PetaVision

Simulating the Thoughts behind the Scenes

The Laboratory's Synthetic Visual Cognition Team is using state-of-the-art supercomputers to learn how the brain sees and does other tasks.

The gift of sight is truly amazing. You "instantly" know what everything is in your field of view without seeming to think about it, without asking yourself, "What am I looking at?"

But to give you this knowledge, your brain must quickly make sense of the huge amounts of visual information constantly gathered by your eyes. How does your brain do it?

Scientists don't know exactly.

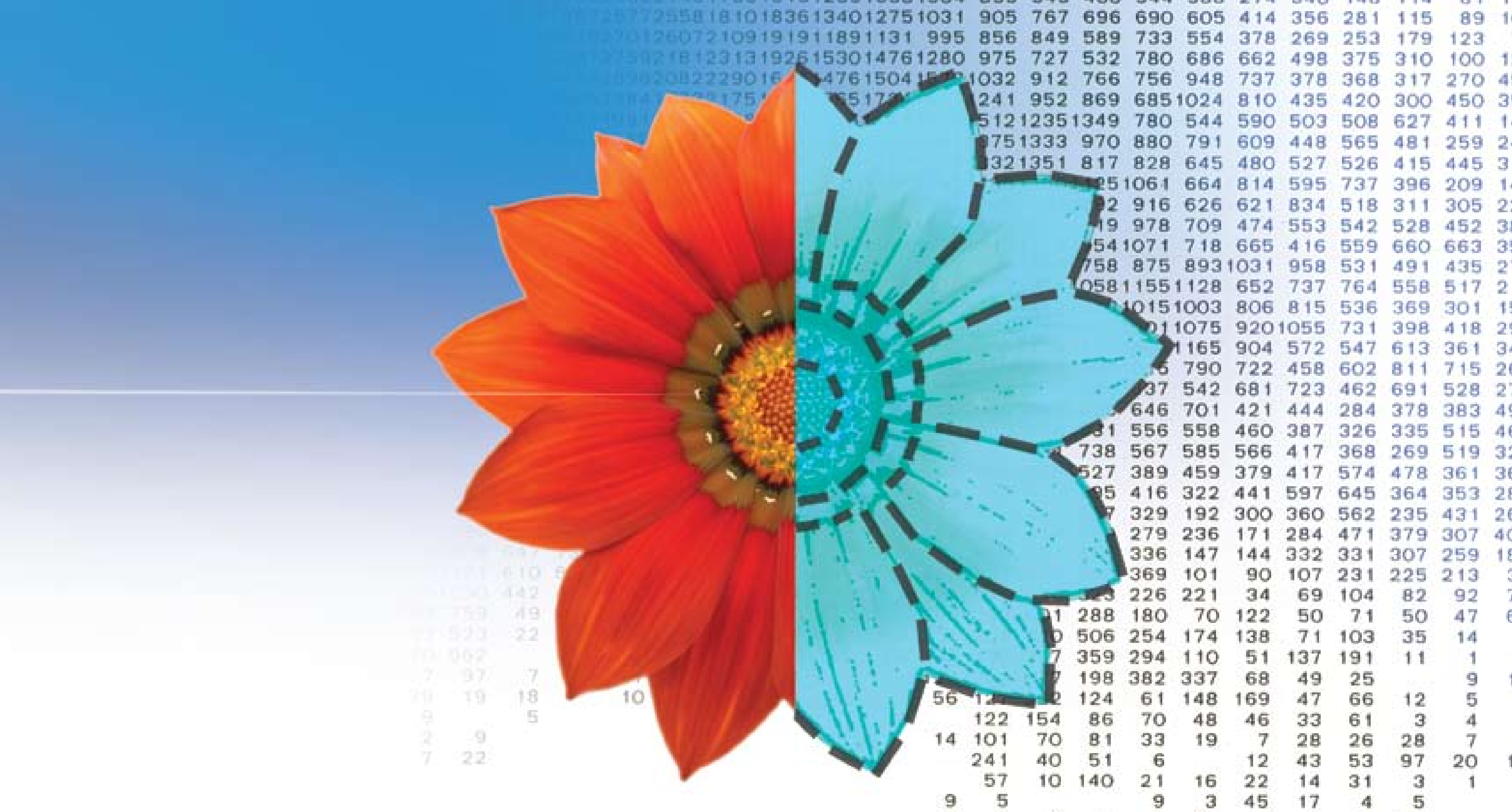
"Brain research is in a pre-paradigm state," says Garrett Kenyon, a Los Alamos neuroscientist and member of the Laboratory's Synthetic Visual Cognition Team. "We know lots of things about the brain, but we don't really know how it works."

