

Supercomputer Math Speeds Up

■ *UNM scientist helps develop a program that makes it possible to solve complex problems faster*

By JOHN FLECK
Journal Staff Writer

David Bader hunts shortcuts.

The University of New Mexico computer scientist has found his share, but none quite as striking as the path he and his colleagues found to the roots of the bluebell family tree.

Supercomputers get faster every year, but there are always scientific problems beyond their reach, that take too long for even the best machines.

The bluebell problem was a doozy.

Bob Jansen at the University of Texas, Austin, was trying to use genetic data to figure out the common flower's

family tree.

"We're interested in understanding the evolutionary history of this flower," Jansen said in a recent interview.

Jansen and his colleagues took genetic data on 13 species of bluebells and cranked up the state-of-the-art in computer codes for comparing genes to find common ancestors.

They let it run. And run. And run.

After two days, they stopped the program to see how far it had gotten. The result — not very far at all.

At the rate it was going, they figured, they'd have their answer in about 200 years.

So Jansen began working with Bader and Bernard Moret, another UNM computer scientist, to find a shortcut.

Using a couple of tricks, the scientists were able to solve the problem in less than two hours.

"How would you like your computational tool run one million times

faster?" Bader asked.

Supercomputers, the giant machines used to solve large scientific problems, are getting faster.

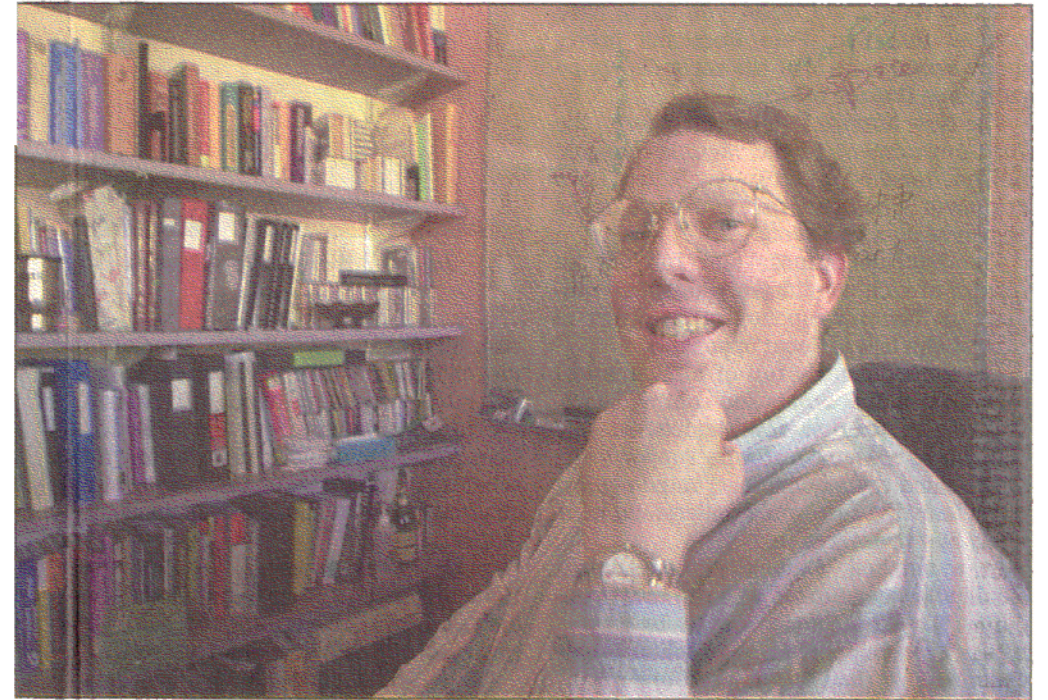
Today's speediest computers are 10 times faster than those just five years ago.

But speeding up by a factor of 10, 100 or even 1,000 won't come close to solving many of the problems facing researchers today, from parsing the human genome to simulating global climate.

That's where computer scientists like Bader come in.

