

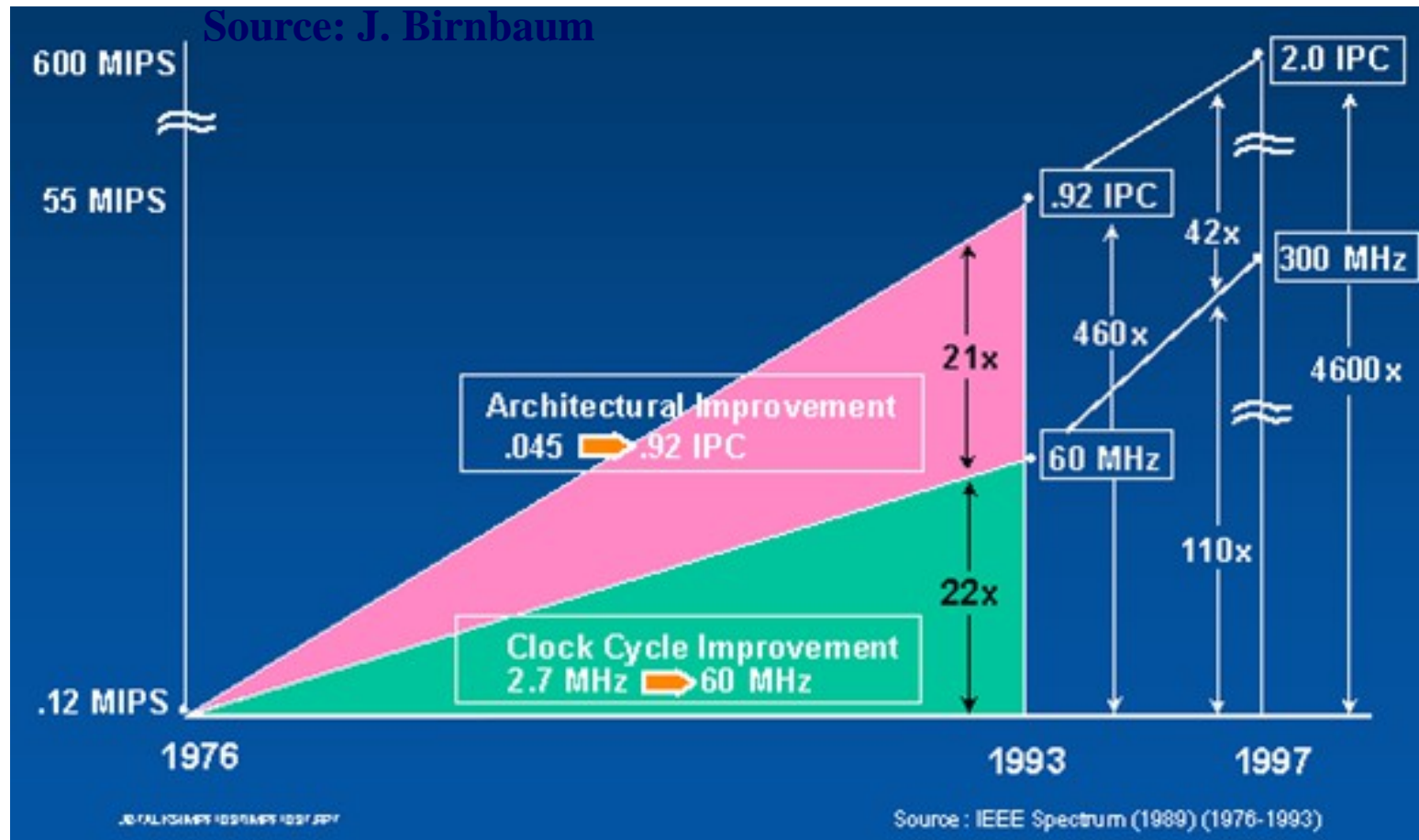
GPU Programming

Rupesh Nasre.