As a result, computer programs designed to emulate the way the brain processes visual information don't begin to approach human levels of performance. For example, an MIT-developed computer-vision program—currently the most-accurate program at identifying objects—misidentifies what it sees 10 percent of the time.

"Imagine that when you crossed the street, 10 percent of the time what you thought was a billboard was actually an oncoming truck," says Luis Bettencourt, leader of the Synthetic Visual Cognition Project. "Clearly this sort of inaccuracy can be lethal in the real world."

So what's missing in the computer programs? What do computers need in order to see as well as people do?

Finding the answer could one day help robots navigate through buildings and cities without running into walls or getting run over and let computers take the wheel of your car in an emergency. It could also allow rapid, automated analysis of the huge volumes of data beamed down each day from reconnaissance satellites or enable computers to identify faces in video



taken at airports—a task at which existing computer methods fail dismally.

Understanding how the brain sees requires a good theory of how the brain works. But neuroscientists disagree about exactly what's needed to formulate such a theory. Research teams all over the world, including the Laboratory's Synthetic Visual Cognition Team, are exploring various possibilities, often aided by advanced supercomputers.

In collaboration with researchers at MIT and elsewhere, the Los Alamos team plans mainly to explore several mechanisms that could improve our understanding of how the brain processes visual information, which should lead to a better understanding of how the brain does all of its tasks. One of the team's major goals is improving the performance of computer-vision software to human levels.

In addition to Laboratory neuroscientists and advanced-computing specialists, the Synthetic Visual Cognition Project features a fairly unique piece of Laboratory hardware—the Roadrunner supercomputer. Roadrunner set the record for supercomputer speed last summer, running software developed by the Synthetic Visual Cognition Team.

Jumping-Off Place

The starting point for several of the team's studies is the MIT computer-vision program. The program implements a model of the primate visual cortex, which is where the brain processes most visual information. The visual cortex is also the best-understood part of the brain. The model is based on experimental studies of how monkeys and humans process visual information.

The team developed a new version of the MIT program to run on "hybrid-architecture" computers such as the Roadrunner supercomputer and a "mini" Roadrunner at the Los Alamos Center for Nonlinear Studies. The new program is called PetANNet, named for the fact that its computer-simulated neurons are connected to compose what's called a neural network, or neural net.

