Accelerated Debugging
Allinea DDT and OpenACC on the XK6

David Lecomber
CTO, Allinea Software
david@allinea.com
Some history

A long time ago (2007) in a galaxy far, far away....

- The CUDA programming model is introduced
  - Powerful, efficient and C based
  - Understood and adopted by new groups of experts
  - Existing codes modified to extract SIMD parallelism and introduce CUDA kernels
  - Performance of codes is optimized by overlapping device and host, or rearranging memory usage inside device
- The first CUDA bug is created
Allinea Software

- HPC development tools company
  - Flagship product Allinea DDT
    - The leading debugger in parallel computing
    - The scalable debugger
      - Record holder for debugging software on largest machines
      - Production use at extreme scale – and desktop
    - Wide customer base
      - Blue-chip engineering, government and academic research
      - Strong collaborative relationships with customers and partners
  - Leaders in performance and usability
Allinea DDT in a nutshell

- Graphical source level debugger for
  - Parallel, multi-threaded, scalar or hybrid code
  - C, C++, F90, Co-Array Fortran, UPC
- Strong feature set
  - Memory debugging
  - Data analysis
- Managing concurrency
  - Emphasizing differences
  - Collective control
May 2011 - CUG 2011

- Petascale debugging becomes real
- Allinea DDT 3.0 – lightning speed – 100,000 cores and beyond
  - Record holder for largest machines
  - Debugging at scale becomes fast, simple and routine
  - Production use at extreme scale on Cray XE and XT systems
Lift-off – beyond petascale

- Fairbanks, May 2011 – Cray XK6 announced
  - Large GPU systems firmly on the agenda
  - Allinea and ORNL collaborate to ensure GPU applications debuggable at scale
    - Petascale debugging, but with GPUs
    - Core needs identified and key features and performance specified
- How would the XK6 be programmed?
  - Candidate pragma languages to remove CUDA burden
How do we fix GPU bugs?

- Print statements
  - Too intrusive
- Command line debugger?
  - A good start:
    - Variables, source code
    - Large thread counts overwhelming
  - Too complex
- A graphical debugger...
Almost like debugging a CPU – we can still:

- Run through to a crash
- Step through and observe
- CPU-like debugging features
  - Double click to set breakpoints
  - Hover the mouse for more information
  - Step a warp, block or kernel
  - Follow threads through the kernel
- Simultaneously debugs CPU code
- CUDA Memcheck feature detects read/write errors
Examining GPU data

- Debugger reads host and device memory
  - Shows all memory classes: shared, constant, local, global, register..
  - Able to examine variables
  - ... or plot larger arrays directly from device memory
**Overviews of GPUs**

- **Device overview** shows system properties
  - Helps optimize grid sizes
  - Handy for bug fixing – and detecting hardware failure!

- **Kernel progress view**
  - Shows progress through kernels
  - Click to select a thread
A New Hope

- Seattle, November 2011
  - CAPS, Cray, NVIDIA and PGI announce new standard for accelerator programming
    - Access CUDA compute power easily
    - A common standard
  - Allinea supports debugging Cray OpenACC compiler
November 2011

- Allinea DDT 3.1 – innovation for all scales
  - Sparklines – automatic data comparison
  - Offline debugging – scalable hands-free
  - Static analysis – automatic detection of coding errors
- Cray XK6 support
- Cray UPC, CoArray and OpenACC support
OpenACC debugging

- On device debugging with Allinea DDT
  - Variables – arrays, pointers, full F90 and C support
  - Set breakpoints and step warps and blocks
- Requires Cray compiler for on-device debugging
  - Other compilers to follow
- Identical to CUDA
  - Full warp/block/kernel controls
XK6 Status

- Interlagos update
  - No problem – tested at scale!
- Last known major GPU issues fixed
  - Driver fixes deployed April 2012
    - 3-way debug sessions to diagnose and fix a device-hang
- Full OpenACC and CUDA support
What's next?

- Large Cray XK6 systems in (or almost in) place
  - ORNL Titan – 300,000 CPU cores
  - NCSA Blue Waters – 380,000 CPU cores
- Allinea DDT chosen for both systems – at scale
  - Watch out for improvements over next 6 months!
  - Kepler support for system upgrades