AMD
Smarter Choice

Graphics Hardware 2008

GPUs vs. Multi-core CPUs

On a Converging Course or Fundamentally Different?

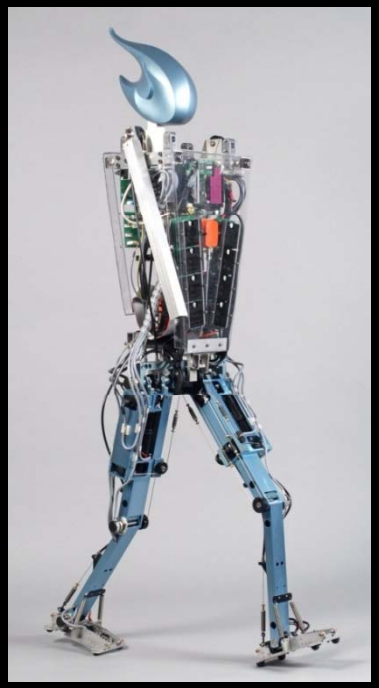
Mike Mantor
AMD Fellow Architect
michael.mantor@amd.com

Many Cores → Disruptive Change

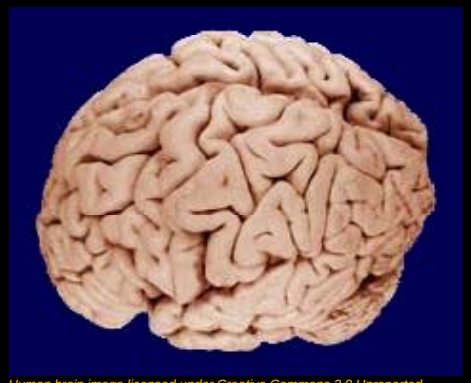
AMD Smarter Choice

- Moore's Law is enabling chaotic change
  - Double transistor density every 18 months
  - Creating need for new software models
- Power/Thermal Constrained World
  - Energy/Cooling Cost Exceed IT equipment Cost
  - Frequency alone is not our friend
- The Complexity Wall/Frequency Wall
  - Single threaded performance maxed
- Customers expect added Visible Value

Models for the Future of Computers



Flame, the robot, walks the way humans walk. (Credit: Image courtesy of Delft University of Technology)



Human brain image licensed under Creative Commons 3.0 Unreported license. See http://en.wikipedia.org/wiki/Image:Brain_090407.jpg

Human Brain

© 2006 Stanford Racing Team Web design by ToTheWeb LLC
http://cs.stanford.edu/group/roadrunner/images/graphics/junior-3.jpg

Autonomous Vehicle

F-16 Fighting Falcon image is a U.S. government work and is in the public domain. See http://en.wikipedia.org/wiki/Image:F-16_Fighting_Falcon.jpg

Fighter Jet

Heterogeneous Systems

- Parallel signal processing of sensor data
- Parallel Classification and Recognition
  Vision, Sound, Contact, Patterns, Object
- Serial Cognitive and Reasoning
- Parallel Search, Scan, Sort

AMD Smarter Choice

Advantages

Optimized for execution of sequential program

- Complex Pipelines to achieve max frequency
- Out Of Order, Super Scalar to achieve max ILP
- Branch Predictors, Speculative execution
- Register Renaming
- Large Caches optimized for low latency access

Multi-Core enables parallel multi-tasks/threads

- Improved user response
- Background task such as OS chores, virus scan, etc

AMD
Smarter Choice

Disadvantage

Fine grain sharing of work between cores/caches
- OS Overhead – spawn, communication, fork
- Finding large number of task to Multi-task is hard

Embarrassing parallel apps
- Limited Speed up (ALU & Bandwidth)
- Limited Power/Flop advantage

AMD
Smarter Choice

Advantages

- Optimized for structured parallel execution
  - Extensive ALU counts & Memory Bandwidth
  - Cooperative multi-threading hides latency
- Shared Instruction Resources
- Fixed function units for parallel workloads dispatch
- Extensive exploitation of Locality

AMD
Smarter Choice

Disadvantages

Ineffective for Single threaded execution

- Divergence
  - Parallelism lost opportunity
  - Loss of efficiency
- Upload/download of data cost
- Requires large workload to fully utilize
- Small Caches are optimized for locality/throughput