<http://www.cse.iitm.ac.in/~rupesh>

IIT Madras
July 2017

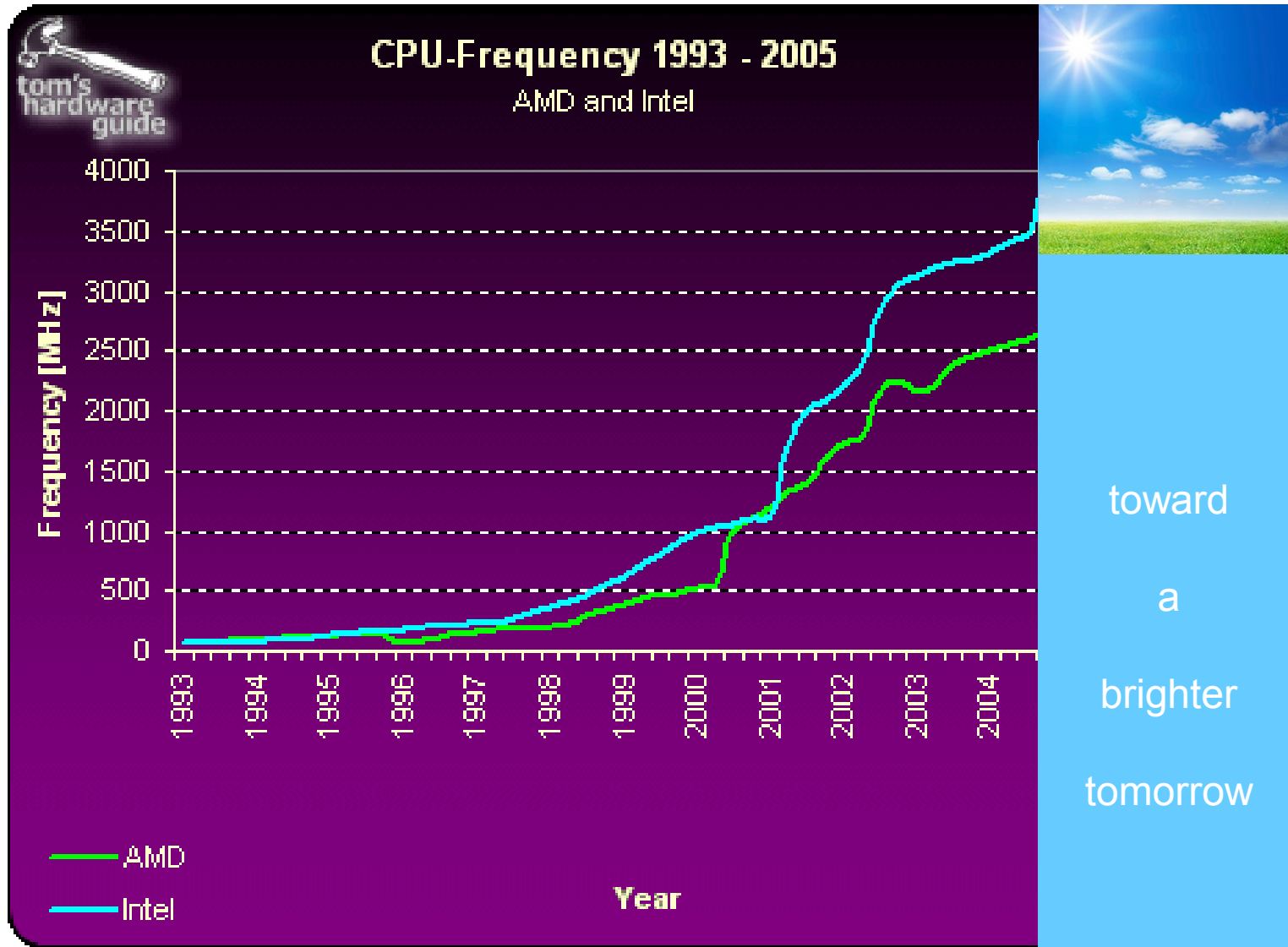
The Good Old Days for Software



Single-processor performance experienced dramatic improvements from clock, and architectural improvement (Pipelining, Instruction-Level-Parallelism)

