

Neuromorphic Microchips



Compact, efficient electronics based on the brain's neural system could yield implantable silicon retinas to restore vision, as well as robotic eyes and other smart sensors

■ ■ ■ ■ BY KWABENA BOAHEN

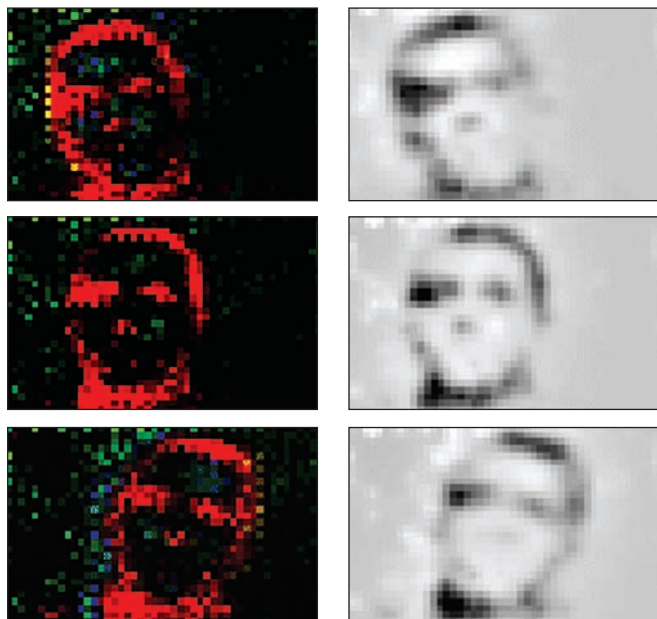
When IBM's Deep Blue supercomputer edged out world chess champion Garry Kasparov during their celebrated match in 1997, it did so by means of sheer brute force. The machine evaluated some 200 million potential board moves a second, whereas its flesh-and-blood opponent considered only three each second, at most. But despite Deep Blue's victory, computers are no real competition for the human brain in areas such as vision, hearing, pattern recognition, and learning. Computers, for instance, cannot match our ability to recognize a friend from a distance merely by the way he walks. And when it comes to operational efficiency, there is no contest at all. A typical room-size supercomputer weighs roughly 1,000 times more, occupies 10,000 times more space and consumes a millionfold more power than does the cantaloupe-size lump of neural tissue that makes up the brain.

How does the brain—which transmits chemical signals between neurons in a relatively sluggish thousandth of a second—end up performing some tasks faster and more efficiently than the most powerful digital processors? The secret appears to reside in how the brain organizes its slow-acting electrical components.

The brain does not execute coded instructions; instead it activates links, or synapses, between neurons. Each such activation is equivalent to executing a digital instruction, so one can compare how many connections a brain activates every second with the number of instructions a computer executes during the same time. Synaptic activity is staggering: 10 quadrillion (10^{16}) neural connections a second. It would take a million Intel Pentium-powered computers to match that rate—plus a few hundred megawatts to juice them up.

Now a small but innovative community of engineers is making significant progress in copying neuronal organization and function. Researchers speak of having “morphed” the structure of neural connections into silicon circuits, creating neuromorphic microchips. If successful, this work could lead to implantable silicon retinas for the blind and sound processors for the deaf that last for 30 years on a single nine-volt battery

IMPLANTABLE SILICON RETINA, shown in this artist's conception, could emulate the eye's natural function, restoring vision for patients with certain types of blindness.



SILICON RETINA senses the side-to-side head movements of University of Pennsylvania researcher Kareem Zaghloul. The four types of silicon ganglion cells on his Visio1 chip emulate real retinal cells' ability to preprocess visual information without huge amounts of computation. One class of cells responds to dark areas (*red*), whereas another reacts to light regions (*green*). A different set of cells tracks leading edges of objects (*yellow*) and trailing edges (*blue*). The gray-scale images, generated by decoding these messages, show what a blind person would see with neuromorphic retinal implants.

or to low-cost, highly effective visual, audio or olfactory recognition chips for robots and other smart machines [see box on opposite page].

Our team at the University of Pennsylvania initially focused on morphing the retina—the half-millimeter-thick sheet of tissue that lines the back of the eye. Comprising five specialized layers of neural cells, the retina “preprocesses” incoming visual images to extract useful information without

the need for the brain to expend a great deal of effort. We chose the retina because that sensory system has been well documented by anatomists. We then progressed to morphing the developmental machinery that builds these biological circuits—a process we call metamorphing.