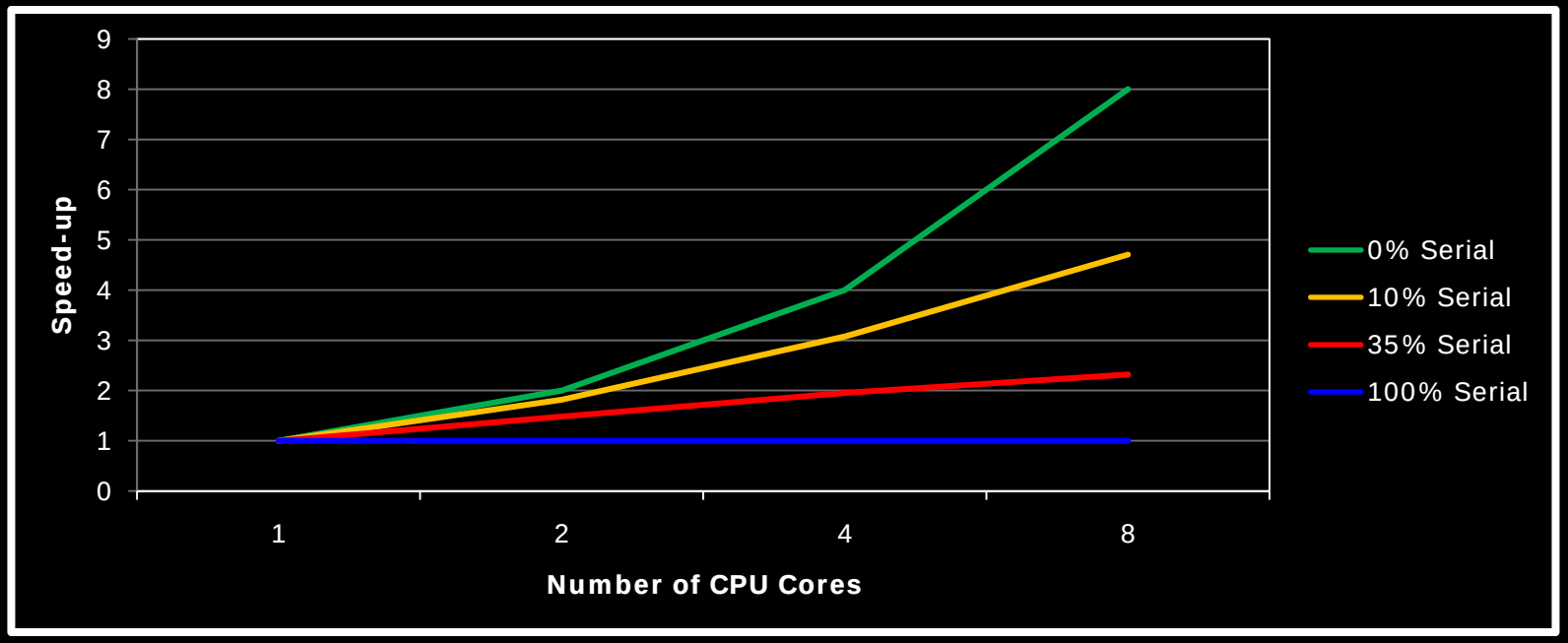
Amdahl's Law

AMD
Smarter Choice

$$\text{Speed-up} = \frac{1}{S_w + (1 - S_w) / N}$$

S_w: % Serial Work

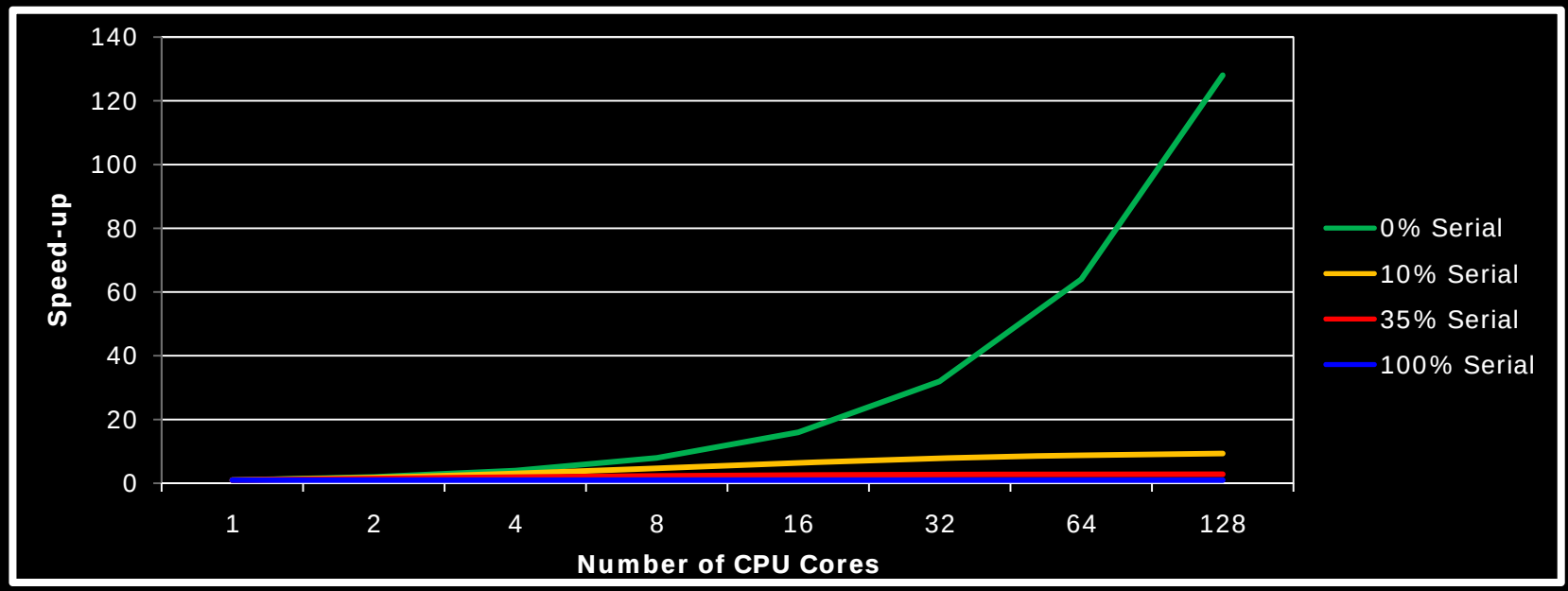
N: Number of processors



Amdahl's Law – *zoom out a bit*

AMD Smarter Choice

- *"Everyone knows Amdahl's Law, but quickly forgets"*
  - Dr. Tom Puzak, IBM Research, 2007



Amdahl's Law seriously inhibits unstructured parallelism …

Performance/(Cost & Watt) Illustration*

AMD Smarter Choice

Example: Assume 200 mm², 80W @ constant freq

Relative Performance per core	**0.25**	**0.5**	**1.0**	**1.4**
mm2/Core	1.56 mm2	6.25 mm2	25 mm2	50 mm2
Power/Core	0.6 W	2.5 W	10 W	20 W
#of Cores (200mm² System)	128	32	8	4
System Performance for parallel Apps	32	16	8	5.6
Relative System Performance/mm²	16%	8%	4%	2.8%
Relative System Performance/Watt	40%	20%	10%	7%

Structured Parallelism enables solutions for more flops less watts

Building Blocks - Best of Both

AMD Smarter Choice

Heterogeneous Cores (Mixed CPU/GPU)

- Lowest Power & Highest Performance Solution
- Serial Single threaded – Leverage Fast OOCs
- Task Parallel – Leverage multi-CPU or GPU cores
- Data Parallel - Leverage GPU or Application Specific Cores

Building Blocks - Best of Both (Con'd)