They develop "algorithms," mathematical recipes to solve problems.

A clever algorithm can go a long way toward solving a problem that would be otherwise too time-consuming to attempt. It's common across many fields, from genetics to the study of



AARON WILSON/JOURNAL

MAKING IT SIMPLER: David Bader, a University of New Mexico computer scientist, has developed a method of simplifying complex problems on supercomputers.

See UNM on PAGE B5

UNM Scientist Speeds Up Math on Supercomputers

from PAGE B1

galaxies.

In 15 years, Los Alamos National Laboratory astrophysicist Mike Warren has seen the computers he uses increase 1,000 times in speed.

Better algorithms have brought another 1,000-fold increase in speed, he said, allowing him to solve problems today 1 million times faster.

From city to city

Mathematicians have a classic example of the kind of problem Bader and Jansen faced — the deceptively simple but maddeningly time-consuming Traveling Salesman Problem.

Give your Traveling Salesman a list of cities and ask him to find the shortest route to visit each city once.

It is easy for a handful of cities.

For three cities, the total number of possible routes is just six (3 times 2 times 1). Calculating the length of each to find the shortest is trivial.

For four cities, that goes up to 24 (4x3x2x1).

"That's not hard," Bader said.

But as you add cities, the number of possible routes to check goes up quickly. How quickly?

Bader sat at his computer recently doing some multiplication, and came up with more than 2 billion billion possible routes among just 20 cities — the number 2 followed by 18 zeroes.

"That's just enormous," Bader said.

Put another way, a computer that solved the four-city problem in 1 microsecond — typical for a fast PC today — would take more than 1,500 years to solve the 20-city problem by testing every possible route, according to Bader.

But real-world science and engineering, from designing telephone networks to analyzing the human genome, are full of problems that are the functional equivalent of the Traveling Salesman Problem with far more cities than 20.

The bluebell problem

The number of genes to be compared in the bluebell problem, for example, is the equivalent of a 210-city Traveling Salesman Problem.

So scientists have to find a shortcut.

That is the sort of problem that Bader, Jansen and their colleagues faced with the genes of the bluebell.

There are 2,000 species of the prolific flowers, growing on every continent but Antarctica.

Because the species are related, their genes are similar. The scien-

tists' job was to try to look at those similarities, and differences, and figure out what the bluebells' family tree looked like.

Comparing each of the parts of the more than 100 genes in a part of the plant called the chloroplast with each of the genes in the other bluebells is very much like the Traveling Salesman Problem. It quickly grows far too large to solve in any reasonable amount of time, Jansen said.

But Bader, working with UNM colleagues Moret and Mi Yan and University of Texas, Austin, researchers Stacia Wyman and Tandy Warnow, came up with an approach that allowed them to drastically trim the number of possibilities they needed to look at.

They also found a way to carve up the problem into smaller pieces that could run simultaneously on multi-chip supercomputers.

The resulting computer program ran on a UNM supercomputer in just one hour and 40 minutes.

Though much work remains to be done, the family tree generated by the program gives the biologists some idea of the possible evolutionary relationships among the flowers, in a reasonable amount of time.

"This is a major accomplishment in terms of speedup," Jansen said.

But the scientists are not resting on their laurels.

From yeast to flies to bluebells to the human genome, the number of organisms whose genes have been mapped is growing rapidly, leaving scientists wading in a deep morass of data and struggling to find ways to use computers to sort through it.

The task of comparing genomes between species makes the bluebell work look like child's play.

"How do we compare genomes between humans and flies?" asked Jansen.

"There is an explosion in biology of complete genomic sequences," Jansen said. "The problem's going to get worse, not better."

But the problems are important enough that Bader recently received a \$385,000 National Science Foundation grant to develop algorithms in genetics, landscape ecology and fluid dynamics.

For scientists like Jansen with intractable problems to solve, the work of Bader and others who develop better algorithms is invaluable.

"It's people like David Bader who have the expertise to develop innovative ways to solve problems," Jansen said.