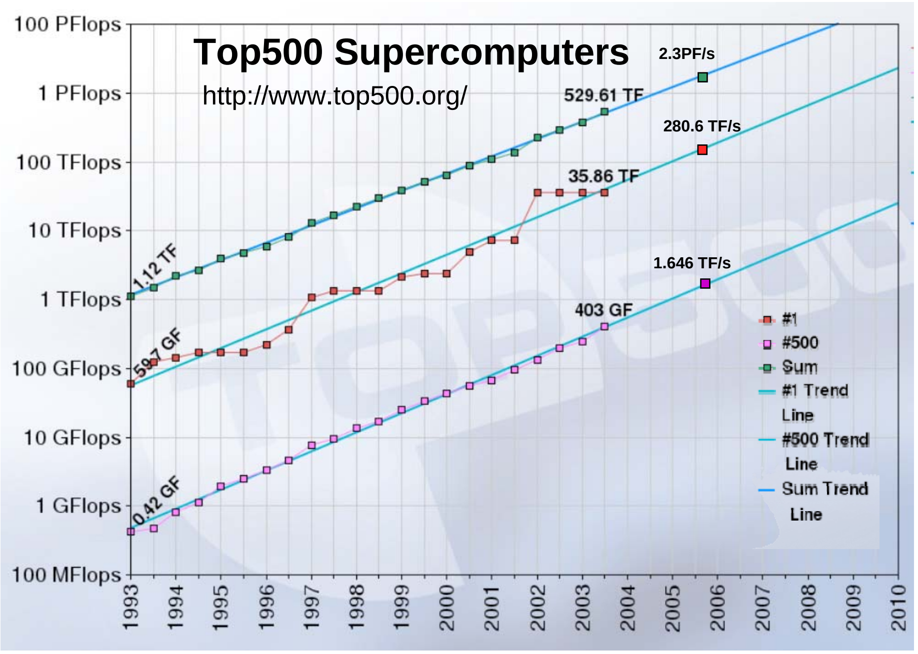
Systems & Technology Group

To PetaFLOPs …. and Beyond

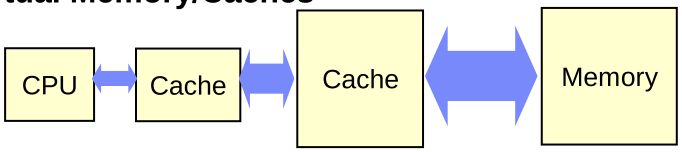
....or, How I Learned to Stop Paying So Much for FLOPS, and Started Having More Fun

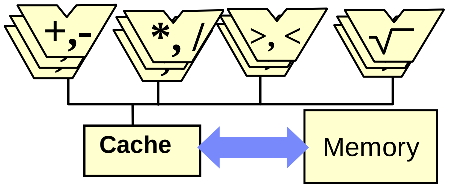
April 2006

"Moore" MegaFLOPS, GigaFLOPS, TeraFLOPS, and …… ***a Lot of Extra Heat !***

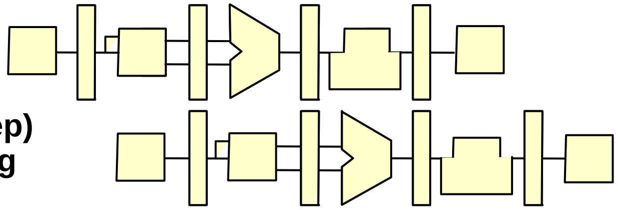


- **Virtual Memory/Caches**

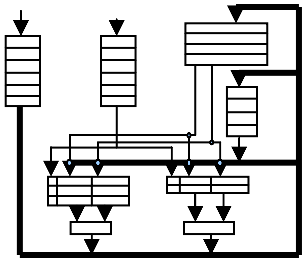




- **Superscalar Micro-architectures**



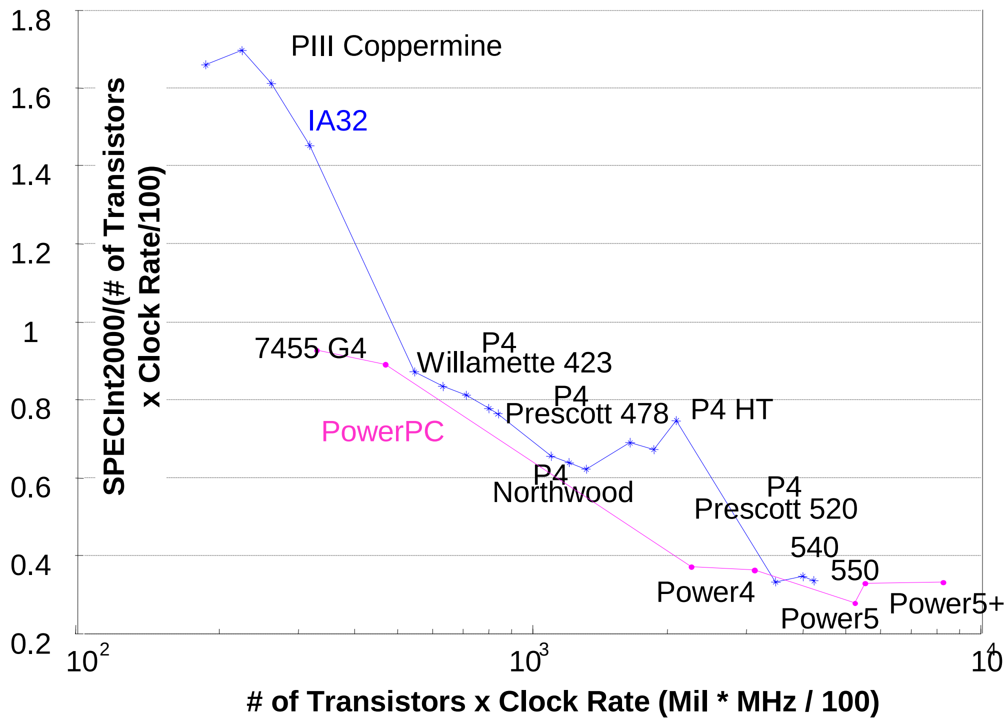
- **Hyper (deep) pipelining**



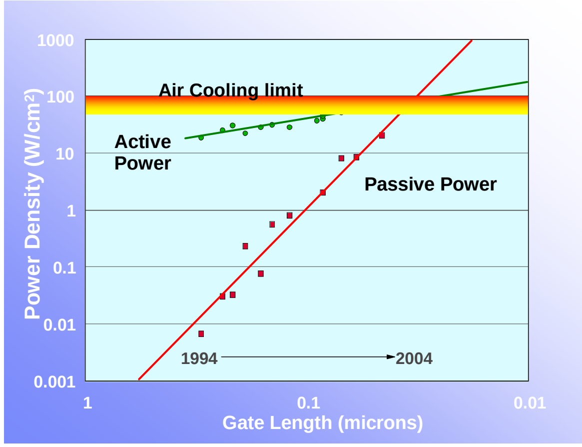
- **Speculative and Out-of-Order Processing**