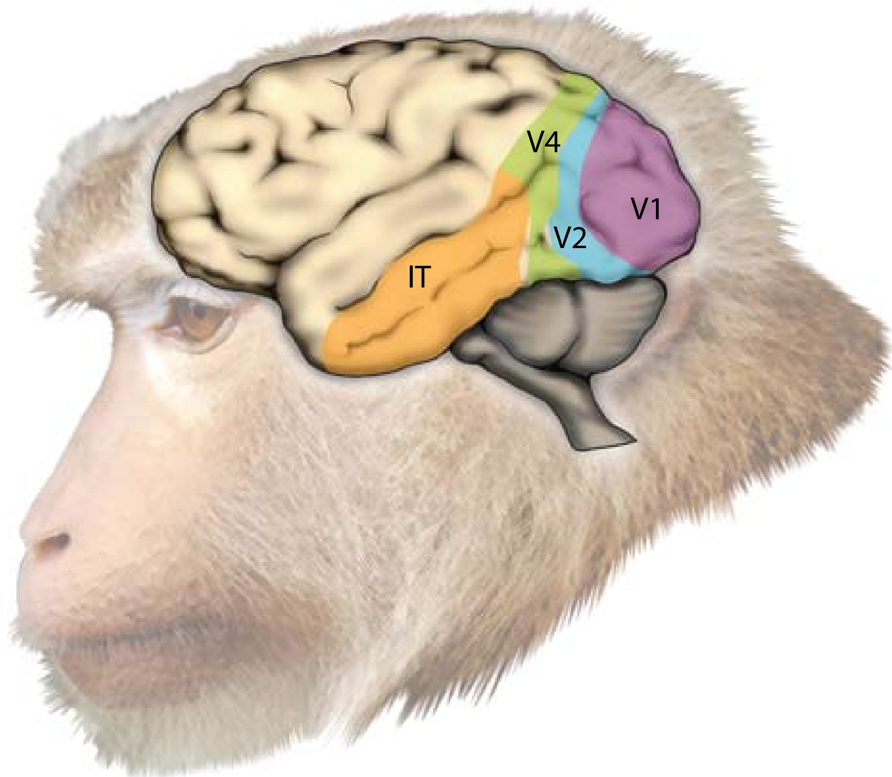
Like the MIT program, PetANNet implements a model of the "what" pathway, the neural pathway that identifies objects in one's field of view. A separate pathway, the "where" pathway, identifies the locations of objects in the visual field. The visual cortex is divided into four major areas from back to front: V1, V2, V4, and IT. The "what" and "where" pathways flow through all four areas, with the "what" pathway on the dorsal side (underside) of the gray matter and the "where" pathway on the ventral (upper) side. In PetANNet, information flows through the "what" pathway almost entirely in the forward direction from V1 to IT, that is, in a "feed-forward" fashion.

Visual information enters the "what" pathway through the lens (cornea) of the eye. The cornea focuses images onto the retina, at the back of the eye, where photoreceptors convert light to the electrical signals the brain's neurons use to communicate with each other. "Roughly speaking," Kenyon says, "your eye has about 500,000 photoreceptors, which is about equal to a half-a-megapixel camera."

The electrical signals from the retina go directly to the back of the brain, to V1, and are then processed through the visual cortex, starting with V1 and ending with IT. The field of view is first characterized in terms of simple visual features present in small square sections of the visual field and then in terms of combinations of simple features that represent more-complex features present in larger sections of the visual field. As the information is processed, individual neurons further up the processing hierarchy recognize features that are more and more complex and present in larger and larger sections of the visual field.

Near the top of the processing hierarchy, in V4, complex features, such as ears and noses, are recognized by individual neurons that view sizable fractions of the visual field—a fact that has been proven through electrophysiology experiments. In IT, individual neurons respond to objects or types of objects that appear anywhere in the visual field, regardless of how they're lit or oriented. "The magic is that an object, say, a face, is identified in IT as belonging to a distinct category regardless of the scene it happens to be part of," says Bettencourt. The V1-to-IT processing hierarchy is illustrated in the figure on the facing page.

Feed-forward processing is thought to determine the minimum time necessary for primates to see and identify objects. Experiments show that when a scene is presented to the visual cortex of a monkey or a human,