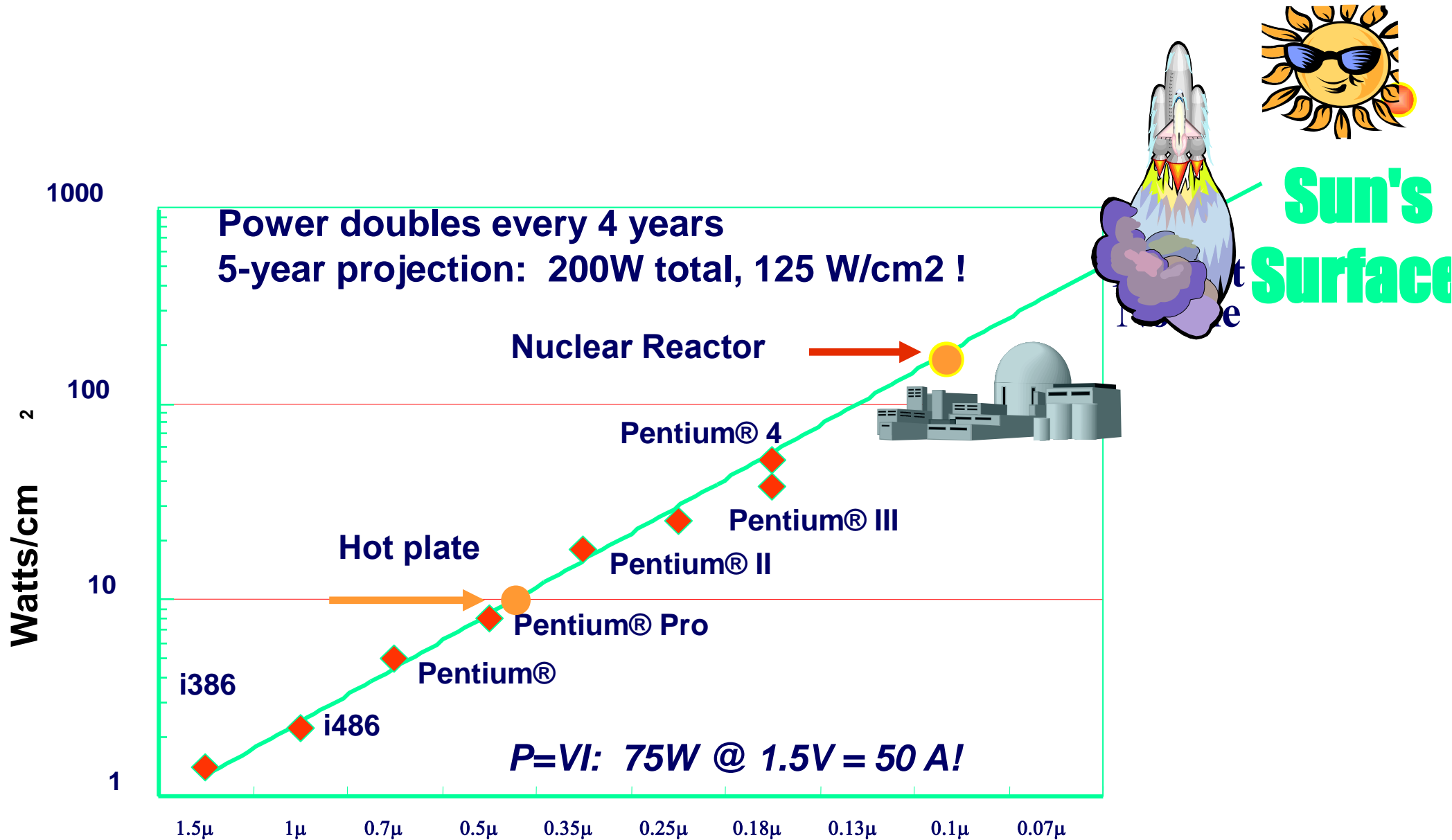
Applications experienced automatic performance improvement

Hitting the Power Wall



http://img.tomshardware.com/us/2005/11/21/the_mother_of_all_cpu_charts_2005/cpu_frequency.gif

Hitting the Power Wall



“New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies” – Fred Pollack, Intel Corp. Micro32 conference key note - 1999.
 Courtesy Avi Mendelson, Intel.

