GP2 Panel Presentation

William Mark, University of Texas at Austin

We're entering an era of disruptive change

- Driven by VLSI technology

  – Too many transistors: CPU performance plateau

  – Heat/Power is now a first-class constraint

  – Possible to fit many processors on a single chip

- Two kinds of change coming:

  – Technical – single-chip parallel computation

  – Industry structure – pressure for vertical re-integration

SIGGRAPH 2004

GPGPU

What do we mean by "CPU vs. GPU"?

- General HW vs. specialized HW

  – GPU's moving towards generality, but not fully there yet

- Sequential vs. Parallel

  – Latency optimized vs. Throughput optimized

- Two separate chips

- Different sets of companies (exception: Intel)

- Raw HW access vs. Managed code

Need at least two parallel programming models

- Stream model

  – Naturally exposes parallelism and communication

  – Easy to use, when problem maps well

- Communicating sequential processes (e.g. pthreads)

  – Explicitly exposes spatial dimension of HW parallelism

  – Efficiently supports data-dependent communication patterns

  – Useful for creating/modifying large irregular data structures

  – Harder to use – e.g. race conditions

  – Hard to get performance portability

GPGPU

HW must satisfy mass-market needs

- Games will continue to dominate
  - Rendering
  - Simulation? – an opportunity
- Maximize impact of research by meeting game needs
  - Chicken/Egg problem: Co-evolve algorithms and architectures
  - Different visibility algorithms – ray casting?
  - Global illumination – shadows, ambient occlusion, reflection, …
  - Parallelize model management, simulation, game behavior, …
- Solving these problems will help other applications

2-year predictions

- CPU's: multi-core trend accelerates

  – Multicore used by games and HPC

- GPU's: More powerful streaming model

  – Scatter, gather, conditional streams, reductions, etc.

  – Start to see more success stories for GPGPU

  – But limits of stream model become apparent

- "Dark Horses" attract increasing attention

  – CELL and others

6-year predictions

- One processing chip for PC's

  - Who makes it?

- Heterogeneous architecture for this chip:

  - Classical CPU

  - Parallel fine-grained shared memory (pthreads)

  - Parallel stream processor (Brook)

- Supports ray-casting visibility

- This architecture emerges in console space first

- This architecture meets many HPC needs