Summary of Results from running on the Cray XT4 System at ORNL

John M. Levesque
Director
Cray Supercomputing Center of Excellence
Member
Cray CTO Office

Outline of Talk

- NSF Track 1 Proposal Effort
- Applications investigated
- Results
  - Projections to the Petascale System
- How does one define the *“FASTEST COMPUTER”* in the World.

NSF Track 1 Proposal Effort

- One of the most exciting projects of my career
  - Work with researchers to identify model problems that would require a sustained Petaflop for 30-40 hours
    - What Break-through Science could be produced?
    - What machine resources would be required?
    - What Software is required?
    - What I/O would be needed?
- NSF required that the bidders project performance for
  - MILC
  - NAMD
  - DNS
- We chose an additional nine applications

Applications Investigated – Science Areas

- MILC
- DNS
- NAMD
- WRF
- POP/CICE/HOMME
- RMG
- PARSEC
- Espresso
- LSMS
- GTC
- Chimera (VH1)
- SPECFEM3D
- GAMESS

- High Energy Physics
- Computational Fluid Physics
- Biology
- Climate/Weather
- Climate/Weather
- Materials
- Materials
- Materials
- Materials
- Fusion
- Astrophysics
- Complex Engineering
- Chemistry

MILC

- NSF asked for High Energy Physics QCD application, quite obvious MILC
- Tremendous support from Steve Gottlieb of Indiana University
  - Steve ran all of the benchmarks
- Able to run a test run that used all of the processors on the XT4
  - Also ran smaller tests that could be scaled to the model problem
- While large test achieved close to 20% of peak, we were able to achieve close to 50% of peak when the sub-grid was able to fit in the core’s cache.
  - This smaller problem required a much more capable interconnect to scale up to larger processor counts

DNS

- NSF proposal called for Fourth order Runge Kutta, Direct Numerical Simulation code using Spectral methods
- Assisted by Mark Taylor of SNL and Ramanan Sankaran of ORNL
- Size of problem was restricted due to the memory of the XT4. Only able to run a 4096^3 grid on 8192 processors
  - Model problem is 12288^3 grid points
- Large majority of the time was spent in the computation and communication of the 3-D FFTs.
  - Additionally the Runge Kutta routine is the computational bottleneck
- For attaining the desired performance on the model problem we need to write the 3-D FFTs in Co-Array Fortran to facilitate the overlap of computation with communication.

NAMD

- NSF proposal asked for NAMD explicitly
- Helped by Prof Sanjay Kale of the University of Illinois and Sadaf Alam and Pratul Agrawal of ORNL
- Ran numerous variations of problems to investigate both weak and strong scaling
- Performance projections with NAMD were difficult due to the currently CHARM++ implementation that is built upon MPI.
- Given the “Global Addressability” of the Gemini interconnect we were able to project to excellent timings for running the NAMD model problem.

WRF

- Not a requested NSF application; however, this application was chosen to represent the Climate modeling area
- All work was performed by Pete Johnson with assistance from John Michalakes of NCAR
- Model problem is a girdle grid around the globe at 12.5 kilometers
- Ran several benchmarks, one a 2.5 kilometer grid covering the continental United States as well as a coarser girdle grid
- For WRF, the limiting factor was the communication required to perform the nearest neighbor updates.

POP/CICE and HOMME

- Another selected application was a more refined coupled climate model that would include
  - 1/10 degree POP
  - 1/10 degree CICE
  - ¼ degree dynamic core built upon HOMME
  - Land model
  - Coupler
- Only applications that were truly well defined were POP and CICE. Benchmarks on each of these were run as well as the HOMME ¼ degree run. We did not have a well defined Land model or coupler. Additionally the performance of the dynamic core will change drastically with the physics that needs to be added.

Material codes

- We had several material codes which is an important NSF area of research.
  - RMG
    - Jerry Bernholc and Miro Hodak of North Carolina State University
  - Espresso
    - Roberto Ansaloni, Bill Shelton and Edo Apra of ORNL and Carlo Cavazzoni of CINECA
  - PARSEC
    - Jim Chelikowsky and Murilo Tiago of University of Texas and Bill Shelton and Edo Apra of ORNL
  - LSMS
    - Markus Eisenback of ORNL
- All of these applications gave excellent performance on the current XT4 system, projections to larger systems was difficult due to the communication required to handle the larger problems

GTC

- Since simulation of the ITER fusion reactor is such an important upcoming area, We selected the most important ORNL Fusion applications
- Nathan Wichmann and Stefan Ethier of PPPL
- Model problem will be a very refined simulation of ITER.
- Had some problems setting up the benchmark to use for projection to the model problem, since the PETSc sparse solver was not working at the time of benchmarking
- Did get some very good performance from the benchmark on XT4

Chimera

- Astrophysics is a very important area for both DoE and NSF.
  - Nathan Wichmann, Bronson Messer, Tony Mezzacappa of ORNL
- Chimera is an application under development combining 3-D VH1 with other codes to enable them to simulate a complete Super Nova collapse.
- Difficult in doing projections, since the application is not really ready yet.
- Benchmark is a weak scaling run of VH1