The Only Option: Use Many Cores

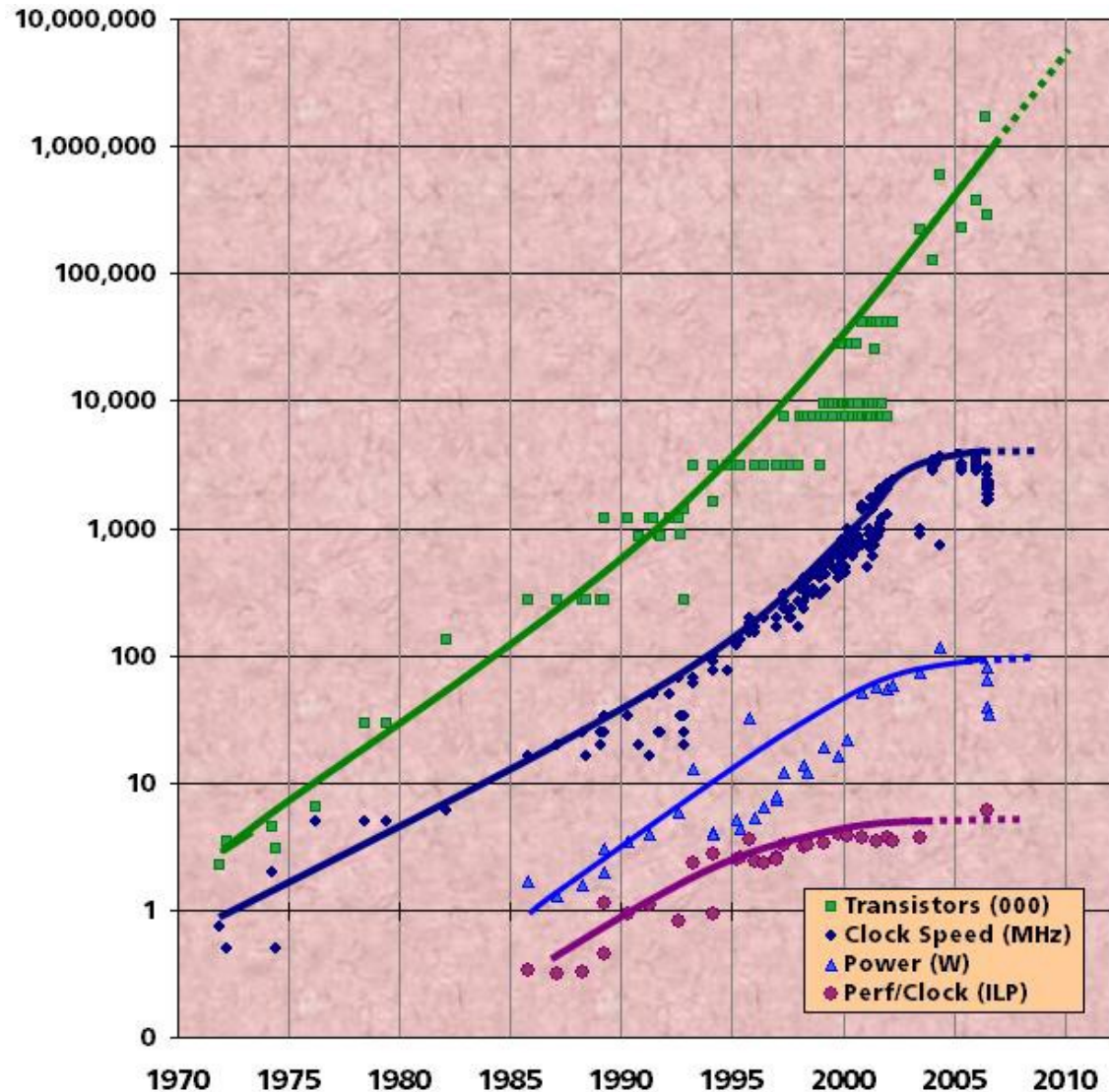
Chip density is continuing increase ~2x every 2 years

