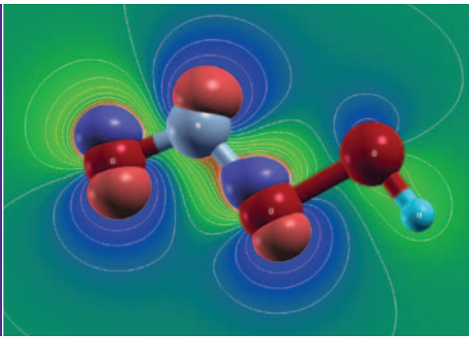
CASE STUDY

# Accelerate Time to Solution with Star-P™ and SGI® Altix®



For scientists and engineers, data is a problem that just keeps growing. Today's models and algorithms are increasingly complicated, and digital simulation has all but replaced physical testing of prototypes. As a result, data sets have swelled in size, and technical computing trends point to an immediate future in which data sets of tens or hundreds of Gigabytes will be commonplace, with some pushing beyond the Terabyte mark.

Users relying on desktop systems typically lack the ability to solve complex problems in their entirety, even when running powerful interactive scientific and engineering tools like MATLAB® from The MathWorks. Some users work around the limitations of their desktop systems by solving smaller pieces of an overall problem – an approach that takes longer and often produces inferior results.

Many others turn to parallel high-performance computing (HPC) resources to process their calculations. But interactive applications like MATLAB don't run on parallel HPC systems, so users are forced to reprogram their algorithms for parallelization, usually reprogramming their code in C, Fortran or MPI. That extra step can stop research in its tracks, consuming up to half the entire time to solution and significantly extend the project lifecycle.

### Automated parallelization transforms workflow

By leveraging the Star-P™ interactive parallel computing platform and SGI® Altix® servers, MATLAB users now can bridge their desktop systems with powerful parallel computing resources without manually reprogramming algorithms
prototyped on the desktop. Star-P automates the process of parallelizing algorithms developed in MATLAB, adding the computational power of scalable, award-winning SGI Altix servers to the desktop interactivity that MATLAB users rely on.

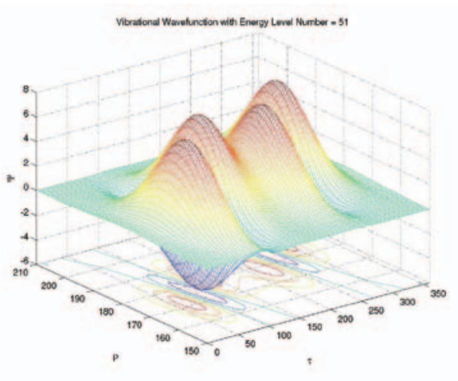
With a few simple modifications to their code, MATLAB users can quickly and easily harness the proven HPC capabilities of SGI Altix systems to compute even the largest models with the most complicated algorithms.

Built on industry-standard components and a 64-bit Linux operating environment, the flexible SGI Altix platform is available in affordable, entry-level configurations of 2 to 16 processors and up to 128GB of shared memory, with larger shared memory and cluster servers available as user needs evolve. SGI Altix servers are powered by industry-leading Intel Itanium 2 processors, and support both fine grained (global) and coarse-grained (distributed) parallelism.

SGI Altix systems are popular in performance-hungry environments, including:

- **Aerospace:** Space vehicle design optimization, airframe optimization
- **Defense:** Control systems, synthetic aperture radar, signal processing
- **Earth sciences:** Air and water current analysis, weather forecasting, earthquakes
- **Energy:** Oil exploration, reservoir modeling
- **Finance:** Risk modeling, neural networks, energy trends
- **Intelligence:** Cryptography, image processing, face recognition
- **Life sciences:** Drug discovery, image processing, molecular simulation, neuroscience, computational biology
- **Manufacturing:** Control systems, post-processing of CAE analyses, FPGA design

The growing need to merge desktop interactivity with HPC compute power has prompted MATLAB users in a range of markets to explore how Star-P and SGI Altix servers can transform their research workflow. This document spotlights four such customers.




**Molecular simulation and quantum physics: calculating thermodynamic properties**

One of the world's most respected universities breaks through a workflow bottleneck to determine the thermodynamic properties of a molecule that plays a key role in the creation and behavior of smog.

