



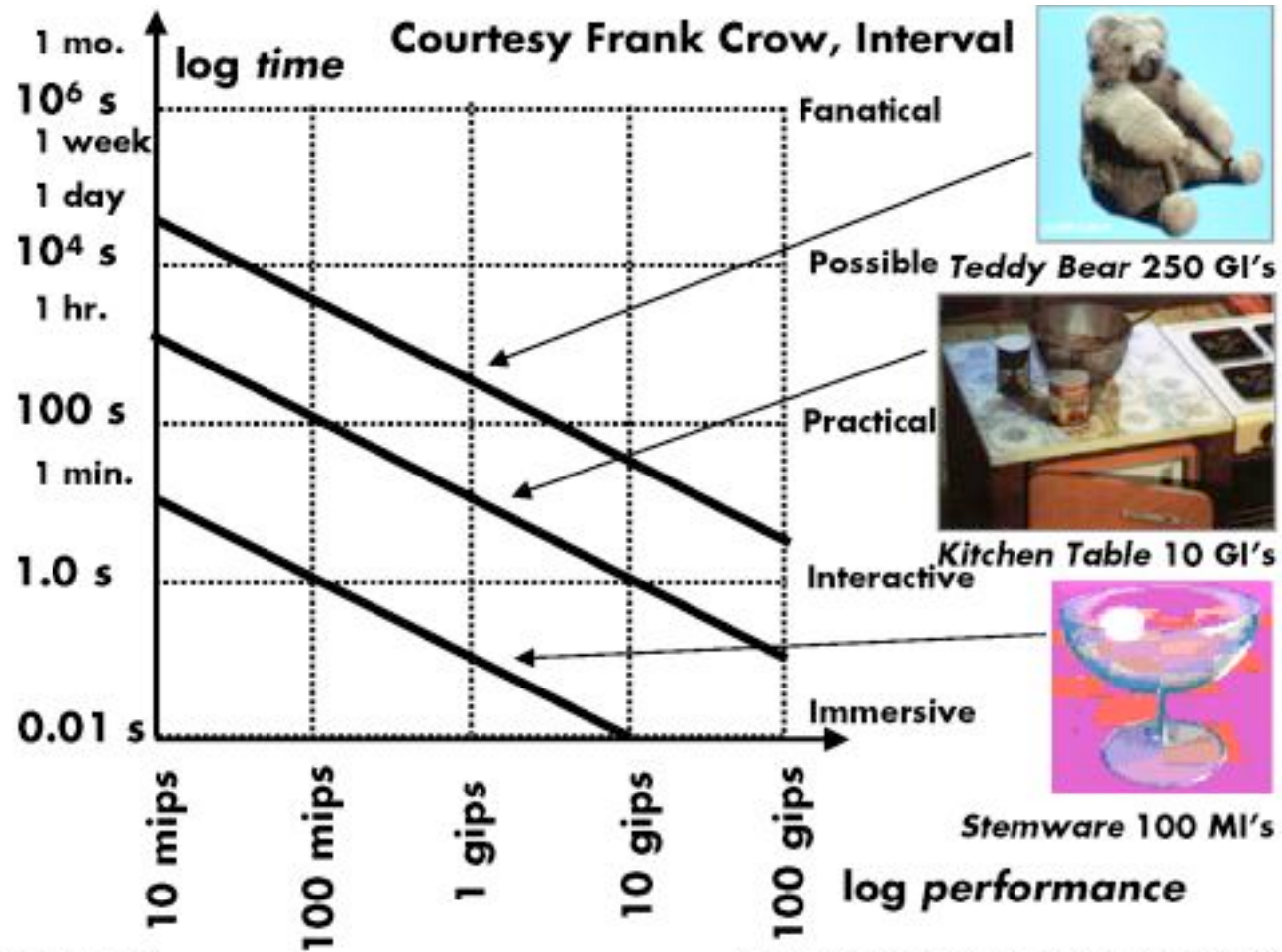
The Future: What's Next for GPUs?