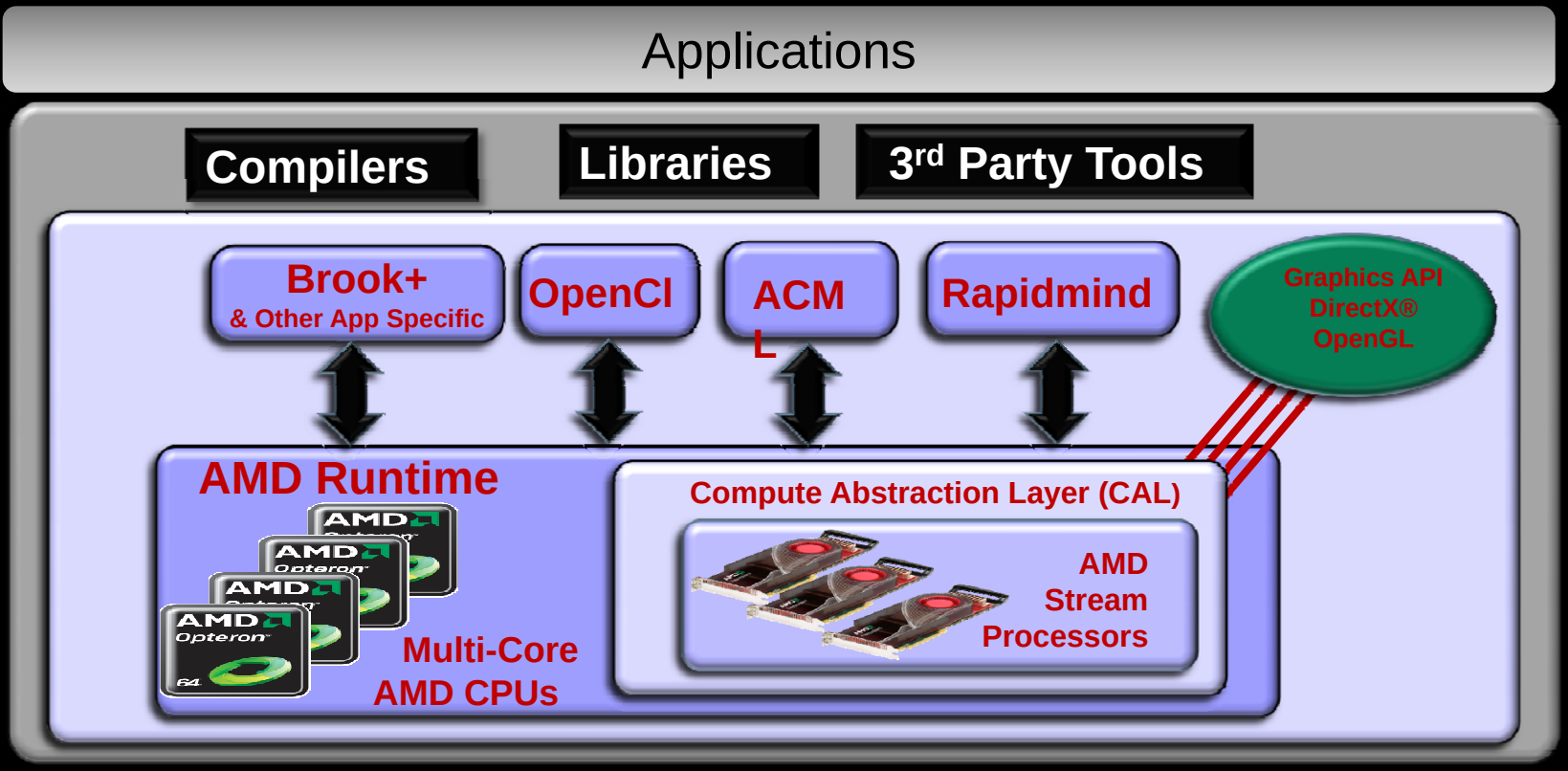
AMD Smarter Choice

Unified Memory Architectures

- Remove large data copies of workload and results
  - Reduce power consumption per computation
  - Enable small job offload
- Remove OS Overhead and latency of communication
- Fast Synchronization Primitives
- Increasing capable memory systems & BW

AMD Stream SDK *Software Development Stack*

AMD
Smarter Choice



Software Solution (Real Challenge)

