

Platforms Design Challenges with many cores

Raj Yavatkar, Intel Fellow

Director, Systems Technology Lab

Corporate Technology Group



Environmental Trends: Cell

BusinessWeek online

AUGUST 25, 2005

NEWS ANALYSIS

By Arik Hesseldahl

Cell: A Chip That's Going Public

The processor already powers PlayStation 3. But IBM, Sony, and Toshiba will release its technical details to stimulate new uses

FORTUNE

GIVING TO GET MORE

The 9-in-1 Wonder Chip

Built for games, IBM's mighty Cell chip could help reshape all of computing.

By David Kirkpatrick

THE ECONOMIC TIMES

IBM eyes more mkts for 'cell' chip

REUTERS [FRIDAY, SEPTEMBER 16, 2005 09:30:54 AM]

SAN FRANCISCO: Two years ago, analysts were calling on IBM to exit the microchip business after production problems and lower-than-expected orders led to losses of more than \$1 billion since 2002 at the division.

BUSINESS 2.0

The playbook for a new generation of leaders™

FEATURES

The Cell of a New Machine

An IBM-led consortium named its radical speed-demon-of-a-chip after the basic building block of life. Will it give birth to a revolutionary era in the electronics industry?

By *Erick Schonfeld*, May 18, 2005

Environmental Trends: GPUs Branching Out

the **INQUIRER** News, reviews, facts and friction

Nvidia warns multicore CPUs could stiff innovation

Is the GPU the real brain of the computer

By [INQUIRER staff](#): Wednesday 17 August 2005, 15:07

[apcmag.com: article](#)

Hijacking the GPU

Thursday 11, August 2005 | By [Dan Warne](#)
From: [Epinions > General](#)

From APC Magazine: GPU hackers are turning the super video cards against Intel. Dan Warne investigates.



