HARDWARE TRENDS
NSF Compiler Workshop
Sept. 6, 2001

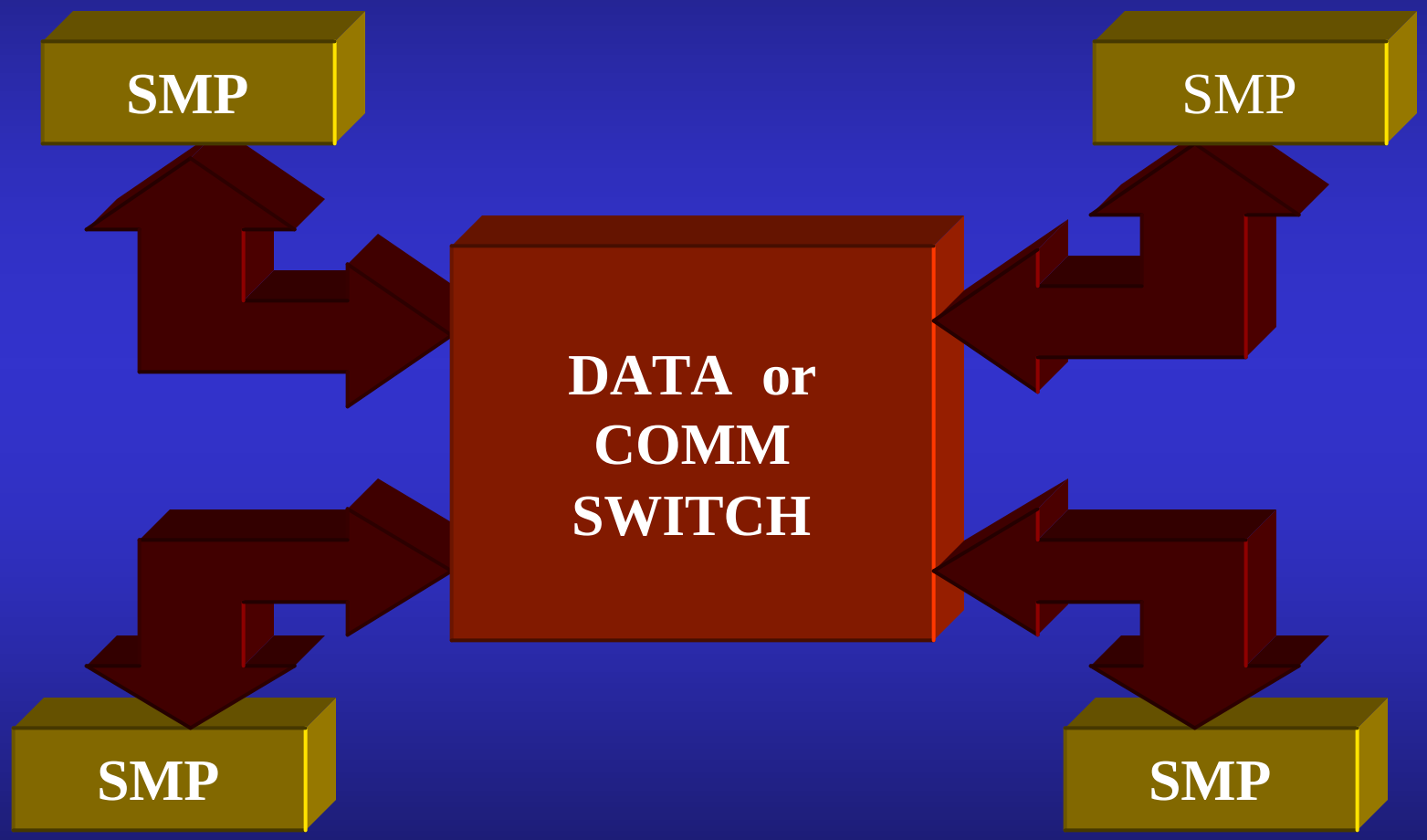
Steve Wallach

swallach@chiaro.com

APPROACH

- Overall System Architecture
- Crude Software Example
- Use SIA roadmap
- Design a COTS PETAFLOP System

SYSTEM ARCHITECTURE



ABOVE/BELOW THE LINE

- 5 Levels (x,y,z,t,iteration)
- DO IT =

EXPLICIT - MPI, THREADS, HPF, OBJECTS

  - DO TIME =
    - DO X=
    - DO Y=
    - DO Z=

AUTOMATIC - FORTRAN/C

SIA STUDY

- Provides a prediction of semi-conductor trends.
- In general, actual is BEATING predicted
- Many semiconductor firms benchmark their own roadmap and progress against the SIA roadmap
- http://public.itrs.net

