

IBM System and Technology Group



IBM Systems and Technology

World Modeling Summit for Climate Prediction ECMWF, Reading England, May 2008

petascale – Capability or Capacity

Kent Winchell, IBM Deep Computing
kentwin@us.ibm.com

System Design - Two Directions

- Where should I spend my \$\$?
- Why should I care? Can't I just buy a balanced system?
- Small system changes reflect large \$\$ amounts at petascale

- Optimize for performance on 100,000 core single job
 - ▶ System wide low latency barriers
 - ▶ Fully meshed interconnect
 - ▶ High reliability required

- Optimize for Throughput on 1000, 100 core jobs
 - ▶ System partitions or modules
 - ▶ Fully meshed within partition
 - ▶ Focus on efficient job management
 - ▶ Low reliability, since each partition is small % of system

Cost Tradeoffs

- 1 PetaFlop/s sustained over a wide range of applications suggests peak performance targets of 3-10 PetaFlops
- 5 PetaFlop/s peak performance implies:
 - ▶ 2.5PB of DRAM memory (.5B per FLOP)
 - ▶ 100PB of Disk Storage
 - ▶ 5TB/s Bandwidth to Disk

DRAM price estimates for 2010 range from 4¢ to 12¢ per MegaByte

- ▶ @4¢/MB, 2.5PB of DRAM costs \$100M
 - We expect memory costs to be twice this
- ▶ Re-think ideas like I need 2 GB/core

5TB/s Storage bandwidth in 2010 is estimated to require ~150,000 drives.

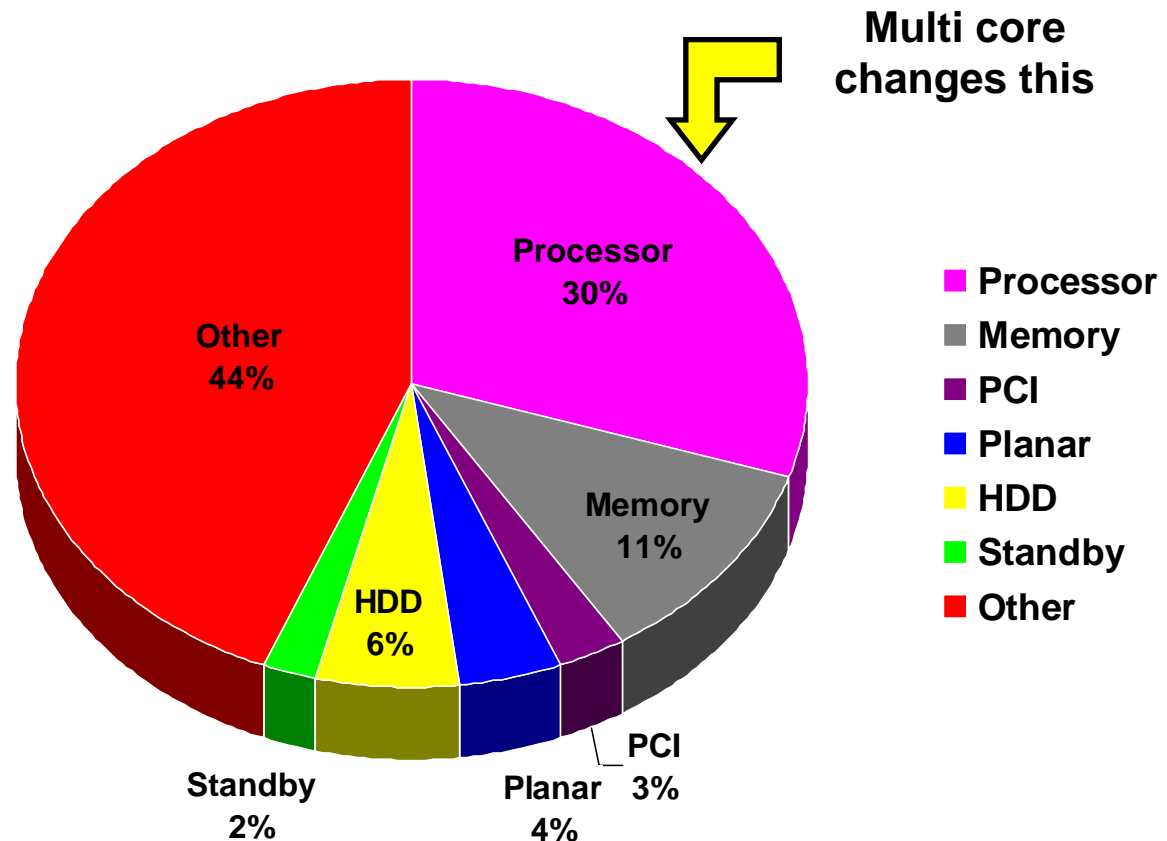
- ▶ Bandwidth requirement dominates the capacity requirement (100PB)
- ▶ Costs could be in the \$60-\$100M range
- ▶ Potential floor space and power implications

