Overview of Scalable Parallel Applications for T3D/E Systems

David Caliga, Applications Division, Cray Research

ABSTRACT: *This paper will review the activities of the general community in making parallel application available on T3D/T3E systems. The review will cover 1) Review of T3D/E application efforts, 2) Performance and scalability of key applications, 3) Parallel application success stories, and 4) Issues confronting developers of scalable parallel applications.*

1 Introduction

A great deal of focus has been made on getting applications parallelized? There have been several pressures on industry that have occurred

- Industry is becoming more dependent on simulation to be competitive

  The development process is using more simulation to cut the costs of experimentation. The results of simulation are providing trusted results and are being used to augment experimentation.

- Competitive pressures require getting results at lower costs

  Industry is seeing that to remain competitive they need to reduce their product development costs. This pressure is increasing the value derived from simulations.

- Computer industry is reacting to a new set of customer pressures

  Cost pressures are forcing customers to look for better price performance out of their computing systems. RISC systems are becoming increasely attractive for performance and cost reasons. There is still the requirement to have the greatly reduce the simulation's time-to-solution. This time-to-solution pressure requires major performance improvements and that can only be achieved if applications run efficiently in parallel.

2 Review of Application Efforts

The application activity in the MPP arena has in general been focused on academic and proprietary codes. The number of commerical industrial codes that are available on MPP systems has greatly increased in the last few years. This activity has been seen in the interest for T3D/T3E systems. The size of the customer base for the T3E systems has increased relative to that of the T3D customers. The requirement for a large number of PEs has been evident by the initial set of T3E customers. The average T3E PE count is roughly the same as T3D systems, even though the application compute speed has increased by a factor of 3-5.

The breakdown by industry of the T3D/T3E customers shown below. The T3E customer set is as of September 1996.

Table 1. T3D/T3E Customers by Sector

Industry Sector	No. Customers T3D	T3E
University	14	17
Research	7	7
Government	7	13
Petroleum	4	1
Energy	2	5
Environmental	3	5
Automotive	2	1
Aerospace	2	1
Electronics	1	0
Finance	1	0
Avg # PE/System	180	158

2.1 Third Party Software Vendors

The efforts of software vendors are mainly driven by the requirements of their user base. Those users that are asking or demanding parallel versions of the code are mainly in the chemistry and petroleum industries.

There is a growing number of codes that are becoming available in the marketplace. The following table shows the application by domain area that are available on the T3D and those that are/will be available on the T3E system.

Table 2. T3D/E Application Availability

Domain Area	Application	T3D	T3E
Structures	MPP-DYNA3D	X	1/97
	SYSNOISE		97
	Spectrum		97
CFD	AIRPLANE	X	1/97
	FLO67	X	1/97
	FLUENT/UNS	X	
	RAMPANT	X	1/97
	STAR-CD	X	10/96
Chemistry	GAUSSIAN94	X	1/97
	CHARM25al	X	10/96
	DISCOVER950	X	10/96
	AMBER 4.1	X	1/97
	GAMESS	X	10/96
	GAMESS/UK	X	2/97
	LAMMPS	X	1/97
	SUPERMOLECULE	X	6/97
	Turbomole	X	97
	UNICHEM	X	1/97
Earth Science	FOCUS	X	1/97
	GEOVECTEUR+	X	10/96
	ECLIPSE 100/E300	X	1/97
	FALCON	X	1/97
	VIP	X	1/98
Environmental Science	HIRLAM	X	1/97
	MM5	X	10/97
	NOAA/GCM		1/97
	PCCM	X	1/97
	POP	X	1/97
	MICOM	X	1/97
	FEMWATER	X	2/97

2.2 HPC Centers

Many of the High Performance Centers (HPC) have ported proprietary and commercial applications to MPP platforms. The Edinburgh HPCI Center is an excellent example of a center that has many activities with both academic and industrial groups to get applications onto the T3D system.

