

Performance Modeling of Jaguar: Pre- and Post-Upgrade

Kevin J. Barker, Kei Davis, Darren J. Kerbyson,
 Adolfy Hoisie, Michael Lang, Scott Pakin, Jose Carlos Sancho
 Performance and Architecture Lab (PAL)
 Computer Science for HPC, CCS-1
 Los Alamos National Laboratory

1. Why Model?

- Predict performance of post-upgrade system on applications of interest to ORNL
- Provide sanity check of actual system performance post-upgrade, diagnostic information.
- Enable exploration of
 - Hypothetical alternative architectural changes
 - Impact of run-time parameters, e.g. process mapping.

2. Methodology

- Develop performance models for the GTC and S3D applications
 - The performance models capture the applications' key computational, communication, and scaling characteristics
- Validate the models using appropriate input decks
 - On a generic Opteron/Infiniband cluster
 - On Jaguar (pre)
- Use performance models to predict performance of Jaguar (post), considering
 - New architectural and topological characteristics
 - New communication and computational performance
- **Because actual hardware representative of the upgraded Jaguar was not available for testing, several model inputs are based on assumed values.**
- Use performance models to explore the performance implications of various process mapping strategies.

3. Applications GTC and S3D

The **Gyrokinetic Toroidal Code (GTC)** is a 3D particle-in-cell code developed by the Princeton Plasma Physics Laboratory to study microturbulence in magnetically confined fusion plasmas. GTC is currently the flagship SciDAC fusion microturbulence code and has been shown to scale to large processor counts; typical runs use on the order of 1024 processor cores with *weak scaling*, i.e. keeping the particle count per processor core fixed as the core count increases. GTC was run with one MPI process per processor core.

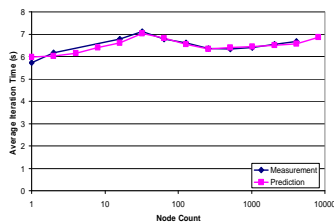
S3D was developed, initially at Sandia National Laboratories, for direct numerical simulation of turbulent combustion. S3D solves full compressible Navier-Stokes, total energy, species, and mass-continuity equations coupled with detailed chemistry. The governing equations are solved on a 3D Cartesian mesh. The computation is parallelized using a 3D domain decomposition with uniform domain decomposition in each dimension, yielding uniform computational load per computational process. S3D is typically run in weak scaling mode where the subgrid size remains constant per computational process, with one process per processor core, typically using weak scaling.

4. Jaguar pre- and post-upgrade Configurations

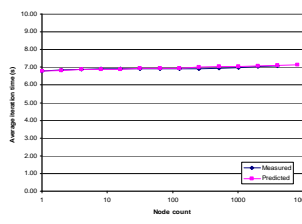
| | Jaguar (pre) | Jaguar (post) |
|---|----------------|----------------------|
| System | | |
| Processor type | Opteron | Opteron |
| System topology | 31x16x24 mesh | 21x16x24 mesh |
| Total node count | 11508 | 8064 |
| Cores per processor | 2 | 4 |
| Total # cores | 23,016 | 32,256 |
| System peak (Tflops) | 120 | 284 |
| Node | | |
| Processors per node | 1 | 1 |
| Cores per node | 2 | 4 |
| Processor speed (GHz) | 2.6 | 2.2 |
| Peak/node (GFlops) | 10.4 | 35.2 |
| Memory per node (GB) | 4 | 8 |
| Memory type/speed | DDR2-667MHz | DDR2-667/800MHz |
| Network | | |
| Network type | Cray SeaStar 2 | Cray Seastar 2+ |
| Peak injection rate (GB/s) | 2.0 | 3.2 |
| Peak link bandwidth (GB/s) | 7.0 | 9.6 |
| Measured MPI bandwidth (single ping-pong 1MB message) | 1670MB/s | 2670MB/s (estimated) |
| Measured MPI latency (0B message) | 6.9µs | 6.9µs (estimated) |

5. Validations of Models on pre-upgrade Jaguar

For the pre-upgrade Jaguar, the model predicts GTC performance with a maximum error of 2.6% (avg. 1.2%), and S3D performance with a maximum error of 1.4% (avg. 0.53%)—very good considering that testing was performed while the machine was in heavy use.



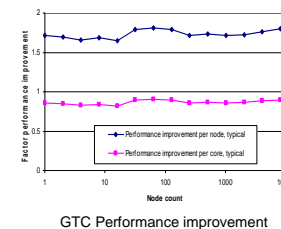
GTC modeled vs. measured performance



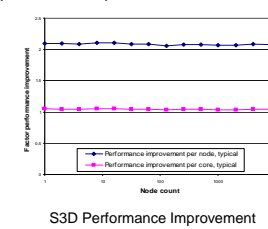
S3D modeled vs. measured performance

6. Post-upgrade Predictions

- We predict approximately equal *per processor core* performance, on a per-node basis, for GTC and S3D, pre- and post-upgrade. This is in spite of reduced clock speed, and increased memory and network contention, post-upgrade.
- Because the upgraded machine will have four cores per node, *per-node* performance will be approximately double.



GTC Performance improvement



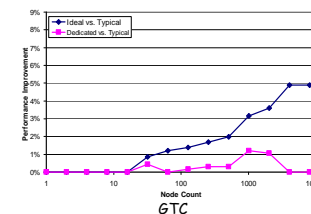
S3D Performance Improvement

7. Process Mapping

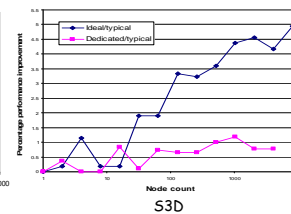
Because the model necessarily accurately models network contention, we may easily explore the performance implications process mapping strategies. Here we define

- *Typical* mapping as one that might be made with multiple concurrent users;
- *Ideal* mapping is the application-specific mapping to give best performance.
- *Dedicated* is allocation in node ID order

Figures below show the performance improvements realizable using ideal or dedicated mappings vs. typical mappings.



GTC



S3D

8. Conclusions and Future Work

- This work
 - Developed performance for models for GTC and S3D
 - Validated the models on Opteron/Infiniband cluster, and Jaguar (pre)
 - Employed these model to predict performance of Jaguar (post)
 - Used the models to explore alternative process mapping strategies
- Future work
 - Will refine predictions with model inputs from measurements of actual hardware
 - Use models and predictions to verify final system performance