Technology Watch and Evaluations

Marc Paterno and V. Daniel Elvira
Fermilab

2nd Annual Concurrency Forum Meeting

- Thanks to all our speakers
- It is not possible to do justice to each talk in a single slide

- But we'll try our best…

Practical Results of the Intel MIC/Xeon Phi Project at CERN openlab (A. Nowak, CERN openlab)

- Ported 3 real-world HEP applications to run on pre-production MIC architecture: ALICE track fitter, MLFit, Geant4-MT prototype
- Porting times from <1 to ~month; tuning times <1 week to several weeks
- Optimized applications surpass dual-socket Xeon performance
- Non-optimized applications approximately match single-core Xeon performance
- For best results, need to think of vectorization, parallelization (threads, MPI) and small memory usage

Brief correlation study on x86 compiler flags and performance events (A. Nowak, CERN openlab)

- Question addressed: Can we combine knowledge about compiler flags and the response they produce in hardware?
- Results of similar experiments were difficult to reproduce
- It is possible to semi-automatically characterize benchmarks, and to establish which compiler flags are likely to reduce a particular [performance] bottleneck
- It is difficult to predict with good accuracy which compiler flags will improve a particular workload.
- Full report at http://openlab.web.cern.ch/sites/openlab.web.cern.ch/files/technical_documents/CompilerFlags_Review2.pdf.

Accelerating Science with Kepler and CUDA 5 (J. Bentz, Nvidia)

- Major new features: SMX, Hyper-Q, Dynamic Parallelism
- SMX cores: 6 times as many, 3x perf/watt
- Hyper-Q: Run up to 32 simultaneous MPI tasks on GPU
- DP: GPU can launch additional threads dynamically
- Up to 255 registers per thread (4x Fermi's limit)
- Variety of math libraries available in CUDA 5 toolkit
- Much additional information available as extra slides, at the workshop Indico site

Programming Models for Intel Xeon Processors and Intel Xeon Phi Coprocessors (S. McMillan, Intel)

- Concentrated on use of Xeon Phi as a coprocessor
- 60 cores, wide vector units
- Different modes of use:
  - as “cluster on chip”
  - Like an accelerator, “many-core hosted”
  - Symmetric use of host Xeon and coprocessor Xeon Phi
- Supports multiple parallel programming technologies, including *Threading Building Blocks*, *MPI*, and *OpenMP*.
- Can port from x86 to Phi fairly cheaply, and then optimize incrementally

Transforming Geant4 for the Future (B. Lucas, USC)

- The US Department of Energy (DOE) charged Bob Lucas (Advanced Scientific Computing for Research – ASCR) and Rob Roser (High Energy Physics – HEP) to co-chair US ASCR/HEP workshop to discuss "Transforming G4 for the Future"
- Final report available at http://science.energy.gov/~/media/ascr/pdf/research/scidac/GEANT4-final.pdf

  48 participants from HEP, ASCR, experiments.
- ASCR and HEP should investigate together
  - Optimize today's Geant4 for immediate impact
  - Refactor and re-engineer Geant4 for future computing systems
  - Address challenges from petabytes of data generated
- The "Concurrency Forum" and the Geant4 Collaboration are the natural communities for this effort to be discussed and integrated to the international effort

Performance Measurement Tools for Parallel Applications (S. Jun, Fermilab)

- Included in requirements: (1) support of multi-threaded applications, (2) support of Linux, (3) no source code instrumentation, (4) advanced analysis (tracing, callgraphs)
- Short list of toolkits: HPCToolkit, Open|SpeedShop, TAU, nvvp [for CUDA profiling]
- Each tool has its strengths, none does everything
- Performance analysis require domain knowledge as well as computing system knowledge
- Expect to benefit from collaboration with ASCR institutes