

## **Characterizing the Performance of Overflow on Linux Beowulf Architectures**

Th. Hauser  
*Utah State University*

R. P. LeBeau, P.G. Huang  
*University of Kentucky*

Thanks to Kentucky NASA EPSCoR, the Kentucky Science  
and Engineering Foundation  
Part of the work was support by the National Science Foundation under  
Grant CTS-0321170

High Performance Computing  
Utah State University

44<sup>th</sup> AIAA Aerospace Sciences Meeting, Jan 9-12 2005

Mechanical Engineering  
University of Kentucky

### **Objectives**

- Characterize performance of Overflow on Linux clusters
- Effects
  - Cache architecture
  - Dual core / dual processor
    - OpenMP performance
    - MPI performance

High Performance Computing  
Utah State University

44<sup>th</sup> AIAA Aerospace Sciences Meeting, Jan 9-12 2005

Mechanical Engineering  
University of Kentucky

## Outline

- Cluster architectures
  - 32 bit
  - 64 bit
  - single/dual core
- Overflow
  - Test cases
- Performance results
  - Single Processor
  - Shared memory parallel
  - Distributed memory parallel

## Kentucky Fluid Clusters



## Kentucky Fluid Clusters - 64Bit architecture

### KFC4 – 32 bit

- 47 nodes
- AMD Barton 2500+
- 512 MB main memory
- 512 KB L2 cache,
- Networking
  - Single-switch 100Mb
  - Single-switch GigaBit

### KFC5 – 64 bit

- 47 active nodes
- AMD Athlon 64 - 3200+
- 512 MB main memory
- 512 KB L2 cache
- Networking
  - GigaBit ethernet
  - 48 port switch

## Utah State (HPC@USU) Cluster 64 Bit architecture

### Uinta

- 64 nodes
- two, dual-core AMD Opterons 265 per node
- 3 different networks
  - Gigabit ethernet
  - Myrinet
  - FNN with Gigabit
- 62 compute nodes 4 GB
- 2 login nodes - 8 GB
- Linux Networx



## Overflow

- Navier-Stokes solver
  - Structured grids
  - Chimera overset system
- Overflow 2.0y
- MPI and OpenMP parallelism
- Test case
  - Flat plate in test subdirectory
  - 2d
    - 11x21x3, 51x61x3, 111x151x3, 161x209x3, 321x419x3, 631x1011x3
  - 3d
    - 51x61x11, 51x61x41, 111x151x41, 111x151x101

## Tools

### Valgrind

- Version 3.1
- Cachegrind tool
- Cache simulation software
- Instruction & Data
- Miss rates for L1, L2
- <http://valgrind.kde.org>

### PGI - compiler

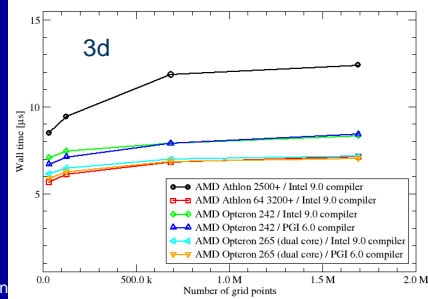
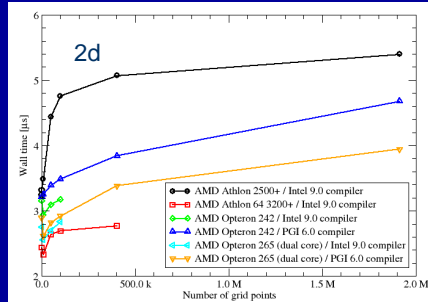
- Version 6.0
- -mp -fastsse -tp k8-64  
Ktrap=fp

### Intel - compiler

- Version 9.0
- -openmp -fpp -ipo -O3 -  
no-prec-div -static -ip -  
pad -Vaxlib -w
- Problems with stack size  
limitations

## Wall time / step / grid point of the Overflow cases

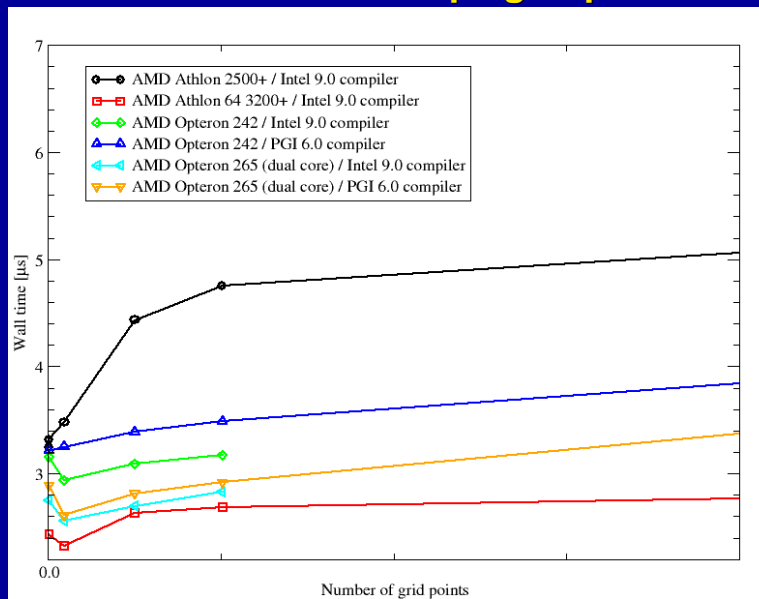
- Wall time increases with grid size
- Performance related
  - Clock speed
  - Memory architecture
- AMD Opteron
  - Clock speed influences performance
  - Improved memory architectures shows performance gains



