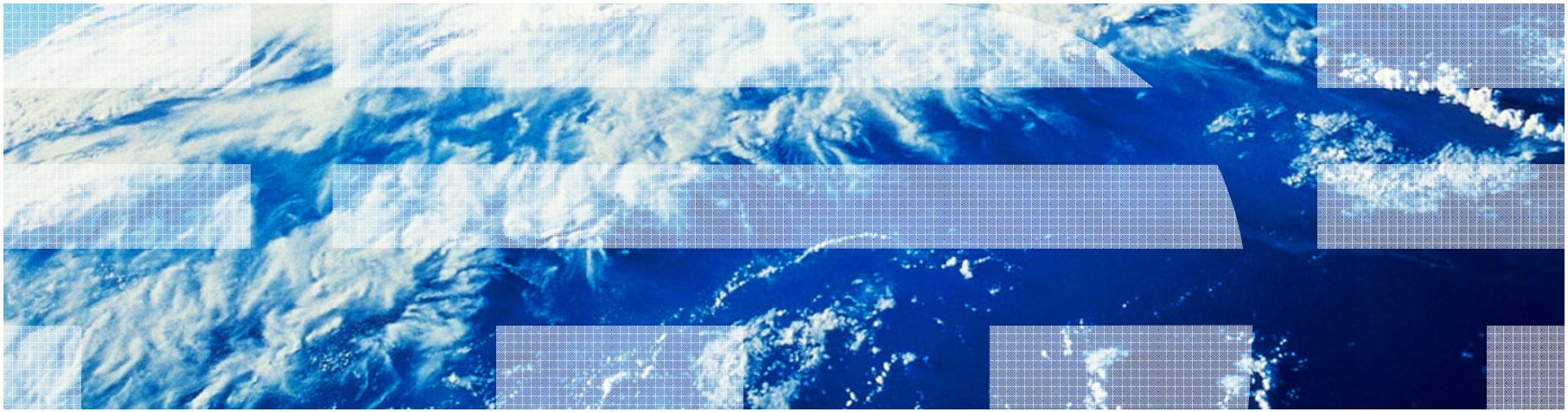
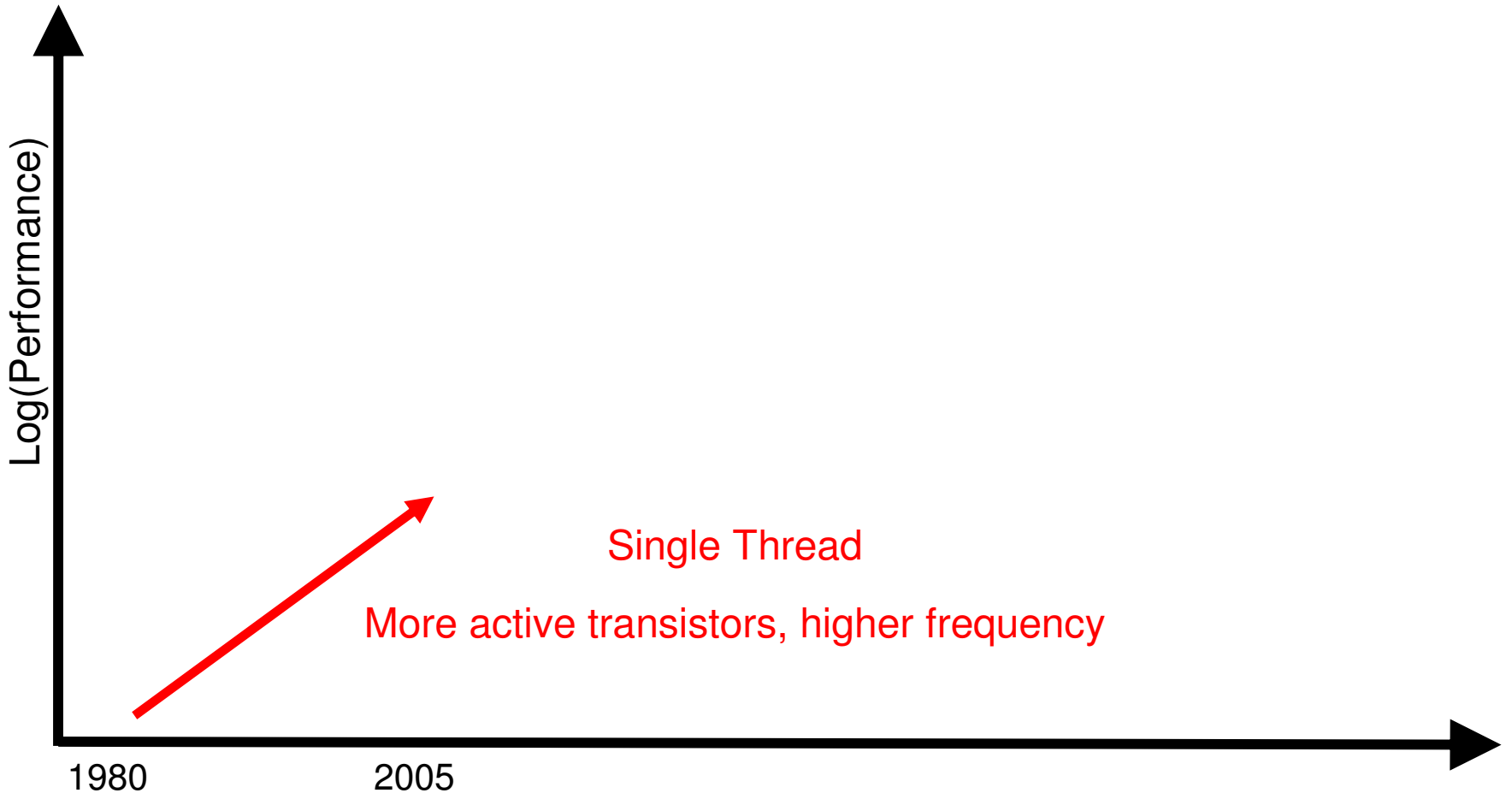
Joshua Friedrich – Senior Technical Staff Member, IBM Server & Technology Group