Neuromorphing the Retina

THE NEARLY ONE MILLION ganglion cells in the retina compare visual signals received from groups of half a dozen to several hundred photoreceptors, with each group interpreting what is happening in a small portion of the visual field. As features such as light intensity change in a given sector, each ganglion cell transmits pulses of electricity (known as spikes) along the optic nerve to the brain. Each cell fires in proportion to the relative change in light intensity over time or space—not to the absolute input level. So the nerve’s sensitivity wanes with growing overall light intensity to accommodate, for example, the five-decade rise in the sky’s light levels observed from pre-dawn to high noon.

Misha A. Mahowald, soon after earning her undergraduate biology degree, and Carver Mead, the renowned microelectronics technologist, pioneered efforts to reproduce the retina in silicon at the California Institute of Technology. In their groundbreaking work, Mahowald and Mead reproduced the first three of the retina’s five layers electronically [“The Silicon Retina,” by Misha A. Mahowald and Carver Mead; *SCIENTIFIC AMERICAN*, May 1991]. Other researchers, several of whom passed through Mead’s Caltech laboratory (the author included), have morphed succeeding stages of the visual system as well as the auditory system. Kareem Zaghloul morphed all five layers of the retina in 2001 when he was a doctoral student in my lab, making it possible to emulate the visual messages that the ganglion cells, the retina’s output neurons, send to the brain. His silicon retina chip, Visio1, replicates responses of the retina’s four major types of ganglion cells, which feed into and together make up 90 percent of the optic nerve [see illustration on this page].

Zaghloul represented the electrical activity of each neuron in the eye’s circuitry by an individual voltage output. The voltage controls the current that is conveyed by transistors connected between a given location in the circuit and other points, mimicking how the body modulates the responses of neural synapses. Light detected by electronic photosensors affects the voltage in that part of the circuit in a way that is analogous to how it affects a corresponding cell in the retina. And by tiling copies of this basic circuit on his chip, Zaghloul replicated the activity in the retina’s five cell layers [see box on page 60].

The chip emulates the manner in which voltage-activated ion channels cause ganglion cells (and neurons in the rest of the brain) to discharge spikes. To accomplish this, Zaghloul installed transistors that send current back onto the same location in the circuit. When this feedback current arrives, it increases the voltage further, which in turn recruits more feedback current and causes additional amplification. Once a certain initial level is reached, this regenerative effect accelerates, taking the

Overview/*Inspired by Nature*

- Today’s computers can perform billions of operations per second, but they are still no match for even a young child when it comes to skills such as pattern recognition or visual processing. The human brain is also millions of times more energy-efficient and far more compact than a typical personal computer.
- Neuromorphic microchips, which take cues from neural structure, have already demonstrated impressive power reductions. Their efficiency may make it possible to develop fully implantable artificial retinas for people afflicted by certain types of blindness as well as better electronic sensors.
- Someday neuromorphic chips could even replicate the self-growing connections the brain uses to achieve its amazing functional capabilities.

NEUROMORPHIC ELECTRONICS RESEARCH GROUPS

Researchers seek to close the efficiency gap between electronic sensors and the body's neural networks with microchips that emulate the brain. This work focuses on small sensor systems that can be implanted in the body or installed in robots.

ORGANIZATION	INVESTIGATORS	PRINCIPAL OBJECTIVES
Johns Hopkins University	Andreas Andreou, Gert Cauwenberghs, Ralph Etienne-Cummings	Battery-powered speech recognizer, rhythm generator for locomotion and camera that extracts object features
ETH Zurich (University of Zurich)	Tobi Delbruck, Shi-Chii Liu, Giacomo Indiveri	Silicon retina and attention chip that automatically select salient regions in a scene
University of Edinburgh	Alan Murray, Alister Hamilton	Artificial noses and automatic odor recognition based on timing of signaling spikes
Georgia Institute of Technology	Steve DeWeerth, Paul Hasler	Coupled rhythm generators that coordinate a multisegmented robot
HKUST, Hong Kong	Bertram Shi	Binocular processor for depth perception and visual tracking
Massachusetts Institute of Technology	Rahul Sarpeshkar	Cochlea-based sound processor for implants for deaf patients
University of Maryland	Timothy Horiuchi	Sonar chip modeled on bat echolocation
University of Arizona	Charles Higgins	Motion-sensing chip based on fly vision

voltage all the way to the highest level, resulting in a spike.