John Owens

University of California, Davis

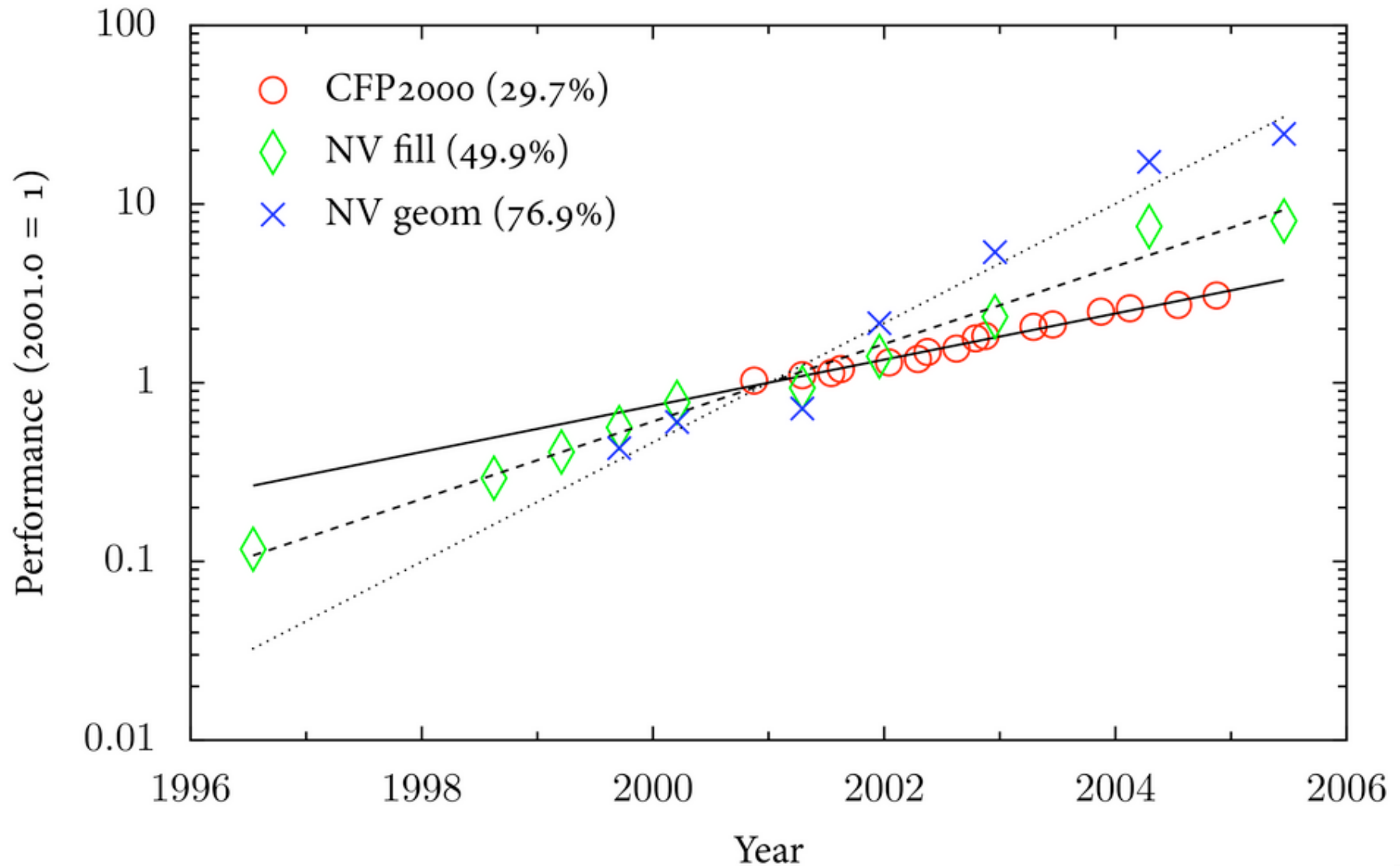
GP GPU

Off-line to on-line to real-time ...