High Performance Computing  
Utah State University

44<sup>th</sup> AIAA Aerospace Sci

## Wall clock / time step / grid point



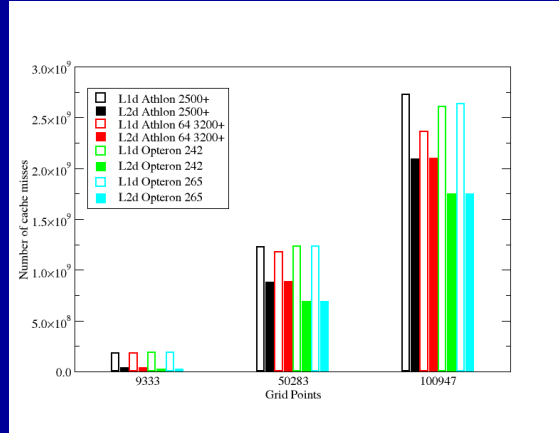
High Performance Computing  
Utah State University

44<sup>th</sup> AIAA Aerospace Sciences Meeting, Jan 9-12 2005

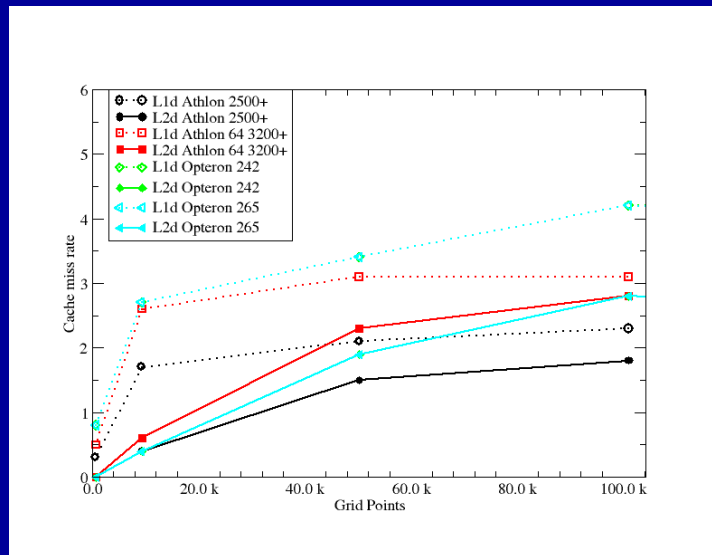
Mechanical Engineering  
University of Kentucky

## L1 and L2 data cache misses

- Valgrind
- Total number of cache misses
  - L1 the same
  - Larger L2 cache on the AMD Opteron



## L1 and L2 data cache miss rate



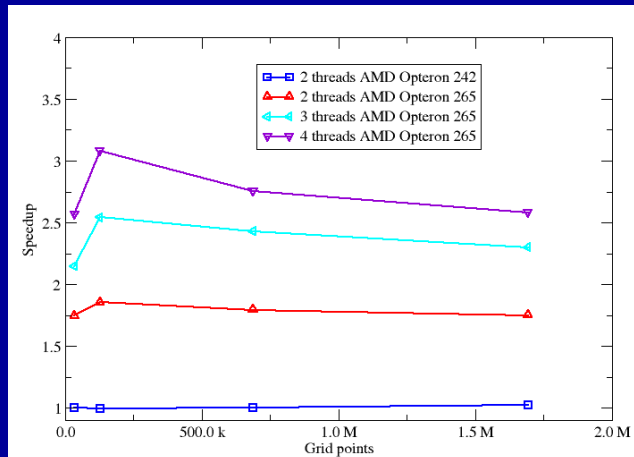
## Profile combined with cache data

- Combination of data for most expensive functions in terms of
  - Wall time (pgi profiler)
  - L2 data misses (valgrind)
  - L1 data misses (valgrind)

Time	L1	L2	
10.4	43.9	39.7	tscale
8.15	23.5	20.2	qadds
7.71	22.8	13.5	absamj
7.69	22.6	11.6	absamf
7.61			qhlk
7.52			qhlj
6.11	31	23.6	vflk
5.94	16.1	14.3	mgupd
5.01	37.1		bc2d
4.63	52.7	21.8	dqib
4.6	27	23.2	flcj
4.57			npinv
	48.4		v2psn
	29.4		v2psn3
	29	22	nrvsup
	27.8	17.8	rsk
	25.8	17	rsj
	25.6	21.4	rhlom

## Shared memory parallelism within a node

- OpenMP parallelism
- Expected Speedup
- Performance penalty with dual core
- 2 cores share one memory link
- Advantages
  - Smaller footprint
  - Standard CPU architecture in the future



## Price Performance

- 1 GB of memory per processor
- Motherboard with GigaBit network card
- Source: [www.monarchcomputer.com](http://www.monarchcomputer.com)
- Opteron 64 (2.0 GHz)
  - Price per node: \$690
- Dual core Opteron 64 (2.0 GHz)
  - Price per node: \$870
  - Price per processor: \$435
- Dual Opteron (2.0 GHz)
  - Price per node: \$1460
  - Price per processor: \$730
- Two dual core Opterons (2.0 GHz)
  - Price per node: \$3090
  - Price per processor: \$773

## Conclusions

- Overflow is memory bound
- Clear performance improvement with AMD 64 bit architecture
- Considerable number of cache misses
- Good shared memory performance using OpenMP
- Parallel performance on dual core nodes strongly dependent on MPI implementation
  - NUMA aware
  - Open MPI

## Future Work

- Instrument Overflow
  - Hardware performance counters
  - Papi
- Optimize selected subroutines
- Evaluate parallel performance
  - Networks
  - Convergence
- Parallel I/O
- Mixed Parallelism
  - OpenMP within a node
  - MPI between the nodes

## Questions?