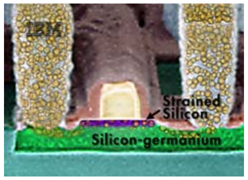
Large On-Chip Cache & Memory and Architectural Improvements Improve Processor Performance

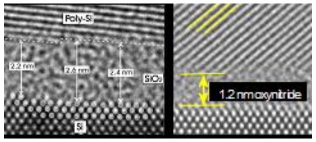
- ***Total Chip Power & Device Leakage Have Increased, Efficiency Has Fallen***
- ***Single Thread Processor Performance Improvements Are Reaching a Limit of Diminishing Returns…***



Data Source:
http://www.spec.org/cpu2000/results/cint2000.html
http://www.geek.com/procspec/procspec.htm
http://www.bayarea.net/~kins/AboutMe/CPUs.html



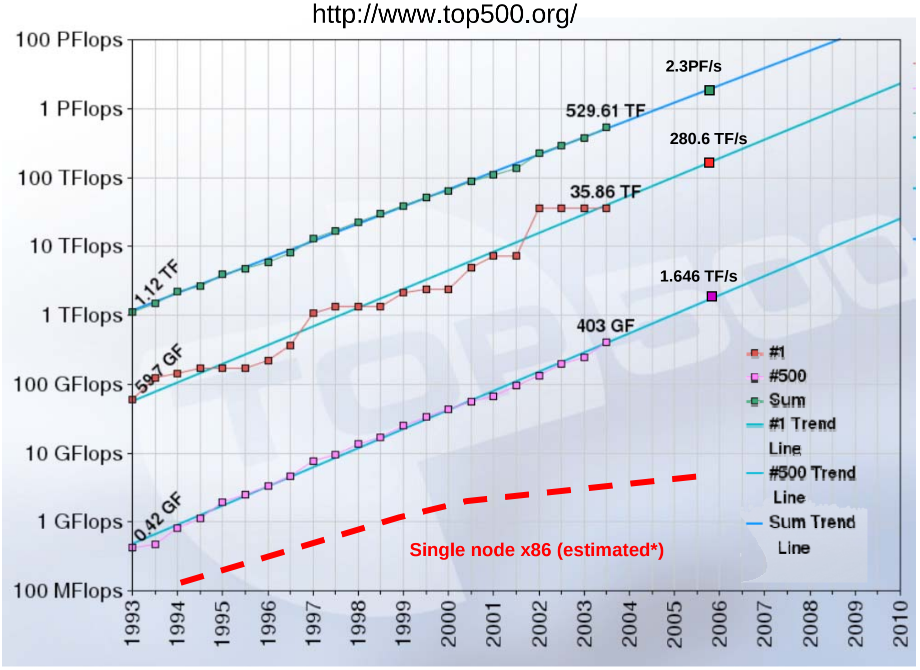




Will Slowing Single Thread Processor Performance Gains and Increasing Power Requirements Impact Future HPC Systems ?

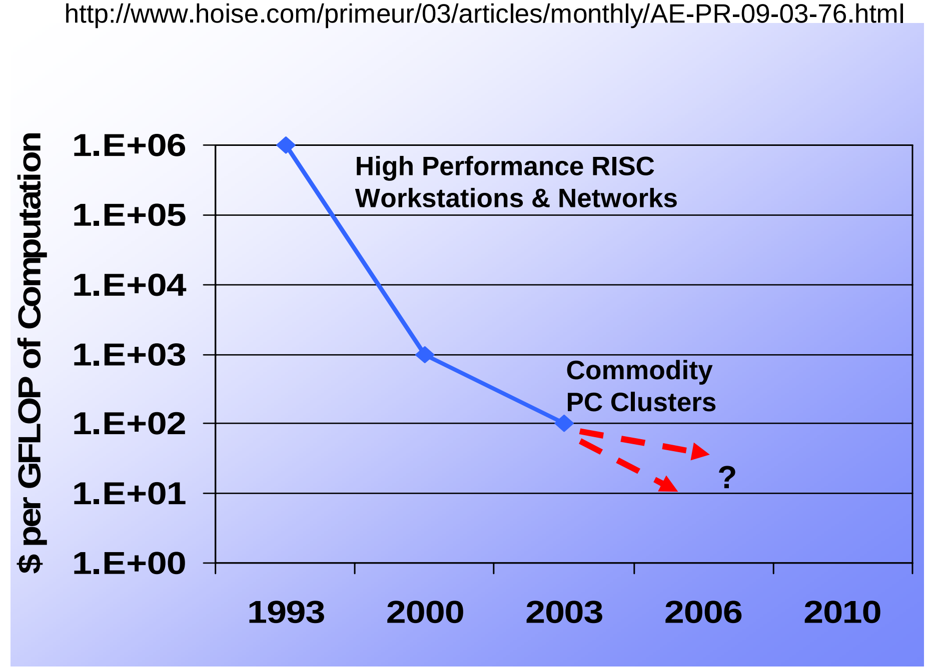
Top500 Supercomputers

http://www.top500.org/



Declining Cost of Supercomputing

http://www.hoise.com/primeur/03/articles/monthly/AE-PR-09-03-76.html



Can We Build a Single Processor Capable of 100GFLOPs ?

How About 10 $US / GFLOP, or Less?

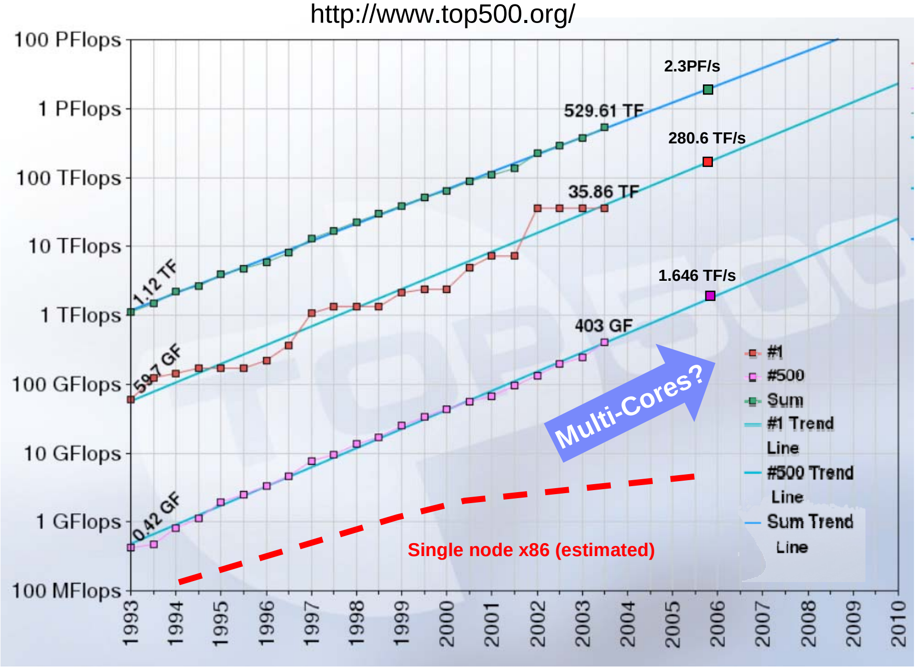
*** Based on 50% of processor peak GFLOPS**

