# Buried in bytes

by Patti Koning

illustration by Chris Brigman

Once upon a time, digital information was scarce. No more. According to IBM we create 2.5 quintillion bytes of data each day. Ninety percent of the digital data in the world right now was created in the past two years.

Sandia applied mathematician Tammy Kolda has a way to shovel through all that information — tensors.

"You can think of a tensor as a hypercube of data," she says. "A tensor lets you characterize data in higher dimensions. Most data analysis is still done in two dimensions, even though the data itself is multidimensional."

Data analysis is often done using matrices, or two-dimensional mathematical arrays. Imagine a spreadsheet with rows and columns. Now consider a medical researcher making multiple observations such as height, weight, blood pressure and glucose level on multiple patients. A multidimensional array, or tensor, allows richer analysis. Add one more dimension, like time, and you've got multi-way data suited for tensor analysis.

Sandia has become a leader in tensor analysis during the past decade. "A problem with tensors is their sheer size, in terms of the computing power needed to work with them," Kolda says. "You aren't going to store a dense 10,000 by 10,000 by 10,000 tensor on a standard computer."

While the blocks of data are huge, there isn't always a lot of data inside. For example, in analyzing transactional data like email traffic, many entries will be zero. "If you look at a group of people, only a few pairs are emailing each other in any given time period," Kolda says. "We don't have to store those zero entries. Instead, we store a sparse tensor that is smaller and easier to analyze."

*As big data gets bigger, a mathematician looks at digging out with multi-dimensional storage cubes. Her work could revolutionize computing as we know it.*

Sandia is the primary source of sparse-tensor software, developed with two Laboratory Directed Research and Development projects and an investment from the U.S. Department of Energy Office of Science that recently was renewed for another three years.

"Tensors are applicable almost everywhere," Kolda says. "If you have big data, tensor decompositions can yield deep insights." Tensor decomposition is a mathematical procedure that converts multidimensional data into sets of correlated observations and is frequently used in exploratory data analysis.

Breakthroughs in medical technology are a major source of big data. Kolda is working with several universities on diffusion tensor imaging (DTI), an advanced magnetic-resonance imaging technique that maps the diffusion process of molecules in biological tissues and distinguishes gray matter from white matter. It is particularly important to neurosurgeons who need to avoid white matter.

"The analysis requires a tensor factorization for every single voxel [the 3-D equivalent of a pixel] of data, so we're talking millions of calculations. You quickly hit a wall with the sheer size of the data," Kolda says. Her group is developing symmetric tensor algorithms that quickly interpret the data. The ultimate goal is real-time imaging during surgeries.

## No ordinary toolbox

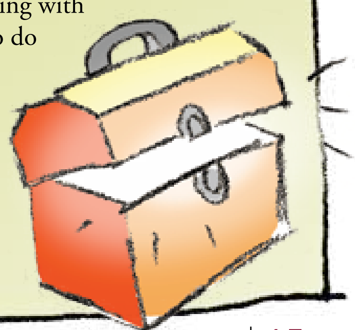
Tammy Kolda and other pioneers in the field of sparse tensors had to invent their own tools, most notably the Tensor Toolbox for MATLAB.

"About eight years ago my group was working on a project to apply tensor analysis to graph data, but the data was too big to store as a tensor," she says. "We made our own data structure to solve this problem, which is now being used by thousands of people worldwide."

MATLAB is a high-level language and interactive environment that lets users perform computationally intensive tasks faster than with traditional programming languages. The Tensor Toolbox is a collection of six object-oriented classes for efficient numerical calculations with large-scale multidimensional arrays. Users can quickly devise prototype advanced algorithms.

"The Tensor Toolbox makes working with tensors as easy as MATLAB makes working with matrices," Kolda says. "The user doesn't have to worry about the low-level details to do complex, high-level operations."

The Tensor Toolbox is most commonly used for data mining, but there are other applications that have surprised Kolda, such as watermarking, music genre classification, air traffic control, continuum mechanics and quantum entanglement. And students are even downloading the Tensor Toolbox to complete homework assignments.