The T3D academic activities are centered around several Grand Challenge Consortia. The domain areas covered by these consortia are:

- CFD
- Ocean Modeling
- Atmospheric Modeling
- Macromolecules
- Materials
- Chemical reactions
- Atomic Physics

Edinburgh HPCI is very aggressive in going after industrial activities for the T3D system. The following are examples of their efforts:

- Quadstone, Commercial traffic simulation
- UK Met, Parallelization of their Unified Model weather forecasting
- British Aerospace, Porting of their irregular multigrid Euler CFD code
- Avro International Aerospace, Parallelized their 3D Euler/Laminar/Reynolds-averaged Navier-Stokes flow solver.

The focus on getting industry to understand the advantage of MPP systems is very important. Many people have felt that the marketing build up for the value of MPP systems is over rated. The problems that much of industry wants to solve require a large amount of computing, yet the cost of achieving the high level of compute has turned people away and forced them to use cheaper computing solutions. The software vendors are reluctant to convert their software to an MPP environment unless there is a large customer base. Therefore, the work of EPCC is extremely valuable to industry in showing the value of MPP.

2.3 University

The University sector has been very proactive in their use of MPP systems for academic research. A major advantage of these research projects is that they are prototypes for future applications that will be used in industry.

There are currently 14 T3D systems at universities. Many of the universities have research grants for MPP applications that are funded by Cray Research. These grants are in the following areas:

Computer Science	9
Education	1
Environmental Sciences	12
Geophysics/Petroleum Engineering	7
Material Science	2
Medical	11
Numerical Analysis	5
Space Research	1
Structural Engineering	1

2.4 Parallel Application Technology Program

The Parallel Application Technology Program, PATP, is an effort that involves scientists and application developers at 5 T3D sites. A major objective of PATP is to create next generation parallel applications.

The program has been very successful because of the use of multi-discipline teams. The teams are composed of specialists in parallelization of algorithms/applications and domain specialists. The team has the ability to look at parallelism in applications from two directions. The domain specialist can view parallelism from the application data layout perspective, whereas the algorithmic specialist can view parallelism from the mathematics perspective. This has been very productive in finding ways to make the application scale on a large number of processors.

The programming model that was predominantly used was fast message passing. PVM and ShMem were initially the message passing the syntax of choice. The applications that needed to be portable to other platforms used PVM. Those applications that needed more performance tended to use ShMem. MPI has taken the place of PVM as the portable message passing protocol of choice. The standardization process and performance of MPI have contributed to the shift of use from PVM.

The application efforts in PATP were in the following domain areas:

Biology and Medical
CFD
Chemistry
Earth Science & Environmental
Electromagnetics
Materials, Structures and Manufacturing
Visualization and Image Processing

3 PATP Application Success

The program has seen a large number of application successes. Several of these applications are reviewed.

3.1 *New Commercial Applications*

Several new commercial applications were developed in the program. There were two major reasons behind why the effort to rewrite or develop and application from scratch was made. The first reason was that there are major problems that industry needs to solve. The existing applications were not providing the desired time-to-solutions. The second is that the existing applications or application techniques are not providing the performance and much less the cost-performance necessary to solve the problems in the time frame that would make the computational solution attractive. The following are examples of new applications developed in PATP.

3.1.1 QChem, Quantum Chemistry

The application can accelerate drug design by using first principle quantum mechanics simulations of structural configurations of biomolecules (750-1000 atoms) in minutes.

3.1.2 FALCON, Reservoir Simulation

This application will provide oil companies better exploitation strategies of oil reservoirs by running multi-million grid block production quality simulations in hours rather than days/weeks. This was an ideal project in that is teamed up a application developer, software vendor and customer to produce the application.

3.1.3 ParFlow, Groundwater

The application will provide the ability to develop economical remediation strategies for contaminated sites by running multi-million spatial zone models with subsurface heterogeneity and using a large number of geophysical realizations.

3.2 *Performance and Scalability*

The two factors in making an application perform well on an MPP platform is the single processor performance and the scalability of the application across a large number of processors.