The PetANNet Model

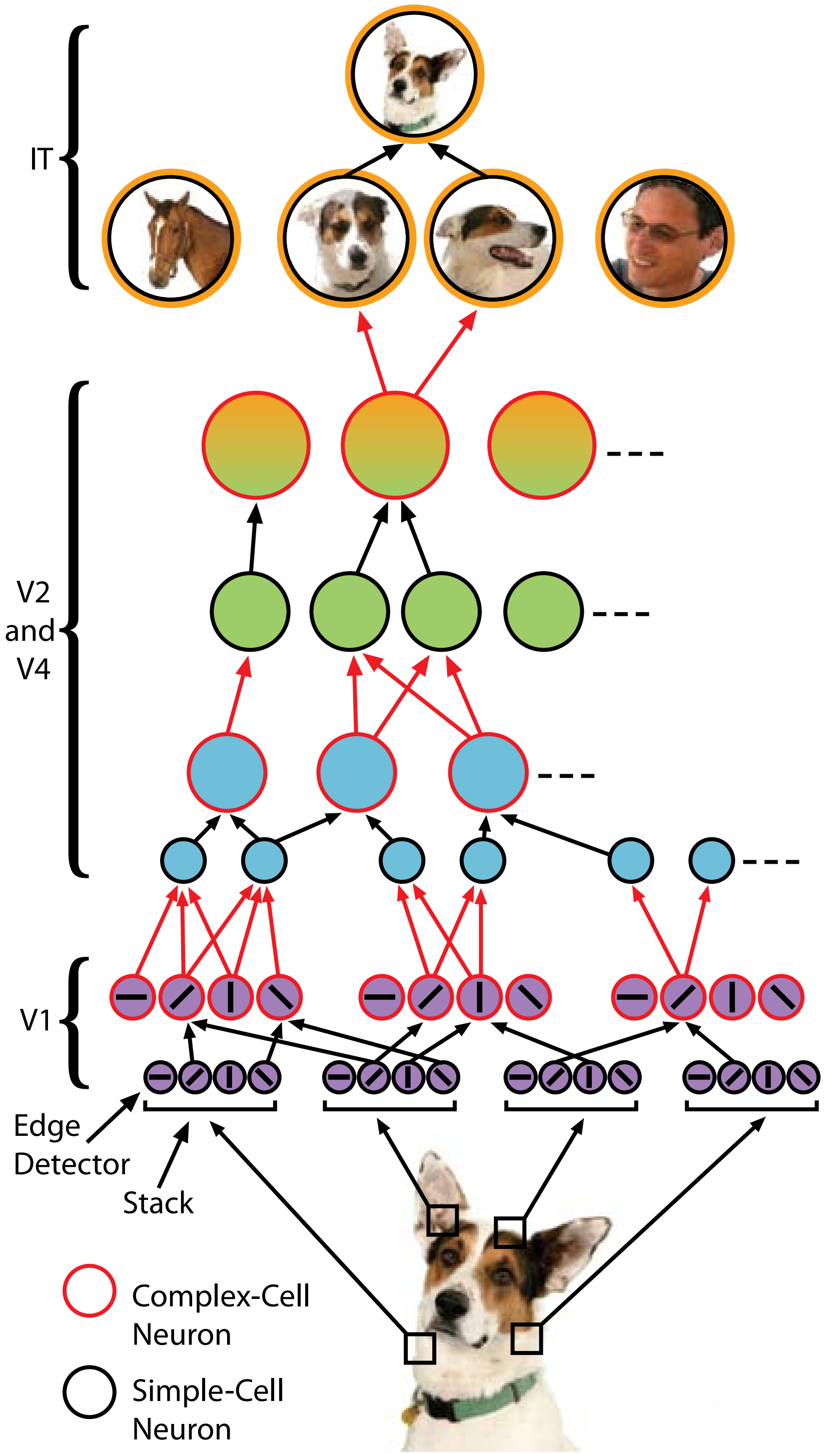
Above: The four major areas in the "what" neural pathway of the primate visual cortex—the pathway that identifies objects in the visual field—are V1, V2, V4, and IT, color coded here to map to the PetANNeT model.

Right: Visual processing begins with V1 dividing the field of view, which in this case contains a dog, into many small square sections, each consisting of about 7 × 7 pixels (photoreceptors). V1 contains "stacks" of neurons, and each stack views one of the small squares. Each neuron in a stack is a "feature detector" that responds to a specific feature in the square viewed by that stack.

The most-prevalent feature detectors in V1 are "edge detectors," each of which is a "simple-cell" neuron sensitive to an edge at a specific angle to the horizontal—exactly horizontal, standing vertical, slanted at 70 degrees, and so on. Other feature detectors are sensitive to, for example, color, spatial frequency, direction of motion, and so on. The visual information next goes to a layer of "complex-cell" neurons, where information is sampled and "pooled." One complex-cell neuron will sample the outputs of several edge detectors sensitive to the same edge angle. By "pooling" information in this way, the complex cell determines if that particular edge angle is present in a larger section of the field of view. Thus begins the process of identifying an object, regardless of where it is in the visual field or how it is oriented or lit.