At 60 milliwatts, Zaghoul's neuromorphic chip uses 1,000 times less electricity than a PC. With its low power needs, this silicon retina could pave the way for a total intraocular prosthesis—with camera, processor and stimulator all implanted inside the eye of a blind person who has retinitis pigmentosa or macular degeneration, diseases that damage photoreceptors but spare the ganglion cells. Retinal prostheses currently being developed, for example at the University of Southern California, provide what is called phosphene vision—recipients perceive the world as a grid of light spots, evoked by stimulating the ganglion cells with microelectrodes implanted inside the eye—and require a wearable computer to process images captured by a video camera attached to the patient's glasses. Because the microelectrode array is so small (fewer than 10 pixels by 10 pixels), the patient experiences tunnel vision—head movements are needed to scan scenes.

Alternatively, using the eye itself as the camera would solve the rubbernecking problem, and our chip's 3,600 ganglion-cell outputs should provide near-normal vision. Biocompatible encapsulation materials and stimulation interfaces, however, need further refinement before a high-fidelity prosthesis becomes a reality, maybe by 2010. Better understanding of how various retinal cell types respond to stimulation and how they contribute to perception is also required. In the interim, such neuromorphic chips could find use as sensors in automotive or security applications or in robotic or factory automation systems.

Metamorphing Neural Connections

THE POWER SAVINGS we attained by morphing the retina were encouraging, a result that started me thinking about how the brain actually achieves high efficiency. Mead was prescient when he recognized two decades ago that even if computing managed to continue along the path of Moore's law (which states that the number of transistors per square inch on inte-

grated circuits doubles every 18 months), computers as we know them could not reach brainlike efficiency. But how could this be accomplished otherwise? The solution dawned on me eight years ago.

Efficient operation, I realized, comes from the degree to which the hardware is customized for the task at hand. Conventional computers do not allow such adjustments; the software is tailored instead. Today's computers use a few general-purpose tools for every job; software merely changes the order in which the tools are used. In contrast, customizing the hardware is something the brain and neuromorphic chips have in common—they are both programmed at the level of individual connections. They adapt the tool to the specific job. But how does the brain customize itself? If we could translate that mechanism into silicon—metamorphing—we could have our neuromorphic chips modify themselves in the same fashion. Thus, we would not need to painstakingly reverse-engineer the brain's circuits. I started investigating neural development, hoping to learn more about how the body produces exactly the tools it needs.

Building the brain's neural network—a trillion (10^{12}) neurons connected by 10 quadrillion (10^{16}) synapses—is a daunting task. Although human DNA contains the equivalent of a billion bits of information, that amount is not sufficient to specify where all those neurons should go and how they should connect. After employing its genetic information during early development, the brain customizes itself further through internal interactions among neurons and through external interactions with the world outside the body. In other words, sensory neurons wire themselves in response to sensory inputs. The overall rule that regulates this process is deceptively simple: neurons that fire together wire together. That is, out of all the signals that a neuron receives, it accepts those from neurons that are consistently active when it is active, and it ignores the rest.