**Challenge:** A renowned US-based university has been working to identify the thermodynamic properties of HOONO, a molecule that is crucial to understanding the dynamics of smog. University researchers use Gaussian to calculate the molecule's energy in a variety of configurations and then use MATLAB to determine the vibrational wave function for the configurations based on quantum mechanics principles. But to see how the molecule would interact within a larger context such as a weather system, scientists must first obtain an accurate model of HOONO's thermodynamic properties. To achieve this, they must employ larger matrices in computing the molecule's properties – larger than their desktop systems' practical limit of 3,000 by 3,000 elements.

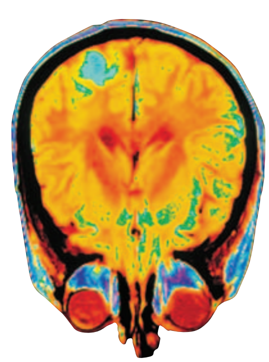
The university's SGI Altix system would be an ideal platform for efficiently performing the calculations, but manually preprogramming the model for a parallel system would take months, putting a stop to ongoing research.

If researchers were to rely on a desktop workaround, they wouldn't fare much better. Attempting to roughly model molecule properties on the desktop could produce significant errors that would be substantially amplified when the model is employed in a weather system or pollution simulation. And with so many possible permutations of molecules, experimentally measuring molecular properties can rapidly prove too costly, time-consuming and complex to be feasible.

**Solution:** Using Star-P, researchers made about 20 minor modifications to the original MATLAB algorithm, identifying the variables and calls that would benefit most from parallelization. Star-P uses simple extensions to the MATLAB language to address large memories and multiple processors, while preserving the visual interface, high-level language and interactivity of MATLAB.

In a matter of seconds, researchers were able to send the large-matrix problems to the university's 16-processor SGI Altix system, which is configured with 80GB of memory. These memory-bound problems particularly benefit from SGI's third-generation NUMAflex™ architecture, a unique global shared-memory architecture that leverages Star-P's support for fine-grained parallelization. NUMAflex enables the server to hold entire data sets in memory, allowing for faster and more interactive data analysis, and resulting in more incisive conclusions.

**Results:** In a matter of minutes, researchers were able to parallelize their algorithm – a process that would have taken months by reprogramming the code in C. Now they can calculate matrices in excess of 10,000 by 10,000 elements, many times the practical limit of the university's desktop systems. Such calculations offer powerful new insights into the role that the HOONO molecule plays not only in smog behaviors, but also in the context of specific weather systems and the environment at large.



**Life sciences: brain scan analysis**

An internationally recognized research hospital overcomes the challenge of interactively analyzing brain scans.

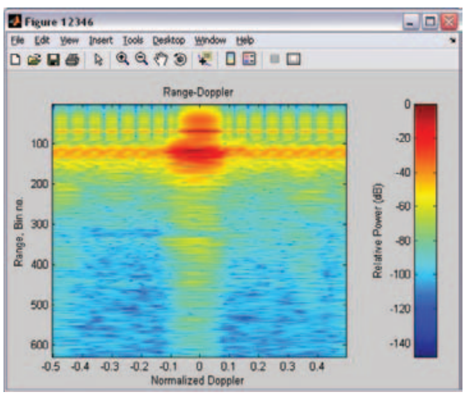
**Challenge:** The task faced by researchers at one of the world's leading research and clinical treatment hospitals is daunting: They seek to pinpoint how new cancer drug treatments impact blood flow in the brain, but the high-resolution Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) techniques that are so helpful in their studies generate so much data that the hospital's desktop systems are now useful only for simple prototyping and testing of analysis algorithms.

CT and MRI imagery is simply growing too large for the desktop systems to accommodate the rapid evaluation of multiple algorithms that researchers require to narrow down the candidates in their search for the optimal image analysis routine. Even with earlier generations of brain scan imagery, analysis of imagery measuring 300 by 300 pixels could take an hour to process. But today's 2D scans are often 5 to 10 times larger than earlier CT and MRI imagery, and increased 3D resolution can boost total data set size by up to 100-fold. If researchers were to rely solely on desktop systems, their time to solution would be unacceptable. Yet because project leaders lack parallel programming expertise, recoding their algorithms in C, Fortran or MPI, as time-consuming as that would be, is not a practical option.