Application-Optimized Designs:

Matching Application Requirements & Architecture to Maximize Performance, Reduce Power

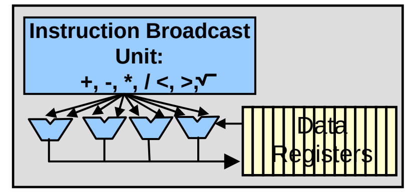
Top500 Supercomputers

http://www.top500.org/

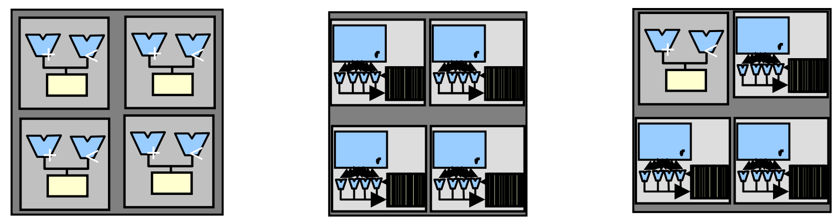


Multi-Core Processors Leading the Way ?

- **Vector/SIMD (data parallel)**



- **Multi-Core Processors**



Multiple, Simplified Superscalar CPUs

SIMD/Vector Processing Units (e.g. GPUs)

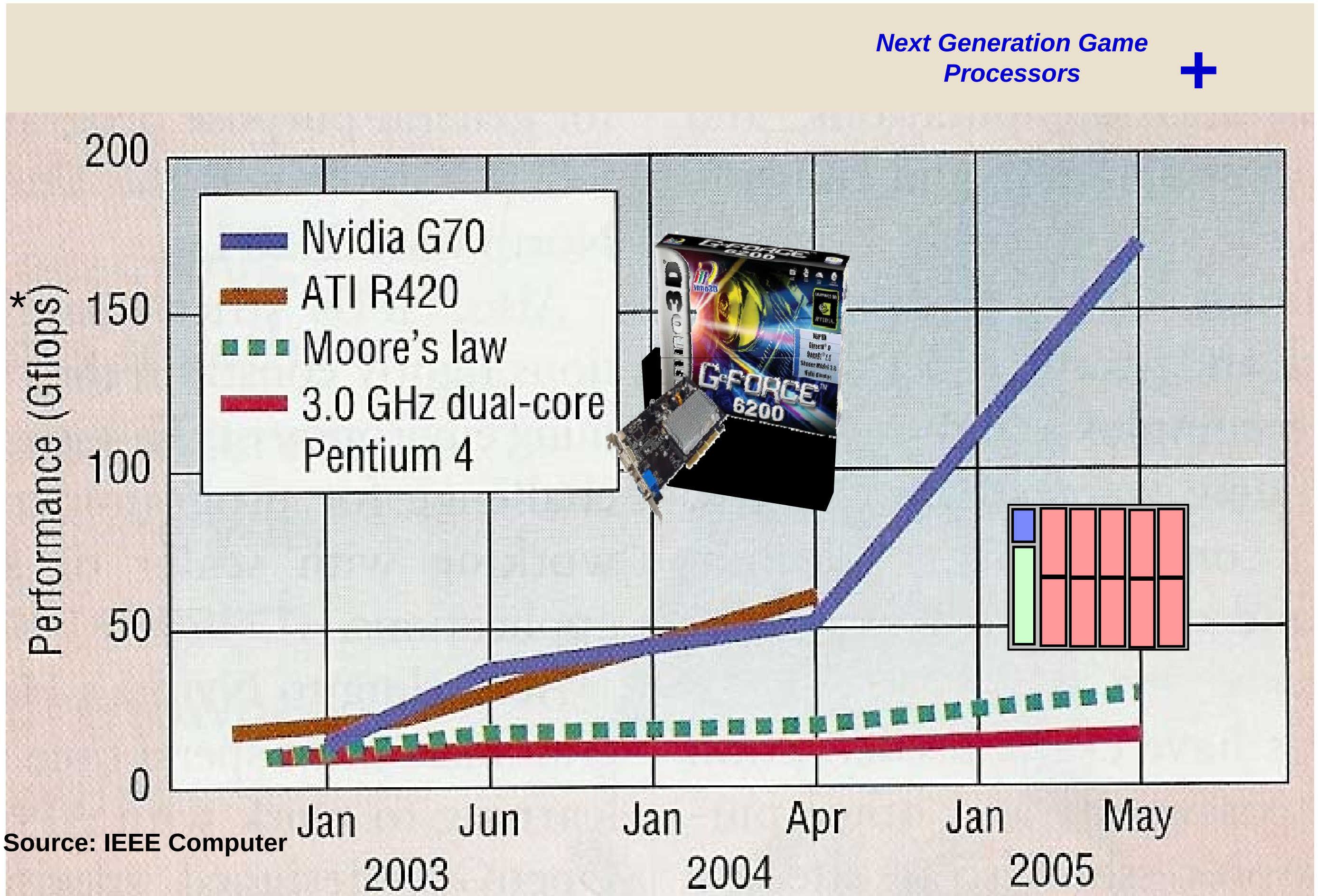
Hybrid Combinations (e.g. Cell BE)

- **New Memory Organization:**

  (increased explicit, programmer control)

Attack of the Computer Game Consoles !!!!

- ***GPUs & Game Processor Architectures Are an Excellent Match for Game Applications***
- ***Performance Has Been Growing <u>Faster</u> Than Moore's Law !***



Source: IEEE Computer

*Single Precision

- Typical Street Price for High End PC Graphics Card: ***300 – 400 US$***

Gen'l Purpose Core	Special Function Accelerator	Cache

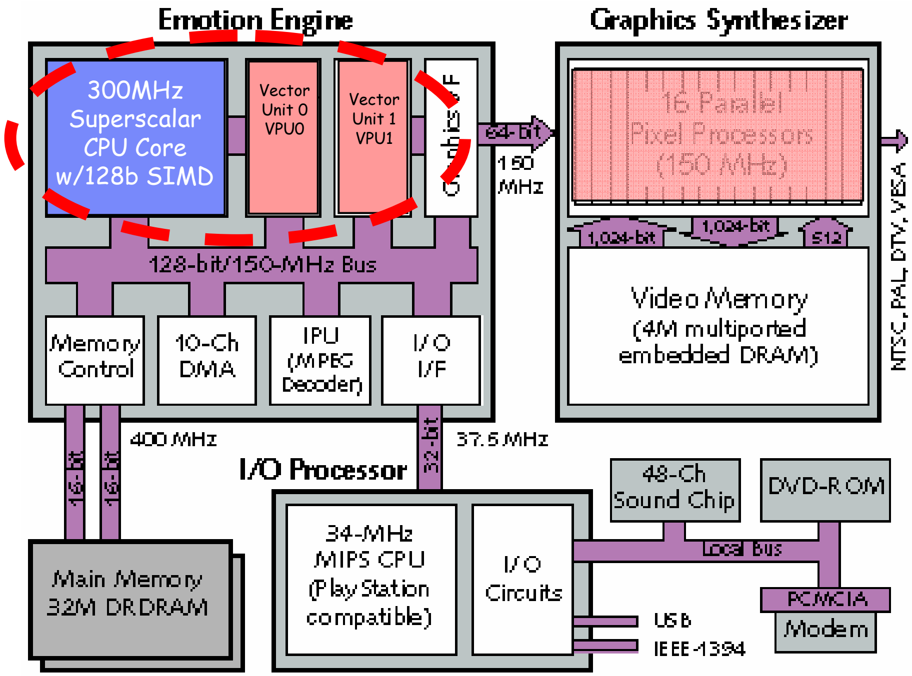
Processor & System Architecture , Implementation "Tuned" to the End User Application

