"APPLIED PARALLEL COMPUTING"
Education and Research Center

3 Days GPU Training Course

At The *Irish Centre for High-End Computing*

Requirements

- Advanced CUDA experience

Outcome

Course will cover selected advanced topics for experienced CUDA developers and HPC centers: programming models, applications analysis and debugging for single and multiple GPUs.

Course Plan

Day 1 (February, 1st)

9.00 – 9.15 – Opening (ICHEC)

9.15 – 10.30

1. CUDA programming model
   - Brief revise of CUDA programming model, new features of 4.0 and 4.1
   - CUDA runtime API
     > Asynchronous execution
     > Handling runtime errors in CUDA
     > Querying GPU capabilities

10.30 – 10.45 – Coffee break

10.45 – 12.00

2. Memory hierarchy
   - Texture memory, data access patterns for texture memory
   - L1/shared vs L2 caching management
   - Unified virtual address space (UVA)
     > Case study of using of UVA: when efficient and inefficient
     > General view of pros&cons of using UVA both CPU/GPU and CPU/multi-GPUs

12.15 – 13.30

3. Thrust
    - Transforms and functors
    - Placeholders and tuples
    - Performance considerations
    - Thrust and CUDA/C interoperability

13.30 – 14.30 – Lunch

14.30 – 15.30

4. Hands-on on Thrust: implement data processing transform, measure performance

15.45 – 17.00

5. Using multiple GPUs
    - CUDA context
    - MPI
    - POSIX-threads
    - OpenMP
    - Boost.Threads
    - Case study: DirectGPU/MPI CUDA aware library

Day 2 (February, 2nd)

9.00 – 10.30

6. CUDA Streams
    - Example: concurrent kernels execution
    - Example: matrix multiplication
    - Example: Multi-GPU Async peer-to-peer Copy
    - Asynchronous time measurement

10.30 – 10.45 – Coffee break

10.45 – 12.00

7. Debugging
    - Principles and terminology
    - gdb
    - cuda-gdb
    - cuda-memcheck

12.15 – 13.30

8. Hands-on: CUDA debugging and diagnostics

13.30 – 14.30 – Lunch

14.30 – 15.30

9. Profiling
    - Meaning of most common used counters, examples
    - Setting up and running CUDA command line profiler

15.45 – 17.00

10. CUDA Low-level & compiler
    - NVCC pipeline
    - PTX assembly
    - CUDA Driver API
    - Compiling & customizing nvopencc
    - Compiler-driven CUDA properties

Day 3 (February, 3rd)

9.00 – 10.30

11. CUDA and generalized operators discretization

10.30 – 10.45 – Coffee break

10.45 – 12.00

12. 3D ADI Method for Fluid Simulation: Scaling to multiple GPUs

12.15 – 13.30

13. Optimization Techniques
    - Types of performance limiters
    - Global memory optimization
    - Instructions optimization
    - Overall latency optimization

13.30 – 14.30 – Lunch

14.30 – 15.30

14. Porting numerical models in Fortran to CUDA
    - General approach of combining CUDA and Fortran
    - Porting WSM5 and RRTM blocks of WRF model

15.45 – 17.00

15. Q&A and projects discussions (?)