In V2, a new set of simple-cell neurons monitors the outputs of combinations of the V1 complex-cell neurons over a larger field of view. Each combination represents a new feature that is more complex than those viewed by the stacks in V1 and that is present in a larger swath of the visual field. This process is then repeated in the higher levels of the processing hierarchy, where features are even more complex and appear over even larger swaths.

Finally, in IT, individual neurons are associated with particular objects and categories of objects. Some of these neurons are activated whenever, say, a specific dog appears anywhere in the entire field of view. Others are activated whenever any kind of dog appears.



information *initially* flows mainly from the back of the brain to the front—that is, in a feed-forward fashion—rather than laterally or backwards (through "feedback" pathways). The slower processes related to lateral and feedback neural connections kick in *after* the feed-forward processes do, and those connections are not represented in PetANNet. So it's not surprising that when a scene is presented to a human for up to about 50 milliseconds, the human brain identifies objects with about the same accuracy as the program does—70 to 90 percent. But when presented with a scene for longer times, humans become nearly perfect—accurate to at least 99.999 percent. So the question is, how can feed-forward programs be improved?

Grow the Program?

Simply making the program much bigger could help. The feed-forward architecture has roots in the 1950s, when MIT's Marvin Minsky first simulated cortical function by hooking together simulated neurons to form neural nets.

In those days, the limited speed and memory of computers could handle only a small number of neurons and neural connections. Consequently, the neural nets were applied only to very simple problems. The performance of these neural nets was not good, or the problems they solved were trivial. But the proponents of neural nets have claimed ever since that scaling-up the size of the system by adding more neurons to include more feature detectors and more connections would help the simulations learn more about the world and thereby improve their performance to the point that it might eventually rival that of biological cortical material.

"With Roadrunner, we can actually test this hypothesis for the first time," says Bettencourt.

Another member of the Synthetic Visual Cognition Team, Steven Brumby, ran PetANNet on a standard workstation and found that it took about 38 seconds to process a black-and-white image of 320 × 240 pixels. If the model's parameters were scaled up to human values—for example by increasing the number of feature detectors (see illustration at left)—it would take about a day to process a color image of a million pixels, which means that such a simulation could process only 300 or 400 scenes per year! And even if scaling-up significantly improved object-identification accuracy, the software would be much too slow to be useful.

Roadrunner, however, is fast enough to simulate the operation of the entire visual cortex in real time. There are about 10 billion neurons in the human visual cortex, and each neuron is connected to about 10,000 others. Each neuron also fires about 10 times per second, which for a computer means about 10 "floating-point operations" per second (called "flops"). Multiplied together, these numbers give a quadrillion flops per second, or one "petaflop" per second. The speed record the team set with Roadrunner last summer was 1.14 petaflop per second.

So, Roadrunner has what it takes to prove whether scaling-up a feed-forward neural net will improve the software's accuracy to human levels. If scaling-up is the answer, Roadrunner will also be able to identify objects as quickly as humans do.

However, the main research challenge in simulating a system as complex as the visual cortex is teaching the simulation about the visual world. Scaling up means that the representations of the visual world, especially in the upper layers of the visual cortex, can be more numerous and more precise. However, these representations are constructed only when the simulation actually observes the visual world. So, to fully realize the potential for creating more representations that have greater precision, the simulation must also be exposed to the visual world as widely as possible. Thus, the "training set" of visual images used to develop those representations must be as large and diverse as possible.

We also note that humans take months to start seeing well and years to understand what they see. Roadrunner will be able to test new ideas of how the human brain learns about the visual world and how it organizes itself, by making neural connections, to recognize features and to abstract meaning from what it sees.

