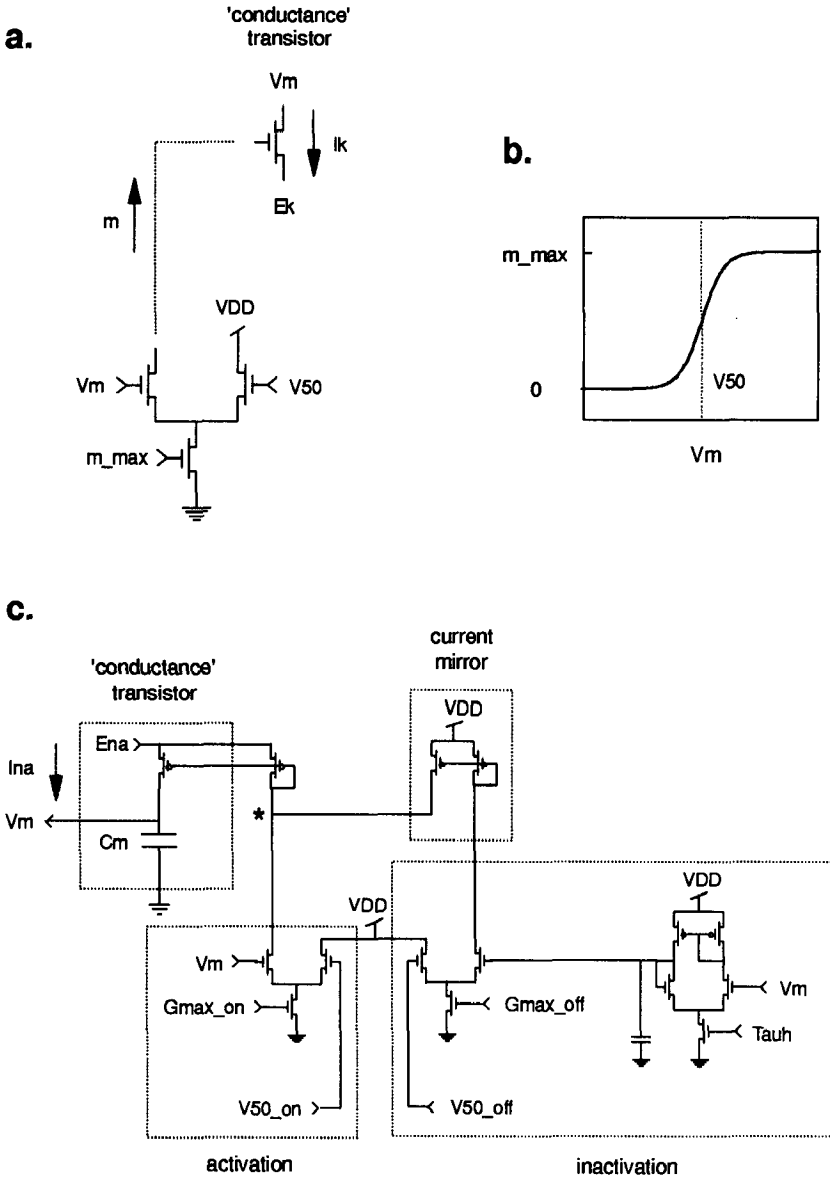
aVLSI neurons are synthesized by combining modular circuits that emulate the behavior of biological conductances. As in the example in Figure 5, the prototypical conductance module consists of an activation and an inactivation circuit that compete for control of an element that emulates the membrane conductance. When CMOS circuits are operating in their subthreshold regime (Mead 1989), the transistors' gate voltages control the heights of the energy barriers over which the charge carriers with Boltzmann-distributed energies must pass in order to flow across the channels. This similarity between the physics of membrane and transistor conductance offers a method of constructing compact circuits that represent biological channel populations.

The usual operation of these analogue circuits is similar to Hodgkin-Huxley descriptions (Hodgkin & Huxley 1952, Hille 1992) of membrane conductance. If the membrane potential is driven into the range of activation, the conductance for that ion will increase and charge the membrane towards its associated reversal potential. During activation, inactivation will grow with a slower time constant, finally quenching the conductance and allowing the membrane to relax back to its resting state.

The prototypical circuits are modified in various ways to emulate the particular properties of a desired ionic conductance. For example, some conductances are sensitive to calcium concentration rather than membrane voltage and require a separate voltage variable representing free calcium concentration. Synaptic conductances are sensitive to ligand concentrations, and these circuits require a voltage variable representing neurotransmitter concentration. The dynamics of the neurotransmitter in the cleft is governed by additional time constant circuits. These special circuits are simple modifications of the general theme of the prototypical circuit. Circuit modules that have been designed and tested so far include the sodium and potassium spike currents, the persistent sodium current, various calcium currents, the calcium-dependent potassium current, the potassium A–current, the nonspecific leak current, the exogenous (electrode) current source, the excitatory synapse, the potassium-mediated inhibitory synapse, and the chloride-mediated (shunting) inhibitory synapse.

