Neural Hardware for Vision

by Carver A. Mead

BIOLOGY HAS ALWAYS BEEN the inspiration for computational metaphor. In the mid-1930s Alan Turing's original model for computation, which we call the sequential process, was based on the way mathematicians proved theorems. Because mathematicians are biological entities, we can say that even Turing's sequential process was inspired by the way biological systems work. But I will be discussing some biological systems that are simpler than mathematicians, since nobody, including mathematicians, can understand the way mathematicians work.

In the last decade or so the knowledge of what goes on in the brain has increased tremendously. When Max Delbrück first interested me in biology 20 years ago, the picture we had of the brain at that time was much more simplistic and much less analog in nature. At the time, neurobiologists were completely preoccupied with nerve impulses and the way they were generated in neurons. Now they are looking more deeply at the principles on which neural computation is based. And there are some surprises here. Nerve impulses, which are quasi-digital, play a surprisingly small role in the actual computation process. Most of the computation is analog, and it's done at the very tips of the dendritic tree of the neuron. Throughout the brain there is distributed feedback from these dendritic tips to the nerves that are driving them.

These new discoveries prompted us to take a fresh look at neural computation to see whether we might be able to synthesize systems that have some of the properties of real neural systems. It turns out that it's probably just the right time to be doing this. What's different today from attempts in the last 30 years to build neurocircuits is that now we have a technology that makes it possible to put a billion transistors on a six-inch wafer and interconnect them all. Conventional digital technology has difficulty using a full wafer, since many transistors are inoperative. Re-creating the brain's distributed analog computation gives us inherent redundancy and robustness under failure. We can actually use a substantial fraction of these billion transistors. So, the technology that was developed for microprocessors and memories has provided us a base on which we can build neural computing systems. These computing systems are based on very different principles from any of the conventional computing

engines, analog or digital, that were built in the past.

The particular system we have been working on is a very simple model of the part of the brain wrapped up behind the eyeball. Although it's quite simple by brain standards, it does a level of computation that even our most powerful computers today can't handle. The lens of the eye focuses an image on the surface of the retina, where the first levels of visual processing occur. When we want to see details of shapes, such as letters, the image gets focused on the fovea, a small area of the retina with tightly packed photoreceptors. But the fovea is responsible for only a fraction of the retina's activity. Most of the action happens at the periphery, where movement of the image produces signals that are transformed into nerve pulses that are transmitted over the optic nerve "cable" to the higher centers in the brain.