June 6, 2012

POWER™ Processor Design & Methodology Directions

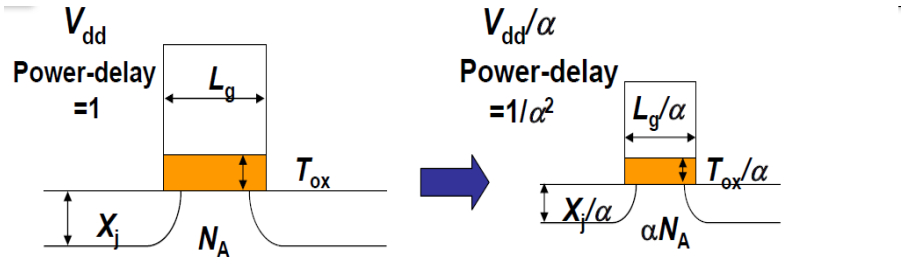


CMOS Microprocessor Trends, The First ~25 Years (Good old days)

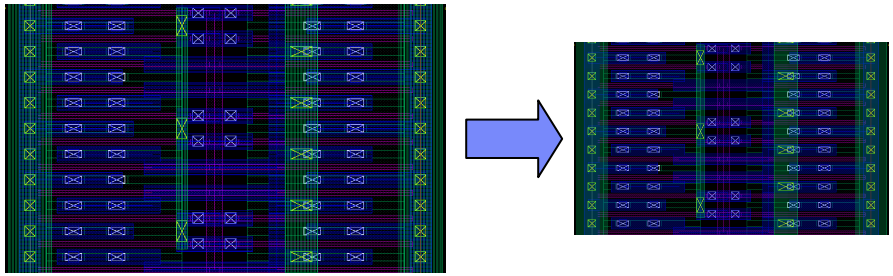


Characteristics of Single Thread Era

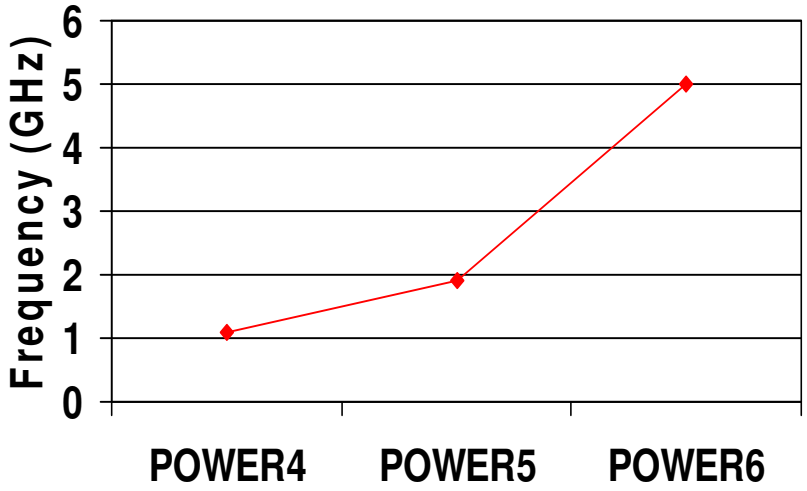
Dennard Scaling



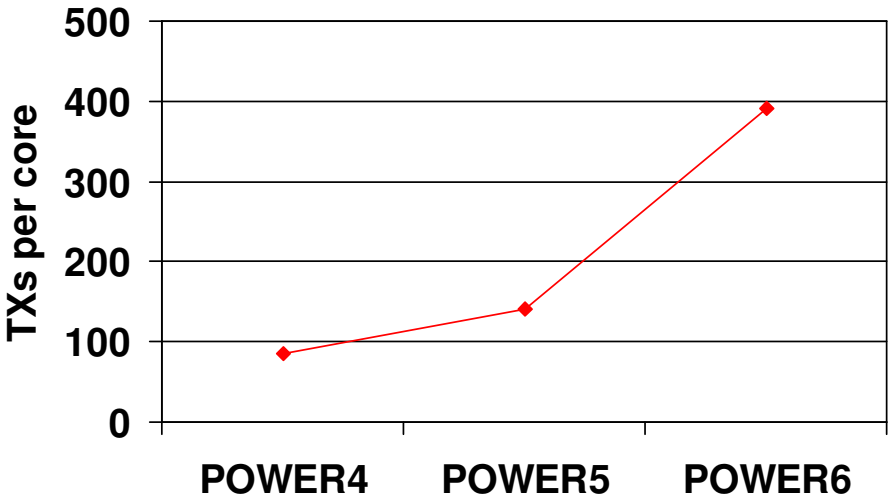
Optical Scaling / Node Migration



Exponential Frequency Growth

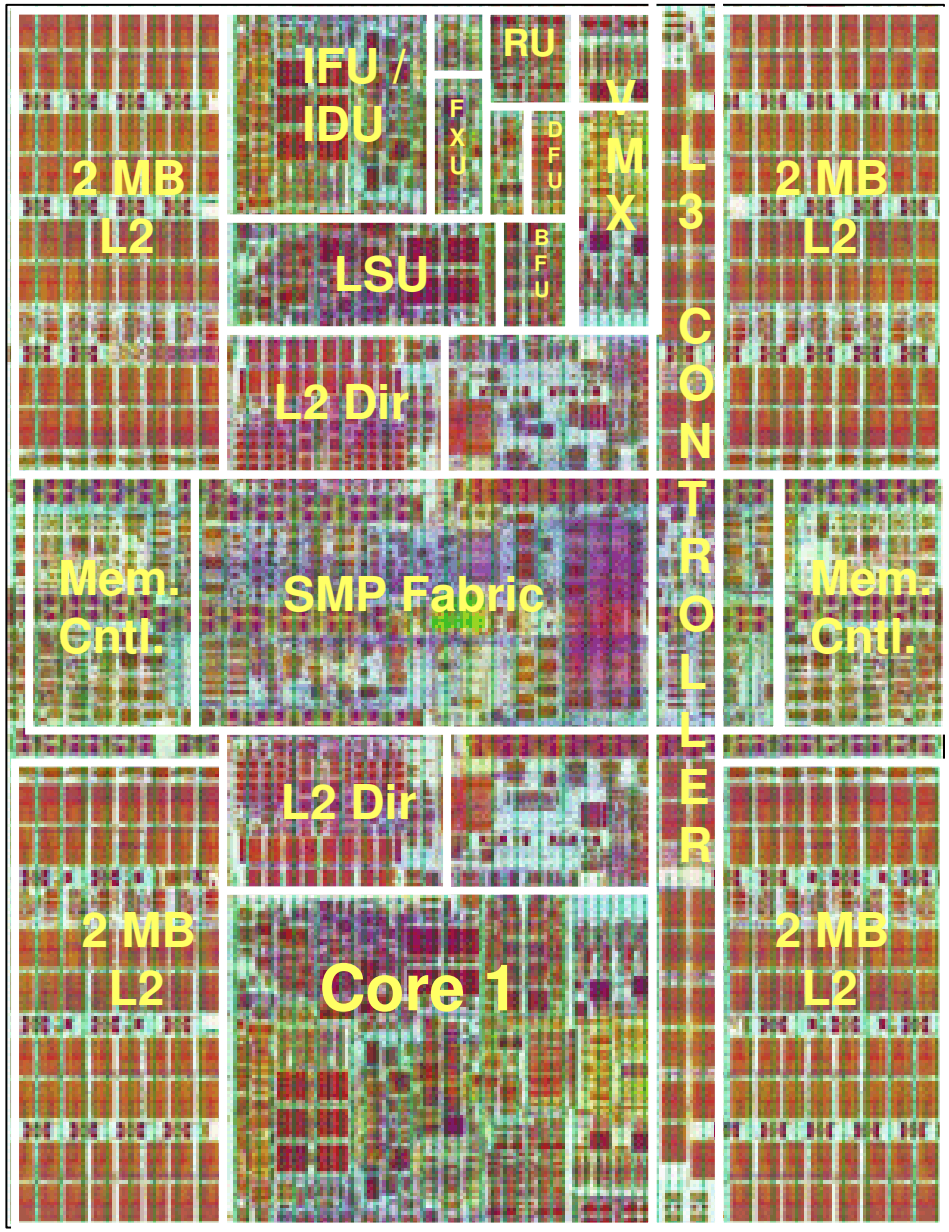


Expanding uArch Complexity



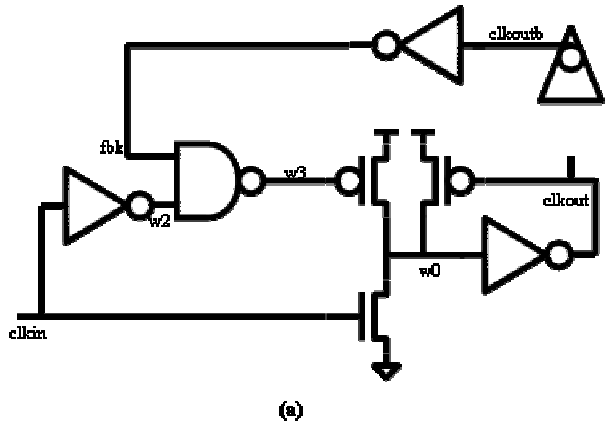