To learn how one layer of neurons becomes wired to an-

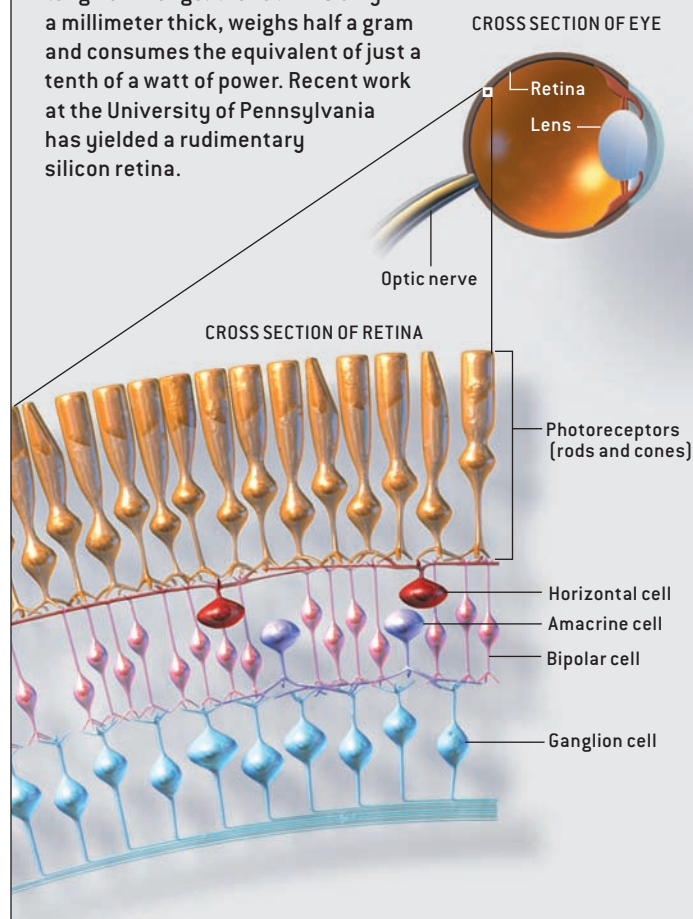
other, neuroscientists have studied the frog's retinotectal projection, which connects its retina to its tectum (the part of the midbrain that processes inputs from sensory organs). They have found that wiring one layer of neurons to another occurs in two stages. A newborn neuron extends projections ("arms") in a multilimbed arbor. The longest arm becomes the axon, the cell's output wire; the rest serve as dendrites, its input wires. The axon then continues to grow, towed by an amoeboid structure at its tip. This growth cone, as scientists call it, senses chemical gradients laid down by trailblazing precursors of neural communication signals, thus guiding the axon to the right street in the tectum's city of cells but not, so to speak, to the right house.

Narrowing the target down to the right house in the tectum requires a second step, but scientists do not understand this process in detail. It is well known, though, that neighboring retinal ganglion cells tend to fire together. This fact led me to speculate that an axon could find its retinal cell neighbors in the tectum by homing in on chemical scents released by active tectal neurons, because its neighbors were most likely at the source of this trail. Once the axon makes contact with the tectal neuron's dendritic arbor, a synapse forms between them and, voilà, the two neurons that fire together are wired together.

In 2001 Brian Taba, a doctoral student in my lab, built a chip modeled on this facet of the brain's developmental process. Because metal wires cannot be rerouted, he decided to reroute spikes instead. He took advantage of the fact that Zaghoul's Visio1 chip outputs a unique 13-bit address every time one of its 3,600 ganglion cells spikes. Transmitting addresses rather than spikes gets around the limited number of input/output pins that chips have. The addresses are decoded by the receiving chip, which re-creates the spike at the correct location in its silicon neuron mosaic. This technique produces a virtual bundle of axons running between corresponding locations in the two chips—a silicon optic nerve. If we substitute one address with another, we reroute a virtual axon belonging to one neuron (the original address) to another location (the substituted address). We can route these "softwires," as we call them, anywhere we want to by storing the substitutions in a database (a look-up table) and by using the original address to retrieve them [see box on page 62].

In Taba's artificial tectum chip, which he named Neurotrope1, softwires activate gradient-sensing circuits (silicon growth cones) as well as nearby silicon neurons, which are situated in the cells of a honeycomb lattice. When active, these silicon neurons release electrical charge into the lattice, which

Biological sensory systems provide compact, energy-efficient models for neuromorphic electronic sensors. Engineers attempting to duplicate the retina in silicon face a tough challenge: the retina is only half a millimeter thick, weighs half a gram and consumes the equivalent of just a tenth of a watt of power. Recent work at the University of Pennsylvania has yielded a rudimentary silicon retina.



Taba designed to conduct charge like a transistor. Charge diffuses through the lattice much like the chemicals released by tectal cells do through neural tissue. The silicon growth cones sense this simulated diffusing "chemical" and drag their softwires up the gradient—toward the charge's silicon neuron source—by updating the look-up table. Because the charge must be released by the silicon neuron and sensed by the silicon growth cone simultaneously, the softwires end up connecting neurons that are active at the same time. Thus, Neurotrope1 wires together neurons that fire together, as would occur in a real growing axon.