Once a repertoire of circuits is available, the desired profile of currents can be inserted into the neuronal compartments that are capacitance nodes separated by axial resistances, just as one might specify a neuron for digital simulation. The behavior of the individual circuits that compose the neuron can be controlled by setting the voltages applied to the gates of various

transistors in the module. These parameters determine, for example, the temporal dynamics of activation and inactivation, the voltage at which activation or inactivation occurs, and the maximum conductance that can be obtained when the ion conductance is fully activated. The parameters can be varied to



emulate neurons with a range of biophysical properties. The effect of changing these parameters is immediate, so the electrophysiological personality of the silicon neuron can be switched rapidly, for example, from a regular adapting to a bursting pyramidal cell (Figure 6). The fact that a range of neuronal subtypes with the same circuitry can be emulated suggests that the control parameters of the silicon neuron are analogous to those available in the development of real neurons. Of course, the range of reconfigurability is limited to the range of interaction of the circuits that are placed in the neurons at the time the chip is fabricated. Modules can be inactivated and so effectively removed from a compartment, but no new modules can be added. If additional properties are required, another silicon neuron with different morphology and different types of channels can be fabricated by using variations of the basic circuit modules.

To produce compact circuits it is often necessary to make approximations in the design of the circuit modules (Douglas & Mahowald 1994a). However, the striking results of the aVLSI emulations are the degree to which qualitative realistic neuronal behavior arises out of circuits that contain such approximations, and the fact that adjustments in the control parameters cause the neuron to change its behavior in a way that is consistent with our understanding of neuronal biophysics.

Communication

The degree of connectivity in neural systems and the real-time nature of neural processing demand different approaches to the problem of interchip communication than those used in traditional digital computers. VLSI designers have adopted several strategies for interchip communication in silicon neural net-

Figure 5 Example of a neuromorphic CMOS aVLSI circuit. (*a, b*) Basic circuit that emulates transmembrane ion currents in the silicon neuron (Douglas & Mahowald 1994a,b; Mahowald & Douglas 1991). A differential pair of transistors have their sources linked to a single bias transistor (*bottom*). The voltage, m_{max}, applied to the gate of the bias transistor sets the bias current, which is the sum of the currents flowing through the two limbs of the differential pair. The relative values of the voltages, *Vm* and V_{50}, applied to the gates of the differential pair determine how the current will be shared between the two limbs. The relationship between *Vm* and the output current, *m*, in the left limb is shown in *b*. The current, *m*, is the activation variable that controls the potassium (in this example) current, *Ik*, that flows through the conductance transistor interposed between the ionic reversal potential, *Ek*, and the membrane potential. (*c*) The circuit that generates the sodium current of the action potential is composed of activation and inactivation subcircuits that are similar to those shown in *a*. The activation and inactivation circuits compete for control of the sodium conductance transistor by summing their output currents at the node marked by the asterisk. The current mirror is a technical requirement that permits a copy of the inactivation current to interact with the activation current. In this example, sodium current, *I*na, flows from the sodium reversal potential, *E*na, onto the membrane capacitance, *Cm*. The transconductance amplifier and capacitor on the right of the inactivation circuit act as a low pass filter, causing the inactivation circuit to respond to changes in membrane voltage with a time constant set by the voltage, τh.

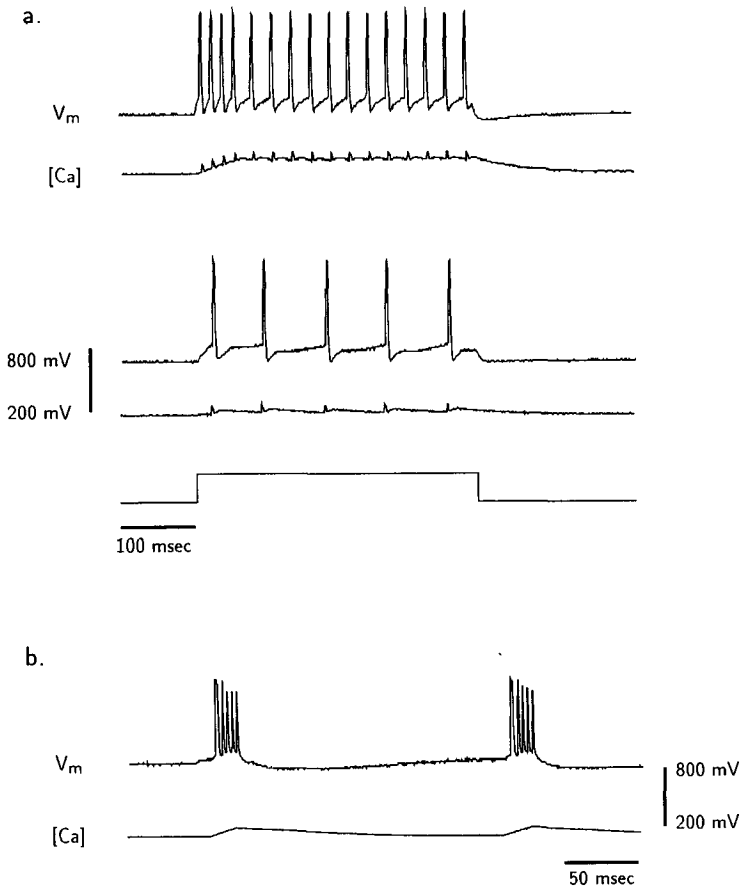


Figure 6 A silicon neuron that emulates the response of a cortical pyramidal neuron. (*a*) Response to two levels of intrasomatic current injection (time course shown below). The neuron is configured as a regular adapting pyramidal cell. Each pair of response traces consists of the somatic membrane potential, *Vm* (800 mV scale bar), and a voltage that is proportional to the intrasomatic free calcium concentration, [Ca] (200mV scale bar). Notice the adaptation of discharge and after-train hyperpolarization in the upper voltage trace. The voltage response of the silicon neuron is ten times larger than that of a real neuron (Douglas & Mahowald 1994a). (*b*) Response to sustained current injection of neuron configured as a bursting pyramidal cell.

works. Each strategy has advantages, and the choice of method depends on which factors are most crucial to the application.