In a cross section through the retina one can see on the surface a layer of photoreceptors, below which lie layers of three different kinds of cells— bipolar, horizontal, and amacrine. Below these cells are the ganglion cells, whose axons form the fibers of the optic nerve. The principal signal flow in the retina runs from the receptors down through the bipolar cells (the horizontal and amacrine cells spread across a large area of the retina in layers transverse to the signal flow) and into the ganglion cells, which turn the signal into nerve pulses. In engineering terms we can say that the process starts by transducing the light energy into an electrical signal. We send that signal on to an amplifier and then off through a cable. The signals in the retina are all analog until they go out the cable as nerve pulses, which are quasi-digital (digital in amplitude but analog in time).
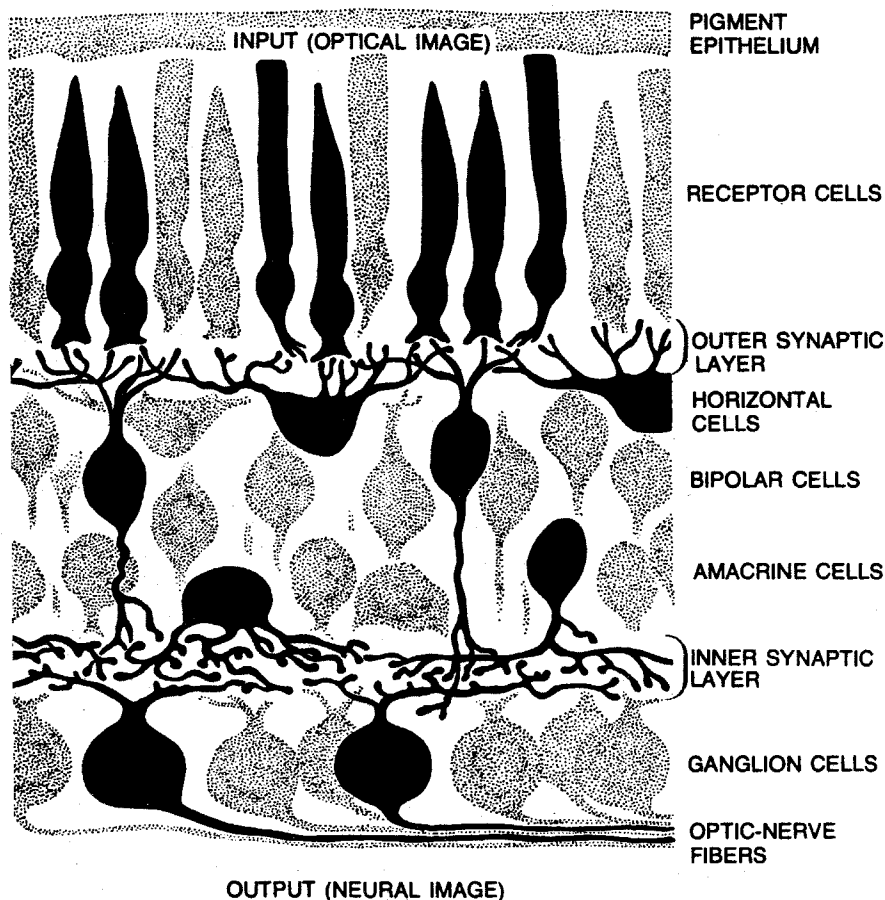
This basic structure (with some diversity in the details) is universal throughout the vertebrates. We can assume that the animals that evolved this eye structure ate any that did not. It is characteristic of biological systems that they are here because they work. An animal didn't live long if it couldn't see the predators that were about to jump on it, and its genes did not have a chance to get represented in the next generation. Because evolution has such a ruthless way of dealing with bad designs, we can view surviving biological structures as highly engineered systems.

The visual system is there to see things

about the world. The scene coming into the eye, however, is not the world. It's a bunch of photons that arrive because there is some light somewhere that shines on objects in the world and gets reflected off them into the eye. The light that falls on the image surface is the product of an illumination function multiplied by the reflectance of the object. But we don't want to see the illumination function; we want to see the object. Nobody ever got jumped on by an illumination function. So we take the logarithm of the intensity, and that factors the problem into the log of the illumination function, which is often a smooth function (except for shadows), plus the log of the reflectance of the object. The computation of the logarithm is done in the receptors or in their interactions with each other.

The visual system also has to make sure that the signals are within range. If they're not, you get blanked out. You have probably noticed this phenomenon, say, watching a baseball game on television. When someone hits a ball up into the stands, the television camera pans from the brightly lit field over to

INPUT (OPTICAL IMAGE)

PIGMENT EPITHELIUM

RECEPTOR CELLS

OUTER SYNAPTIC LAYER

HORIZONTAL CELLS

BIPOLAR CELLS

AMACRINE CELLS

INNER SYNAPTIC LAYER

GANGLION CELLS

OPTIC-NERVE FIBERS

OUTPUT (NEURAL IMAGE)

3

the stands in the shade. The camera has an elaborate automatic gain control system, but in such a mixed scene you see a pure white field and pure black stands; one signal is above range, and the other is below range, so you don't see anything at all. If an animal did that, its visual system would not be around in the next generation because the predators would simply jump from places that were half in the shade and half in sunlight.

But in the visual system, unlike the television camera, there is a measure of the local average intensity of the light; this value is used as the midpoint for the acceptable range of input levels. Basically this is a mechanism for deciding whether the pixel we are looking at is sufficiently different from the pixels around it to be reported. This level-normalization computation is performed by the horizontal cells. The horizontal cells look at the potentials on a bunch of photoreceptors and then take a spatial average. Then the difference between that spatial average and the local receptor is computed in the synaptic complex in the foot of the receptor. The resulting spatial derivative gets shipped on to the bipolar cells.