AMD
Smarter Choice

Development of Software often exceeds that of hardware

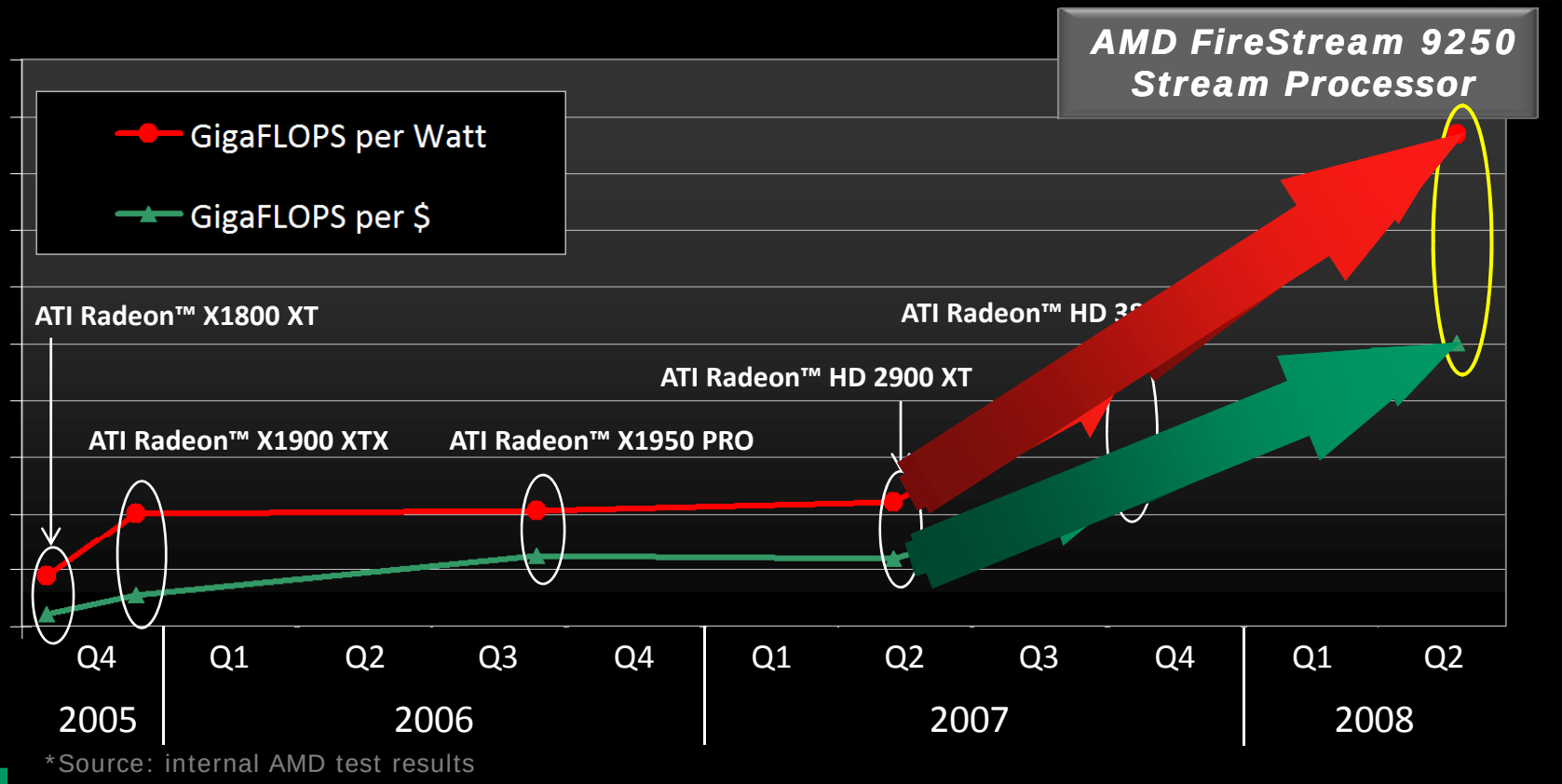
- Determine and implement long term solutions
- Enable highly coherent machines with heterogeneous accelerators
- Enable control of memory hierarchies for system scaling
- Communication/Messaging ➔ Producer/Consumer relationships

Simplifying Programming Model

- Build on emerging multi-core models
- Enable the Masses to Programming with parallel abilities
- Enable Application Specific Libraries
- Enable open standards

Processing Efficiency*

AMD
Smarter Choice



*Source: internal AMD test results

Disclaimer and Attribution

DISCLAIMER

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

ATTRIBUTION