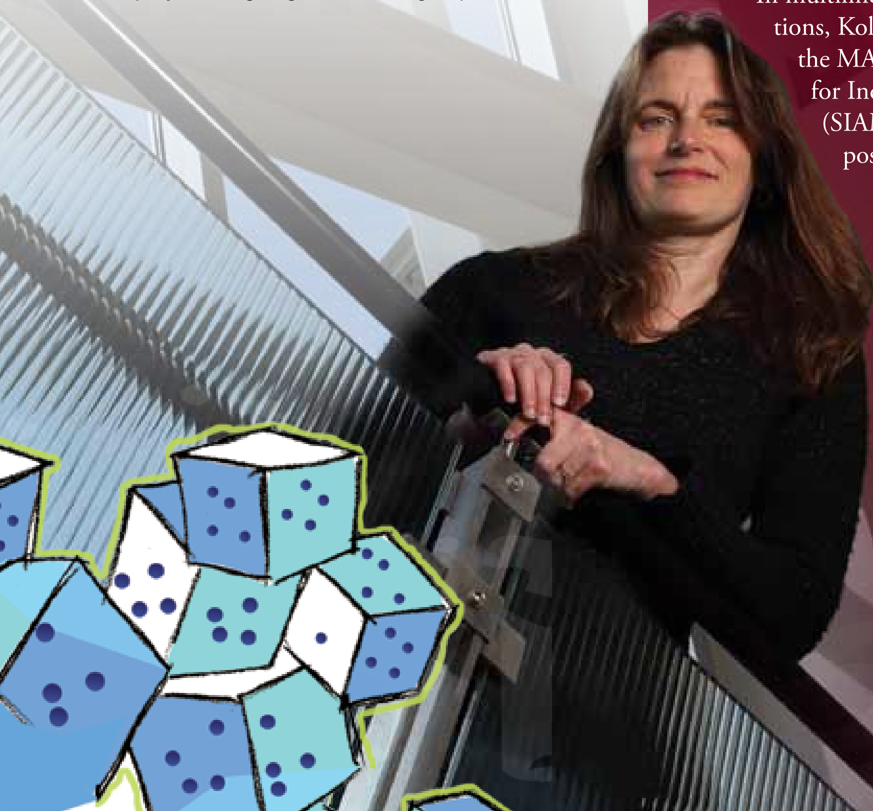


Tensors also may be key to reducing communications in standard computational kernels, the main component of a computer's operating system. Gray Ballard, who worked with Kolda for two summers as a graduate student at the University of California, Berkeley, will return to Sandia this fall as a Truman Fellow to tackle the problem.

Ballard worked on tensors, but he and Kolda also spent time talking about his research in communication-avoiding algorithms. "Moving data on a computer has a big cost in terms of both time and energy. Finding ways to avoid moving data reduces the expense and time for calculations," Kolda says.

To see how this relates to tensors, you have to go back to 1969 when Volker Strassen developed a new method for matrix multiplication that reduces both computation and communication costs. "Tensor factorizations hold the key to improving on Strassen's method by expressing matrix multiplication as a tensor and then decomposing it," Kolda says.

"It's kind of crazy that this 50-year-old math concept can be expressed as a tensor decomposition. It's also really exciting, because we can apply our knowledge of tensor decomposition to a problem that could revolutionize computing as we know it. Even the incremental gains we make in this project are going to have a big impact." ■



MEET

# Tammy Kolda

**Kolda teaches a weekly flow yoga class at Sandia. She began practicing yoga 12 years ago as a new exercise challenge. "To me yoga is much more interesting and challenging than other exercise classes. There is a lot of emphasis on form and breath. Yoga is endlessly deep. Whenever you think you have mastered something, there is always more to learn," she says.**

## STATS

- Bachelor of science in mathematics from University of Maryland, Baltimore County
- Master of science and Ph.D. in applied mathematics from University of Maryland, College Park
- Householder Postdoctoral Fellow in Scientific Computing, Oak Ridge National Laboratory
- Joined Sandia in 1999 and works in the Informatics and Systems Assessments group.
- Kolda's current research interests include network modeling and analysis, multilinear algebra and tensor decompositions and compressed sensing.
- In multilinear algebra and tensor decompositions, Kolda is best known for her work on the MATLAB Tensor Toolbox and a Society for Industrial and Applied Mathematics (SIAM) Review article on tensor decompositions and applications.
- Kolda is a Distinguished Member of the Association for Computing Machinery, an elected member of the SIAM Board of Trustees, section editor for the Software and High Performance Computing section of the SIAM Journal on Scientific Computing and an associate editor for the SIAM Journal on Matrix Analysis.