The outputs of the bipolar cells feed the amacrine layer, which is responsible for computing the time derivative of the signal. Rising edges of the bipolar waveform are turned into peaks, which in turn cause ganglion cells to fire. In rough terms, the amacrine layer is extracting motion information from the incoming retinal image. In some animals, like the frog, very elaborate motion computations are performed. A visual scene of the frog's natural habitat moving as a whole elicits no response. When a small, dark spot is moved relative to the background, however, a large response results. In higher vertebrates, much of this kind of complex motion calculation has migrated to visual cortex, and the retina computes a simple time derivative.

How much does something have to be moving for us to see it? The answer depends on how much the rest of the image is moving. Another level of gain control mechanism makes sure that, if we are going to report a derivative event, that event is significant relative to the rest of the scene. If we are looking at a tree, and the leaves are all blowing in the wind, something has to move significantly before we will report it. Otherwise, our higher levels of information processing would get overloaded by reports about all those little fluttering leaves. For a primate it usually takes something bigger than a leaf to jump on you and hurt you very much.

A derivative signal with respect to time is taken by the interaction of the bipolar, amacrine, and ganglion cells. Exactly how biological systems do this is not known. The local derivative with respect to time is compared to the derivatives that are being taken in the surrounding area. If the local signal is significantly larger, it gets reported.

We might wonder why so much of the information in the optic nerve is derivative. After all, we could just ship all the intensity information about the scene up the optic nerve. The optic nerve has a bandwidth approaching that of a television signal. People who design machine vision systems usually start with a television signal; they take one frame and compare it with the succeeding frame, and so on. Motion is characterized as something in one position in the first frame that is in a different position in the second frame.

It would be easy for a living system to do gain control in the camera, like television does, and then send the intensity information up to the brain to extract the motion information where there is a lot more horsepower to do so. So why go to all the trouble of building this elaborate derivative processing machinery down at the camera level? The answer is a straightforward one: A television camera samples every point on the image once every 1/30 of a second. But a predator in the visual field can move a distance of many pixels in 1/30 of a second. So what we have done is to take a simple problem — taking a directional derivative with respect to time — and transformed it into a complicated one. Now we have an image at time t and an image at $t + 1/30$, and we have to decide what point in the first image corresponds to what point in the succeeding image. So sampling transforms the processing task into the extremely difficult *correspondence problem*. People use supercomputers to try to solve that problem. Living systems didn't have supercomputers; they solved the problem the easy way and just took the derivative.

So when we built our rudimentary electronic retina, we built it to just take the derivative also. We based our system on the following four insights from biology:

1. It's important to take a logarithm of the signal, because logarithms factor the scene into the illumination function and the prop-
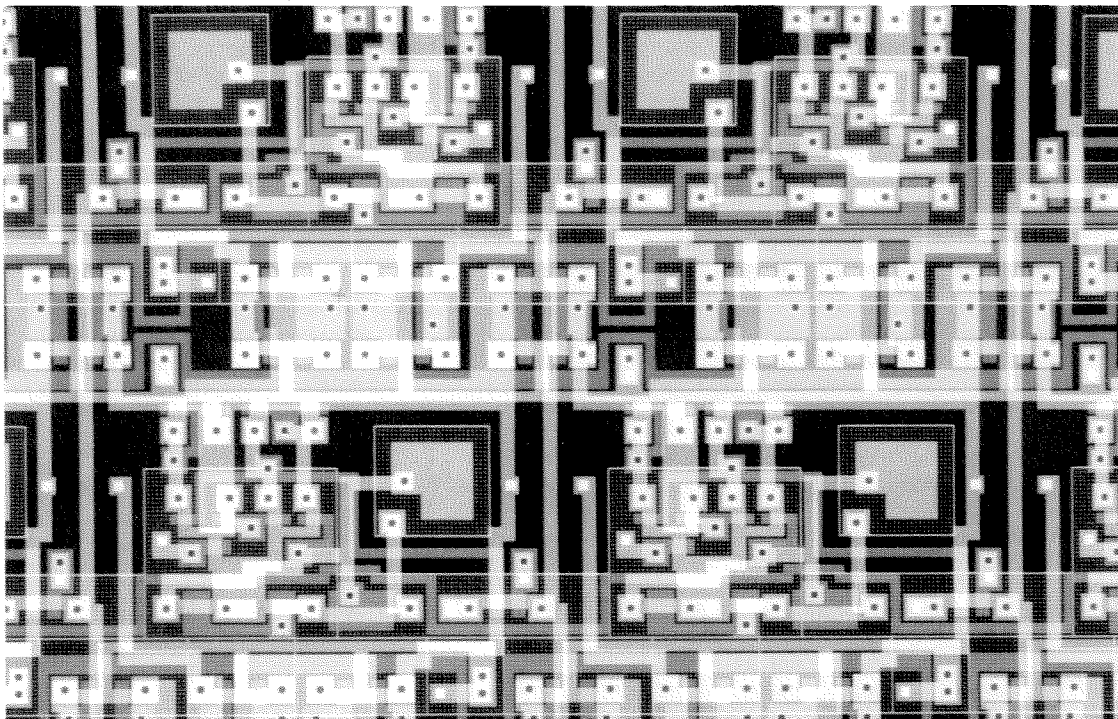
This computer drawing of a small group of pixels (one pixel appears on the cover) from the center of the retina shows how the individual cells are composed to form the processing array. The entire chip, shown on the following page, contains a 48 × 48 array of these pixels.

erties of the objects.

2. It's important to keep the signals in range.

3. Normalization should be done on a local basis; there is information in the shade and in the sunlight.

4. It's important to take time derivatives before we have sampled the image with respect to time. Otherwise, we would be throwing away the single most important piece of information in the image.

We have designed a simple retina and have implemented it on silicon in a standard, off-the-shelf CMOS (complementary metal-oxide semiconductor) process. The basic component is a photoreceptor, for which we use a bipolar transistor. In a CMOS process this is a parasitic device, that is, it's responsible for some problems in conventional digital circuits. But in our retina we take advantage of the gain of this excellent phototransistor.

There's nothing special about this fabrication process, and it's not exactly desirable from an analog point of view. Neurons in the brain don't have anything special about them either; they have limited dynamic range, they're noisy, and they have all kinds of garbage. But if we're going to build neural systems, we'd better not start off with a better process (with, say, a dynamic range of 10^5), because we'd simply be kidding ourselves that we had the right organizing principles. If we build a system that is organized on neural principles, we can stand a lot of garbage in the individual components and still get good information out. The nervous system does that, and if we're going to learn how it works, we'd better subject ourselves to the same discipline.

As in a biological eye, the first step is to take the logarithm of the signal arriving at the photoreceptor. To do this, we use the standard trick of electrical engineers, that is, to use an exponential element in a feedback loop. The voltage that comes out is the logarithm of the current that goes in. We think this operation is similar to the way living systems do it, although that is not proven. The element that we use to make this exponential consists of two MOS transistors stacked up. A nice property of this element is that the voltage range of the output is appropriate for subsequent processing by the kinds of amplifiers we can build in this technology. When we use the element to build a photoreceptor, the voltage out of the photoreceptor is logarithmic over four or five orders of magnitude of incoming light intensity. The lowest photocurrent is about 10^{-14} amps, which translates to a light level of 10^5 photons per second. This level corresponds approximately to moonlight, which is about the lowest level of light you can see with your cones.

There are two kinds of receptors in the eye — cones and rods. We use the cones under all normal circumstances and the rods

only in very low-illumination conditions. The rods are more sensitive, but they don't have good contrast sensitivity. Our silicon photoreceptor can't compete with the rods, but its intensity range is approximately that of the cones. It's a good photoreceptor, and it's logarithmic over the right range.