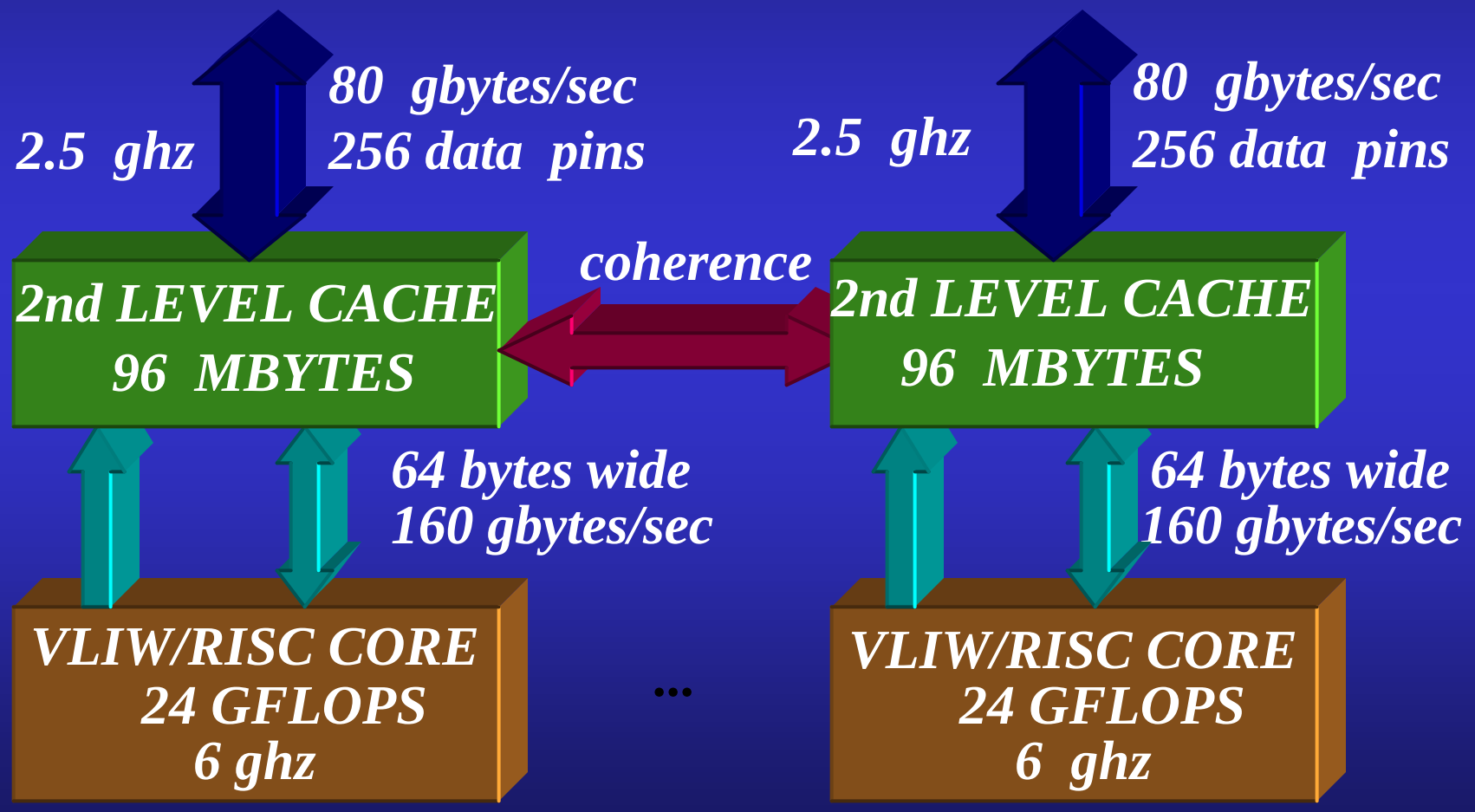
ARCH. - LONG TERM- 2009

- **THE SIA STUDY TEACHES US:**
  - 64 gbits of dram - (8 gbytes)
  - 8 gbits of sram
  - 520 million MPU transistors
  - 70 nm lithography, 2.54 cm on-a-side
  - 6 ghz clock within vliw/risc core
  - 2.5 ghz across die
  - 2500 external signal pins

