# High Performance Computing Systems

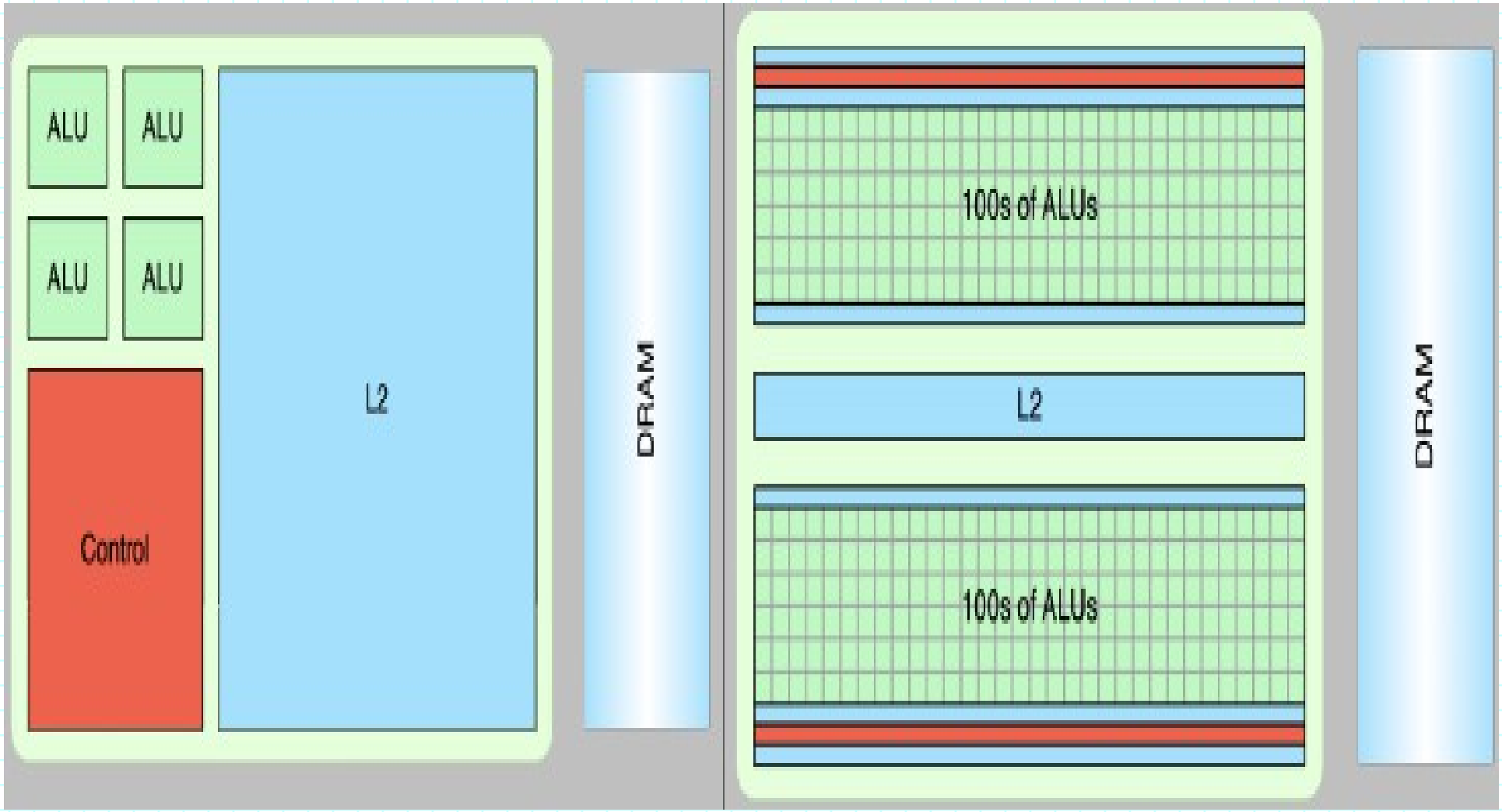
- GPUs

*Doug Shook*

# Presentations

- 20-25 minutes + Q&A
- Grading:
  - 75% content/knowledge of topic
  - 25% style
- Style tips:
  - Sparse Slides
  - At least 1 minute per slide
  - Practice!