Starting with scrambled wiring between the Visio1 chip and the Neurotrope1 chip, Taba successfully emulated the tendency of neighboring retinal ganglion cells to fire together by activating patches of silicon ganglion cells at random. After stimulating several thousand patches, he observed a dramatic change in the softwiring between the chips. Neighboring artificial ganglion cells now connected to neurons in the silicon tectum that were twice as close as the initial connections. Be-

THE AUTHOR

KWABENA BOAHEN is a neuromorphic engineer and associate professor of bioengineering at the University of Pennsylvania. He left his native Ghana to pursue undergraduate studies in electrical and computer engineering at Johns Hopkins University in 1985 and became interested in neural networks soon thereafter. Boahen sees a certain elegance in neural systems that is missing in today's computers. He seeks to capture this sophistication in his silicon designs.

BRYAN CHRISTIE DESIGN; KAREEM ZAGHLOUL (silicon retina)

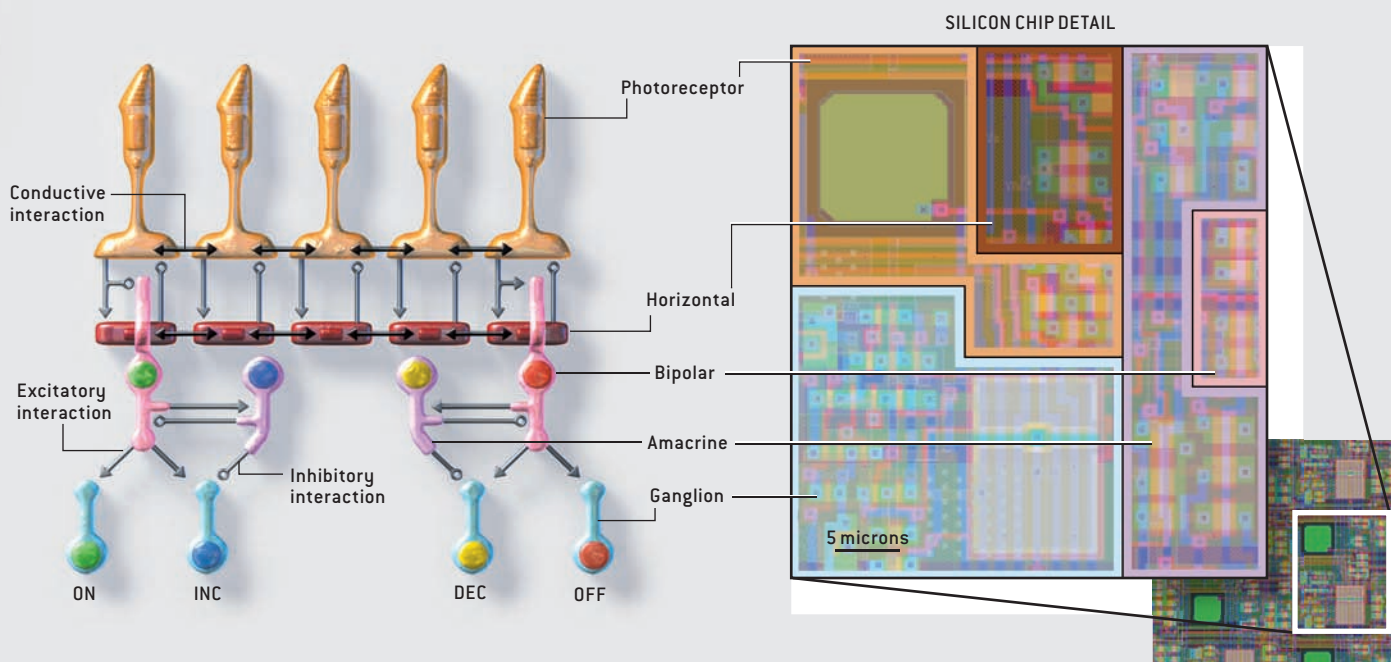
NEURONS AND NEUROMORPHIC VISION CHIPS

BIOLOGICAL RETINA

The cells in the retina, which are interconnected, extract information from the visual field by engaging in a complex web of excitatory (*one-way arrows*), inhibitory (*circles on a stick*), and conductive or bidirectional (*two-way arrows*) signaling. This circuitry generates the selective responses of the four types of ganglion cells (*at bottom*) that make up 90 percent of the optic nerve's fibers, which convey visual information to the brain. On (*green*) and Off (*red*) ganglion cells elevate their firing (spike) rates when the local light intensity is brighter or darker than the surrounding region. Inc (*blue*) and Dec (*yellow*) ganglion cells spike when the intensity is increasing or decreasing, respectively.