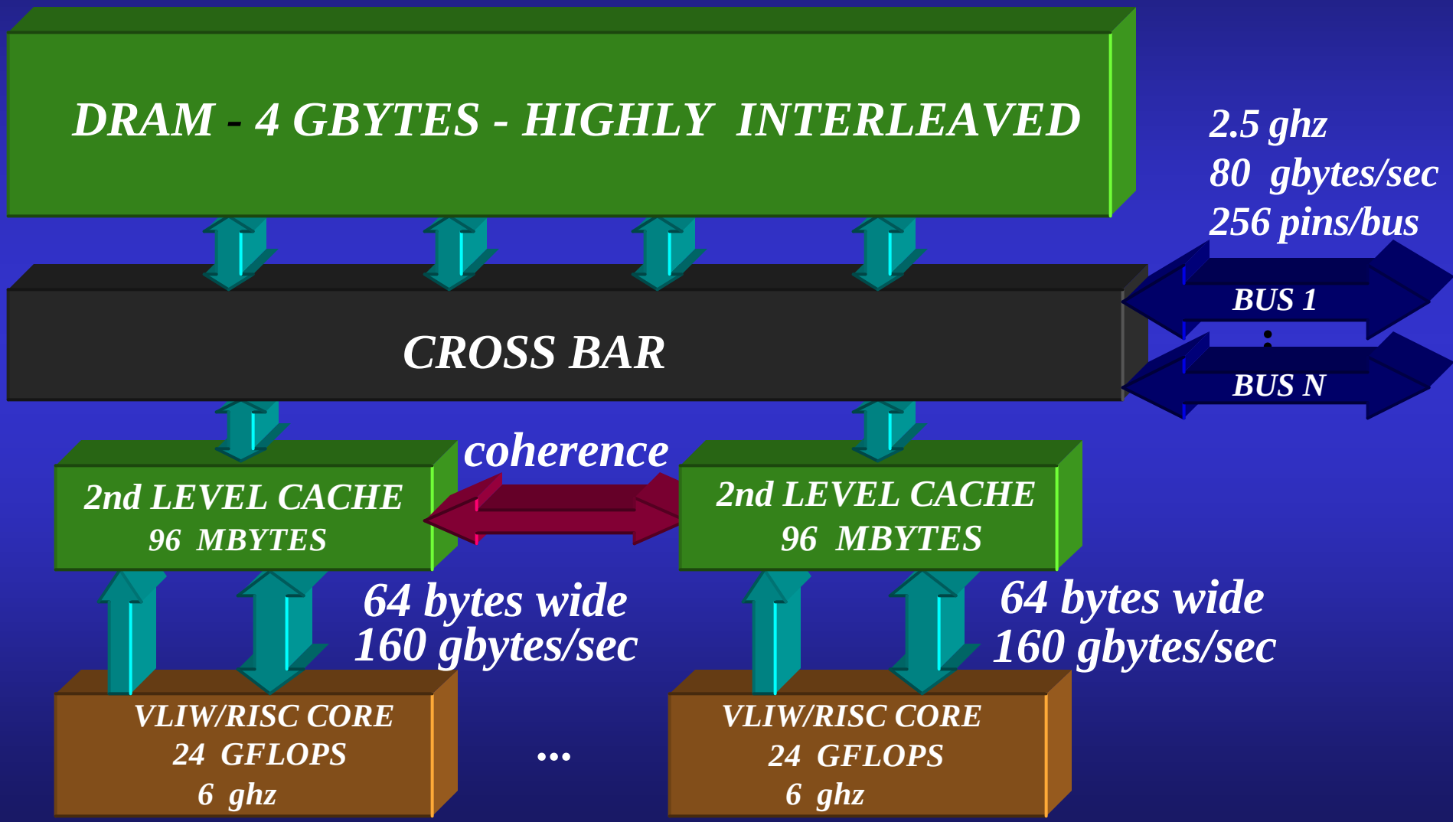
ARCH. - LONG TERM - 2009 DESIGN ASSUMPTIONS

- 9 million transistors - vliw/risc core with first level cache.
- 2nd. Level cache - rule of thumb. 1/4 to 1/2 mbyte per 100 mflops peak.
- 96 mbyte 2nd. Level (6 Inst, 90 data)
- 170 watts
- .6 to .9 volts power supply