- Clock speed is not
- Number of processor cores may double

There is little or no more hidden parallelism (ILP) to be found

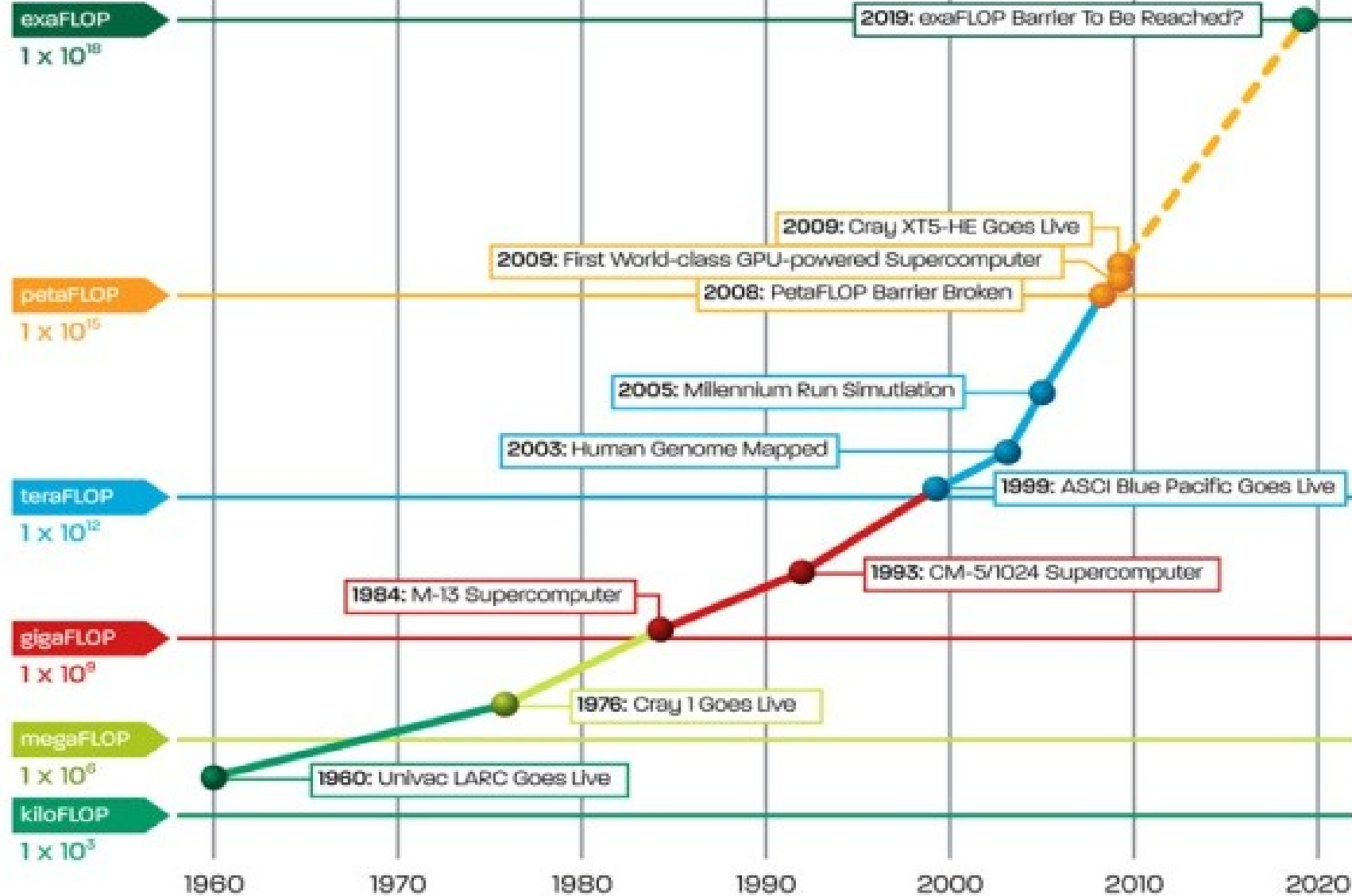
Parallelism must be exposed to and managed by software

Source: Intel, Microsoft (Sutter) and Stanford (Olukotun, Hammond)



High-Performance Computing Milestones (1960–2019)