**Solution:** Accurately identifying the most effective algorithm requires attacking the problem in a dynamic, iterative fashion and using real-world input data. With Star-P, the hospital's researchers efficiently parallelized their algorithm on a four-processor SGI Altix 350 mid-range server equipped with 8GB of memory. As research progresses, they can transparently off-load large-image analysis to the SGI Altix system, while avoiding the kind of half-day or overnight batch processing sessions that send workflow grinding to a halt.

The hospital's SGI Altix 350 server provides leading price/performance and scalability for departmental and mid-range computing deployments. Enabling productivity breakthroughs that traditional Linux clusters or repurposed UNIX servers can't achieve, SGI Altix 350 systems accommodate as many as 32 Intel Itanium 2 processors and 384GB of memory in a single system. So as CT and MRI resolutions continue to increase, the SGI Altix 350 server provides hospital research staff with the compute headroom to meet that challenge.

**Results:** With Star-P and their SGI Altix server, hospital researchers are able to execute their algorithms more than 300 percent faster than ever before. This gives researchers more time to test and validate more algorithms, with an eye toward helping drug developers deliver safe, effective treatments to cancer patients.




### Defense: Radar analysis and system design

An advanced defense systems laboratory meets the data explosion head-on by leveraging HPC resources to achieve both fine- and coarse-grained parallelization in exploring multiple radar analysis algorithms.

**Challenge:** For years, most radar systems have gathered information primarily from land-based facilities and reconnaissance aircraft. But recently, defense organizations have expanded their reliance on satellite-based radar systems. Orbiting high above the earth's surface, these radar imaging satellites use microwave signals to peer through cloud cover and capture photographic-quality images that help military analysts conduct surveillance, communications, navigation, and missile warning activities.

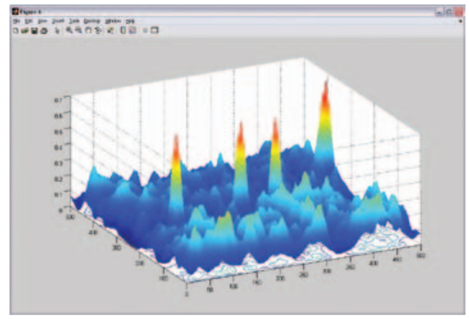
But they also present problems for defense system researchers working to develop new radar analysis systems. Traditional radar images might measure 10MB in size. But today, it isn't unusual for the research facility to receive Terabytes of satellite radar data in a single day, dramatically complicating the task of analysis.

Adept at evaluating radar analysis algorithms using MATLAB, the researchers sought a way to preserve the familiarity and interactivity of their desktop environment while bringing the power of parallel computing to analyze increasingly larger data sets. Due to the time-sensitive nature of their work, however, they couldn't afford the time it would take to reprogram their algorithms in Fortran.

**Solution:** By deploying Star-P, the researchers were able to leverage their eight-processor SGI Altix 3000 server equipped with 20GB of memory to meet this new computational challenge. With Star-P, standard MATLAB commands and functions perform in a parallel manner transparently to the user. By tagging just a handful of variables within their algorithm codes, researchers were able to repurpose existing MATLAB scripts to run their larger problems in parallel and preserve their workflow.

The researchers also benefit from mutual support for both fine- and coarse-grained parallelization. Where some problems are memory-bound and require large-scale memory access and inter-processor communication, others can be segmented to run on a series of independent processors. Ease of use is maintained for both approaches in Star-P, and SGI Altix systems provide a uniquely flexible platform that excels in both fine- and coarse-grained parallel computing, delivering genuine interactivity no matter the task at hand.

**Results:** Researchers now can preserve their mission-critical workflow while working with data sets that are orders of magnitude larger than those they can process on the desktop. With a faster time to solution, they can efficiently parallelize their algorithms to identify the most effective tools for identifying targets from satellite radar data.



### Medical: Blood cell matching

A team of hematologists dramatically accelerate the process of detecting pattern matches among white blood cell samples.

**Challenge:** Working to develop algorithms that may have implications well beyond the scope of their primary research, a team of hematologists is developing a method for finding a specific white blood cell within patient blood samples. While spotting white cells is easy – in healthy patients, a single drop of blood may contain 7,000 to 25,000 white cells – the problem of sorting through blood sample images to find matches of a specific cell structure is far more difficult. The process of finding a match grows even more computationally intensive when more white cell counts increase; in Leukemia patients, for example, 50,000 white cells may exist in one drop of blood.