One of the most literal approaches to interconnecting silicon neurons has been adopted by Mueller et al (1989a,b). They use a direct physical connection between nodes on different chips through a cross-bar switching array. A major

advantage of this approach is that it allows continuous-time communication between nodes. In addition, the switching arrays provide flexible connectivity and can be programmed digitally by a host computer. The system is able to handle larger connectivities than are possible on a single chip because the dendrites of a single artificial neuron are extended over multiple chips. However, this approach requires many chips to model even a small number of neurons. The number of artificial neurons on each output chip is limited to less than 100 by the number of output pins available.

To increase the number of neurons that can be placed on a single chip, a strategy must be adopted that relies on the speed advantage of VLSI technology over biological neurons. Because neuromorphic systems are intended to interact with the real world, their dynamics should evolve concurrently with external events. To achieve speeds comparable to biological neurons, neuromorphic aVLSI circuits operate about 10^5 times more slowly than the maximal speed of CMOS. The excess bandwith is available for communication by multiplexing a single connection path in time. This communication strategy attempts to achieve in time what the axonal arborizations achieve in space.

A number of multiplexing techniques are currently in use in artificial neural systems (Brownlow et al 1990; Mahowald 1994; Murray et al 1989, 1991; Murray & Tarrasenko 1994; Sivilotti et al 1987). One communication strategy that is closely related to the action potential coding of biological neurons is an address-event representation (AER) (Mahowald 1992, 1994).

The address-event protocol works by placing the identity (a digital word) of the neuron that is generating an action potential on a common communications bus that is an effective universal axon. The bus broadcasts this address to all synapses, which decode the broadcast neuronal addresses. Those synapses that are connected to the source neuron detect that their source neuron has generated an action potential, and they initiate a synaptic input on the dendrite to which they are attached. Many silicon neurons or nodes can share a single bus because, like their biological counterparts, only a small fraction of the silicon neurons embedded in a network are generating action potentials at any time and because switching times in the bus are much faster than the switching times of neurons. At present, neuronal events can be broadcast and removed from the data bus at frequencies of 1 MHz. Therefore, about one thousand address-events can be transmitted in the time it takes 1 neuron to complete a single 1-ms action potential. And if, say, 10% of neurons discharge at 100 spikes/s, a single bus could support a population of about 10^5 neurons. Of course, many more neurons will be required to emulate even one cortical area. Fortunately, the limitation imposed by the bandwidth of the AER bus is not as discouraging as it may seem. The brain has a similar problem in space. If every axon from every neuron were as long as the dimensions of the brain, the brain would explode in size as the number of neurons increased. The brain

avoids this fate by adopting a mostly local wiring strategy in which the average number of axons emanating from a small region decreases at least as the inverse square of their length (Mead 1990, Mitchison 1992, Stevens 1989). If the action potential traffic on the AER bus were similarly confined to a local population of neurons, the same bus could repeat in space and thus serve a huge number of neurons.

The AER communication protocol has a number of advantages. It preserves the temporal order of events taking place in the neural array as much as possible. Events are transmitted as they occur. AER has better temporal accuracy than standard sequential multiplexing because the bandwidth of the communication channel is devoted to the transmission of significant signals only. For example, an AER silicon retina generates address-events only at regions of the image where there is spatial or temporal change. In areas of uniform illumination, the neurons are generating only a few action potentials, so they do contribute much to the communication load. Finally, the AER scheme has the advantage that the mappings between source neurons and recipient synapses can be reprogrammed by inserting digital circuitry to map the output addresses to different synaptic codes.

Synapses

The synaptic couplings between neurons determine the function of the network. They are critical elements of artificial neural networks because it is believed that, in addition to establishing connections between neurons, they are the site of learning. Most VLSI synapse circuits have been developed for artificial neural networks and simply express a coupling weight, without concern for the temporal behavior of their biological counterparts.

Synaptic circuits fall into two broad classes: those that store their coupling weights locally and those that receive their weights together with each presynaptic input. Local weight storage facilitates local learning, which can occur in parallel at each synapse. The synaptic weight can be stored digitally[2] (Säckinger et al 1992) or in analogue form as the charge on a capacitor. Analogue storage may be volatile (Churcher et al 1993), in which case it must be refreshed periodically by some external source, or it may be a permanent charge stored on a well-insulated capacitor called a floating gate (Benson & Kerns 1993, Vittoz et al 1991). Permanent storage is usually preferable to

[2]Digital storage can be more precise than analogue storage, depending on the number of bits required. But for moderate precision (about six bits), digital storage requires a larger silicon surface area than analogue storage. Furthermore, digital arithmetic rules for synaptic update are cumbersome. Digital hardware is more appropriate for functions that are multiplexed, rather than those that occur massively in parallel.

volatile storage, but methods for updating the state of the floating gate are not yet flexible enough to implement many common learning rules.