SPECFEM3D

- Finite Element modeling of the Earth
  - Jeff Larkin, Jerome Thromp of CalTech
- Ran the largest model ever run on the XT4
- Nearest neighbor communication
- Benchmark achieve very high percentage of peak

GAMESS

- Very important area that we have not been able to penetrate with the XT3, due to the lack of efficient one-sided messaging on Portals
- Benchmark currently runs very poorly
- Once again when we get the Baker system with Gemini this as well as other Chemistry applications will perform significantly better

RESULTS Benchmarks on the XT4

Table 3.2. Additional Applications show excellent Scaling and Performance on Cray XT4

Class	Application	TF	Cores	Sust (%)	Sca (%)
Physics	MILC *	7.7	8,192	18	95
CFD	DNS *	2.8	4,096	13.2	50
Biology	NAMD *	9.3	12,000	15	50
Geosciences	WRF *	7.1	12,500	11	50
	POP *	1.447	10,705	2.6	30
	HOMME	5.750	6,912	16	50
Materials Nanosciences	LSMS *	37.3	10500	68	98
	RMG *	7.78	4,096	36.5	50
	Espresso *	7.68	4,096	36	50
	PARSEC *	7.45	4,096	35	50
Chemistry	GAMESS				
	NWCHEM				
Astrophysics	Chimera *	10.48	10,400	19	80
Fusion	GTC *	7.1	10,240	13.3	50
Engineering	SPECFEM3D *	4.68	4,096	22	80

Strong Scaling

Fastest in the World

PROJECTIONS
Application Needs Drive System Design

Node Characteristics:
- Which Applications impact what compute node characteristics

Interconnect
- What is needed in the form of interconnect latency, bandwidth

I/O
- What is important for I/O

	Node Performance				Interconnect				I/O	
	Memory Size	Memory BW	SMP enabled	Cache Opt	Injection BW	Global BW	Latency	Global Addressability	Parallel I/O	
DNS	Now ☺		Rewrite	Rewrite		Now ☺		Rewrite		
MILC		Now	Rewrite	Rewrite	Now ☺			Rewrite		
NAMD			Now	Rewrite		Now	Now ☺	Rewrite		
WRF		Now	Now	Rewrite				Rewrite		
POP		Now	Now	Rewrite				Rewrite		
HOMME			Now	Rewrite				Rewrite		
CICE			Now	Rewrite				Rewrite		
RMG	Now ☺		Rewrite	Rewrite		Now				
PARSEC	Now ☺		Rewrite	Rewrite		Now				
Espresso	Now ☺		Rewrite	Rewrite		Now				
LSMS			Rewrite	Rewrite			Now			
SPECFEM3D			Rewrite	Rewrite	Now			Rewrite		
Chimera			Rewrite	Rewrite				Rewrite		
GTC			Now	Rewrite				Rewrite		
GAMESS	Now ☺			Rewrite				Rewrite		

Now (blue) Rewrite (yellow)

☺ Architectural driver

PROJECTIONS
Application Needs Drive Software Development

System considerations:
- As with the hardware the applications drive the software delivered with the system.

Application libraries
- Surveys of many applications to make sure all the math, IO, and communication libraries are optimal.

Programming models
- Support for existing and future programming model needs.

	Programming Model						Scaling Aids					Performance Libraries								I/O		
	Fortran 77	Fortran 90	Fortran 95	C/C++	MPI	OpenMP	OpenMP	Pthreads	CAF/UPC	SHMEM	MPI2	Assembler	BLAS	FFTW	LAPACK	PBLAS	PetSC	Scalapack	Cray Scilib	HDF5	MPI-IO	netCDF
DNS			1		1		1		1			1		1					1		1	
MILC				1	1			1	1			1										
NAMD				1	1			1	1					1					1			
WRF			1		1	1			1												1	1
POP			1		1	1			1												1	1
HOMME			1		1	1			1												1	1
CICE			1		1	1			1												1	1
RMG			1		1				1				1	1	1	1		1	1		1	
PARSEC			1		1				1				1					1	1		1	
Espresso			1		1				1					1	1				1		1	
LSMS			1		1						1				1				1		1	
SPECFEM3D			1		1				1			1									1	
Chimera			1		1				1						1				1	1	1	
GTC			1		1				1								1		1		1	
GAMESS			1							1			1		1				1			

How does one define the "FASTEST" Computer in the World

- Surely it is not the Top 500
  - Although the next one will have numerous XT4s in the top 20
- Currently our only competitor is IBM's BG/L
- A better measurement would be to select the most important applications in the world and benchmark those
  - The only apps that BG/L is performing well on are those that IBM develops or funds, specifically to run on BG/L
    - Blue Matter instead of MILC
    - QBOX instead of CPMD, Espresso, etc
    - IBM has had problems running WRF on BG/L
- Of the application that we do have apples to apples comparisons, we get BG/L by a significant margin. Much more than the difference in the cost of the system.