GPGPU

Example: Black-Scholes options pricing

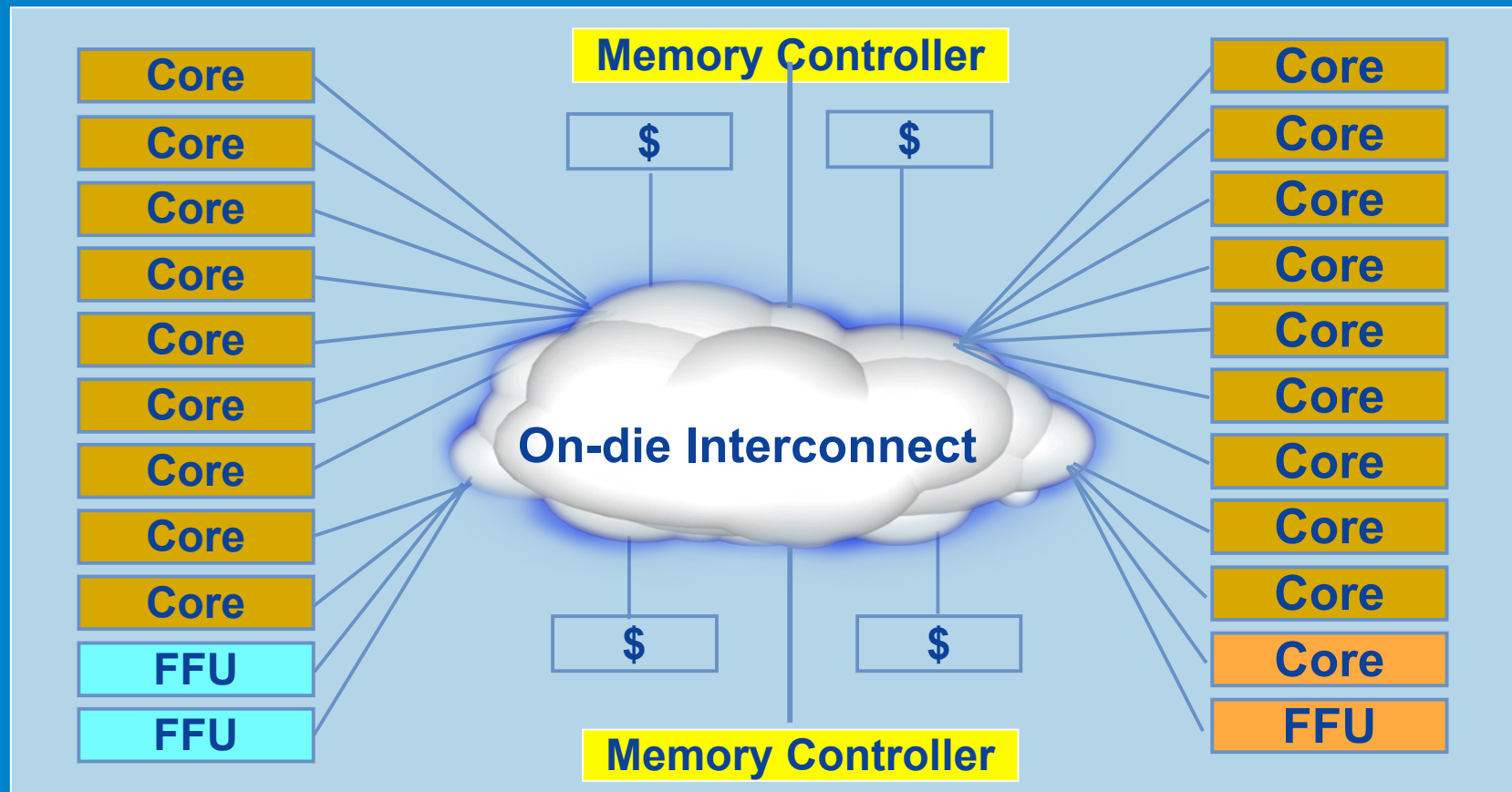
- Widely-used model for pricing call/put options
- Implemented in ~15 lines of Cg, use combinations of input parameters as separate simulations (fragments)
- Perf:
 - Fast (~3GHz) P4, good C++: ~3.0 MBSOPS (1X)
 - Quadro FX 3000, Cg: ~2.8 MBSOPS (~.9X)
 - Quadro FX 4400, Cg: ~14.4 MBSOPS (4.8X)
 - Quadro FX 4400, Cg, 100 runs: ~176.0MBSOPS (59X)
(remove test/data transfer bandwidth overhead)
- How?
 - CPU: ~11GFLOPS, slow exp(), log(), sqrt(), fast mem access
 - GPU: ~65GFLOPS, fast exp(), log(), sqrt(), slow mem access
 - Black-Scholes has high ratio of math to memory access
 - GPU has Parallelism

David B. Kirk, NVIDIA

IEEE Hot Chips 2005



Many-Core by Example



Parallel extension of IA

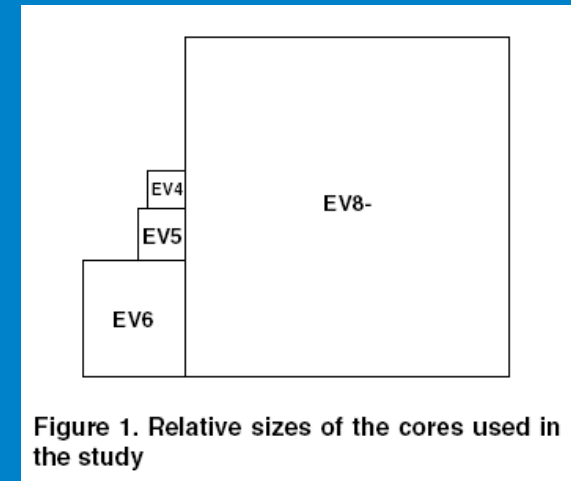
- array of cores
- Fixed-function units
- Coarse- and fine-grained data- and thread-level parallelism