Single Thread Era Icon: POWER6

- 5+ GHz operation, >790M transistors, 341mm^2 die
- 65nm SOI with 10 levels of Cu interconnect
- 2 superscalar, SMT cores with 7 instruction dispatch & 9 execution units
- 8 MB Level-2 cache + support for 32MB L3
- 2 memory controllers, Two-tier SMP Fabric
- Same pipeline depth & power at 2x frequency versus POWER5

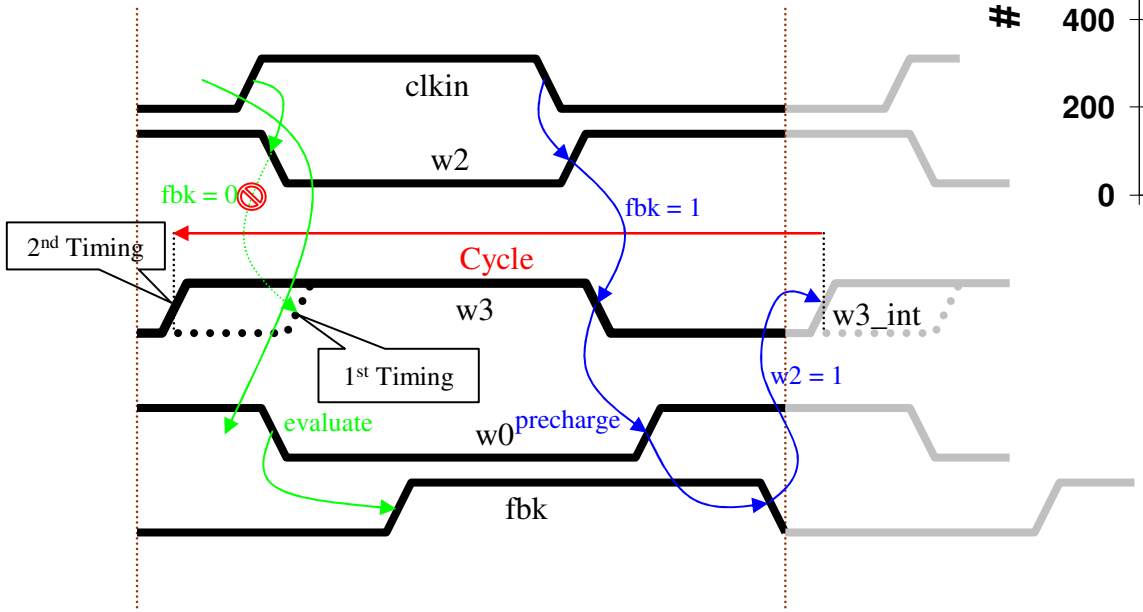


Single Thread Era EDA: Transistor Analysis & Optimization

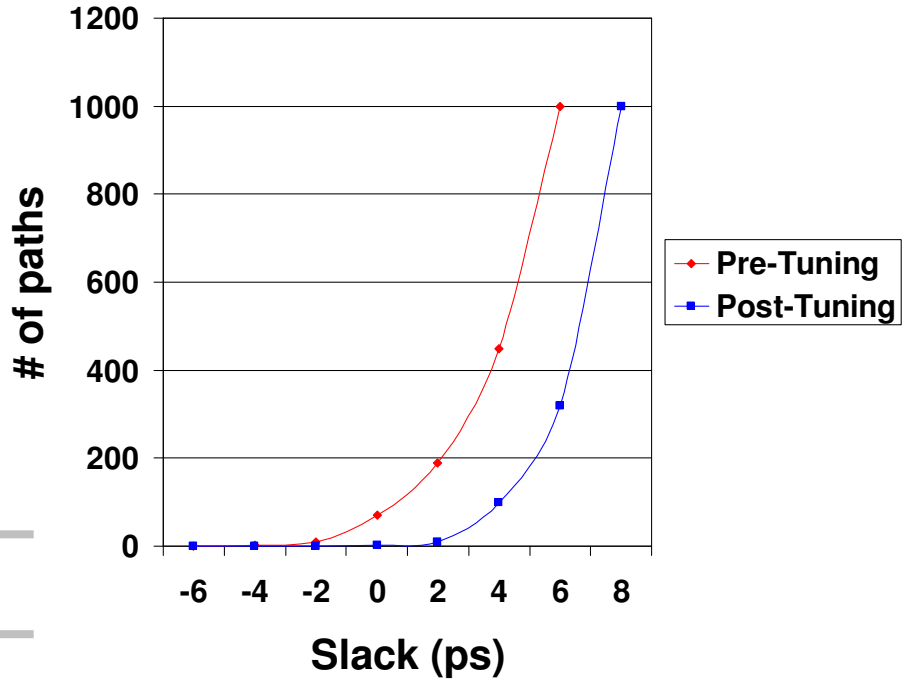
EinsTLT: Static timing analysis of complex circuits