Emotion Engine	Description
Frequency	300 MHz
CPU Core:	MIPS III, MIPS IV subset + 128b SIMD
Registers	32 × 128-bit
Microarchitecture	2-issue, two 64-bit integer units, 1 FPU
CPU Pipeline	6 stages
Instruction Cache	16K, two-way set-associative
Data Cache	8K, two-way set-associative
Scratchpad RAM	16K
TLBs	48-entry combined instruction/data TLB
Vector Unit 0:	4 FMACs, 1 FDIV
Memory	4K instruction, 4K data
Vector Unit 1:	5 FMACs, 2 FDIV
Memory	16K instruction, 16K data
Image Processing Unit	MPEG-2 macroblock decoder
DMA	10 channels
On-Chip Bus Bandwidth	2.4 **GB/s** peak, 2.0 **GB/s** effective
Main Memory:	32M, two DRDRAM channels
Bandwidth	3.2 GBytes/s peak
Performance:	
Floating-Point Peak	6.2 GFLOPS *
Perspective Transform	66 Mpolygons/s
With Lighting & Fog	36 Mpolygons/s
Bezier Surface Patches	16 Mpolygons/s
Image Decompression	150 Mpixels/s
Process:	0.25 µm (0.18 µm L_{e}), 4-layer-metal
Size	240 mm², 10.5 million transistors
Power	15 W at 1.8 V
Package	540-contact PBGA

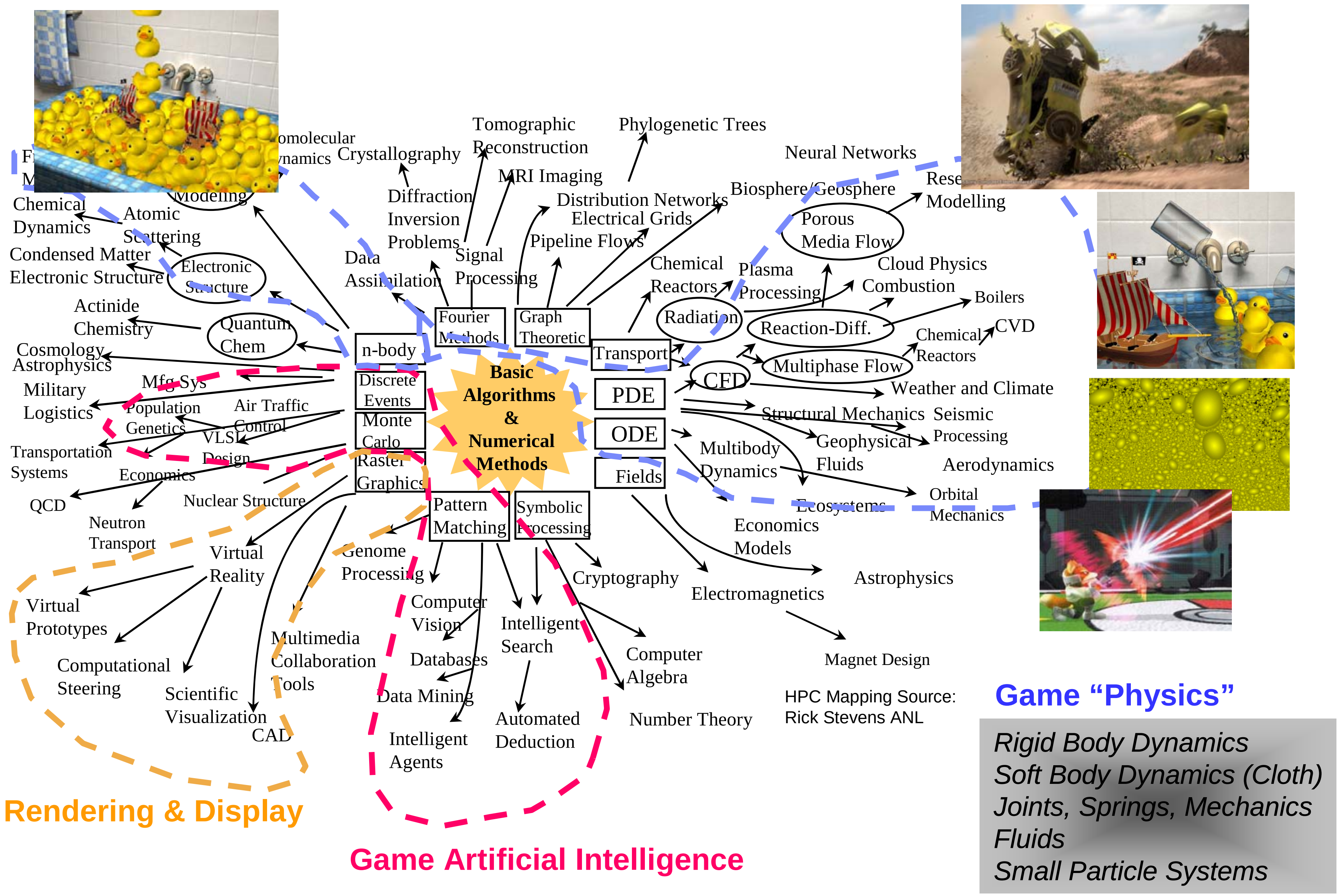
•Single Precision

~ $21 / GFLOP
0.124 GFLOPS/Watt

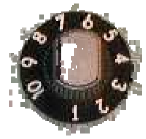
42 GFLOPS/ Cu. Ft.
1.16 GFLOPS / LB.



- **Low Cost, Simple**
- **Lots of FLOPS**
- **Small Memory Capacity, but High BW for Game Graphics**
- **Not Really "Scalable", & tough to program ☺**