Because the number of synapses per neuron is large ($5 - 10 \cdot 10^3$), they are the limiting spatial consideration in neuromorphic design. The synapses of individual neurons can be distributed across dendrites on multiple chips (Lansner & Lehmann 1993, Mueller et al 1989a), but this approach raises the problem of accurately transmitting the analogue voltages and currents at the compartment boundaries between chips. Alternatively, entire silicon neurons can be fabricated on the same chip so that only robust, action potential–like events need be transmitted between chips. In this case the synapses of the neuron are also limited to a single chip, and their number becomes a trade-off between space and biological versimilitude.

MULTICHIP SYSTEMS

Multichip neuromorphic systems are still in the early stages of development. The address-event protocol has been used to communicate between retinae and receiver chips, which contain an array of receiver cells that are scanned out onto a video monitor. Similar unidirectional connections have been developed for outputting address events from a digital computer to a receiver array (Mahowald 1992, 1994) and for receiving AER data from a silicon cochlea (Lazzaro et al 1993). These simple systems have been used to test the performance of the AER and explore the possibility of using neuromorphic preprocessing for sensory input to digital computers. The first steps toward more complex sensory-motor systems and multichip processing are described below.

Stereopsis

A functional multichip neuromorphic system has been described by Mahowald (1992, 1994). It is a stereopsis system consisting of two silicon retinae that transmit their image data to a third, stereocorrespondence, chip by using the address event protocol. This system computes the stereodisparity of objects continuously as the objects move around in the field of view of the retinae.

Stereodisparity arises from the projection of visual targets onto different positions in the images formed by the eyes. The stereodisparity of a visual target is proportional to the distance between the target and the viewer. To determine the distance of a target, the viewer must find the image features that correspond to each other and calculate their stereodisparity. Finding corresponding image features is difficult if many image features are similar to each other. The wallpaper illusion, in which a surface covered with identical, periodically arranged targets is perceived at the wrong depth, exemplifies the difficulty of determining stereocorrespondence.

The algorithms used by the cortex to compute stereodisparity are not known,

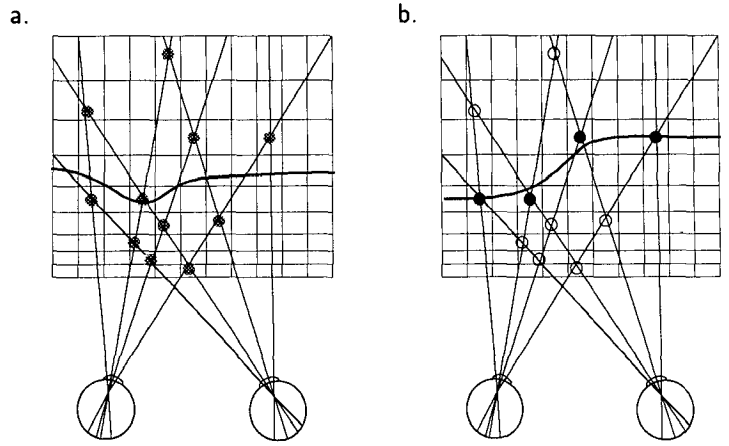


Figure 7 The cooperative multiresolution approach to the problem of stereocorrespondence. (*a*) The one-dimensional retinae receive disparate images from the projection of identical targets (*gray circles*) located in a two-dimensional space. The pattern of all possible targets consistent with the retinal images is shown. A two-dimensional array of correlators is indicated by the grid. A correlator is stimulated when a target lies within its grid position. The outputs of the activated correlators are averaged to provide an analogue estimate of the distance of the targets from the viewer (*heavy line*). (*b*) Positive feedback from the analogue distance estimate to the correlators most consistent with the analogue estimate, combined with competition between correlators, results in the enhancement of the correlators associated with some targets (*solid circles*) and the suppression of correlators associated with others (*open circles*). The analogue estimate of distance converges to a solution close to the enhanced correlators as the suppressed correlators are removed from the average. Targets associated with the enhanced correlators are sufficient to explain the images on the two retinae. They reflect the stereocorrespondence chip's estimate of the true state of the two-dimensional world.

but the response characteristics of certain elements that compute disparity have been recorded from cortex. For example, Poggio has described five distinct types of disparity-tuning reponses in cortical neurons of the primate (Poggio 1984, Poggio et al 1988). The principle innovation of the cooperative multiresolution algorithm (Figure 7) is to introduce cooperative feedback interactions between disparity representations at different spatial scales. The high spatial frequency representation of disparity, analogous to Poggio's tuned-zero neurons, is spatially precise but susceptible to false correspondences. The low spatial frequency representation, analogous to Poggio's tuned-near and tuned-far neurons, makes an analogue estimate of disparity that averages out false correspondences at the expense of spatial blurring. Positive feedback between these two representations, coupled with nonlinear surround inhibition, allows the system to converge to a precise and correct determination of stereodisparity.