Partitioned array

- Application domains
- Isolated communication traffic
- Fault tolerance

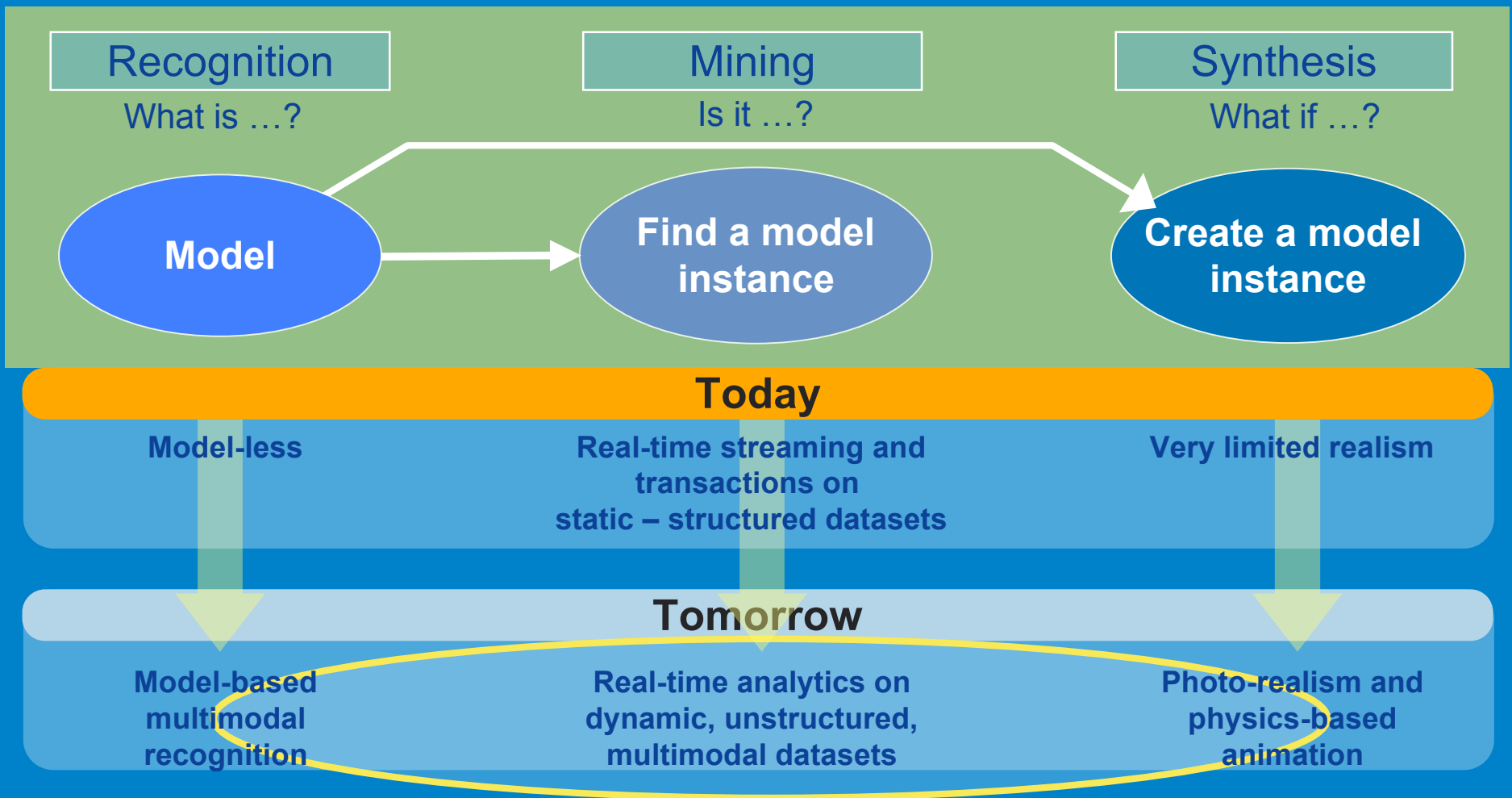
Rethinking the Core

- Need efficient cores
 - Reduced energy per instruction
 - More TLP, less ILP
- Potential benefit – full ISA plus:
 - 0.25 area and power
 - 0.5X legacy performance
 - 0.6-0.8X performance for MT apps
 - 1.0X FP performance