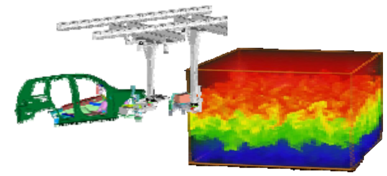

Basic Algorithms & Numerical Methods
n-body
Discrete Events
Monte Carlo
Raster Graphics
Fourier Methods
Graph Theoretic
Pattern Matching
Symbolic Processing
Transport
PDE
ODE
Fields
Quantum Chem
Electronic Structure
Radiation
Reaction-Diff.
CFD
Multiphase Flow
Porous Media Flow
Chemical Dynamics
Atomic Scattering
Condensed Matter Electronic Structure
Actinide Chemistry
Cosmology Astrophysics
Military Logistics
Mfg Sys
Population Genetics
Air Traffic Control
VLSI Design
Transportation Systems
Economics
QCD
Neutron Transport
Nuclear Structure
Virtual Reality
Virtual Prototypes
Computational Steering
Scientific Visualization
CAD
Multimedia Collaboration Tools
Genome Processing
Computer Vision
Databases
Data Mining
Intelligent Agents
Intelligent Search
Automated Deduction
Cryptography
Computer Algebra
Number Theory
Electromagnetics
Magnet Design
Astrophysics
Economics Models
Ecosystems
Orbital Mechanics
Multibody Dynamics
Geophysical Fluids
Aerodynamics
Structural Mechanics
Seismic Processing
Weather and Climate
Chemical Reactors
CVD
Boilers
Cloud Physics
Combustion
Plasma Processing
Chemical Reactors
Biosphere/Geosphere
Neural Networks
Modelling
Distribution Networks
Electrical Grids
Pipeline Flows
Phylogenetic Trees
Tomographic Reconstruction
MRI Imaging
Crystallography
Diffraction Inversion Problems
Signal Processing
Data Assimilation
Rendering & Display
Game Artificial Intelligence
Game "Physics"
Rigid Body Dynamics
Soft Body Dynamics (Cloth)
Joints, Springs, Mechanics
Fluids
Small Particle Systems
HPC Mapping Source:
Rick Stevens ANL

Where Would You Place the Characteristics of Video Games and "Real" World Simulation Physics on a Scale of 1 --- 10 ? Today ? …... Tomorrow?



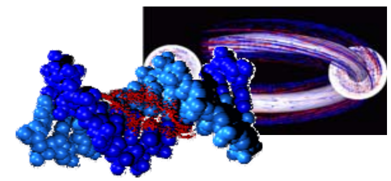
Fidelity

accuracy & faithful representation of the world



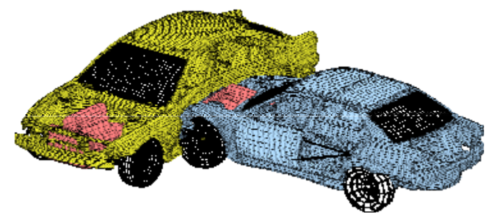
Scale

Number of physical objects simulated
or level of detail of the effect



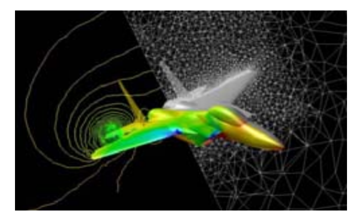
Interaction & Responsiveness

pervasive interactivity in which every object is capable
of acting or responding to every other object thru
defined physical relationships



Sophistication

types of physical effects and objects simulated



Consumer price point, human perception & comprehension demands for game machines have impacted their system balance, total compute, memory BW and capacity, storage interconnect capabilities versus large HPC machines….*but…*

Human Perception & Comprehension

"You Can't Fool Mother Nature…."

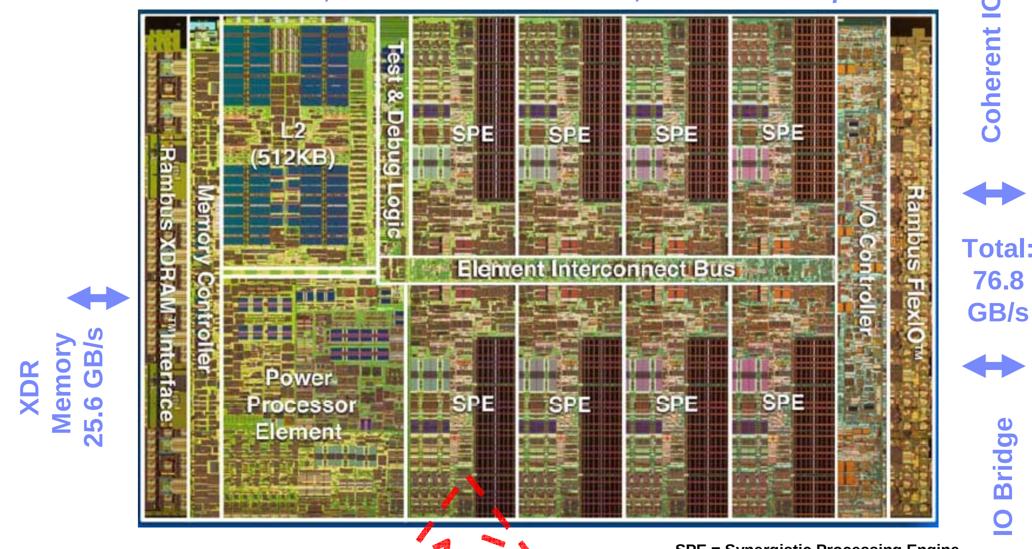
Supercomputing Capability for High Volume, Consumer Systems

- Multi-core, multi-thread , "cluster-on-a-chip"
  - **64bit PowerPC Control Processor**
  - **+ 8 Tightly integrated *accelerators* (SPE)**
    - **128 bit SIMD/Vector, MAC**
    - **256KB Embedded Memory**
  - **Integrated I/O and memory interfaces**

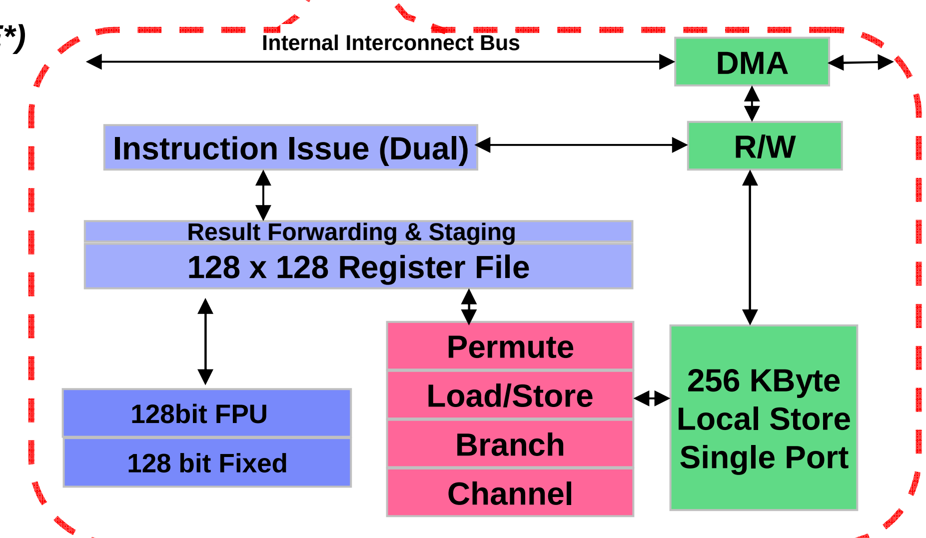
- *High* Performance
  - **3.2 GHz clock frequency**
  - **205 GFLOPs/s peak, single precision**
    (dual issue, in-order execution, 25.6 GFLOPS per SPE)
  - **~20 GFLOPs/S peak, double precision**
    (2 DP instructions every 7 cycles, 1.83 GLOPS per SPE*)
  - **205 GB/s internal interconnect bandwidth**
  - **~100 GB/s BW for memory, external IO**