Now we can build a network of resistive elements patterned after the horizontal cells in the eye. The horizontal cells take the out-puts of all the receptors and average them spatially. They take a weighted average that is a function of distance from the local receptor; the farther away an input is, the less weight it is given. It's an extremely simple mechanism, and it's used in many places in peripheral sensory systems. We want to have something to compare our signal to, but we don't want that something to be global. A television camera blanks out because it compares each

The retina chip enlarged below is 5.4 millimeters by 4.8 millimeters in size and contains about 100,000 transistors.
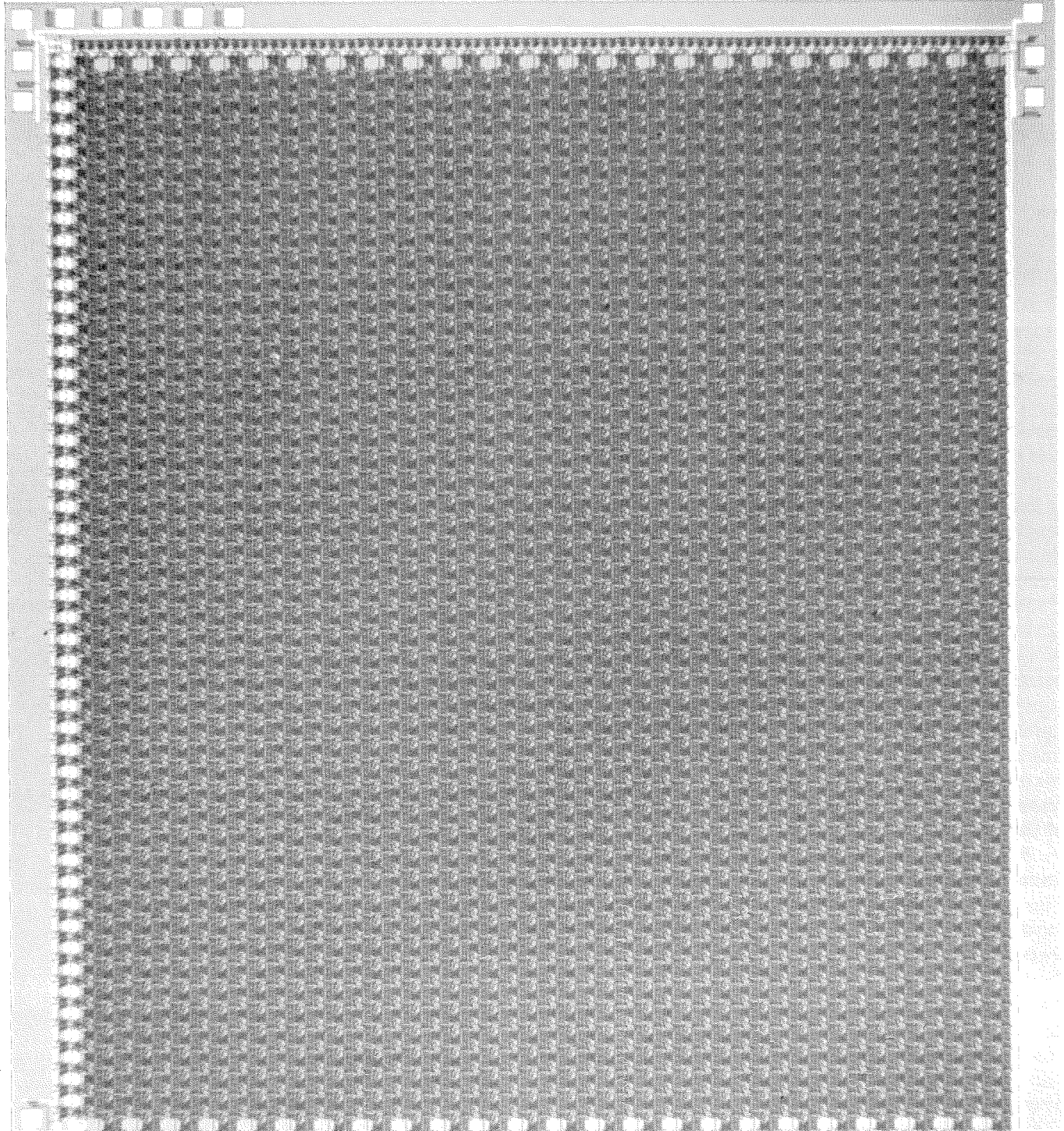
local signal with the average level over the entire scene. A biological visual system is more intelligent. It takes a local average, which gives progressively less weight to inputs that are farther away.