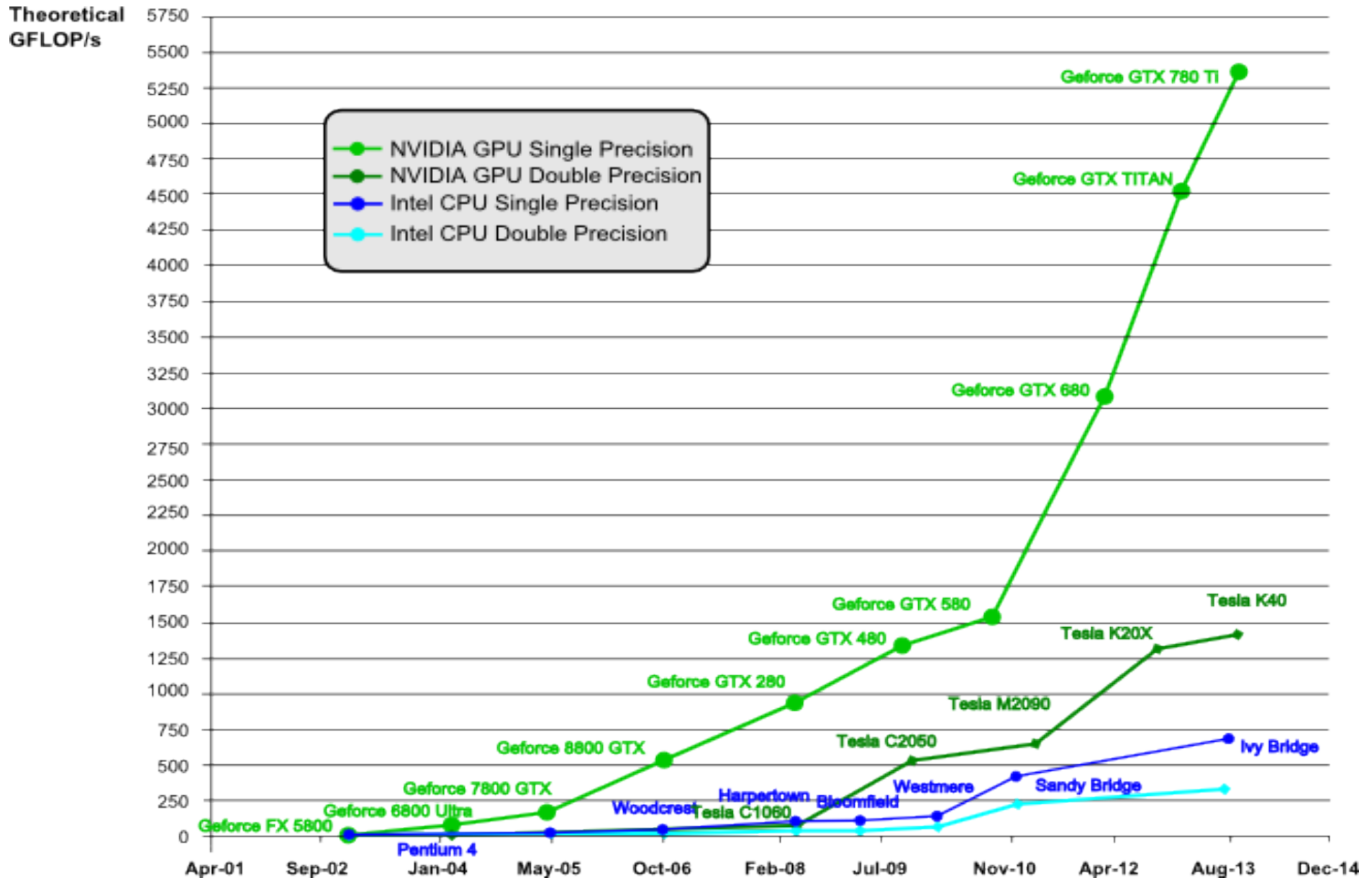
Floating point operations per second



Parallel Platforms

- Shared memory systems (multi-core)
- Distributed systems (cluster)
- Graphics Processing Units (many-core)
- Field-Programmable Gate Arrays (configurable after manufacturing)
- Application-Specific Integrated Circuits

GPU-CPU Performance Comparison



In this course...

- Basic GPU Programming
 - Computation, Memory, Synchronization, Debugging
- Advanced GPU Programming
 - Streams, Heterogeneous computing, Case studies
- Topics in GPU Programming
 - Unified virtual memory, multi-GPU, peer access

Logistics

- TAs
 - Somesh Singh, Jyothi Vedurada, Shouvick Mondal
- Evaluation
 - Five assignments (7 + 7 + 13 + 13 + **20**)
 - MidSem (20) + EndSem (20)
 - Absolute grading (95, 80, 70, ..., 40)
 - You have this week to suggest changes to dates.
- Attendance ($57 * 0.85 = 48.85$)
 - I don't shy to give W grades.
- Moodle
 - Your responsibility to subscribe to it.

Logistics

- Tutorials and lectures would be intermixed.
 - But we will have separate doubts / practice sessions.
- You need a login on GNR Cluster in CC.
 - You can run assignments on your laptop, but make sure they run on the GNR Cluster.
 - GNR Cluster would be slow.