- Linux OS
  - **Simultaneous multiple OS support**
  - **+ *Real-time* support**

221 mm^2 , 234 million Transistors, 90 nm SOI chip



SPE = Synergistic Processing Engine



Cell BE Performance Characteristics

Performance comparison of Cell BE versus other processors for different applications

Type	Algorithm	3.2 GHz GPP	3.2 GHz Cell BE	Perf Advantage
HPC	Matrix Multiplication (S.P.)	25.6 Gflops* (w/SIMD)	200 GFlops (8SPEs)	8x (8SPEs)
	Linpack (S.P.) 4kx4k	25.6 GFlops* (w/SIMD)	156 GFlops (8SPEs)	6x (8SPEs)
	Linpack (D.P.) 1kx1k	7.2 GFlops (3.6GHz IA32/SSE3)	9.67 GFLops (8SPEs)	1.3x (8SPEs)
graphics	TRE	.85 fps (2.7GHz G5/VMX)	30 fps (Cell BE)	35x (Cell BE)
	transform-light	128 MVPS (2.7GHz G5/VMX)	217 MVPS (one SPE)	1.7x (one SPE)
security	AES ECB encryp. 128b key	1.03 Gbps	2.06Gbps (one SPE)	2x (one SPE)
	AES ECB decryp. 128b key	1.04 Gbps	1.5Gbps (one SPE)	1.4x (one SPE)
	TDES ECB encryp.	0.13 Gbps	0.17 Gbps (one SPE)	1.3x (one SPE)
	DES ECB encryp.	0.43 Gbps	0.49 Gbps (one SPE)	1.1x (one SPE)
	SHA-1	0.9 Gbps	2.12 Gbps (one SPE)	2.3x (one SPE)
video processing	mpeg2 decoder (sdtv)	354 fps (w/SIMD)	329 fps (one SPE)	0.9x (one SPE)

* assuming 100% compute efficiency, achieving theoretical peak of 25.6GLOPS, in its single precision MatrixMultiply & Linpack implementation

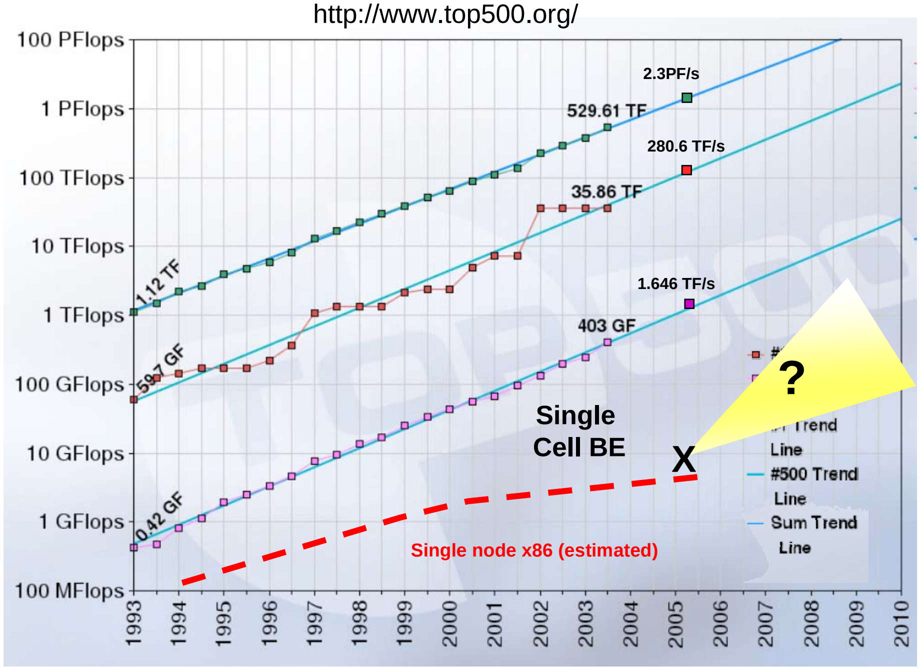
http://www-128.ibm.com/developerworks/power/library/pa-cellperf/

16 Million Point FFT	2 GHz GPP	1.65 GHz Power5	3.2 GHz CellBE
GFLOPS	1.2	1.55	46.8

http://www-3.ibm.com/chips/techlib/techlib.nsf/techdocs/0AA2394A505EF0FB872570AB005BF0F1/$file/GSPx_FFT_paper_legal_0115.pdf

Top500 Supercomputers

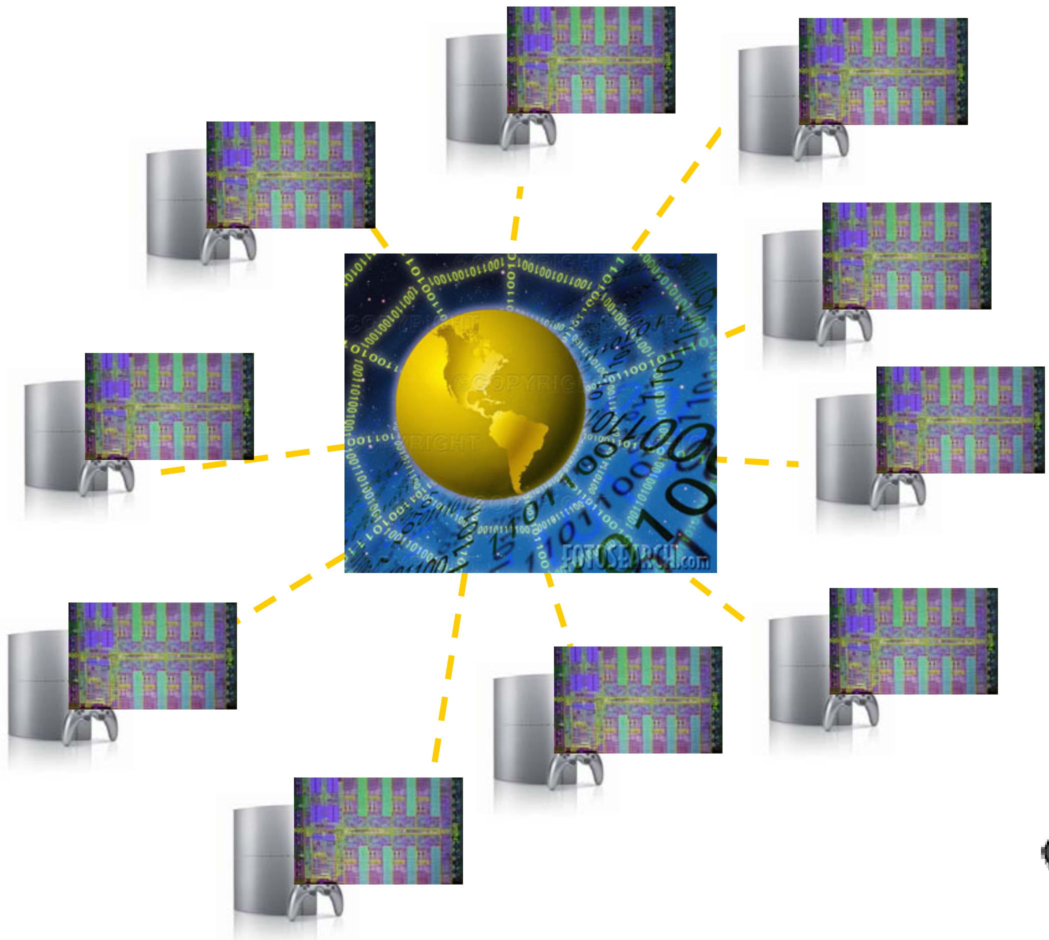
http://www.top500.org/



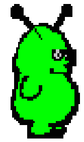
Can game system technologies play a bigger role in more general purpose computing in the future? What about $ / GFLOP ?