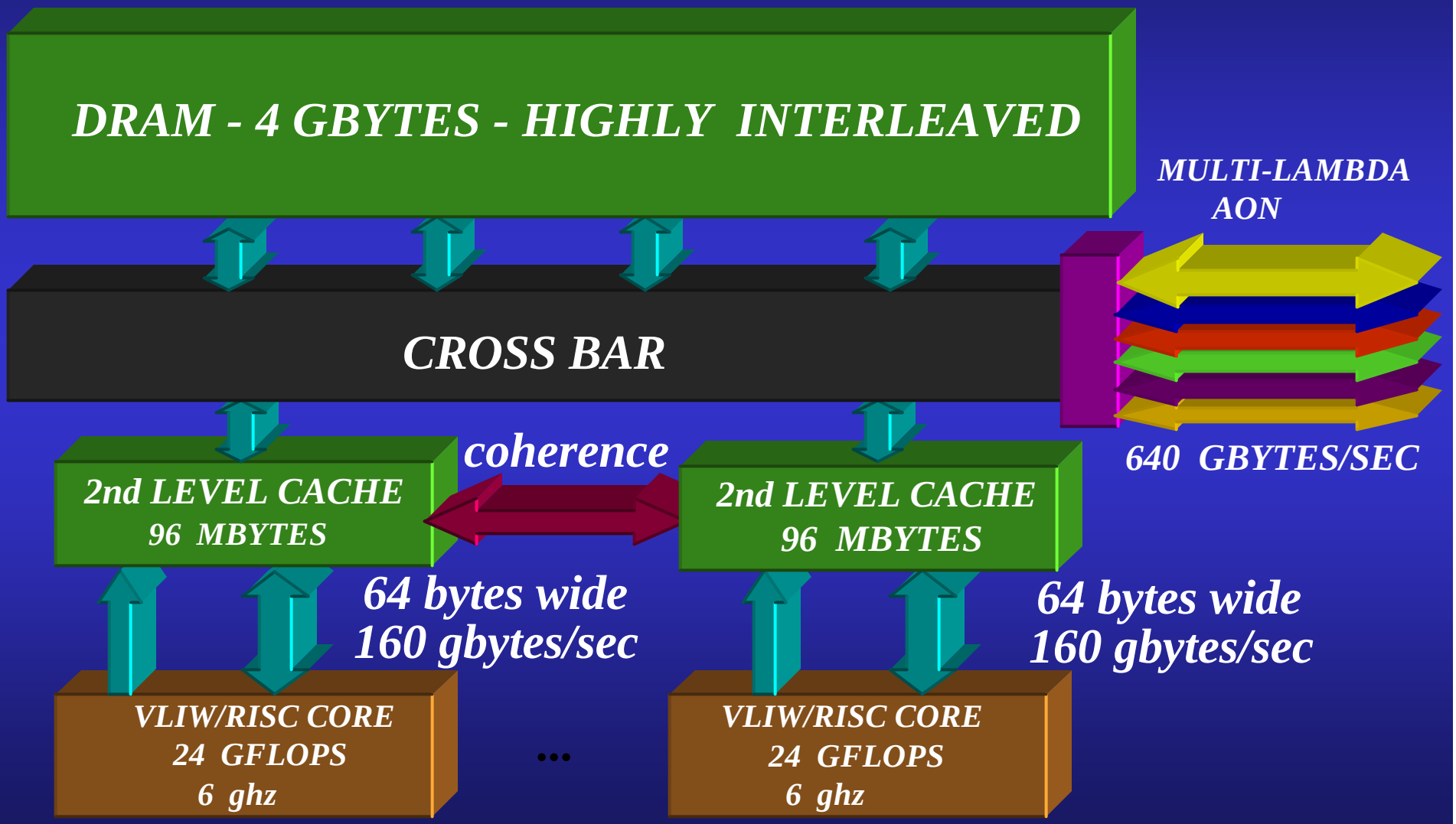
MAXIMUM PIN-USE EXTERNAL SMP- 6/8 CPU'S



INTEGRATED SMP - 4 CPU



INTEGRATED SMP - WDM



COTS PetaFlop System

- 8192 Dies (4 CPU/die-minimum)
- Each Die is 120 GFlops
- 1 PetaFlop Peak
- Power 8192 x200 Watts = 1.6 MegaWatts
- Extra Main Memory yields in excess of 3 MegaWatts (512 TBytes)

COTS PetaFlop System

- 15.36 TFlops/Rack (128 die)
- 30 KWatts/Rack - thus 64 racks - 30 inch
- Common System I/O
- 2 Level Main Memory
  - local on chip
  - off-chip within same rack
  - same bandwidth/longer latency
  - meets Byte/Flop metrics
  - reduces external rack bandwidth by a factor of 10

