Large Matrix-Matrix Multiply on PS3 clusters

15 September 2010

Mark Barnell, AFRL RITB
Mark.Barnell@rl.af.mil

Dennis Fitzgerald, ITT
dennis.fitzgerald@itt.com

DISTRIBUTION STATEMENT A. Approved for public release; distribution unlimited. (Approval given by Public Affairs Office (September 2010).)
Description

- Matrix-Matrix multiplication of large matrices
- $> 100k \times 100k$
- Parallelized over a number PS3s
- Maintained near peak performance on each Cell BE
Challenges

• Near peak computation rate on the Cell BE for small matrix sizes
• Data and thread coordination between PowerPC and Cell BE with near zero overhead
• Balanced IO with Cell BE’s peak FLOPS to keep PS3 computationally busy
• Network performance sufficient to deliver enough data to many PS3s
Approach

• Core MM algorithm > 99% efficient (128x128)
  – Daniel Hackenberg – Dresden
• PowerPC code to coordinate larger rectangular matrices – Miriam Leeser – Northeastern
• Multi-buffering & semaphors to reduce wait time
• Blocked sub-matrix distribution with data sized to balance compute and IO
Results

Matrix-Matrix Multiply GFLOPS

- 48k x 48k
- 48k x 240k
- PS3 Max GFLOPS (153)

Number of PS3s

GFLOPS

UNCLASSIFIED