The chip is able to fuse random dot stereograms and to solve other hard stereocorrespondence problems such as the fusion of surfaces tilted in depth.

The chip demonstrates that a network can use positive feedback to perform a significant computation, the suppression of false targets, while remaining continuously sensitive to retinal input. The biological relevance of the chip is that it expresses the stereofusion problem as just one instance of a general class of constraint-satisfaction problems in sensory perception and shows how this class of problems can computed with neuron-like elements.

Tectum

The real-time performance of neuromorphic systems suggests that the natural application of aVLSI is the investigation of neural computation in the context of behavioral functions. One of the behavioral functions that has been well studied in biological systems and that complements the aVLSI systems described above is the control of eye movement. Eye movement allows the visual system to actively seek new data based on previous data. It changes the fundamental character of the sensing process from a passive one to an active one (Ballard 1991).

The first steps towards synthesis of an aVLSI eye control system have been taken by DeWeerth, who constructed a primitive tecto-oculomotor system for tracking the brightness centroid of a visual image (DeWeerth 1992, Horiuchi et al 1994). Visual input to the system is provided by a two-dimensional retina with aggregation circuitry that computes the centroid of brightness in the horizontal and vertical dimensions. The magnitudes of the visual input signals are normalized before they are aggregated. This prevents changes in global illumination from affecting the localization of the centroid. The superior colliculus employs a similar computational principle to integrate visual information for oculomotor control (Sparks & Mays 1990). The silicon tectum resolves one example of a more general problem in neuroscience and artificial intelligence, that of reducing the dimensionality of complex data, such as audio and visual images, to low dimensional representations that can be used for motor control or decision making.

The aVLSI tecto-oculomotor system converts the output of the centroid computation into a motor signal and uses this signal to update the position of the retina. The aggregated output currents are converted into pulse trains that are directly applied to drive antagonistic motors in a bidirectional mechanical system that models the ocular muscles. Two motors are placed in an antagonistic configuration such that they produce torque on the eye in opposite directions. Each pulse output generated by the chip controls one of the motors. If the chip is stimulated by a bright spot that is not at the center of the photoreceptor array, the motors will produce a differential torque that rotates the eye so that the center of the photoreceptor array moves towards the stimulus. Because of the background firing rates of the neuron circuits, the motors will produce equal and opposite torques when the stimulus is centered.

Negative feedback is generated by the mechanical system to move the retina chip so that it faces the stimulus.

This simple system is a foundation for future development. The present tectum has photodetectors integrated in the two-dimensional array that computes the centroid. But, with the AER communication protocol, a two-dimensional image derived from a separate retina or an auditory localization chip could be transferred to the tectum to drive eye movements. The oculomotor system also presents an opportunity to explore hierarchical structures for motor control, such as the coordination of vergence and conjugate eye movements for pointing the eyes at targets in depth.

CONCLUSION

CMOS aVLSI has been used to construct a wide range of neural analogues, from single synapses to sensory arrays, and simple systems. The circuits in these examples are not completely general; rather, they are specialized to take advantage of the inherent physics of the analogue transistors to increase their computational efficiency.

These analogues emulate biological systems in real time, while using less power and silicon area than would an equivalent digital system. Conservation of silicon resource is not a luxury. Any serious attempt to replicate the computional power of brains must confront this problem. The brain performs about 10^{16} operations per second. Using the best digital technology, this performance would dissipate over 10 MW (Mead 1990) by comparison with the brain's consumption of only a few watts. Subthreshold analogue circuits are also no match for neuronal circuits, but they are a factor of 10^4 more power efficient than their digital counterparts. In the short term, small, low-power neuromorphic systems are ideal for prosthetic applications. Development of these applications is just beginning. For example, silicon retina projects have inspired work on a VLSI retinal implant (Wyatt et al 1993), and Fromherz et al (1991a,b) are studying silicon-neural interfaces. The development of other useful applications, such as the silicon cochlea for cochlear implants, will depend on the collaboration of neuromorphic silicon designers and biomedical engineers.

The specialized but efficient nature of neuromorphic systems causes analogue emulation to play a different role in the investigation of biological systems than does digital simulation. Analogue emulation is particularly useful for relating the physical properties of the system to its computational function because both levels of abstraction are combined in the same system. In many cases, these neuromorphic analogues make direct use of device physics to emulate the computational processes of neurons so that the base level of the analysis is inherent in the machine itself. Because the computation is cast as

a physical process, it is relatively easy to move from emulation to physiological prediction.

Analogue VLSI systems emphasize the nature of computation as a physical process and focus attention on the need for computational strategies that take account of fundamental hardware characteristics. For example, the inherent variability between transistors operating in their subthreshold regime means that analogues cannot depend on finely tuned parameters. Instead, like biological systems, they must rely on processing strategies that minimize the effects of component variation. Biological systems operate with exquisite sensitivity over a wide range of physiological and environmental conditions by adjusting their operating points and combining multiple outputs to extract and amplify meaningful signals without amplifying component noise. Consequently, biological strategies such as signal aggregation, adaptation, and lateral inhibition also improve the performance of analogue silicon systems. By designing neuromorphic systems, we enlarge our vocabulary of computational primitives that provide a basis for understanding computation in nervous systems. This circuit vocabulary can be written into silicon systems to perform practical tasks and test our understanding. This is the quest of neuromorphic aVLSI research.