Pittsburgh Supercomputer Center, PSC, has performed a set of tests where they looked and the performance of the application and then looked at the cost-performance of the applications on a singe processor C90 system.

Table 3. Single PE Performance and Cost-Performance

Application	Mflops/PE	C90 Equiv.
PJAC - Eigensolver	80	
NMR Density Matrix	77	8
QMD - Car Parinello	51	2.3
Vortex Simulation	40	4
QCD Conguient Gradient	38	2
PARSA - Simulated Annealing	27	1.8
Compiler Driven Communication, 2D FFT	26	4
Ab-initio Materials	25	2.1
AMBER	16	2
PFEM, Finite Element Solver	15	1
Quasigeostrophic Multigrid	15	1.4
Compressive Turbulence, PPM	13	1.7
Shake ‘N Bake	10	8
CHARMM	9	1.4
GAMESS	8	5
Multiscale & Turbulence	7	5

The cost-performance comparison was derived by comparing the performance of the application on a 32 PE T3D system and the performance on a single processor C90 system. The cost of these two systems were roughly the same at the time of the study. The table shows the number of C90 processors that would be required to give the same performance on the 32 PE T3D system.

Scalability of application performance to a large number of processors has been a challenge for many applications. Many applications experience a scalability wall of 64 or 128 PEs. The scalability tends to fall off dramatically at this point. PSC has put a great deal of effort to get a wide variety of applications to scale beyond 128 PEs.

3.3 *Application Success Stories*

Many of the application projects in PATP have delivered parallel versions of applications that can deliver performance on problems that provide the ability to tackle problems in a production environment. The following are examples of applications that are providing this new capability.

3.3.1 MICOM -- Miami Isopycnic Coordinate Ocean Model

This is work of Rainer Bleck of University of Miami and Matthew O'Keefe and Aaron Sawdey of University of Minnesota. The goal of the project is to get high resolution global ocean modeling. Initial results of the project has shown that in order to correctly model the Gulf Stream, the grid resolution had to be at most .08 degree. Figure 1 shows the results of several grid sizes. It was not until the grid size was .08 degree that the Gulf Stream was properly modeled. The model of circulation in the Atlantic Ocean has correctly predicted the course of the Gulf Stream. No other circulation model of the entire Atlantic Ocean has done this. "This degree of realism," says Bleck, "is very much a step up from previous simulations." These results prove the feasibility of a revised approach to ocean modeling.

3.3.1.1 Current Simulation -- North Atlantic Modeling

The simulations that are currently run are on the North Atlantic. The problem size is:

Grid Resolution	.08 degree
Vertical Layers	11
Grid Size	1437 x 1437
Memory	10 Gbytes

These problems take 10 days execution using 256 PEs on T3D systems.

3.3.1.2 Ultimate Goal -- Global Ocean Modeling

The goal is to be able to global ocean modeling with the appropriate fine grid size. The problem sizes would be:

Grid Resolution	.08 degree
Vertical Layers	11
Grid Size	4500 x 2000
Memory	40 Gbytes

The performance improvement that is required to make this a production job is a factor of five. This improvement will be almost achieved by running the application on a T3E system.

3.3.2 ARPS - Advanced Regional Prediction System

This is a project conducted at the Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma. The goal is to be able to predict tornados 4 hours in advance. American Airlines is working with CAPS to make this a production tool. They want to know whether flights need to be diverted and how to then schedule flights in order to reduce the inconvience to passengers and reduce possible airline's losses.