What If ~ 10M Cell BE Game Consoles Were Connected to the Internet .....
2 E 17 FLOPS/s – Double Precision, or 2 E 19 FLOPS/s - Single Precision !!!!!

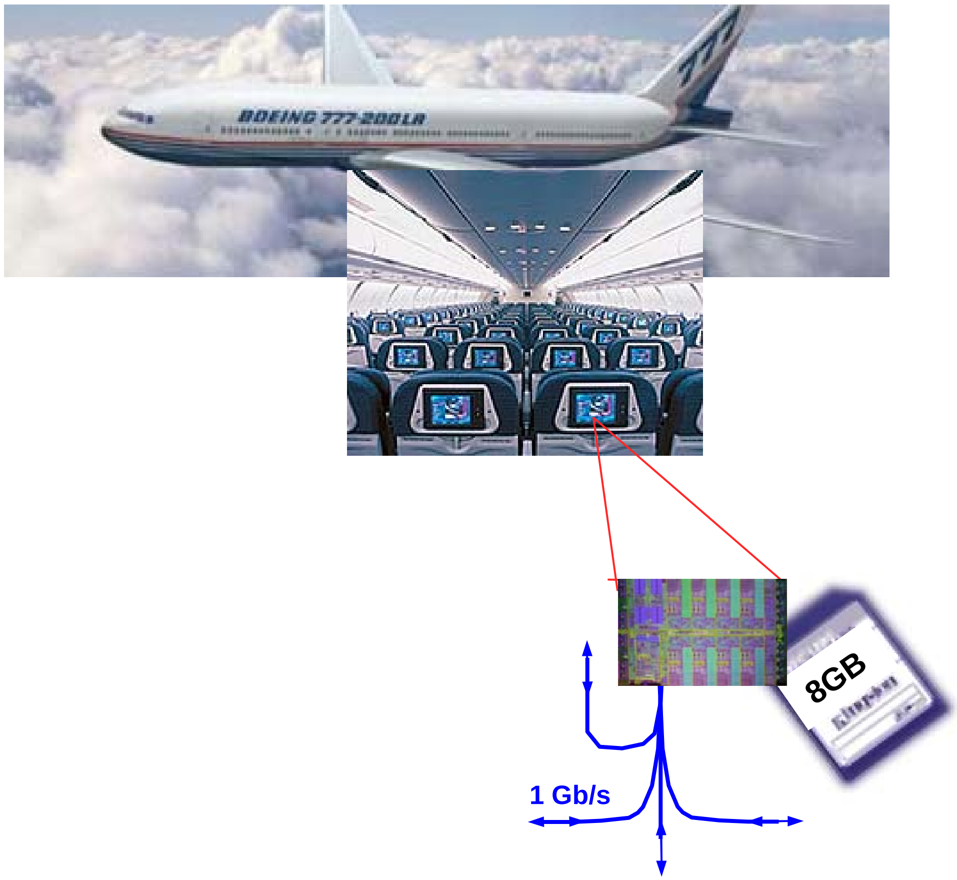
Wow !



PlayStation 2:
6.2 FLOPS (SP) & 129 $US
➔ $21 /GFLOP



In-Flight service for the Computationally Intense Frequent Traveler



300 Seat Wide Body Aircraft

Total Compute Power:
300 * 200GF = 60 TFLOPS/s (SP)

Memory Capacity:
300 * 8GB = 2.4 TBytes

Mesh/Grid Interconnect

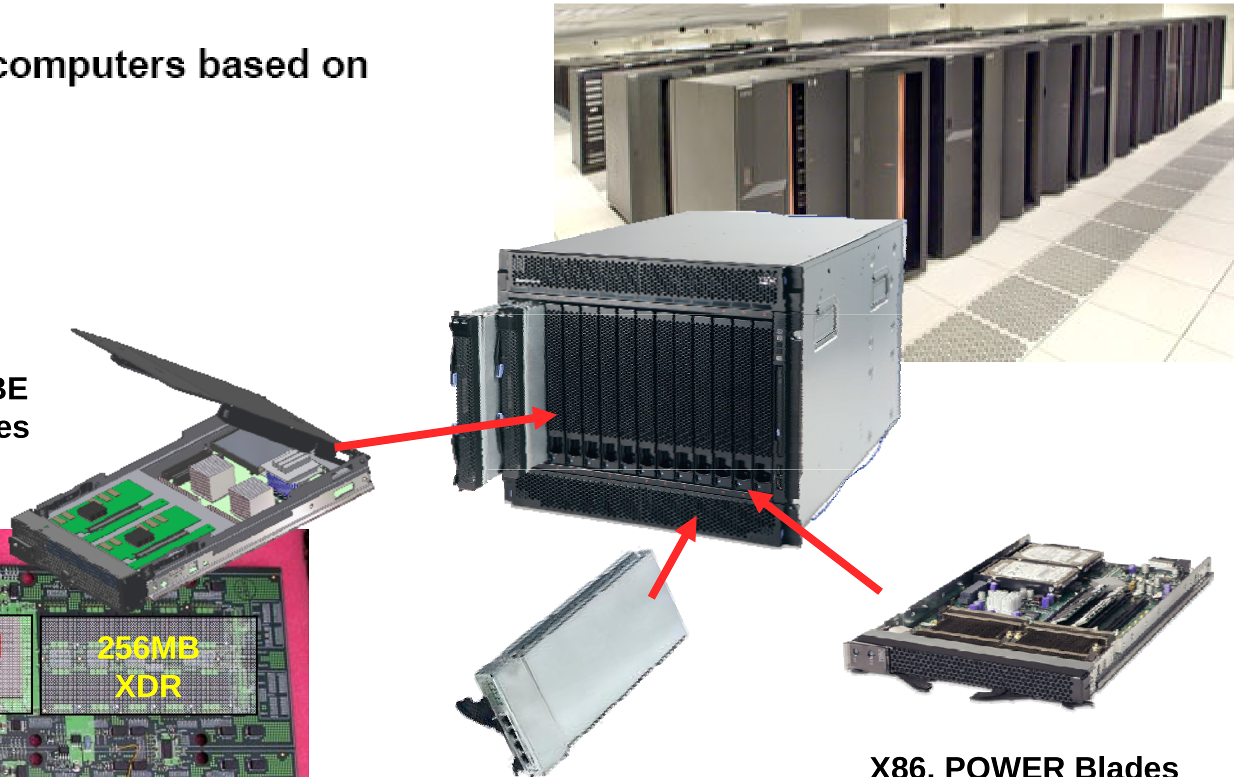
Hundreds of current DVD movies, music titles, games …. and computing to spare ☺