What's using the power 2005?

- The processor power growth is the largest single contributor but there are many other areas- the more you pack into a server the more power it needs!

OTHER?

- AC to DC Transitions
- DC to DC Deliveries
- Fans and air movement

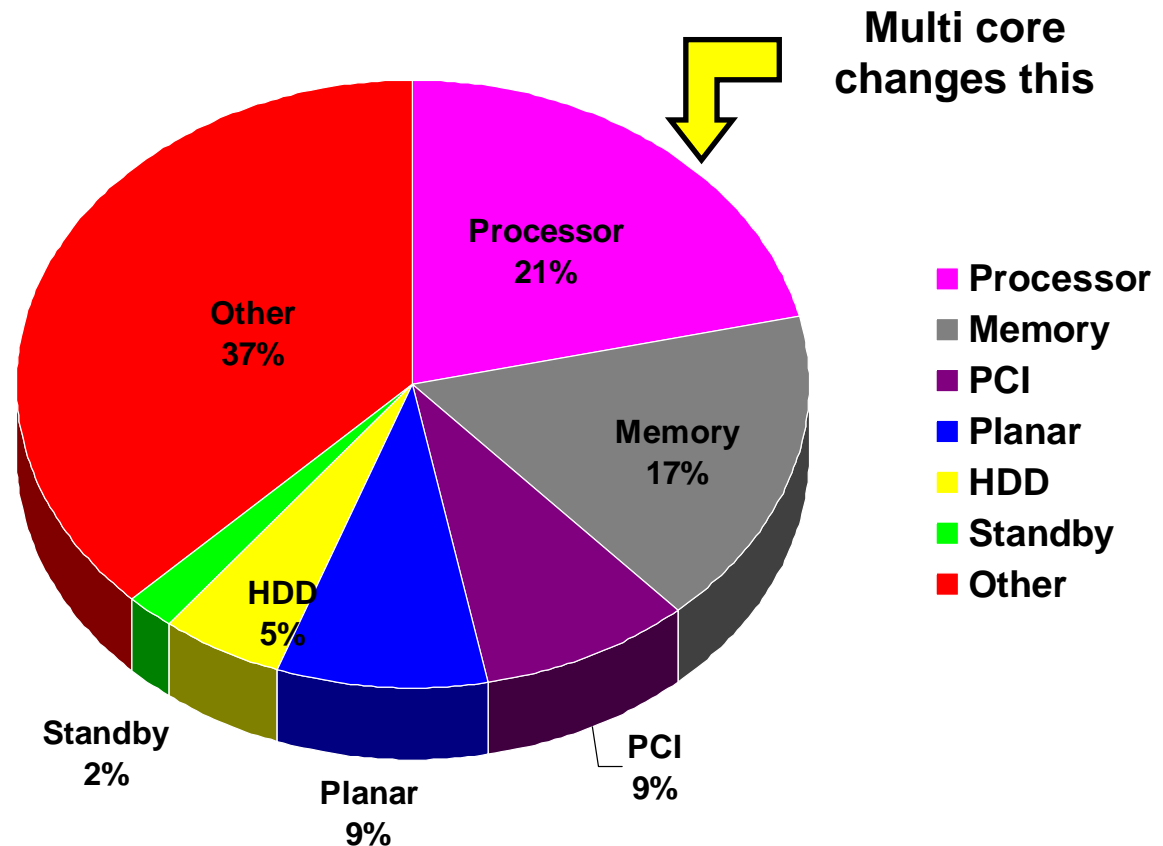


What's using the power 2008?