SILICON RETINA

Neuromorphic circuits emulate the complex interactions that occur among the various retinal cell types by replacing each cell's axons and dendrites (signal pathways) with metal wires and each synapse with a transistor. Permutations of this arrangement produce excitatory and inhibitory interactions that mimic similar communications among neurons. The transistors and the wires that connect them are laid out on silicon chips. Various regions of the chip surface perform the functions of the different cell layers. The large green squares are phototransistors, which transduce light into electricity.



cause of noise and variability, however, the wiring was not perfect: terminals of neighboring cells in the silicon retina did not end up next to one another in the silicon tectum. We wondered how the elaborate wiring patterns thought to underlie biological cortical function arise—and whether we could get further tips from nature to refine our systems.

Cortical Maps

TO FIND OUT, we had to take a closer look at what neuroscience has learned about connections in the cortex, the brain region responsible for cognition. With an area 16 inches in diameter, the cortex folds like origami paper to fit inside the skull. On this amazing canvas, “maps” of the world outside are drawn during infancy. The best-studied example is what scientists call area V1 (the primary visual cortex), where visual messages from the optic nerve first enter the cortex. Not only are the length and width dimensions of an image mapped onto V1 but also the orientation of the edges of objects therein. As a result, neurons in V1 respond best to edges oriented at a

particular angle—vertical lines, horizontal lines, and so forth. The same orientation preferences repeat every millimeter or so, thereby allowing the orientations of edges in different sectors of the visual scene to be detected.

Neurobiologists David H. Hubel and Torsten N. Wiesel, who shared a Nobel Prize in medicine for discovering the V1 map in the 1960s, proposed a wiring diagram for building a visual cortex—one that we found intimidating. According to their model, each cortical cell wires up to two groups of thalamic cells, which act as relays for retinal signals bound for the cortex. One group of thalamic cells should respond to the sensing of dark areas (which we emulate with Visio1's Off cells), whereas the other should react to the sensing of light (like our Visio1's On cells). To make a cortical cell prefer vertical edges, for instance, both groups of cells should be set to lie along a vertical line but should be displaced slightly so the Off cells lie just to the left of the On cells. In that way, a vertical edge of an object in the visual field will activate all the Off cells and all the On cells when it is in the correct position. A horizontal

edge, on the other hand, will activate only half the cells in each group. Thus, the cortical cell will receive twice as much input when a vertical edge is present and respond more vigorously.

At first we were daunted by the detail these wiring patterns required. We had to connect each cell according to its orientation preference and then modify these wiring patterns systematically so that orientation preferences changed smoothly, with neighboring cells having similar preferences. As in the cortex, the same orientations would have to be repeated every millimeter, with those silicon cells wired to neighboring locations in the retina. Taba's growth cones certainly could not

cope with this complexity. In late 2002 we searched for a way to escape this nightmare altogether. Finally, we found an answer in a five-decade-old experiment.