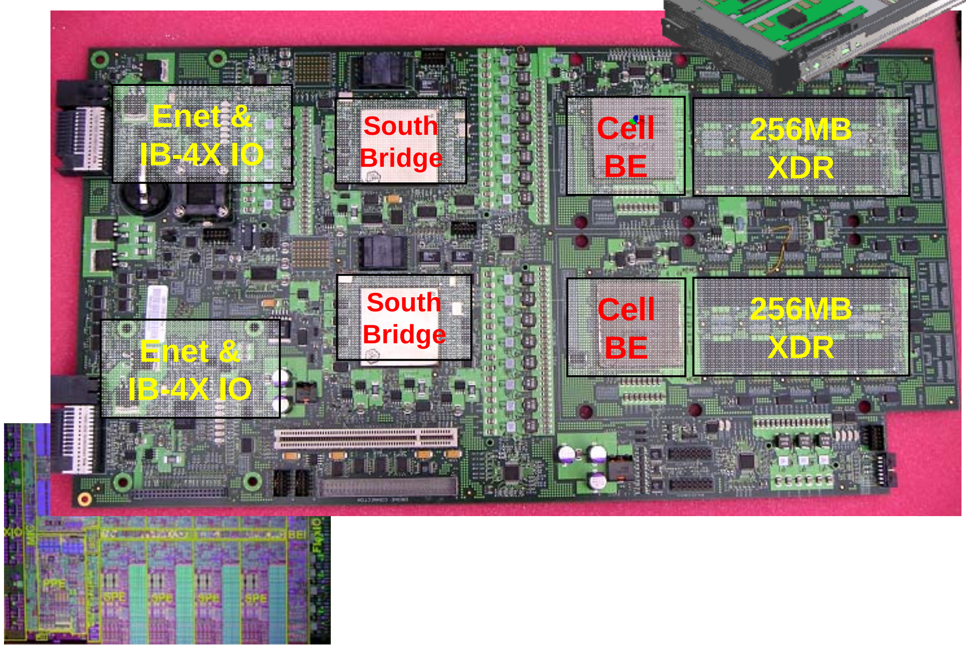
Update: IBM launches business computers based on game chip

News Story by Patrick Thibodeau

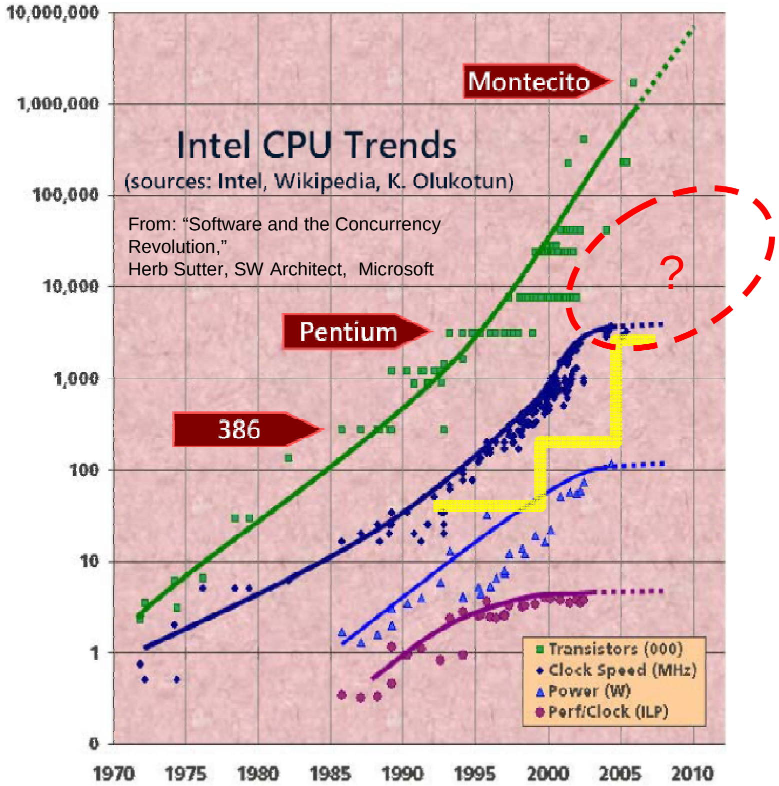
FEBRUARY 08, 2006

(COMPUTERWORLD) - NEW YORK



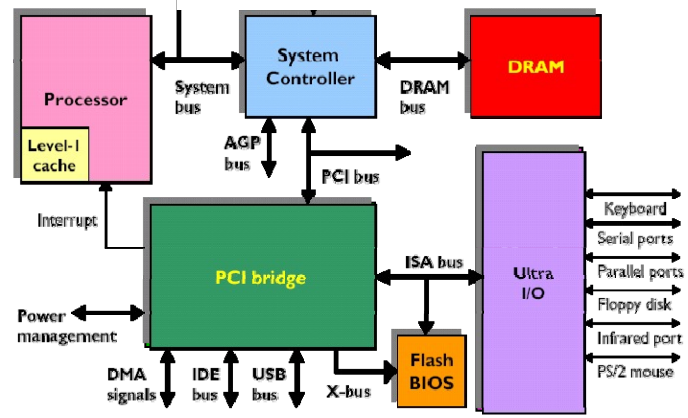


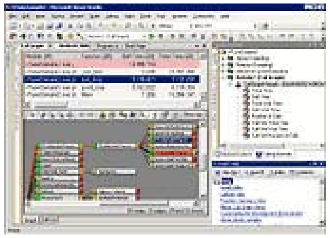
Challenges for Game Processors in the “Real” World

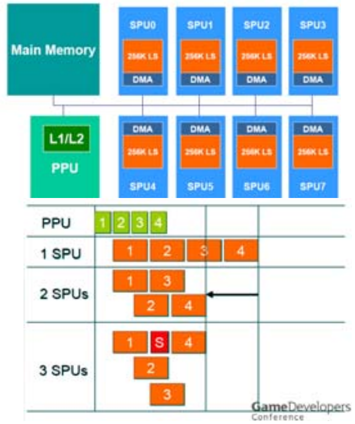


Robust Surrounding Hardware “Ecosystem”

-Memory BW & Capacity, IO, Interconnect …







Game Processor’s Technology
“Time Constant” Has Been
Much Longer than General Purpose PCs

- Drive Cost Down vs Steady Ramp Up of Clock Speed
- Is a ”Convergence” of the Two
  Worlds Approaching with Multi-Cores?

Shifting the Ratio SP to DP in Game Processors
24 x 7 x “Months” of Failure Free Operation

Robust Software “Ecosystem”:
-Explicit Parallel Programming

Conclusions

Disney Productions

- Game processors are performance leaders !
  - Single precision now…. double precision, just around the corner
- Application-optimized processor designs will allow system performance to continue to grow, in spite of single thread performance slowdown
- Next generation game system technologies are poised to break through the old “pre-computed” physics behaviors of video games, enabling exciting new, interactive levels of realism.
- A broad, common computational foundation exists between games and many areas of scientific and high performance computing, and this is expanding to include future processor and system technologies.