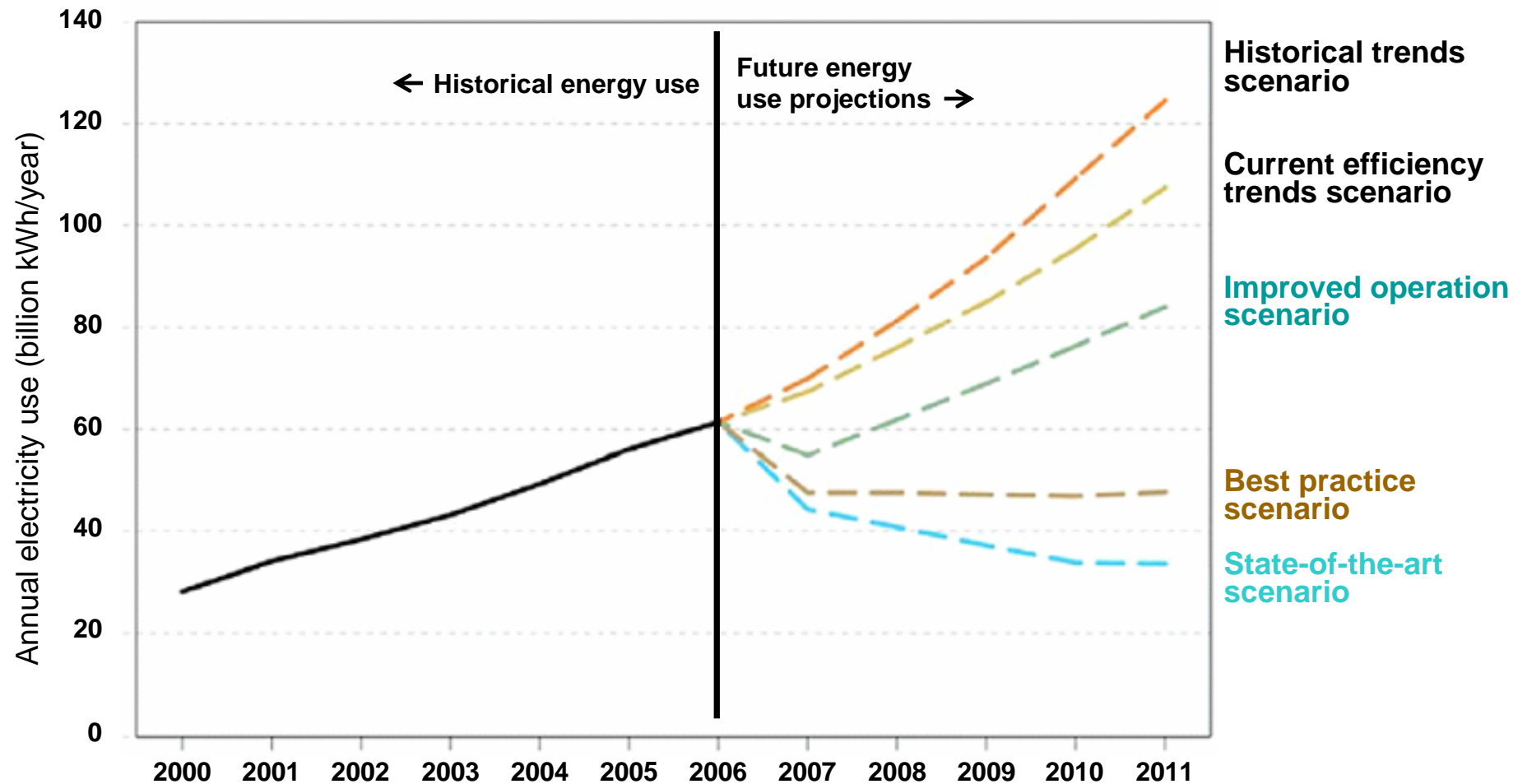
- The processor power growth is the largest single contributor but there are many other areas- the more you pack into a server the more power it needs!