Plans B

However, team members doubt that scaling-up alone will do the trick. So they are developing other schemes in parallel with the scaling-up approach.

One tack is based on the facts that lateral and feedback neural connections kick in after the feed-forward processes do and that humans identify objects more accurately when scenes are presented to them for at least 50 milliseconds. If the second fact follows from the first, including lateral and feedback connections could improve the model.

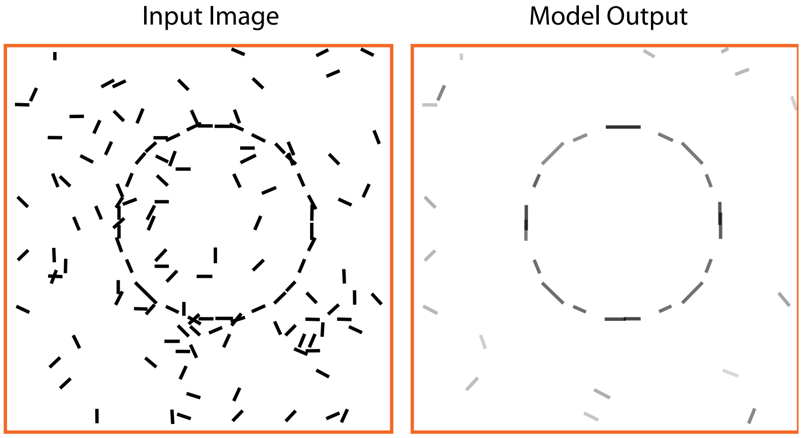
In fact, last summer's Roadrunner speed record was set by including lateral connections in a team-developed program called "PetaVision."

PetaVision simulated only area V1—not the entire

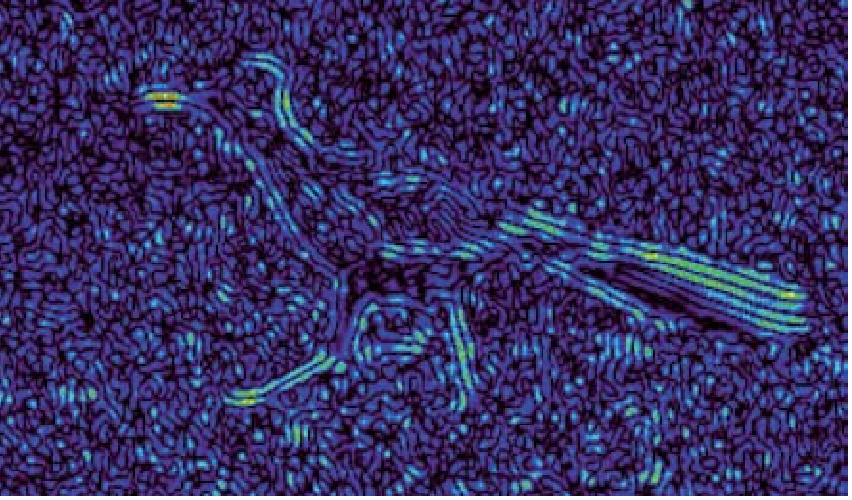
COURTESY JUDY HEDDING, ABOUT.COM:PHOENIX

"what" pathway. Nor did PetaVision include feedback connections. So, PetaVision couldn't test whether lateral and feedback connections can together or separately improve the performance of the entire visual-cortex model. However, PetaVision did show that lateral connections can be important in processing visual information, and the code paved the way for testing the effects of lateral connections in models that include all four major areas of the visual cortex. It also tested some other promising approaches.

PetaVision's neurons were edge detectors whose lateral connections to other neurons were "weighted" to detect smooth curves in the simulated visual field; the weighting was derived from the results of experiments. First, as in PetANNet, each small square of pixels in the visual field was analyzed by a stack of feature detectors—in this case, only edge detectors (see illustration on previous page). However, unlike in the feed-forward model, a PetaVision neuron that detected its targeted edge orientation sent lateral signals to other neurons. If a nearby neuron detected an edge that made a small angle with the edge detected by the first neuron, the weighting caused both neurons to send more signals to each other, generating a local feeding frenzy of neural activity. In this way, the neurons corresponding to segments of a smooth curve became highly active, while neurons corresponding to squares that were blank or contained edges with comparatively large angles were suppressed and became listless.