ACKNOWLEDGMENTS

The preparation of this manuscript, and much of the research discussed in it, were supported by grants to R Douglas, M Mahowald, and K Martin by the Office of Naval Research, and to C Mead by the Office of Naval Research and the Defense Advanced Research Projects Administration. Fabrication facilites were provided by MOSIS.

Literature Cited

Atick J, Redlich A. 1990. Towards a theory of early visual processing. *Neural Comput.* 2: 308–20

Atick J, Redlich A. 1992. What does the retina know about natural scenes? *Neural Comput.* 4:196–210

Ballard DH. 1991. Animate vision. *Artif. Intell.* 48:57–86

Barlow HB. 1981. Critical factors limiting the design of the eye and visual cortex. The Ferrier lecture. *Proc. R. Soc. London Ser. B* 212: 1–34

Benson RG, Kerns DA. 1993. UV-activated conductances allow for multiple time scale learning. *IEEE Trans. Neural Networks* 4: 434–40

Boahen K, Andreou A. 1992. A contrast sensitive silicon retina with reciprocal synapses. See Moody 1992, pp. 764–72

Brownlow M, Tarassenko L, Murray A, Hamilton A, Han I, Reekie HM. 1990. Pulse-firing neural chips for hundreds of neurons. See Touretzky 1990, pp. 785–92

Carr CE. 1993. Processing of temporal information in the brain. *Annu. Rev. Neurosci.* 16:223–43

Churcher S, Baxter DJ, Hamilton A, Murray AF, Reekie HM. 1993. Generic analog neural

computation—the EPSILON chip. See Hanson 1993, pp. 773–80
Cowan J, Tesauro G, Alspector J, eds. 1994. *Advances in Neural Information Processing.* San Mateo, CA: Kaufmann
Delbrück T. 1993. Silicon retina with correlation-based, velocity-tuned pixels. *IEEE Trans. Neural Networks* 4:529–41
DeWeerth SP. 1992. Analog VLSI circuits for stimulus localization and centroid computation. *Int. J. Comp. Vision* 8:191–202
DeYong MR, Findley RL, Fields C. 1992. The design, fabrication, and test of a new VLSI hybrid analog-digital neural processing element. *IEEE Trans. Neural Networks* 3:363–74
Douglas R, Mahowald M. 1994a. A constructor set for silicon neurons. In *An Introduction to Neural and Electronic Networks,* ed. SF Zornetzer, JL Davis, C Lau, T McKenna. San Diego, CA: Academic. 2nd ed. In press
Douglas R, Mahowald M. 1994b. Silicon neurons. In *The Handbook of Brain Theory and Neural Networks,* ed. M Arbib. Cambridge, MA: Bradford. In press
Evans E. 1982. Functional anatomy of the auditory system. In *The Senses,* ed. H Barlow, J Mollon, pp. 251–63. Cambridge, UK: Cambridge Univ. Press
Fromherz P, Offenhäusser A, Vetter T, Weis J. 1991a. A neuron-silicon junction: a Retzius cell of the leech on an insulated-gate field-effect transistor. *Science* 252:1290–93
Fromherz P, Schaden H, Vetter T. 1991b. Guided outgrowth of leech neurons in culture. *Neurosci. Lett.* 129:77–80
Haken H. 1978. *Synergetics.* Berlin: Springer-Verlag. 2nd ed.
Hanson SJ, Cowan JD, Giles CL, eds. 1993. *Advances in Neural Information Processing Systems,* Vol. 5. San Mateo, CA: Kaufmann
Hertz J, Krogh A, Palmer RG. 1991. *Introduction to the Theory of Neural Computation* Redwood City, CA: Addison-Wesley
Hille B. 1992. *Ionic Channels of Excitable Membranes.* Sunderland, MA: Sinauer. 2nd ed.
Hodgkin AL, Huxley AF. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117:500–44
Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79:2554–58
Hopfield JJ. 1990. The effectiveness of analogue "neural network" hardware. *Network* 1:27–40
Hoppensteadt F. 1986. *An Introduction to the Mathematics of Neurons.* Cambridge, UK: Cambridge Univ. Press
Horiuchi T, Bair W, Bishofberger B, Moore A, Koch C, Lazzaro J. 1992. Computing motion using analog VLSI chips: an experimental comparison among different approaches. *Int. J. Comp. Vision* 8:203–16
Horiuchi T, Bishofberger B, Koch C. 1994. An analog VLSI saccadic eye movement system. See Cowan et al 1994, pp. 5821–89
Horiuchi T, Lazzaro JP, Moore A, Koch C. 1991. A delay-line based motion detection chip. See Lippman et al 1991, pp. 406–12
Julesz B. 1971. *Foundations of Cyclopean Perception.* Chicago, IL: Univ. Chicago Press
Keener J. 1983. Analogue circuitry for the FitzHugh-Nagumo equations. *IEEE Trans. Syst. Man Cybern.* 13:1010–14
Koch C, Li H, eds. 1994. *Vision Chips: Implementing Vision Algorithms with Analog VLSI Circuits.* Cambridge, MA: MIT. In press
Koch C, Poggio T. 1987. Biophysics of computation: Neurons, synapses, and membranes. In *Synaptic Function,* ed. GM Edelman, WE Gall, WM Cowan, pp. 637–97. New York: Wiley
Lansner JA, Lehmann T. 1993. An analog CMOS chip set for neural networks with arbitrary topologies. *IEEE Trans. Neural Newtorks* 4:441–44
Lazzaro J. 1991. Silicon model of an auditory neural representation of spectral shape. *IEEE J. Solid State Circuits* 26:772–77
Lazzaro J, Mead C. 1989. Silicon modeling of pitch perception. *Proc. Natl. Acad. Sci. USA* 86:9597–9601
Lazzaro J, Wawrzynek J, Mahowald M, Sivilotti M, Gillespie D. 1993. Silicon auditory processors as computer peripherals. *IEEE Trans. Neural Networks* 4:523–28
Lippman R, Moody J, Touretsky D, eds. 1991. *Advances in Neural Networks Information Processing Systems,* Vol. 3. San Mateo, CA: Kaufmann
Lyon RF, Mead CA. 1988. An analog electronic cochlea. *IEEE Trans. Acoustics Speech Signal Process.* 36:1119–34
Mahowald M. 1992. *VLSI analogs of neuronal visual processing: a synthesis of form and function.* PhD thesis. Calif. Inst. Technol., Pasadena, CA. 237 pp.
Mahowald M. 1994. *An Analog VLSI System for Stereoscopic Vision.* Boston, MA: Kluwer Academic
Mahowald M, Douglas R. 1991. A silicon neuron. *Nature* 354:515–18
Mahowald M, Mead C. 1989. Silicon retina. See Mead 1989, pp. 257–78
Marr D. 1982. *Vision.* San Francisco: Freeman
Marr D, Poggio T. 1976. Cooperative computation of stereo disparity. *Science* 194:283–87
McCullough W, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5:115–33