OTHER?

- AC to DC Transitions
- DC to DC Deliveries
- Fans and air movement



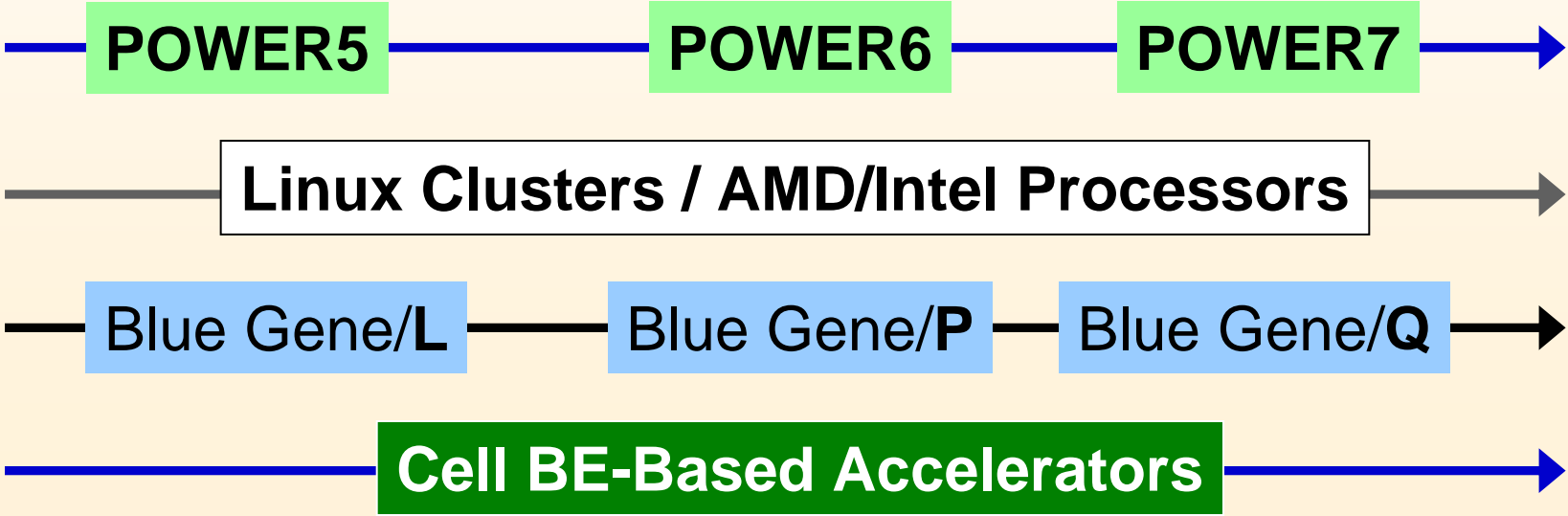
Comparison of Projected Electricity Use, 2007 to 2011



Source: EPA Response to Congress for Public Law 109-431, 08/07/07

Deep Computing Conceptual Roadmap

Petaflop systems with ultra scalability
 Hybrid architectures
 System level accelerators
 Maturing software stack
 Rich application infrastructure



Trends

- Cores are getting cheaper and more plentiful
- Liquid cooling can reduce electrical up to 50%
- Lower power cores are only good if application scales close to linear for large task counts
- Re-visit memory ratios and I/O ratios
 - ▶ 2GB of memory attached to slower core not as economical