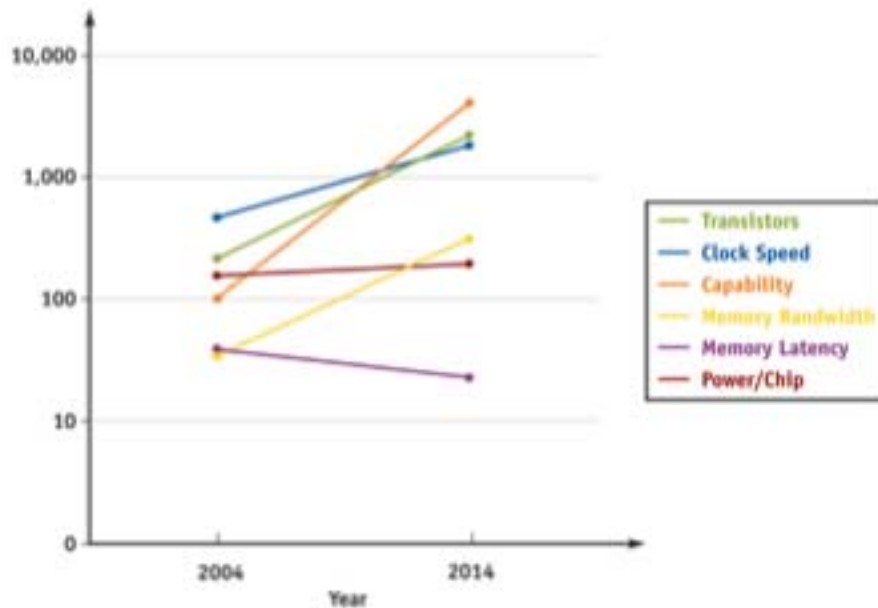


[Courtesy of Crow/Hanrahan/Akeley]

Motivation: Computational Power



10-Year Projection (2004-2014)



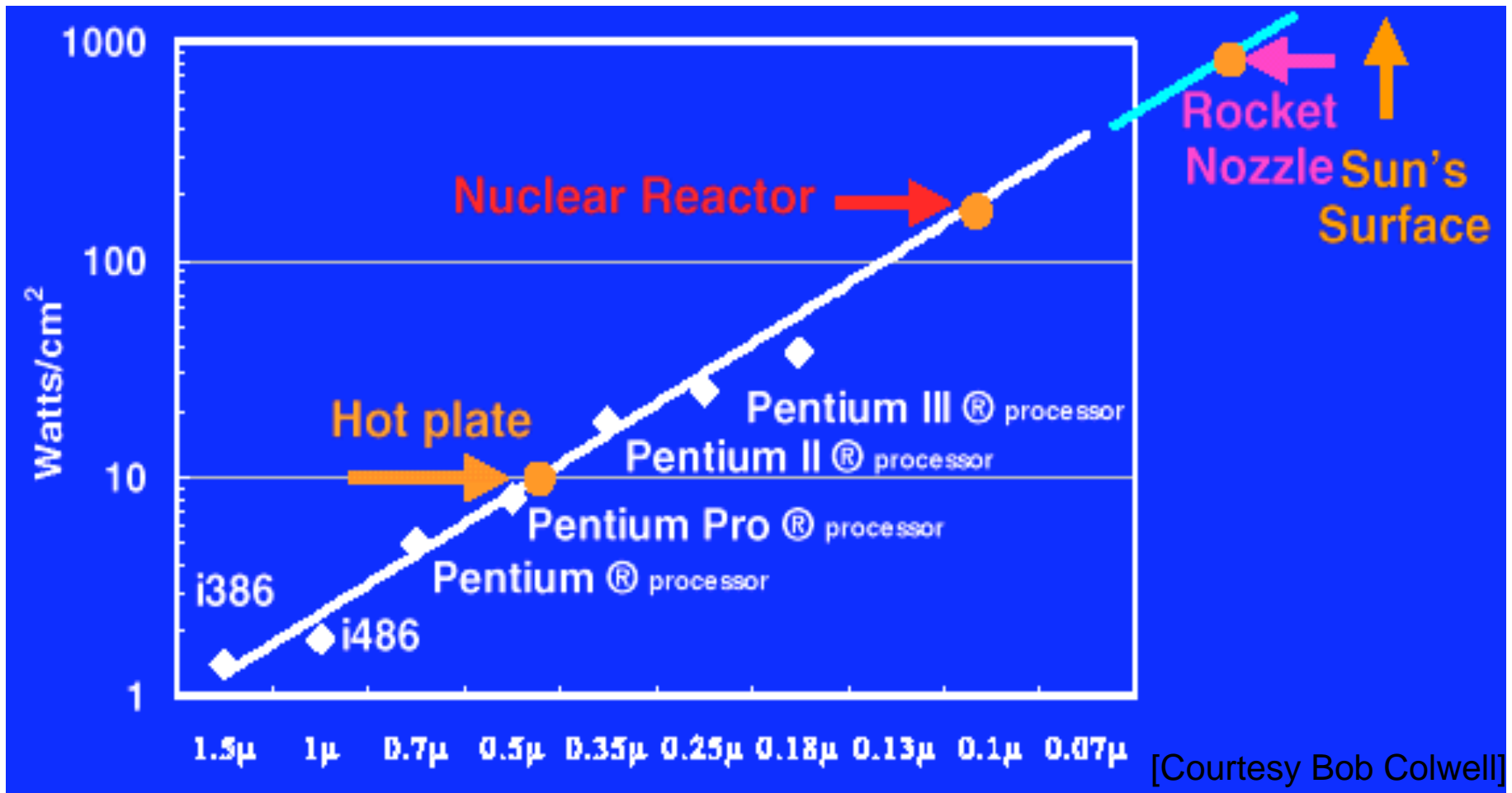
- Transistors (NV40): 222M/2237M
- Clock speed (NV40, MHz): 475/1890
- Capability: 105B/4228B
- Memory bandwidth (NV40, GB/s): 35/322
- Memory latency (RAS, ns): 40/23
- Power/chip (maximum, W): 158/198