3.3.2.1 96 Prototype Code

The project was running validation tests that were reviewed for their simulation accuracy. These tests are now complete. The problem sizes run were:

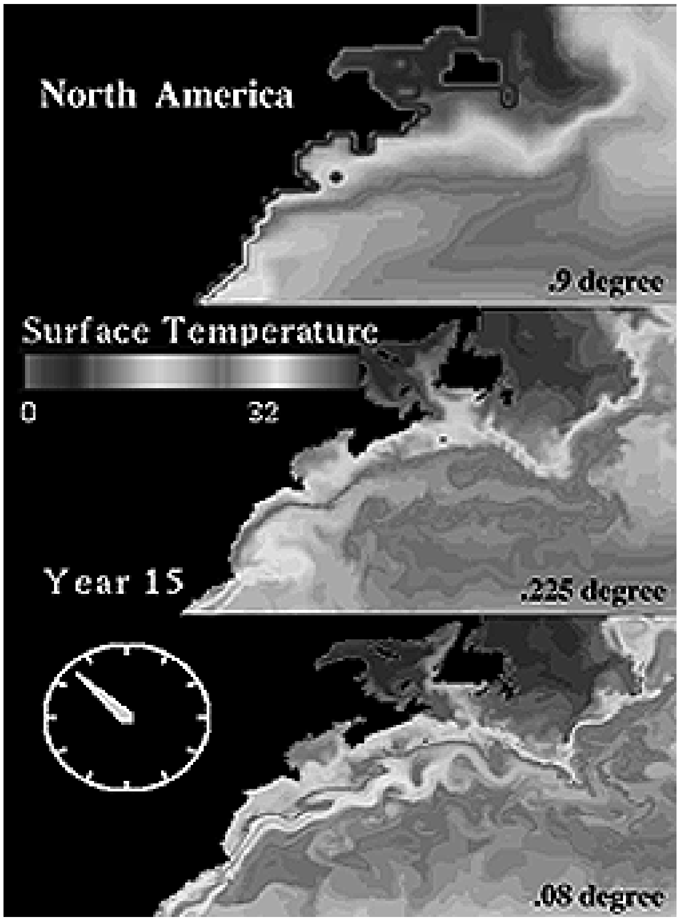


Figure 1: MICOM Simulation at different grid cell sizes

Grid Resolution	3 km x 3 km
Vertical Layers	43
Grid Size	288 km x 288 km
Time Step	4 sec
Number of Flops	10 ^ 13

The execution time for these problems on a 256 PE T3D system was 80 minutes.

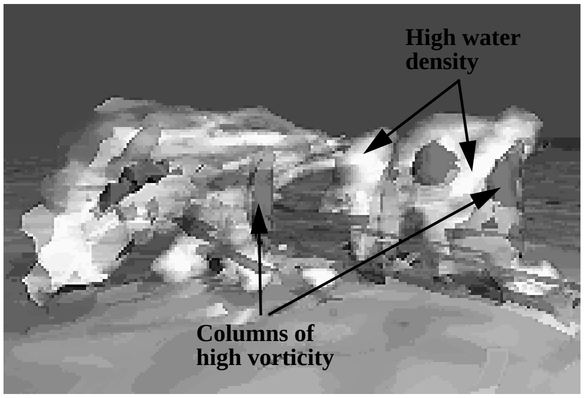
3.3.2.2 Ultimate Goal

The problem sizes that are required for a true production situation are much larger than the prototype cases. These sizes are:

Grid Resolution	1 km x 1 km
Vertical Layers	43
Grid Size	1000 km x 1000 km
Time Step	4 sec
Number of Flops	10 ^ 15

The desired execution time for these jobs is 30 minutes. In order to achieve this goal, the job will have to be able to sustain 675 Gflops. The T3E system at PSC will be used in 1997 for these jobs.

Figure 2: ARPS simulated storm structure



3.3.3 Protein Folding

This project was conducted by Charles Brooks III, Scripps Research Institute and William Young, Pittsburgh Supercomputing Center. The objective was to get a better understanding of the mechanism of helical formation and folding in proteins.

Proteins will assume millions of different shapes in milliseconds before it settles into its final shape. Experimentation suggested that the coils meet as strands and form helices while wrapping around each other to create the coiled coil. CHARMM was used for simulating the strand interactions. Simulation confirmed that the two helical strands are all stretched out, except for one or two turns. This small amount of helical structure falls below the level of experimental detection. This underscores the interplay between experimentation and simulation. The simulation for this experiment required 40000 PE hours on a T3D system!