The hematologists follow a classic pattern-matching methodology to find matches of the target cell from within a virtually endless string of search images. The process is repetitive and iterative: They load the target cell image and search images; separate the image to be searched into hue/saturation/value (HSV) color space variables; apply Fast Fourier Transforms (FFT) algorithms to find matches based on each color variable; and use thresholding on correlation plots to narrow the field of high-probability matches. Because the target cell and search image can vary significantly, researchers must dynamically adjust parameters. This leads to multiple attempts to find the most effective combination for a specific cell type and patient sample.

Conclusions come faster when the team can analyze larger search images, but increasing image sizes quickly renders the team's desktop systems (dual-processor PCs stocked with 2GB of memory) incapable of providing the interactivity required for running multiple detection algorithms. Basic search images, measuring 400 by 400 pixels and amounting to 1.2MB of data, are easily computed on the desktop, but once they grow to 1,600 by 1,600 pixels, the PCs reach their limit. Still, the team wants to scale its analysis to images measuring 12,800 by 12,800 pixels and representing 1.2GB of data, and even larger over time. To achieve this, the team must compute the algorithms on an HPC resource. But for these researchers, whose primary expertise is in the study of blood and blood-forming tissues, manually parallelizing the detection algorithms would be a costly, time-consuming distraction, if not a practical impossibility.

**Solution:** Armed with Star-P software and a 32-processor SGI Altix 350 server with 128GB of memory, the team is able to scale their compute capabilities as their needs grow while still relying on MATLAB features and commands. Fast and easy modification of algorithm code allows the problems to run transparently on the SGI Altix 350 server, whose flexible architecture gives researchers access to 4, 8, 16, or all 32 of the system's processors. The memory-bound image processing routines also take advantage of the SGI Altix system's global shared-memory architecture and Star-P's support for fine-grained parallelization. With more processors and memory availability, researchers can tweak their algorithm settings in real time and immediately view the results. Such interactivity allows them to test a series of algorithms and approaches – an advantage over traditional programming, which requires users to "commit" to a particular algorithm and its settings through the end of the analysis process.

As the team's needs scale, so can their server resources. The SGI Altix family leverages the built-in SGI® NUMAlink™ interconnect fabric, which delivers data up to 200 times faster than conventional interconnects. SGI Altix systems are uniquely capable of independently scaling processors, shared memory and I/O on a single, standard chassis with different expansion modules, providing optimal resource usage for demanding technical applications. SGI Altix systems support up to hundreds of processors when necessary, and researchers worldwide regularly share access to SGI Altix systems that rank among some of the world's most powerful computing resources.

**Results:** Parallelizing the detection algorithms for the SGI Altix system has produced dramatic results. A 3,200 by 3,200 image array – effectively impossible to compute on the desktop – takes only 14 minutes on four of the SGI Altix system's processors. Across eight processors, solution time drops to six minutes. On all 32 processors, the detection algorithms are computed in only 78 seconds. By tapping its HPC resource, the hematology team is able to explore more detection algorithms while reducing its overall time to solution.

Learn how Star-P and SGI Altix systems can transform your workflow and extend your desktop MATLAB environment.

For more information on the Star-P interactive parallel computing platform, visit *www.sgi.com/products/software/starp/*. Or visit Interactive Supercomputing at *www.interactivesupercomputing.com*.

For more information on flexible and scalable SGI Altix servers, visit *www.sgi.com/altix*.

**Flexible parallelization with Star-P and SGI Altix**

Most production-level HPC applications call for both fine-grained, global parallelization and coarse-grained, distributed parallelization. Star-P and SGI Altix servers free MATLAB users from constraints of most solutions by enabling users to work in both formats.

The combined solution's fine-grained parallelization supports algorithms requiring large-scale memory access and inter-processor communication, such as those found in matrix manipulation and signal processing applications. And its coarse-grained mode is ideally suited for parallelization of distributed algorithms, such as computations that can be naturally broken up into largely independent processes. These include such routines as Monte Carlo simulation, or parallelization of FOR loops.

Corporate Office
1500 Crittenden Lane
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5509.1455
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000



3900 [01.24.2006] J15105