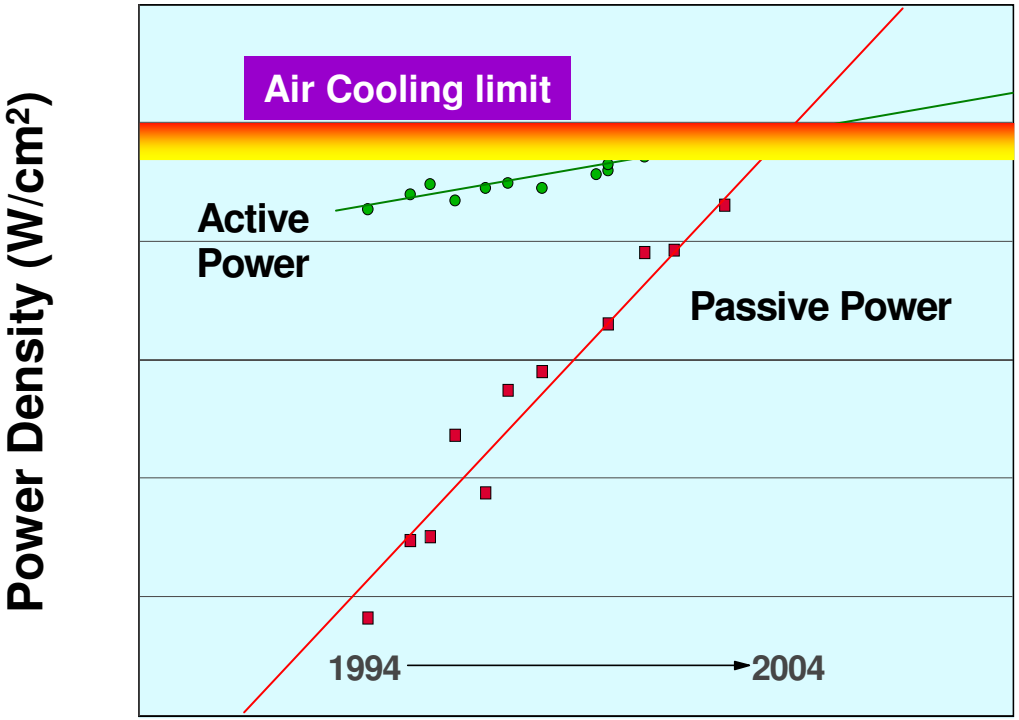
(a)



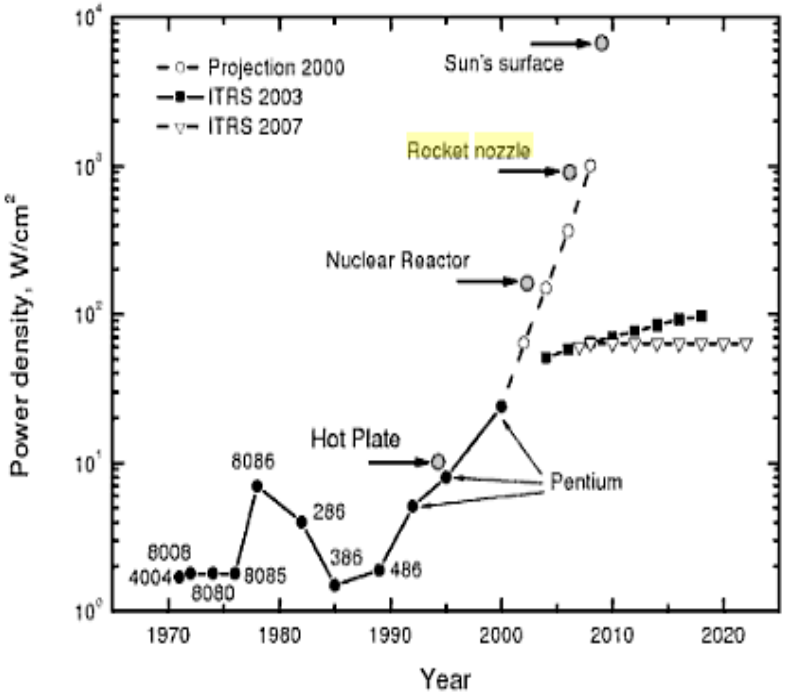
EinsTuner: Transistor Level timing optimization