Figure 3 shows the initial coils. The folding process begins at the “purple ribbon” and “orange cords”. The folding process continues and the coiled-coil dimer starts to form, Figure 4. The “orange cord” indicates the still unstructured portion of the chain.

Figure 3: Initiation of protein folding

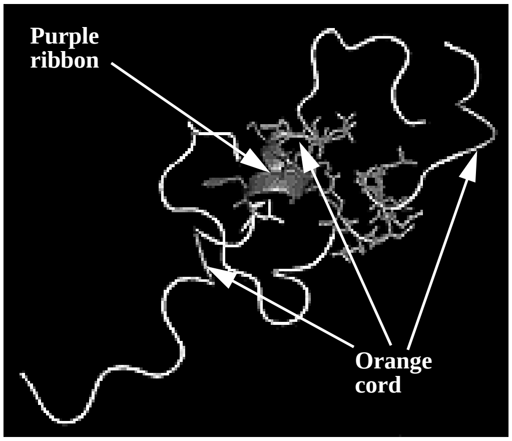
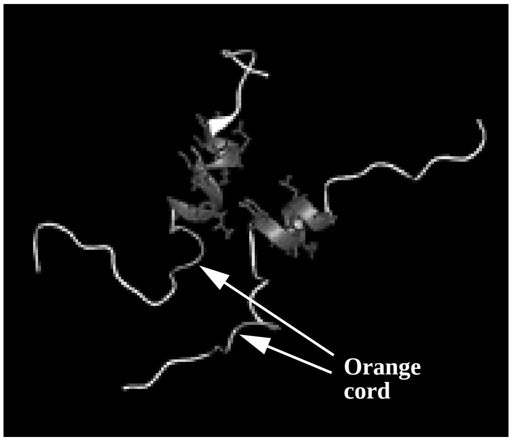


Figure 4: Folding process



3.3.4 ParFlow - Groundwater remediation

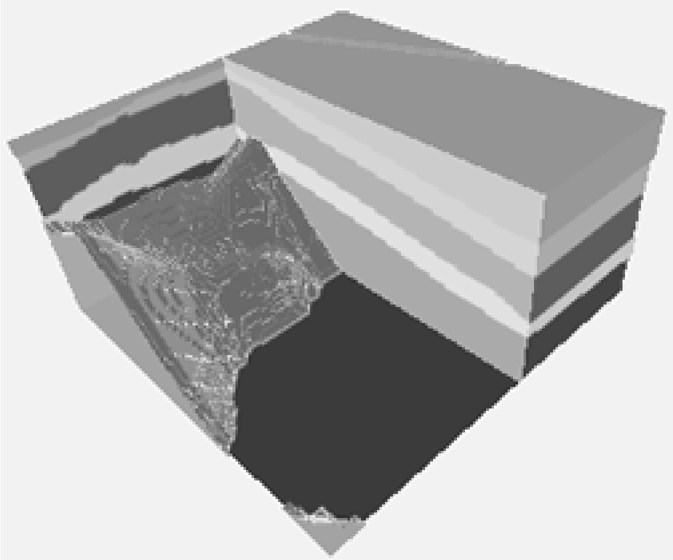
This application was developed by Steve Ashby’s team at LLNL. The goal of the ParFlow Project is to apply high performance computing techniques to the three-dimensional modeling of fluid flow and chemical transport through heterogeneous porous media to enable more realistic simulations of subsurface contaminant migration. These simulations will be used to improve the design, analysis, and management of engineered remediation strategies.

Subsurface heterogeneities must be taken into account if one wishes to draw reliable conclusions about a given remediation strategy. In order to resolve these heterogeneities adequately, especially on large sites, one must use millions of spatial zones.