Mead C. 1989. *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley
Mead C. 1990. Neuromorphic electronic systems. *Proc. IEEE* 78:1629–36
Mead CA, Mahowald MA. 1988. A silicon model of early visual processing. *Neural Networks* 1:91–97
Mitchison G. 1992. Axonal trees and cortical architecture. *TINS* 15:122–26
Moody JE, Hanson SJ, Lippmann RP, eds. 1992. *Advances in Neural Information Processing Systems*, Vol. 4. San Mateo, CA: Kaufmann
Mueller P, Van der Spiegel J, Blackman D, Chiu T, Clare T, et al. 1989a. A programmable analog neural computer and simulator. See Touretzky 1989, pp. 712–19
Mueller P, Van der Spiegel J, Blackman D, Chiu T, Clare T, et al. 1989b. Design and fabrication of VLSI components for a general purpose analog neural computer. In *Analog VLSI Implementation of Neural Systems*, ed. C Mead, M Ismail, pp. 135–69. Boston, MA: Kluwer Academic
Murray A, Del Corso D, Tarassenko L. 1991. Pulse-stream VLSI neural networks mixing analog and digital techniques. *IEEE Trans. Neural Networks* 2:193–204
Murray A, Hamilton A, Tarassenko L. 1989. Programmable analog pulse-firing neural networks. See Touretzky 1989, pp. 712–19
Murray A, Tarasenko L. 1994. *Analogue Neural VLSI*. London: Chapman & Hall
Newell A, Shaw J, Simon H. 1958. Elements of a theory of human problem solving. *Psychol. Rev.* 65:151–66
Poggio G. 1984. Processing of stereoscopic information in primate visual cortex. In *Dynamic Aspects of Neocortical Function*, ed. GM Edelman, WE Gall, WM Cowan, pp. 613–35. New York: Wiley
Poggio G, Gonzales F, Krause F. 1988. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J. Neurosci.* 8:4531–50
Ramacher U, Ruckert U, eds. 1991. *VLSI Design of Neural Networks*. Boston, MA: Kluwer Academic
Rumelhart D, McClelland J. 1986. *Parallel Distributed Processing*. Cambridge, MA: MIT Press
Säckinger E, Boser BE, Jackel LD. 1992. A neurocomputer board based on the ANNA neural network chip. See Moody 1992, pp. 773–80
Scribner D, Sarkady K, Kruer M, Caulfield J, Hunt J, Colbert M, Descour M. 1993. Adaptive retina-like preprocessing for imaging detector arrays. In *IEEE International Conference on Neural Networks*, 3:1955–60. San Francisco: IEEE
Sivilotti MA, Mahowald MA, Mead CA. 1987. Real-time visual computation using analog CMOS processing arrays. In *1987 Stanford Conference on Very Large Scale Integration*, ed. P Losleben, pp. 295–311. Cambridge, MA: MIT Press
Sparks D, Mays L. 1990. Signal transformations required for the generation of saccadic eye movements. *Annu. Rev. Neurosci.* 13: 309–36
Stevens C. 1989. How cortical connectedness varies with network size. *Neural Comput.* 1: 473–79
Suga N. 1988. Auditory neuroethology and speech processing: complex-sound processing by combination-sensitive neurons. In *Auditory Function*, ed. G Edelman, W Gall, W Cowan, pp. 679–720. New York: Wiley
Touretzky DS, ed. 1989. *Advances in Neural Information Processing Systems*, Vol. 1. San Mateo, CA: Kaufmann
Touretzky DS, ed. 1990. *Advances in Neural Information Processing Systems*, Vol. 2. San Mateo, CA: Kaufmann
Vittoz E, Oguey H, Maher M, Nuys O, Dijkstra E, Chevroulet M. 1991. Analog storage of adjustable synaptic weights. See Ramacher & Ruckert 1991, pp. 47–63
von Neumann J. 1958. *The Computer and the Brain*. New Haven, CT: Yale Univ. Press
Watts L. 1993. *Cochlea mechanics: analysis and analog VLSI*. PhD thesis. Calif. Inst. Technol., Pasadena, CA
Werblin FS. 1974. Control of retinal sensitivity. *J. Gen. Physiol.* 63:62–87
Winston P, Shellard S. 1990. *Artificial Intelligence at MIT. Expanding Frontiers*, Vol. 1. Cambridge, MA: MIT Press
Wyatt J, Edell D, Raffel J, Narayanan M, Grumet A, et al. 1993. Silicon retinal implant to aid patients suffering from certain forms of blindness. *Interim Progr. Rep.*, Mass. Inst. Technol., Boston, MA
Yates G. 1990. The basilar membrane non-linear input-output function. In *The Mechanics and Biophysics of Hearing*, ed. P Dalos, C Geisler, J Matthews, M Ruggero, C Steele. Berlin: Springer-Verlag
Zaghoul ME, Meador JL, Newcomb RW, eds. 1994. *Silicon Implementation of the Pulse Coded Neural Networks*. Boston, MA: Kluwer Academic