End of an Era: The Power Wall



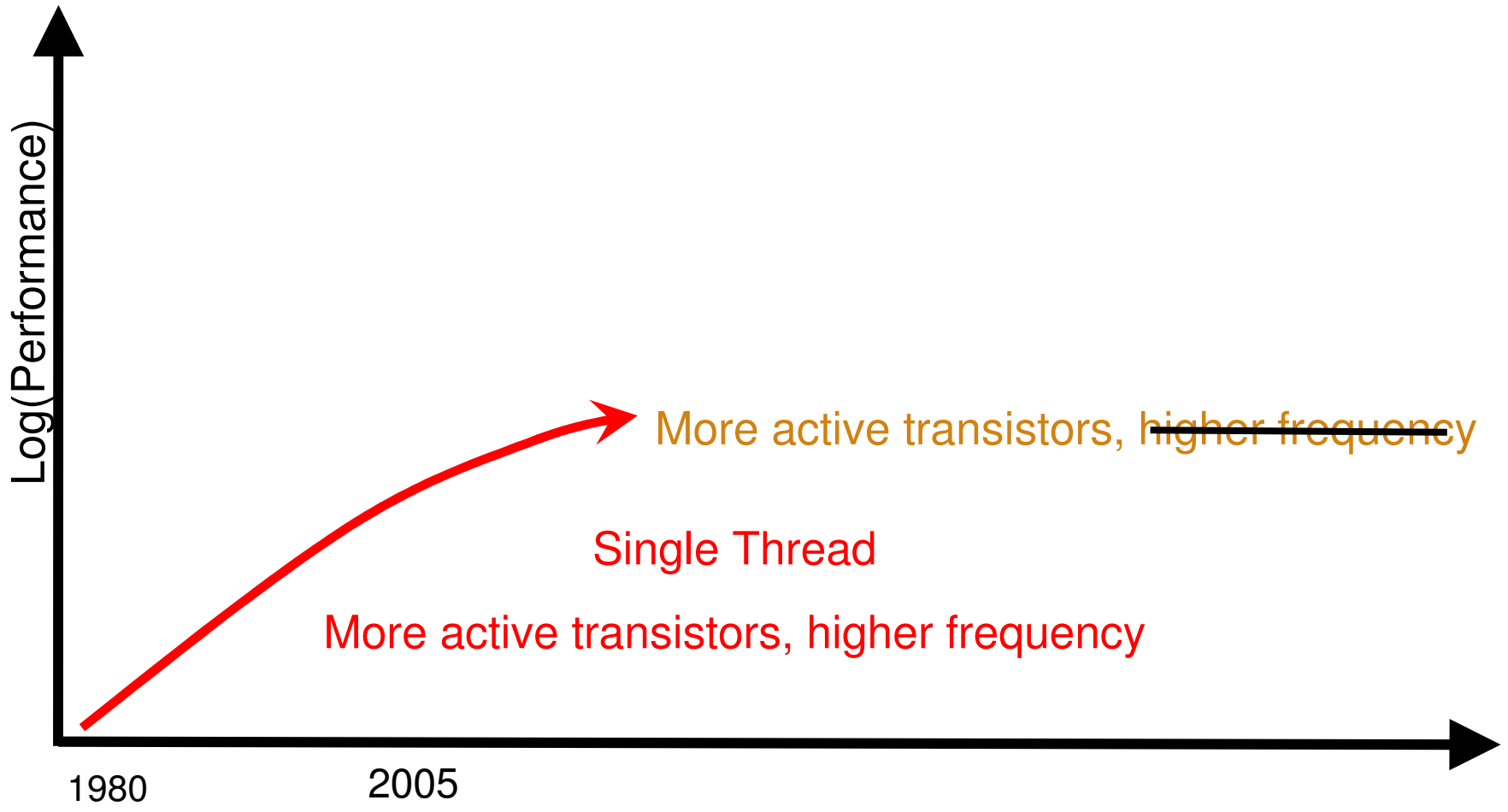
Source: B. Meyerson (IBM)



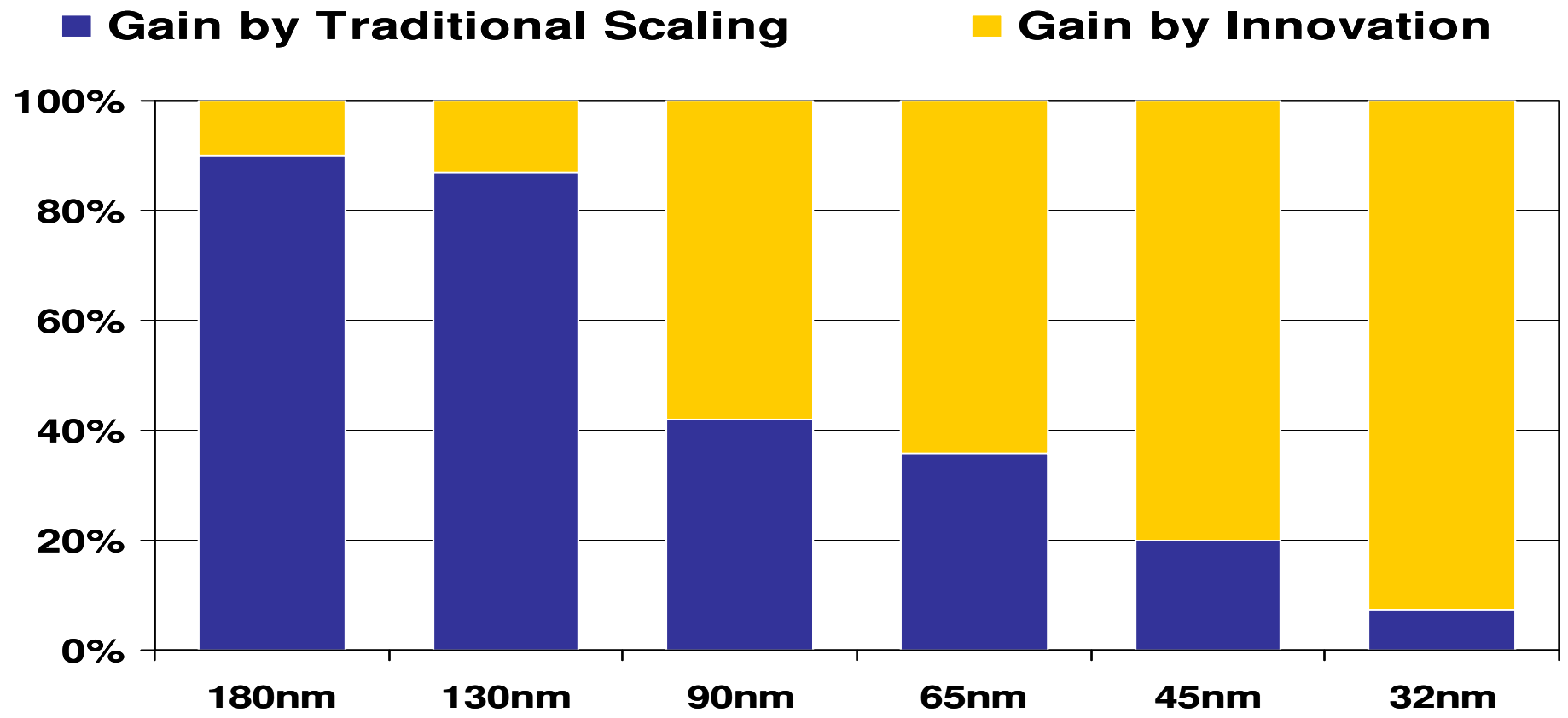
Source: S. Borkar (Intel)

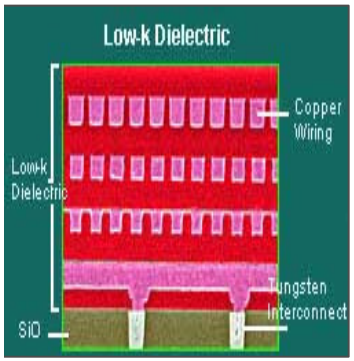
Inability to scale Tox & lower voltage resulted in a power wall for single thread performance