This study looked at the differences between models that had homogeneous and heterogeneous layers. A hypothetical subsurface realization, see Figure 5, with five homogeneous layers, an impermeable clay region (orange), and a fault zone (red diagonal). The top layer is the most permeable, then the second, and so on. The same subsurface realization as before, but with heterogeneous layers. The variability within each layer was reproduced via a turning bands algorithm. The homogeneous model had the following run characteristics.

Grid Cells	257 x 257 x 129
Iterations	18
# flops	2 x 10 ^ 10
Run time	316 sec, 64 PEs

Figure 5: Homogenous model



Subsurface heterogeneities give rise to preferential flow channels. These channels, which are not representable in homogeneous codes, lead to fingering in the contaminant migration, resulting in much more rapid dispersion of the contaminant. The heterogeneous model had the following run characteristics.

Grid Cells	257 x 257 x 129
Iterations	37
# flops	4 x 10 ^ 10
Run time	883 sec, 64 PEs

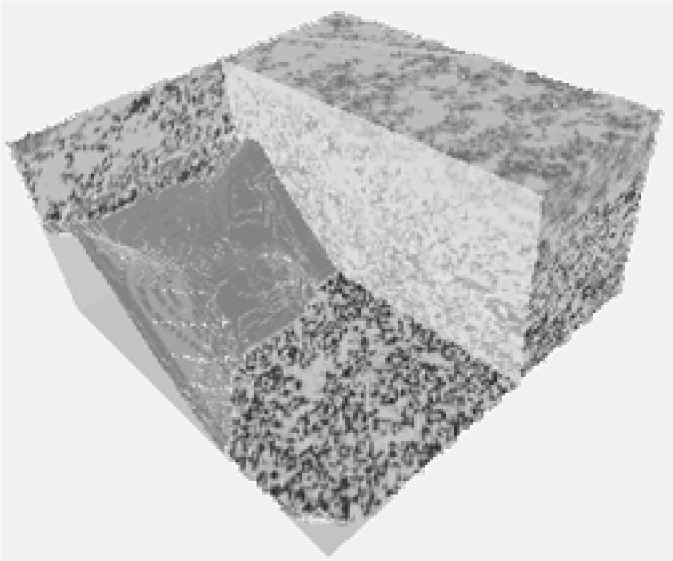
A snapshot in time, Figure 7, of contaminant migration through the subsurface (hypothetical, multilayered, homogeneous realization). The clay layer is now shown in brown and the fault zone is represented by the blue plane. The isosurface is colored according to the underlying hydraulic conductivity. Note the effect of the layers. Figure 8 is a snapshot (at the same point in time as before) of contaminant migration through the same subsurface but with heterogeneous layers. The isosurface is colored according to the underlying hydraulic conductivity. Note the dramatic effect of including heterogeneities within each layer.

The advantage of running on MPP systems is that the time to solution for these heterogeneous layer models is short enough so that the user can run a large number of geophysical realizations to get better understanding of the flow characteristics.

4 Future

The movement of codes to the T3E system has been very straight forward. The T3E system is showing excellent performance improvements.

Figure 6: Hetrogeneous Model



4.1 T3E Performance

The following table shows the performance improvement for a set of applications.

Applications	Performance Improvement	Number of PEs
AMBER	3.7	32
CHARMM	3.41	64
	2.7	256
DISCOVER 950	3.6	8
GAMESS	7.6	128
	6.3	256
LAMMPS	3.3	32
MPP-DYNA	3.89	32
STAR-HPC	3.4	32

4.2 Current Challenges

Application developers have made great progress is getting their applications onto T3D systems. The initial focus of attention was on getting single processor performance and to then have the application performance scale well on a large number of processors.