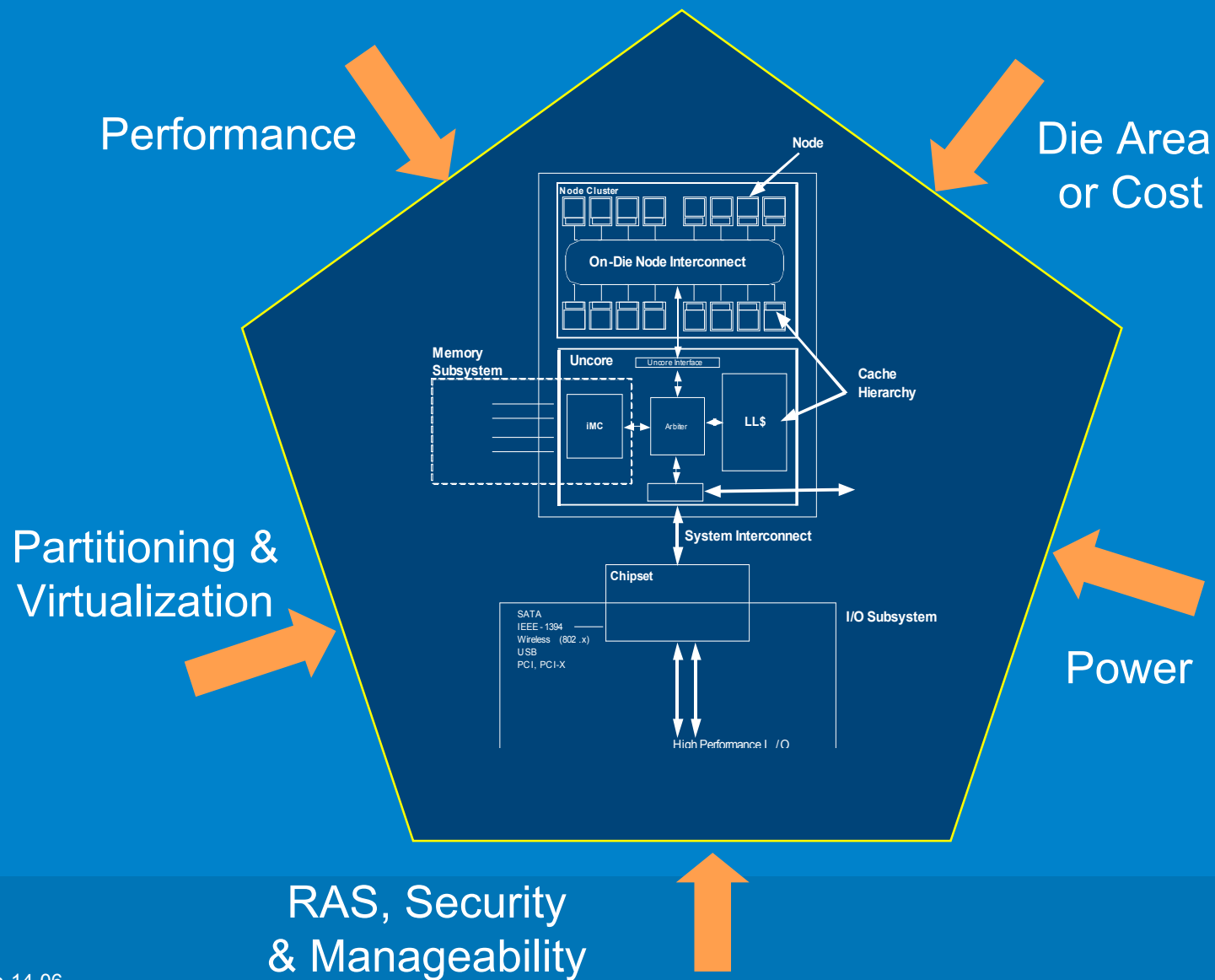


N. Jouppi, Computer Architecture Letters, July 2003

Tomorrow's Killer Applications



Platform Factors or Constraints



“Balanced” Node

- # cores? interconnect?
- memory controllers
- Co-proc / accelerators
- N-Cache: priv or shared