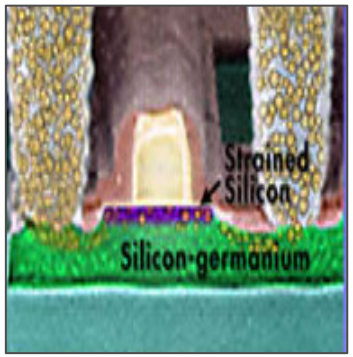
Microprocessor Trends

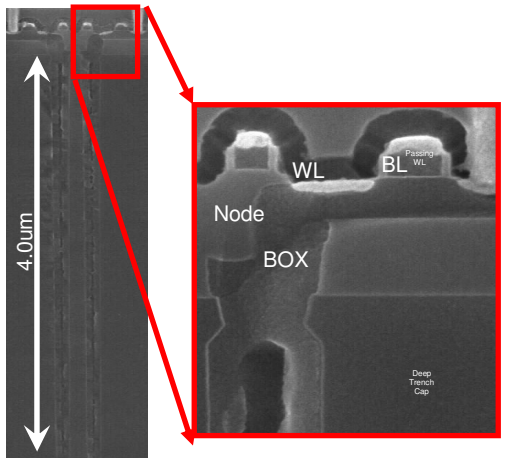


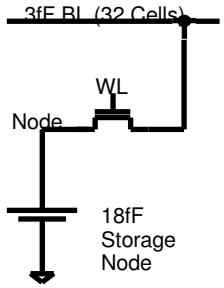
CMOS Scaling Continued

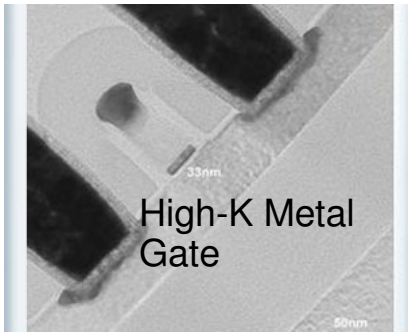




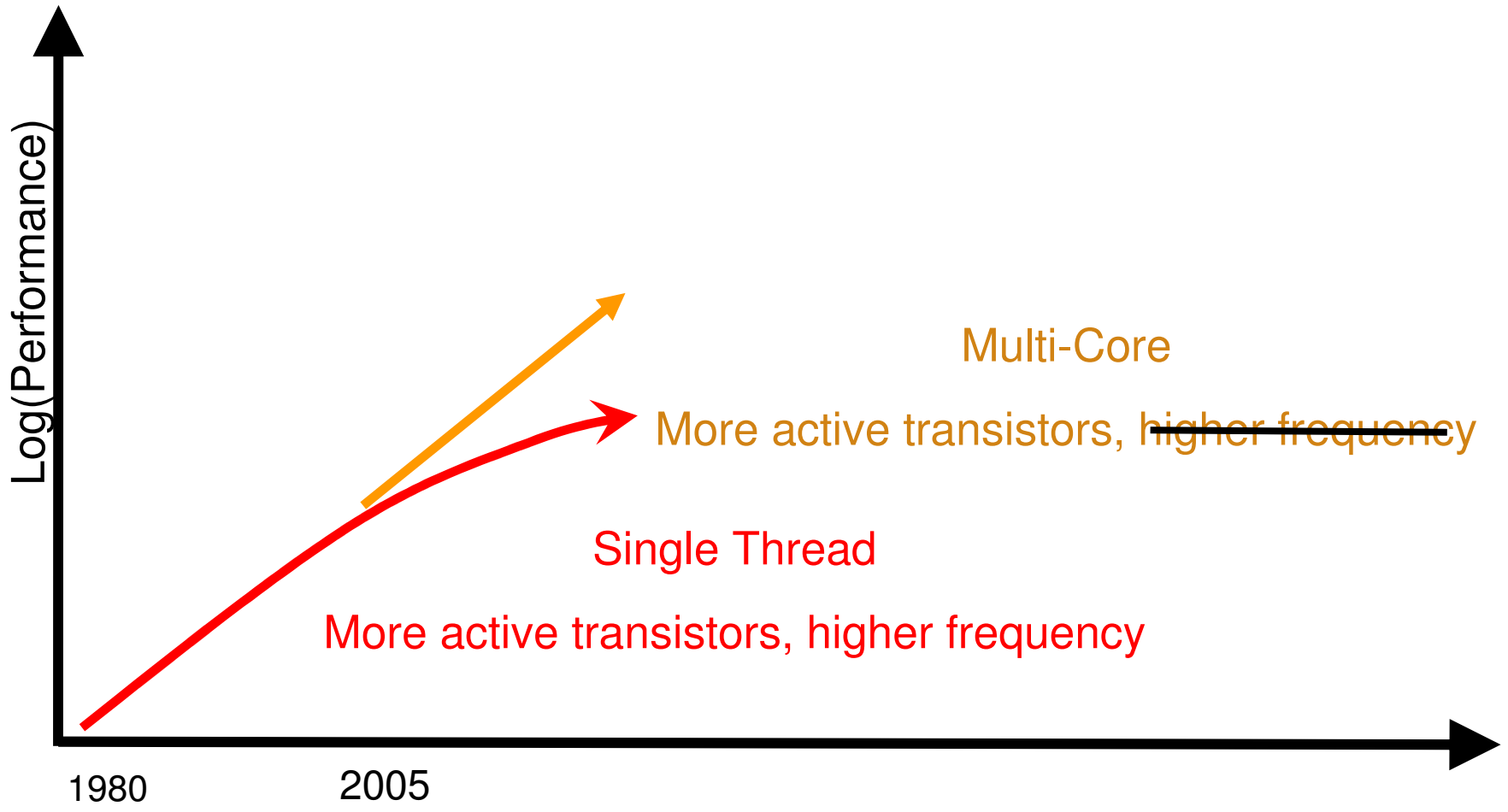




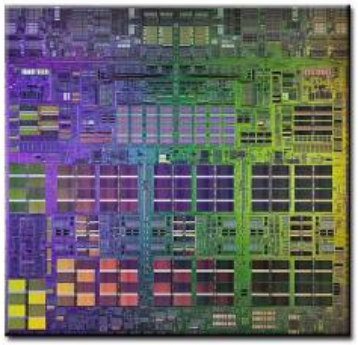




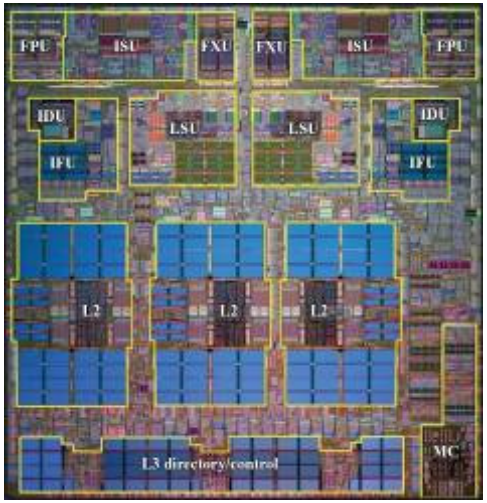