Annual Review of Neuroscience
Volume 18, 1995

CONTENTS

GENE TARGETING IN ES CELLS, *Phillipe Soriano* 1

CREATING A UNIFIED REPRESENTATION OF VISUAL AND AUDITORY SPACE IN THE BRAIN, *E. I. Knudsen and M. S. Brainard* 19

BIOLOGY AND GENETICS OF HEREDITARY MOTOR AND SENSORY NEUROPATHIES, *Ueli Suter and G. Jackson Snipes* 45

TRIPLET REPEAT EXPANSION MUTATIONS: The Example of Fragile X Syndrome, *Stephen T. Warren and Claude T. Ashley, Jr.* 77

THE NEUROBIOLOGY OF INFANTILE AUTISM, *Andrea L. Ciaranello and Roland D. Ciaranello* 101

DYNAMIC REGULATION OF RECEPTIVE FIELDS AND MAPS IN THE ADULT SENSORY CORTEX, *Norman M. Weinberger* 129

ISOLATION, CHARACTERIZATION, AND USE OF STEM CELLS FROM THE CNS, *Fred H. Gage, Jasodhara Ray, and Lisa J. Fisher* 159

NEURAL MECHANISMS OF SELECTIVE VISUAL ATTENTION, *Robert Desimone and John Duncan* 193

FUNCTIONAL INTERACTIONS OF NEUROTROPHINS AND NEUROTROPHIN RECEPTORS, *Mark Bothwell* 223

NEUROMORPHIC ANALOGUE VLSI, *Rodney Douglas, Misha Mahowald, and Carver Mead* 255

SIGNAL TRANSDUCTION IN *DROSOPHILA* PHOTORECEPTORS, *Rama Ranganathan, Denise M. Malicki, and Charles S. Zuker* 283

LONG-TERM SYNAPTIC DEPRESSION, *David J. Linden and John A. Connor* 319

LOCALIZATION OF BRAIN FUNCTION: The Legacy of Franz Joseph Gall (1758–1828), *S. Zola-Morgan* 359

(*Continued*)

MECHANISMS OF NEURAL PATTERNING AND SPECIFICATION IN THE DEVELOPING CEREBELLUM, *Mary E. Hatten and Nathaniel Heintz* 385

LEARNING AND MEMORY IN THE VESTIBULO-OCULAR REFLEX, *Sascha du Lac, Jennifer L. Raymond, Terrence J. Sejnowski, and Stephen G. Lisberger* 409

THE ROLE OF AGRIN IN SYNAPSE FORMATION, *Mark A. Bowe and Justin R. Fallon* 443

MOLECULAR MECHANISMS OF DRUG REINFORCEMENT AND ADDICTION, *David W. Self and Eric J. Nestler* 463

THE ROLE OF THE FLOOR PLATE IN AXON GUIDANCE, *Sophia A. Colamarino and Marc Tessier-Lavigne* 497

MOLECULAR NEUROBIOLOGY AND GENETICS OF CIRCADIAN RHYTHMS IN MAMMALS, *Joseph S. Takahashi* 531

VISUAL FEATURE INTEGRATION AND THE TEMPORAL CORRELATION HYPOTHESIS, *Wolf Singer and Charles M. Gray* 555

INDEXES

Subject Index 587

Cumulative Index of Contributing Authors, Volumes 11–18 597

Cumulative Index of Chapter Titles, Volumes 11–18 601