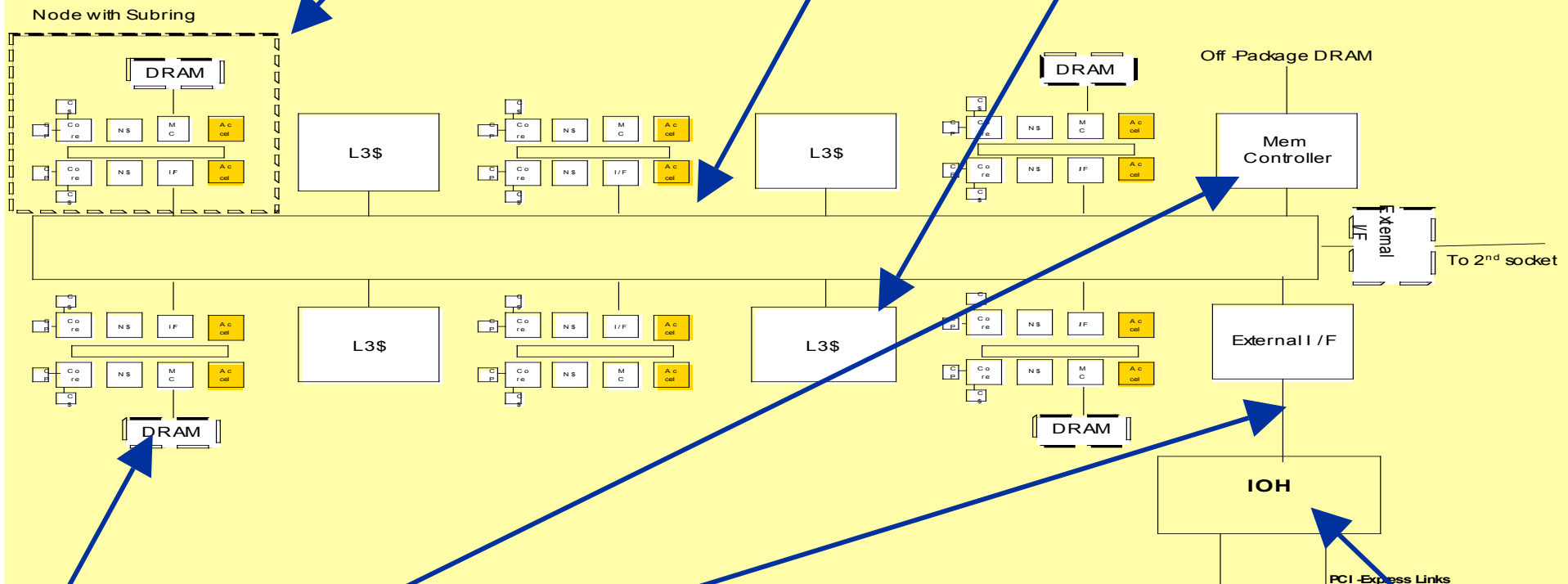
“Scalable” On Die Interconnect

- LowL / HiBW interconnect
- Routing and Interface
- Message passing support
- Partitioning

“Distributed” Cache

- NW of Cache Slices
- NUCA – affinitized?
- Dir-Based Coherence
- Inclusion / Exclusion?

Many core platform



“Large Hi-Bw” Memory

- Distrib or Centr iMC
- On-Pkg (special) DRAM
- Capacity Issues
- DRAM cache?