- Take-home point:
Capability > mem bw > mem latency

Hardware Considerations

- **“Memory wall”**
 - Caching and recomputation vs. communication
 - Continued migration of functionality onto GPU
 - Higher-level graphics functionality
 - Physics & simulation
 - GPGPU?
- **Size of design teams**
 - Intel design teams increase in size 40% / generation
 - Validation for increasingly complex designs
- **Power ...**

Power Considerations



Architecture/Microarchitecture

- **Current programming model:**
 - MIMD or SIMD for vertex processing?
 - SIMD for fragment processing?
- **Can we share units between the stages?**
- **To what will the instruction sets converge?**
- **Are these the only stages that will be programmable?**
- **How will the CPU interact with the GPU?**
- **How can we extend to multiple GPUs and multiple CPUs?**

What *should* we map to GPUs?

- Problems with *high compute requirements*
- Problems with *regular structure*
- Problems with *predictable communication needs*
- Problems that require *interaction with the graphics system*

- Other domains: biology, statistics, chemistry, finance ...

- Enormous opportunity at frontiers of applications, software, and hardware!

GPGPU Top Ten

- The Killer App
- Programming models and tools
- GPU in tomorrow's computer?
- Data conditionals
- Relationship to other parallel hw/sw
- Managing rapid change in hw/sw (roadmaps)
- Performance evaluation and cliffs
- Philosophy of faults and lack of precision
- Broader toolbox for computation / data structures
- Wedding graphics and GPGPU techniques

For more information ...

- **Course web page:** <http://www.gpgpu.org/vis2005/>
- **GPGPU home:** <http://www.gpgpu.org/>
 - Mark Harris, UNC/NVIDIA
 - Research, forums, developer tools, ...
- ***GPU Gems* (Addison-Wesley)**
 - Vol 1: 2004; Vol 2: 2005
- **Recent survey paper (Eurographics 2005 STAR, Owens/Luebke/Govindaraju/Harris/Krüger/Lefohn/Purcell)**
 - http://graphics.idav.ucdavis.edu/publications/print_pub?pub_id=844
- **Conferences: Siggraph, Graphics Hardware, GP²**
 - Course notes: Siggraph '05, IEEE Visualization '05

