

**TECHNOLOGY**

# *Writing the Fastest Code, by Hand, for Fun: A Human Computer Keeps Speeding Up Chips*

**By John Markoff**

Nov. 28, 2005

SEATTLE - There was a time long ago when the word "computer" was a job description referring to the humans who performed the tedious mathematical calculations for huge military and engineering projects.

It is in the same sense that Kazushige Goto's business card says simply "high performance computing."

Mr. Goto, who is 37, might even be called the John Henry of the information age.

But instead of competing against a steam drill, Mr. Goto, a research associate at the Texas Advanced Computing Center at the University of Texas at Austin, has bested the work of a powerful automated system and entire teams of software developers in producing programs that run the world's fastest supercomputers.

He has done it alone at his keyboard the old-fashioned way -- by writing code that reorders, one at a time, the instructions given to microprocessor chips.

At one point recently, Mr. Goto's software -- collections of programs called subroutines -- dominated the rarefied machines competing for the title of the world's fastest supercomputer. In 2003 his handmade code was used by 7 of the 10 fastest supercomputers. (The Japanese Earth Simulator, which was then the world's fastest machine, however, did not use his software.)

In the most recent ranking of supercomputers, I.B.M. machines overtook a number of supercomputers using Mr. Goto's software to capture the top three spots in the fastest computer rankings. Still, the Goto Basic Linear Algebra Subroutines, or BLAS, as his programs are known, were used by 4 of the world's 11 fastest computers.

Mr. Goto has become a legend in the supercomputing community because of his solitary crusade. And he shows no signs of flagging in the contest to wring every ounce of computing speed from the world's fastest microprocessor chips.

But for all the acclaim he has received, Mr. Goto is a relative newcomer to the supercomputing field, having made his breakthrough about a decade ago.

"At first I didn't know anything," he said in an interview at the annual supercomputing conference held in Seattle in mid-November. "This was all trial and error, but now I have experience."

The value of his work goes far beyond setting speed records. Because his programs can more efficiently solve complex linear equations, they can offer better solutions to virtually every computational science and engineering problem. For example, the subroutines are used in simulation programs to model the flow of air over the surface of a plane or a car more precisely.

One of Mr. Goto's principal rivals is a software project known as Atlas, created by a group of researchers working with Jack Dongarra, a computer scientist at the University of Tennessee. Atlas is an automated effort to find the most efficient way to solve linear algebra functions for specific microprocessors -- a task that Mr. Goto does meticulously by hand.

Like chess-playing software, the Atlas project tries to overcome the shortcomings of different kinds of computer designs by systematically testing thousands of solutions for each chip to find the most efficient one for each type of microprocessor.

By contrast, Mr. Goto uses only a program called a software debugger that allows him to track how data moves among different components of a microprocessor.

He then reorganizes the individual software instructions so that his subroutines perform crucial algebraic functions more quickly to gain small amounts of processing speed from a specific type of computer chip.

Typically these are highly repetitive operations that can consume vast amounts of computing capacity. For example, one challenging type of calculation requires the microprocessor to multiply numbers from two tables stored in memory together.

Mr. Dongarra acknowledges that Mr. Goto's hand-tuned programs are more efficient and can still outperform Atlas.

"I tell them that if they want the fastest they should still turn to Mr. Goto," said Mr. Dongarra, who is one of the researchers who maintains the Top 500 listing of the world's fastest-performing computers from a computing speed race held twice a year.

Mr. Goto came to his passion for supercomputing almost by accident. Educated in power engineering at Waseda University in Tokyo, he worked as an employee of the Japanese Patent Office, doing research on early inventions like video recorders.

To help in his work, Mr. Goto purchased a Digital Equipment workstation based on the Alpha microprocessor in 1994 to perform a simulation.

But when it arrived he could not understand why it was performing so slowly. So he explored the Alpha's design to see where the performance bottlenecks were.

He later purchased a second Alpha-based computer and by rewriting the crucial subroutines was able to improve its performance to 78 percent of its theoretical peak calculating speed, up from 44 percent.

Although he was not formally trained in computer or software design, he perfected his craft by learning from programmers on an Internet mailing list focusing on the Linux operating system for the Alpha chip. His curiosity quickly became a passion that he pursued in his free time and during his twice daily two-hour train commute between his job in Tokyo and his home in Kanagawa Prefecture.

"I would frequently work on these problems until midnight," he said. "I did it to relax."

As a teenager, Mr. Goto developed a passion for electronic design, building his own stereo equipment from the most basic components.

His current interest, he says, is not in the pure mathematics of the linear equations, but rather in finding clever ways to overcome the shortcomings of the architecture and internal organization of microprocessors that are used in every kind of computer, from hand-held devices to supercomputers.

Modern computers are organized to offer the programmer a hierarchical series of data storage areas that range from the computer's disk drive DRAM memory, as well as relatively small temporary memory areas called caches. Typically, the fastest memories are also the smallest.

One of the simplest ways to speed a program is to keep the calculation in the memory unit, which is closest to the microprocessor's calculating engine.

Every time the calculation engine is required to stop what it is doing to get new data from a more distant memory area, processing speed slows. But in some cases, keeping data in the closest memory cache may not be as efficient as keeping it in a larger cache that is farther away.

Robert A. van de Geijin, a computer scientist who works with Mr. Goto at the Texas Center, said that Mr. Goto's special skill was in the step-by-step reordering of software instructions to take the greatest advantage of the performance trade-offs offered by each type of chip.

"He combines both scientific insight and engineering skills," Mr. van de Geijin said.

They met in 2002 when Mr. Goto took a sabbatical from his job at the patent office to spend a year at the Texas center. (He has since resigned from the patent office.)

Once Mr. Goto arrived in Texas, he turned his attention to optimizing the speed of the Pentium 4 microprocessor. When computer scientists at the University at Buffalo added Goto BLAS to their Pentium-based supercomputer, the calculating power of the system jumped from 1.5 trillion to 2 trillion mathematical operations per second out of a theoretical limit of 3 trillion.

The increase was so astounding that the record keepers for supercomputing Top 500 called the researchers in Buffalo because they did not think such a speed was credible.

"I teased them and suggested that the speed of light was faster in Buffalo than it was in Tennessee," Mr. van de Geijin recalled.

Recently there has been a quiet controversy around the Goto BLAS because Mr. Goto has been slow to offer his work as open-source software, the free model of software distribution.

Some programmers have suggested that Mr. Goto has not joined the open-source movement because he wants to protect his secrets and strategies from competitors.

That is not so, he said recently, noting that the Goto BLAS software is freely available for noncommercial use. And he said he was preparing an open-source version.

He said his next big challenge was to expose chip designers to his ideas to help speed their processors.

"Computer architects are stubborn," he observed. "They have their own ideas." His ideas on computing efficiency, he said, speak for themselves.

---

A version of this article appears in print on , Section C, Page 1 of the National edition with the headline: TECHNOLOGY; Writing the Fastest Code, by Hand, for Fun: A Human Computer Keeps Speeding Up Chips