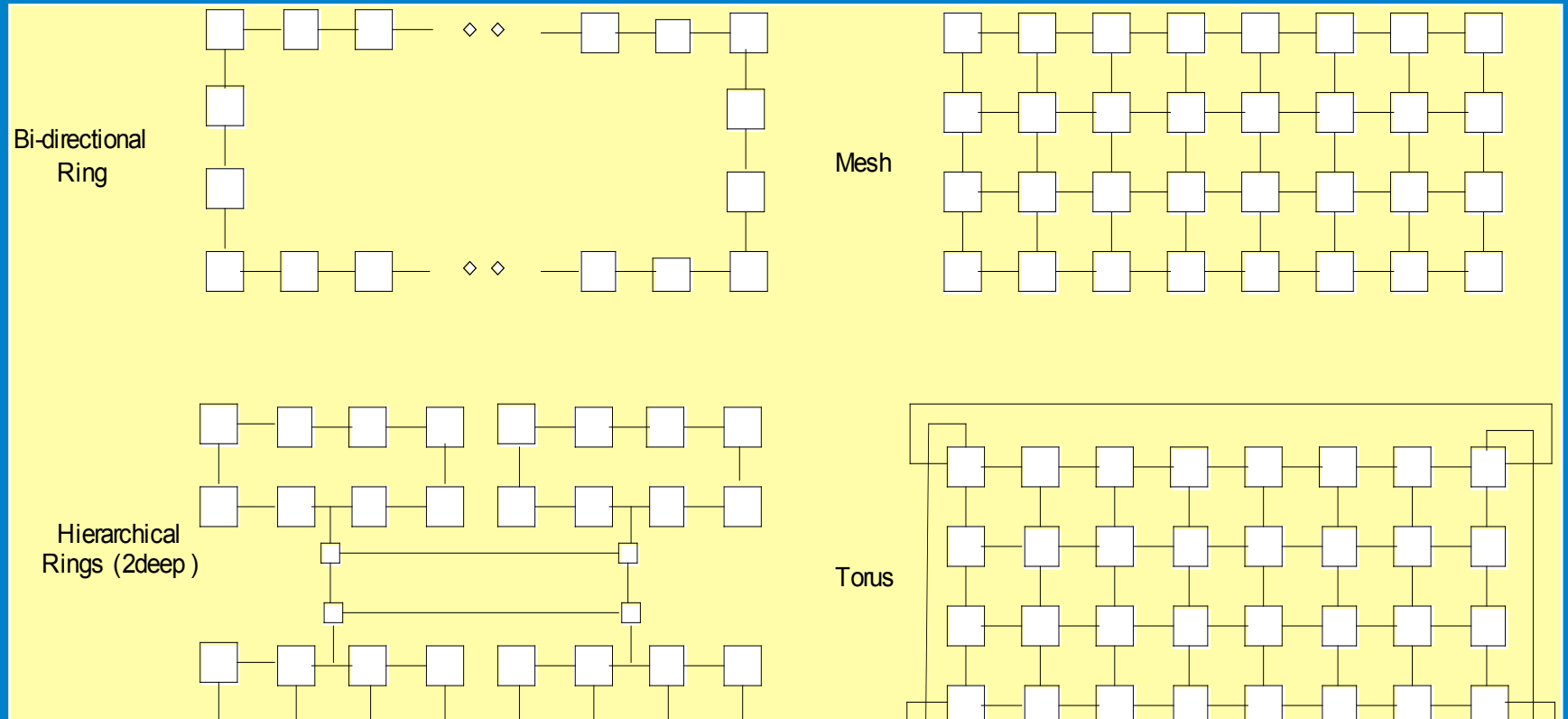
“Efficient” Sys Interconnect

- single socket?
- message passing support
- interconnect scalability

“Scalable” Chipset+I/O

- Virtualization
- Partitioning
- IO-MMU
- New Device models

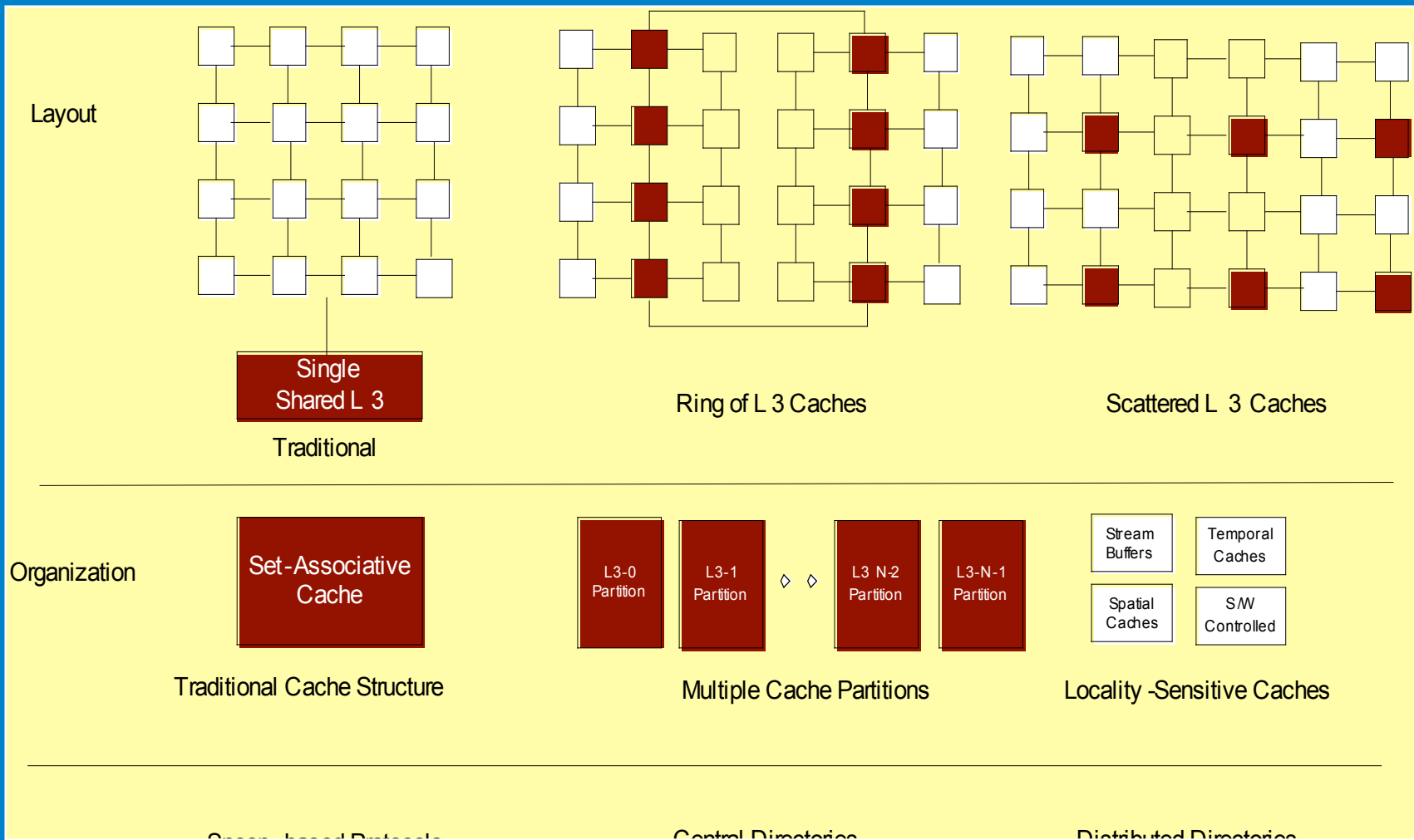
On-Die Interconnect Options - Examples



Interconnect Trade offs:

- Average distance, degree of connectivity, Bi section bandwidth
- Routing and coherency Protocol
- Message semantics

Cache Hierarchy Options - Examples



Some Potential Implications

System memory architecture

- How much DRAM can be placed one hop away from MC silicon
 - signaling, layout, routing, etc. considerations
- Will thermal considerations limit this amount of DRAM yet further?
- As we reach the upper limit of electrically close DRAM capacity
 - Where will additional DRAM capacity reside in the platform ? (NUMA)

System operational models

- If additional memory hierarchy, with NUMA latency, is required then several aspects are affected
 - Caching and cache protocols
 - Address Translation
 - Memory Allocation
 - Buffering for I/O
- Configurability of memory
 - What is the base configuration of DRAM for the machine?
 - Under what workloads considerations is the base configuration to be sized?
 - What are the implications of multiple DRAM hierarchy layers for System Partitioning and resource sequestering ?

Multi-core Platform Effects on Software

Scalability of O/S Data Structures and Policies

- Synchronization and locking
- Scheduling
- Process management
- Data structure sizing and management limitations
- Threading granularity and primitives

Memory Hierarchy Awareness

- Impact of coherency policy
- Efficiency of Data-sharing and Process migration effects
- SW visibility to High speed on-die interconnect
- SW control of Cache hierarchy, NUCA Awareness

High Bandwidth I/O Support

- Light weight Interrupts
- Data movement and transformation engines
- I/O Affinity

Multi-core architecture will have a profound impact on system software and application performance

Many Cores Platform Physical design

Strong interrelationships among the following

- Power Delivery
- Thermal Management
- Power Management
- Packaging
- Physical system organization and concept
 - Compute
 - Memory
 - I/O
 - Off-die fabric
 - Mechanicals and Volumetric Placement

Potential bounds on System Configurability & Performance

Thanks for listening!