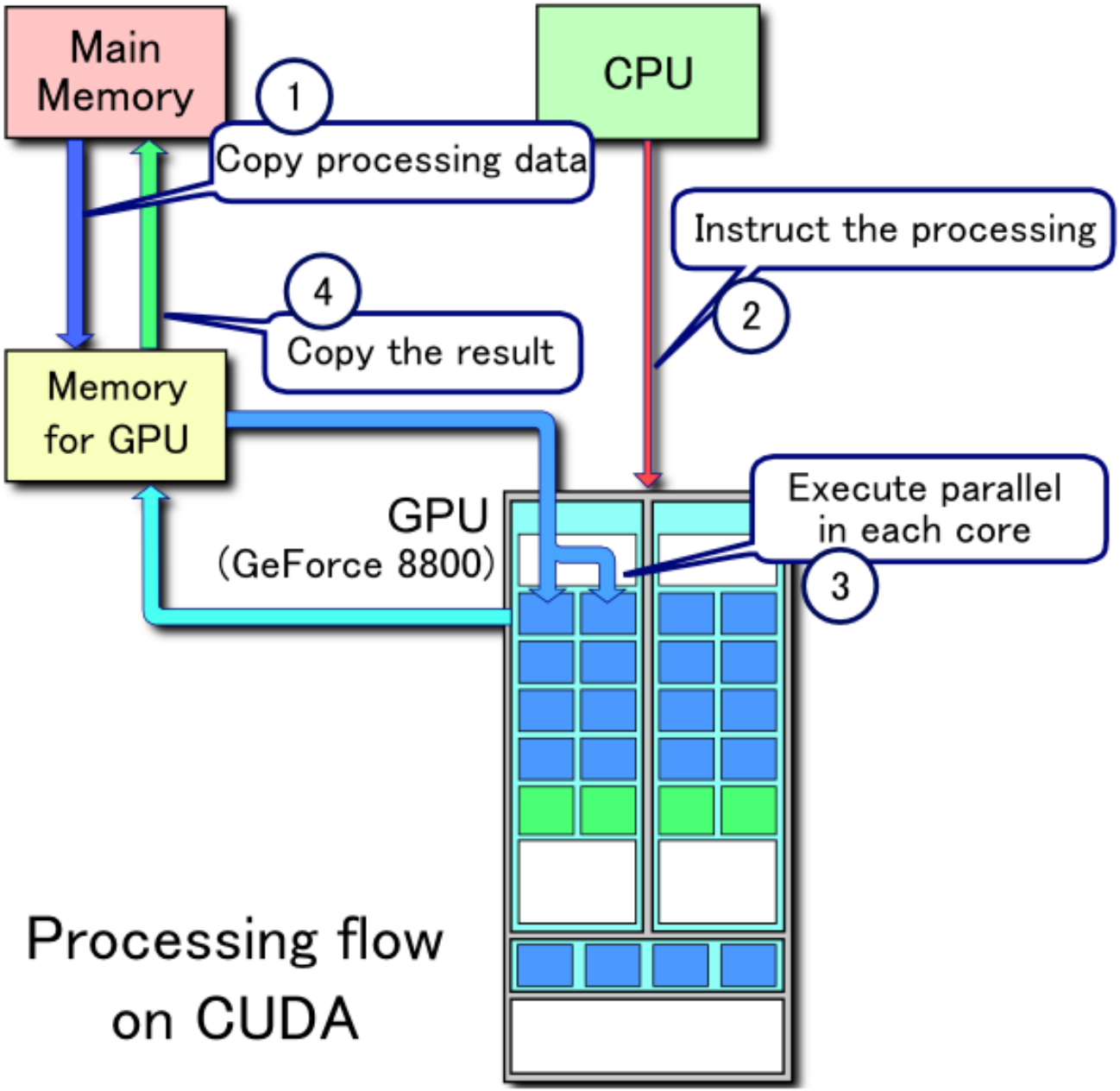
# GPU Architecture



# CUDA

- Developed by NVIDIA
  - 2007
- Works with C/C++/Fortran
- Allows for heterogeneous computing

# Workflow


Main Memory
CPU
1
Copy processing data
Instruct the processing
2
4
Copy the result
Memory for GPU
GPU
(GeForce 8800)
Execute parallel in each core
3
Processing flow on CUDA

# CUDA Examples – Device Info

# CUDA Examples – Hello, World!

- __global__

- Kernel<<<1, 1,>>>

# CUDA Examples – Add

- Let's make __global__ do something…
  - Notice that everything in global is a pointer
- Memory management

# CUDA Examples – BlockAdd

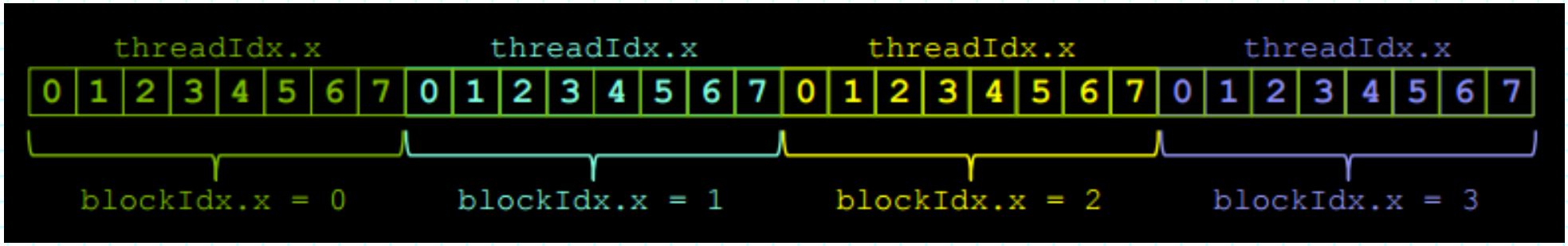
- Let's do some work in parallel

# CUDA Examples – Blocks and Threads

- We can also use blocks and threads together

- What do blocks provide for us?
  - What do threads provide for us?

# Blocks and Threads



# CUDA Examples – Dot Product

- __shared__
- There's something wrong with this….

# CUDA Examples – Multi Block Dot Product

- There's also something wrong with this….

# CUDA Examples – Heated Plate

# Profiling CUDA

- nvprof