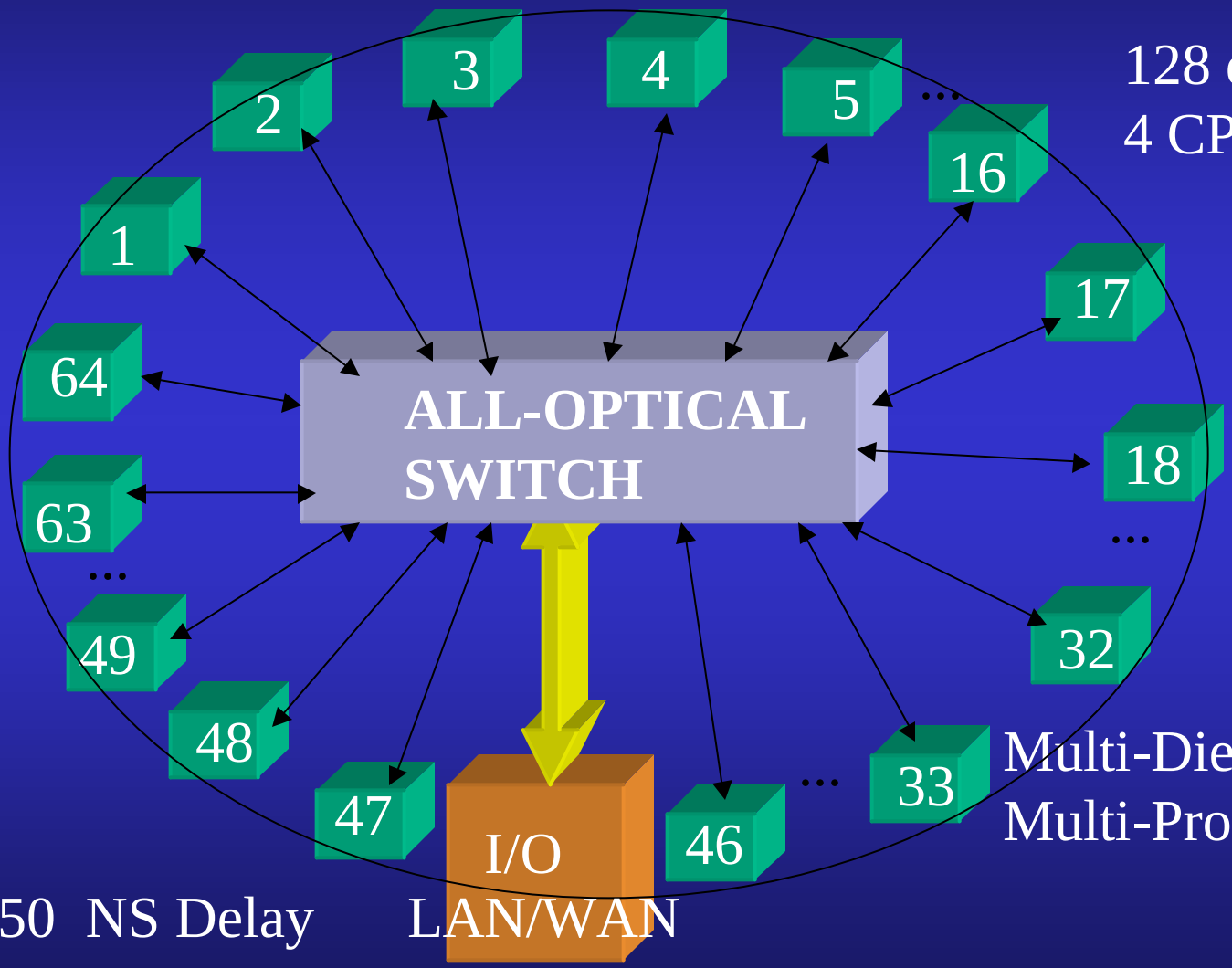
COTS PetaFlop System

- Optical Interconnect
  - OC768 Channels (40 GHz)
  - 128 Channels per Die (DWDM)-5.12 THz
  - 2004/2005 Commercial Transmission
  - ALL Optical Switching
- Bisection Bandwidth of 50 TBytes/sec
  - 15 TFlops/rack*.1bytes/flop/sec*32 racks
- Rack Bandwidth - 15 TFlops*.1= 12 THz
- 2-4 OPTICAL MT-RJ Connectors or equivalent per rack

COTS PetaFlop System Memory Hierarchy

- Physical Memory
  - Cache
  - Local Memory (Max 2 levels)
  - Global Memory
- I/O
- Substantial Bandwidth (like a vector system)

COTS PetaFlop System



128 die/box
4 CPU/die

Multi-Die
Multi-Processor

10 meters= 50 NS Delay

CONCLUSIONS

- Processors micro architectures will get **MORE** complicated not less.
  - VLIW – Instruction Scheduling & ILP
  - Multiple levels of caching
  - Speculation
  - Multi Threading
  - FPGA islands – General Purpose –Special Purpose
  - Power Aware (-o watts)
- Multiple Processors on a die
- Multiple Memory Hierarchies
- Multiple MP approaches