Microprocessor Trends



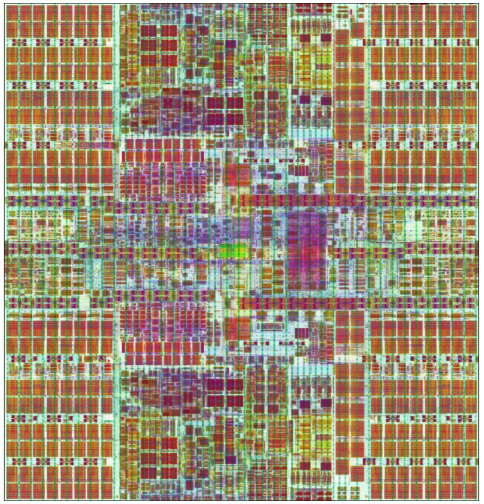
POWER Processors Began the Multi-Core / Multi-Thread Era



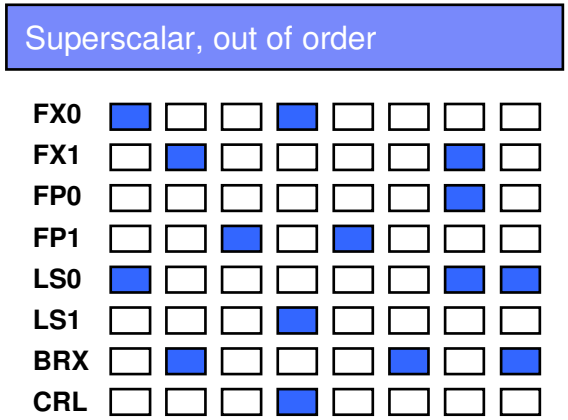
Power 4
2001
Introduces Dual core

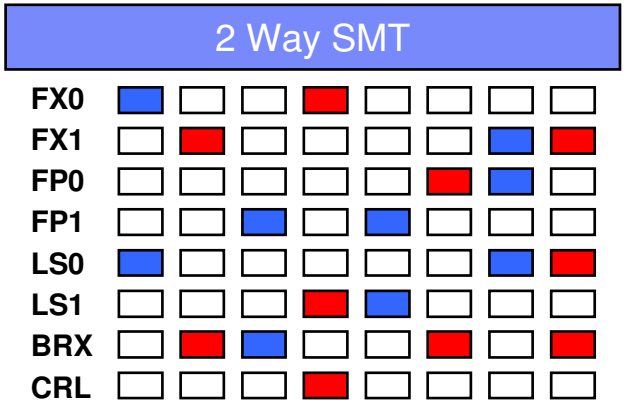


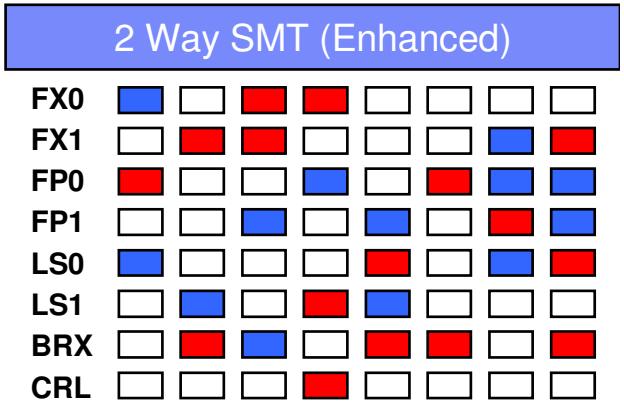
Power 5
2004
Dual Core
Introduces SMT (4 threads)



