COMP 605: Introduction to Parallel Computing
Lecture : GPU/CUDA Programming

Mary Thomas

Department of Computer Science
Computational Science Research Center (CSRC)
San Diego State University (SDSU)

Posted: 04/10/17
Last Update: 04/10/17

Table of Contents

Reading List for GPU/CUDA Topic

- Primary Textbook: CUDA By Example, by Sanders & Kandrot

  https://www.physics.drexel.edu/~valliere/PHYS405/GPU_Story/CUDA_by_Example_Addison_Wesley_Jul_2010.pdf

  - Ch2: Getting Started
  - Ch3: Introduction to CUDA C
  - Ch4: Parallel Programming in CUDA C using Threads
  - Ch5: Thread Cooperation

- Also using: Kirk & Hwu, *Programming Massively Parallel Processors*:
  - Ch3: Intro to Data Parallelism & CUDA
  - Ch4: Data Parallel Execution Model
  - Ch5: CUDA Memories
  - Ch6: Performance considerations (if time permits)

- CUDA example files on tuckoo in /COMP605/cuda

Additional NVIDIA GPU/CUDA Refs

- Tutorials
  - CUDA C Programming Guide
    http://docs.nvidia.com/cuda/cuda-c-programming-guide/
  - CUDA Training:
    https://developer.nvidia.com/cuda-education-training#1
  - CUDA API:
    http://docs.nvidia.com/cuda/cuda-runtime-api/index.html
  - CUDA SDK:
    https://developer.nvidia.com/gpu-computing-sdk
  - Cuda Tutorial: Volume 1 Exercises & Instructions:
        https://developer.nvidia.com/cuda-training/

- GPU Architectures:
  - References: NVIDIA online documents
 - and lecture notes by S.Weiss
    http://www.compsci.hunter.cuny.edu/~sweiss/course_materials/csci360/lecture_notes/gpus.pdf

Additional GPU/CUDA Reading

- http://www.hpcwire.com/hpcwire/2008-10-08/compilers_and_more_programming_gpus_today.html
- http://www.hpcwire.com/hpcwire/2008-10-30/compilers_and_more_optimizing_gpu_kernels.html
- http://www.admin-magazine.com/HPC/Articles/Parallel-Programming-with-OpenMP
- http://people.math.umass.edu/~johnston/PHI_WG_2014/OpenMPSlides_tamu_sc.pdf
    - Matrix Multiplication with CUDA — A basic introduction to the CUDA programming model. Robert Hochberg, August 11, 2012
    - K. Fatahalian, J. Sugerman, and P. Hanrahan, "Understanding the Efficiency of GPU Algorithms for Matrix-Matrix Multiplication," Graphics Hardware (2004)

History

- 1980s: Silicon Graphics promoted 3D graphics processing, which fueled CGI in movies and gamestem Key CGI Features/Approach:
    - model objects with tiny triangles/polygons
    - calculate changes in shape, color, lighting, etc. from frame to frame.
    - synch to visual/movie frame rates
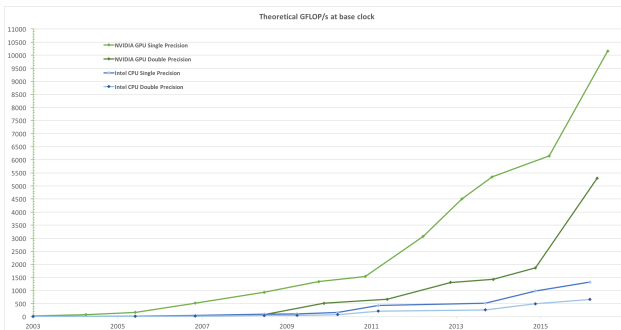   - model images as slices in time
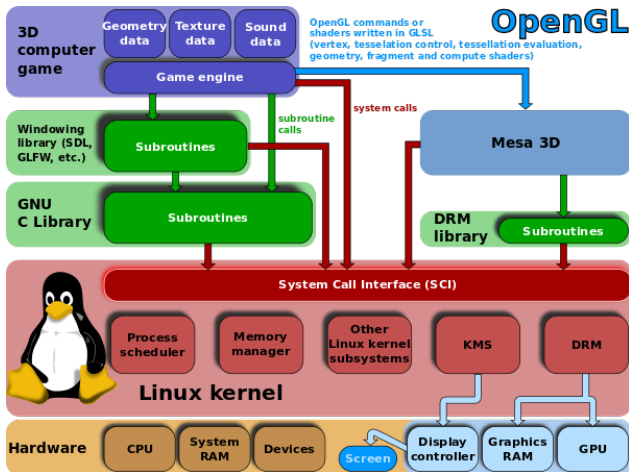
NVIDIA / GPU History

- 1990s: NVIDIA and other companies were developing graphics accelerators
- 2001 release of NVIDIA GE Force 3 series brought in visualization and rendering standards.
- This motivated scientists and researchers to look at using GPU for science applications.
- 2006 NVIDIA releases Ge Force 8800 with CUDA architecture
  - uniform shader architecture
  - ALU
  - designed for general purpose computing

Source: http://docs.nvidia.com/cuda/cuda-c-programming-guide/graphics/floating-point-operations-per-second.png

Cuda Overview: Online tutorial

http://edoras.sdsu.edu/~mthomas/docs/cuda/
cliff-woolley-nvidia-cuda-overview.pdf