Our resistive networks turn out to be extremely good at calculating the spatial average. CMOS technology does not have a resistor of sufficiently high value as an inherent part of the process. All of our circuit components — resistors, capacitors, etc.— are made out of transistors. We have to build a little circuit that functions like a resistor, except that it has a mechanism to control the resistance. Each photoreceptor is hooked up to six neighbors in a hexagonal array linked by the resistive network that calculates the spatial average. The circuit is actually better than a regular resistor because, if the voltage between the two sides gets too big, the current that can go through it is limited. So, for example, if one of our pixels gets stuck, it doesn't take down the whole network. In a network of linear resistors, one stuck input could create damage for a large distance.

We made our amacrine cells out of a couple of amplifiers and a capacitor — again, all made out of transistors. Analogous to the amacrine cells' task in the visual system, this little circuit takes the derivative with respect to time. What it does is take the input signal, which corresponds to that coming out of the bipolar cell in the retina, compares it with a temporally smoothed version of the signal, and reports the difference as a finite time constant derivative circuit. The output represents the difference between the local signal and the time average of the surrounding signals. You can think about the computation that's done locally as taking the amplified difference between the local input and the space-and-time-averaged input, which is weighted over the surround in some way that dies off as it gets to farther neighbors. What our circuit does not have, which the amacrine cells do have, is a motion gain control. It will not turn down the gain if an object in the surround is moving. We have not yet evolved that level of processing.

Compared to an animal's eye, this is all very low-level. It's not the kind of thing that could recognize your grandmother or even locate tanks on a battlefield. But it's the first step in simulating the computation that your brain does to process a visual image. It's done in a smooth analog manner completely analogous to the way the eye does it. And it does indeed have tremendous advantages in the preservation of information compared with any kind of system that starts with a standard TV-type front end.

In a small way, we have embarked upon a second evolutionary path — that of a silicon nervous system. As in any evolutionary endeavor, we must start at the beginning. Our first systems have been simple and stupid. But they are, no doubt, smarter than the first animals were. We are, after all, endowed with the product of a few billion years of evolution with which to study them.

The constraints on our silicon systems are very similar to those on neural systems: Wire is limited, power is precious, robustness and reliability are essential. We may therefore expect the results of our second evolution to bear fruits of biological relevance. The effectivenesss of our approach will be in direct proportion to the attention we pay to the guiding biological metaphor. We use the term metaphor in a very deliberate and well-defined way. We are in no better position to "copy" biological nervous systems than we are to create a flying machine with feathers and flapping wings. But we can use biological organizing principles as a basis for our silicon systems in the same way that a soaring bird is an excellent model of a glider.

It is my conviction that our ability to realize simple neural functions is strictly limited by our understanding of their organizing principles and not by difficulties in realization. If we *really* understand a system, we will be able to build it. Conversely, we can be sure that a system is not fully understood until a working model has been synthesized and successfully demonstrated.

The silicon medium can thus be seen to serve two complementary but inseparable roles:

1. To give computational neuroscience a synthetic element allowing hypotheses concerning neural organization to be tested.

2. To develop an engineering discipline by which real-time collective systems can be designed for specific computations.

The success of this venture will create a bridge between neurobiology and the information sciences and bring us a much deeper view of computation as a physical process. It will also bring us a fresh new view of information processing and the enormous power of collective systems to solve problems that are completely intractable by traditional computer techniques. □