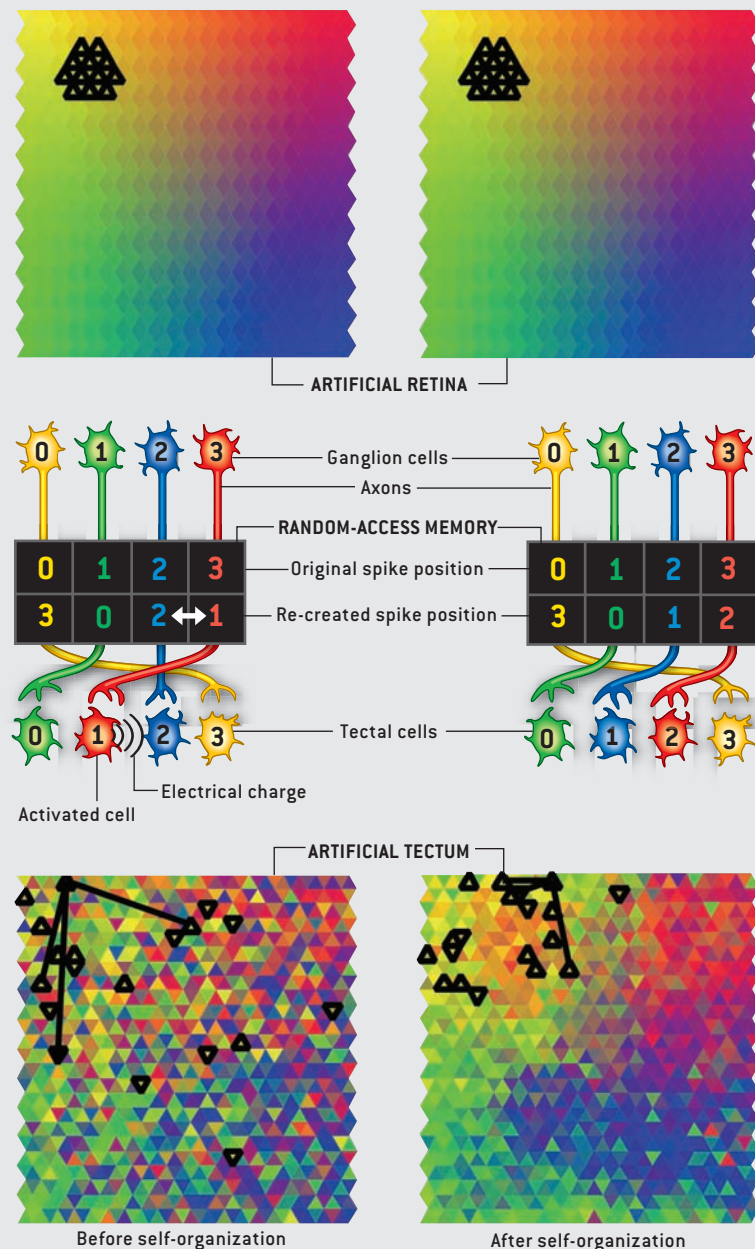
In the 1950s famed computer scientist Alan M. Turing showed how ordered patterns such as a leopard's spots or a cow's dapples could arise spontaneously from random noise. We hoped we could use a similar technique to create neighboring regions with similar orientation patterns for our chip. Turing's idea, which he tested by running simulations on one of the first electronic computers at the University of Manchester, was that modeled skin cells would secrete "black dye" or

MAKING CONNECTIONS (BIOLOGICAL OR SILICON)

In the early stages of the eye's development, ganglion cells in the retina project axons into a sensory center of the midbrain called the tectum. The retinal axons home in on chemical trails released by neighboring tectal cells that are activated at the same time, so neurons that fire together wire together. Ultimately, a map of the retinal sensors' spatial organization forms in the midbrain.

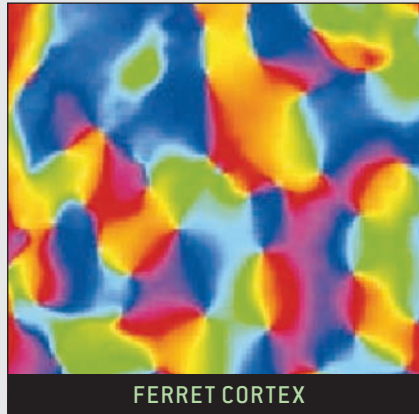
To emulate this process, University of Pennsylvania neuromorphic engineers use "softwires" to self-organize links between cells in their silicon retina chip, Visio1 (*top*), and those in their artificial tectum chip, Neurotrope1 (*bottom*). Electrical output pulses called spikes are "routed" from the artificial ganglion cells to the tectal cells using a random-access memory (RAM) chip (*middle*). The retinal chip supplies the address of the spiking silicon neuron, and the tectal chip re-creates that pulse at the corresponding location. In this example, the artificial tectum instructs the RAM to swap address entries 1 and 2. As a result, ganglion cell 2's axon terminus moves to tectal cell 1, bumping ganglion cell 3's axon from that location. The axons "sense" the gradient of electrical charge released by an activated silicon tectal cell, which helps to guide the connections.

After engineers repeatedly activated patches of neighboring silicon neurons in the artificial retina (*outlined triangles, top left*), the tectal cells' axon end points—which were initially widely distributed (*outlined triangles, bottom left*)—grew closer, yielding more uniform swaths on a colorized map (*bottom right*).