In the process to get the application to scale, several obstacles have been encountered. The stability/robustness of the parallel application has been found in some cases. This often arises because of the timing of numerical calculations on the processors can be different relative to the original serial code. The I/O requirements of the applications can be a problem when the calculations demand an input data rate that cannot be sustained by the original I/O strategy. The solution has been to develop parallel I/O methods. These methods have pushed the performance problem onto the peripheral devices used. Another major issue is the load balancing of the work performed on each of the processors. The scalability of the application requires that the

Figure 7: Homogeneous flow field snapshot

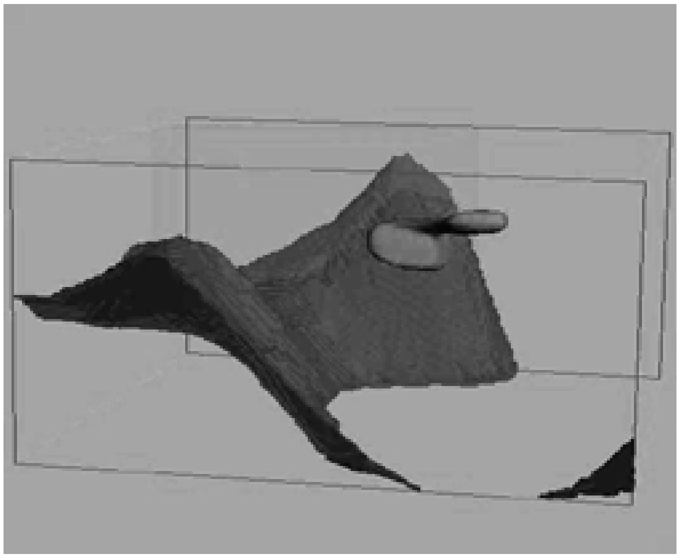
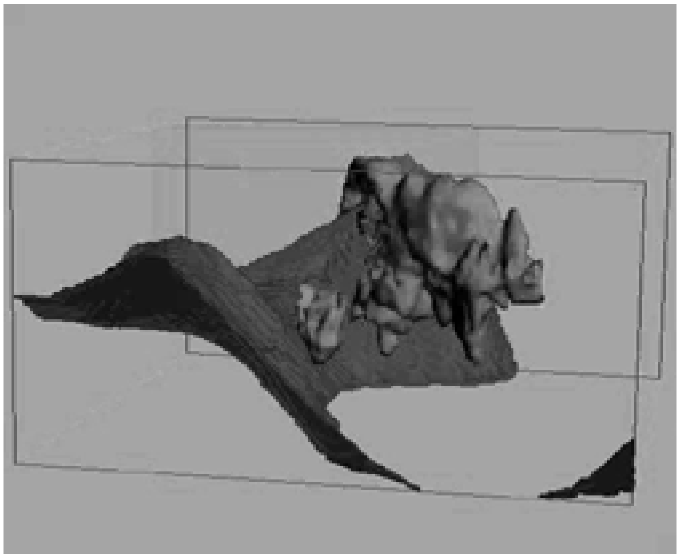


Figure 8: Heterogeneous flow field snapshot



processor utilization is maximized. The application needs to be able to ascertain when it has good load balancing and when it does not and then review whether a different data layout on the processors will achieve better load balancing.

4.3 Building Blocks for the Future

The application developer wants better tools to improve the productivity of making the application parallel. Cray Research is committed to providing tools that will take advantage of the power of the hardware

The programming models that will be supported are a Distributed Memory Model (DMM) using PVM, MPI, ShMem, etc., a Shared Memory Model (SMM) using compiler generated parallelism. A new approach that will be supported is a Hybrid Memory Model that is a combination of DMM and SMM.

The support these programming models will require very high performance parallelization tools. The tools are PVM, MPI, ShMem and parallelizing compilers. The high performance compilers will be Fortran 90, C/C++, HPF and possibly F-- (see March 1996 CUG Proceedings, Bob Numrich).

5 Next Challenge

The challenge is to make simulation relevant to a broader audience. Cray Research is committed to providing the application developer the ability to run large problems fast by maximizing the potential of a large number of processors and the developer does not have to become an expert in computer architecture and “assembly” programming. High performance applications opens the door to new uses of interactivity. This requires tools that will automatically maximize the power of the microprocessor and parallelization of the architecture and the developer does not have to vastly modify the application.