The "Input Image" at right was input to the PetaVision program, which was run on Roadrunner last summer. Using only the simple-cell and complex-cell neurons described on pp. 4 and 5, with lateral connections between the simple-cell neurons that were weighted to emphasize smooth curves, PetaVision found the circle embedded in the input image, as shown in the "Model Output." Such "segmentation" tasks are key to determining the positions of objects in the visual field.



Left: A photo of a roadrunner, the New Mexico state bird, was used as input to the PetANNet program, running on Los Alamos' Roadrunner supercomputer. Right: An interpretation of what PetANNet's complex-cell neurons "saw" in response to the input photo.

This weighting of lateral connections allowed PetaVision to do something PetANNet could never do in its present state: find the border of a circle (see figure on the bottom of the facing page). This may not seem particularly earthshattering, but finding the borders of an object in one's visual field—which is called "segmentation"—is essential to identifying the object's location. Simulations of this sort could pave the way for exploring the more poorly understood "where" pathway of the visual cortex.

Bettencourt also points out that PetaVision found the circle using a *prescribed* weighting. However, the team plans to modify the software so the neurons learn how to find smooth contours *on their own*, just as we do.

Finally, in contrast to the very-simple neuron model in PetANNet, the electrical signals sent between PetaVision's neurons—and the neurons' responses to those signals—were modeled in biological detail.

Biological neurons talk to each other by sending out impulses, "spikes," of voltage. Each spike lasts about a millisecond. PetaVision modeled each spike's amplitude and duration, along with the spike's precise placement in time.

Precise spike timing is known to be used by the cortical tissue that processes auditory information, for example, in bats, "who are geniuses of sound," Bettencourt says. Some of the neurons in the auditory cortical tissue of bats locate sounds by measuring the difference between the placements in time of two spikes to a precision as small as 10 percent of a spike's duration. (The distance from a source of sound is usually slightly different for each ear, so the associated neural signals are slightly displaced from each other in time.) Moreover, Kenyon has studied spike timing in cat retinas, where its importance has also been shown. PetaVision's accurate spike-timing model could help the Synthetic Visual Cognition Team see if spike timing could be important in other cortical activities as well.

Playing Off Each Other

Both Roadrunner and the brain quickly and efficiently process huge amounts of information. There are striking similarities—and differences—in how they do so.

For example, each of Roadrunner's microprocessors performs about one billion operations per second, whereas a neuron performs about a thousand operations per second. However, Roadrunner—even though it is a "green" supercomputer—consumes about 2.341 megawatts of power, enough to run two thousand homes. (The imposing stack of Roadrunner's giant cooling towers, which dissipate the huge amounts of heat generated by the supercomputer's thousands of superfast chips, is a distinctive feature near the Los Alamos building that Roadrunner calls home.) However, because the neurons in the brain operate much more slowly than do a supercomputer's microprocessors and because the brain is far more parallel than a supercomputer is, the brain uses only 20 to 30 watts!

As research programs such as the Synthetic Visual Cognition Project help us learn how cortical circuits work, we may one day be able to build hardware that can do what the brain does with much less power than existing supercomputers need—or that can operate much faster than existing brains do! Meanwhile, PetANNet and PetaVision are proof that the computational limitations of supercomputers are no longer major obstacles to studying the brain as an integrated system.❖

—Brian Fishbine

Garrett Kenyon (left) and Luis Bettencourt (right) lead the two major projects at Los Alamos that deal with simulating brain function. Kenyon heads a project funded by the National Science Foundation to develop a "neural workbench" that can be used to simulate various kinds of cortical tissue. Bettencourt heads the Synthetic Visual Cognition Team, which is funded by the Laboratory Directed Research and Development Office and which focuses on simulating the cortical tissue that processes visual information.