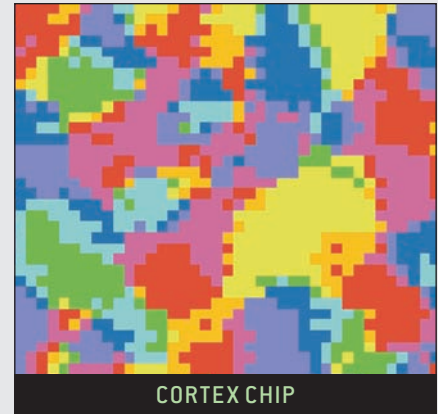


ORIENTATION PREFERENCES IN THE BRAIN AND IN SILICON

In both the visual cortex of a ferret (*left*) and a neuromorphic cortex chip (*right*), researchers have mapped the location of cells that respond preferentially to object edges of a certain orientation (*key, below*). In both maps, neighboring cells tend to have similar orientation preferences, which shows that the cortex chip emulates the biological system.



FERRET CORTX



CORTX CHIP

“bleach” indiscriminately. By introducing variations among the cells so that they produced slightly different amounts of dye and bleach, Turing generated spots, dapples and even zebra-like stripes. These slight initial differences were magnified by blotting and bleaching to create all-or-nothing patterns. We wondered if this notion would work for cortical maps.

Four years ago computational neuroscientist Misha Tsodyks and his colleagues at the Weizmann Institute of Science in Rehovot, Israel, demonstrated that, indeed, a similar process could generate cortex-like maps in software simulations. Paul Merolla, another doctoral student in my lab, took on the challenge of getting this self-organizing process to work in silicon. We knew that chemical dopants (impurities) introduced during the microfabrication process fell randomly, which introduced variations among otherwise identical transistors, so we felt this process could capture the randomness of gene expression in nature. That is putatively the source of variation of spot patterns from leopard to leopard and of orientation map patterns from person to person. Although the cells that create these patterns in nature express identical genes, they produce different amounts of the corresponding dye or ion channel proteins.

With this analogy in mind, Merolla designed a single silicon neuron and tiled it to create a mosaic with neuronlike excitatory and inhibitory connections among neighbors, which played the role of blotting and bleaching. When we fired up the chips in 2003, patterns of activity—akin to a leopard’s spots—emerged. Different groups of cells became active when we presented edges with various orientations. By marking the locations of these different groups in different colors, we obtained orientation preference maps similar to those imaged in the V1 areas of ferret kits [see box above].

Building Brains in Silicon

HAVING MORPHED the retina’s five layers into silicon, our goal turned to doing the same to all six of the visual cortex’s layers. We have taken a first step by morphing layer IV, the cortex’s input layer, to obtain an orientation preference map in an im-

mature form. At three millimeters, however, the cortex is five times thicker than the retina, and morphing all six cortical layers requires integrated circuits with many more transistors per unit area.

Chip fabricators today can cram a million transistors and 10 meters of wire onto a square millimeter of silicon. By the end of this decade, chip density will be just a factor of 10 shy of cortex tissue density; the cortex has 100 million synapses and three kilometers of axon per cubic millimeter.

Researchers will come close to matching the cortex in terms of sheer numbers of devices, but how will they handle a billion transistors on a square centimeter of silicon? Thousands of engineers would be required to design these high-density nanotechnology chips using standard methods. To date, a hundredfold rise in design engineers accompanied the 10,000-fold increase in the transistor count in Intel’s processors. In comparison, a mere doubling of the number of genes in flies to that of humans enabled evolutionary forces to construct brains with 10 million times more neurons. More sophisticated developmental processes made possible the increased complexity by elaborating on a relatively simple recipe. In the same way, morphing neural development processes instead of simply morphing neural circuitry holds great promise for handling complexity in the nanoelectronic systems of the future.

SA

MORE TO EXPLORE

Analog VLSI and Neural Systems. Carver Mead. Addison-Wesley, 1989.

Topographic Map Formation by Silicon Growth Cones. Brian Tabak and Kwabena Boahen in *Advances in Neural Information Processing Systems*, Vol. 15. Edited by Suzanna Becker, Sebastian Thrun and Klaus Obermayer. MIT Press, 2003.

Optic Nerve Signals in a Neuromorphic Chip. Kareem A. Zaghloul and Kwabena Boahen in *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 4, pages 657–675; 2004.

A Recurrent Model of Orientation Maps with Simple and Complex Cells. Paul Merolla and Kwabena Boahen in *Advances in Neural Information Processing Systems*, Vol. 16. Edited by Sebastian Thrun, Larry Saul and Bernhard Scholkopf. MIT Press, 2004.

The author’s Web site: www.neuroengineering.upenn.edu/boahen