Power 6
2007
Dual Core – 4 threads
Enhances SMT Efficiency







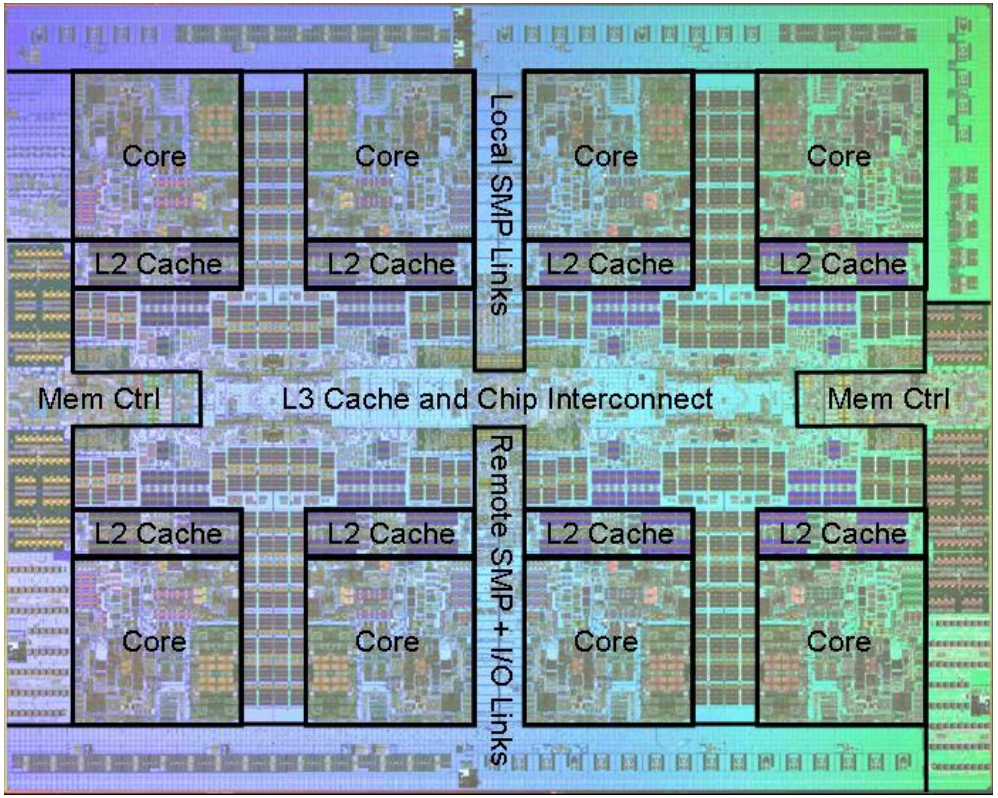
Thread 0 Executing

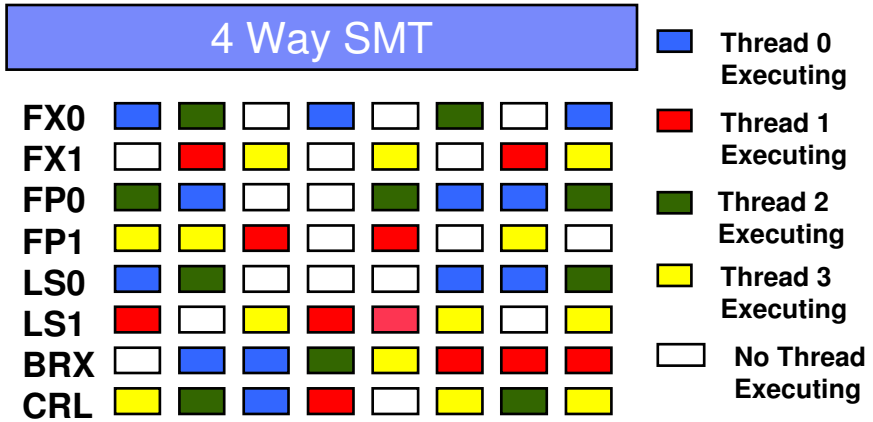
Thread 1 Executing

No Thread Executing

POWER7 Processor Chip

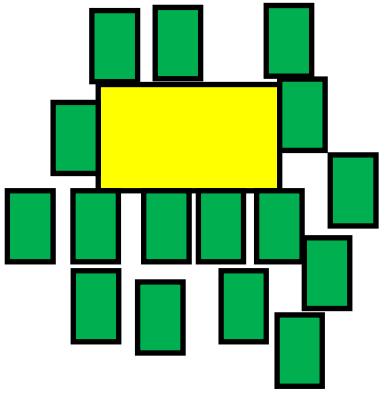
- 567mm^2, 45nm SOI w/ eDRAM
- 1.2B transistors
  - Equivalent function of 2.7B (eDRAM)
- Eight processor cores w/ 4 way SMT
  - 32 threads / chip
- 32MB on chip eDRAM shared L3
- Dual DDR3 Memory Controllers w/ 100GB/s sustained Memory bandwidth
- Scalability up to 32 Sockets
  - 360GB/s SMP bandwidth/chip
  - 20,000 coherent operations in flight





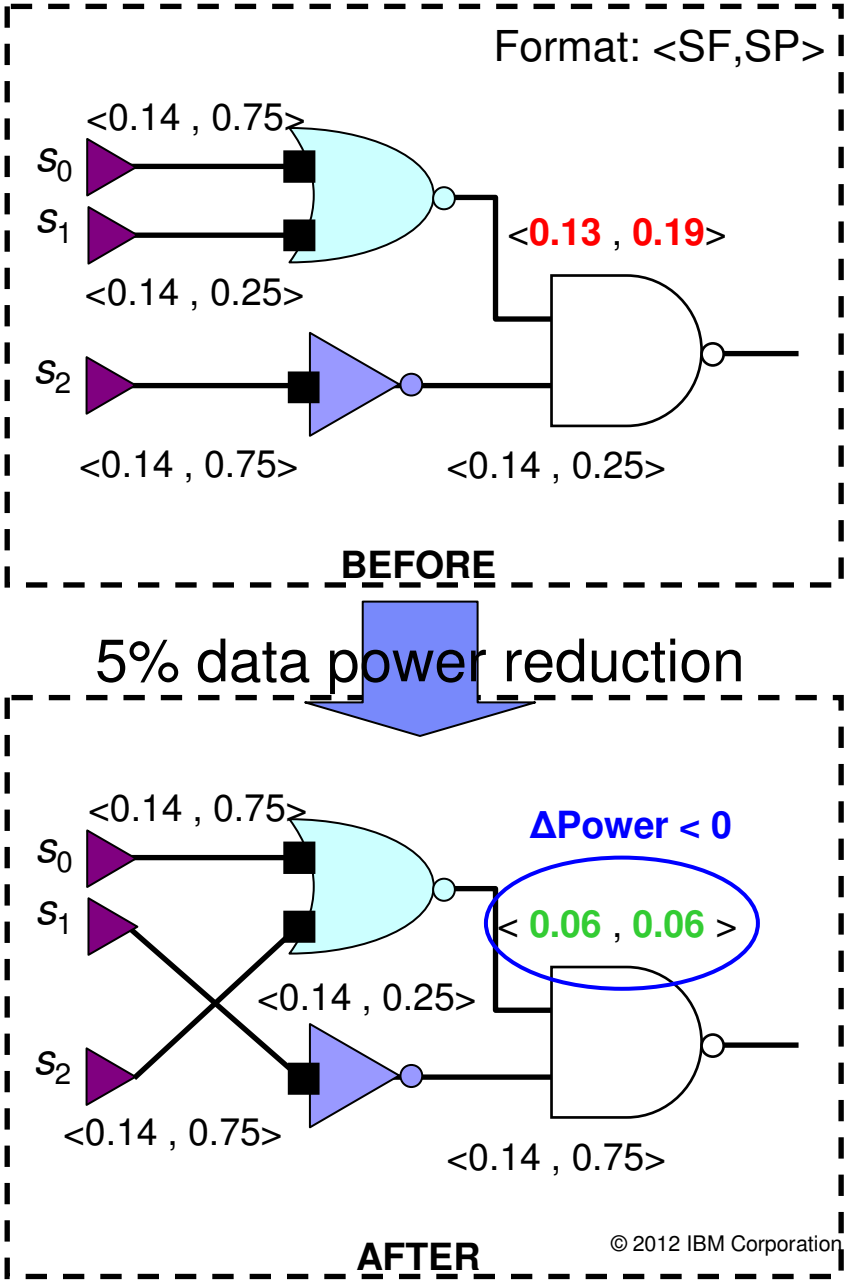
Multi-core Era EDA: Power Analysis & Reduction

Unstructured Clock Buffer & Latch

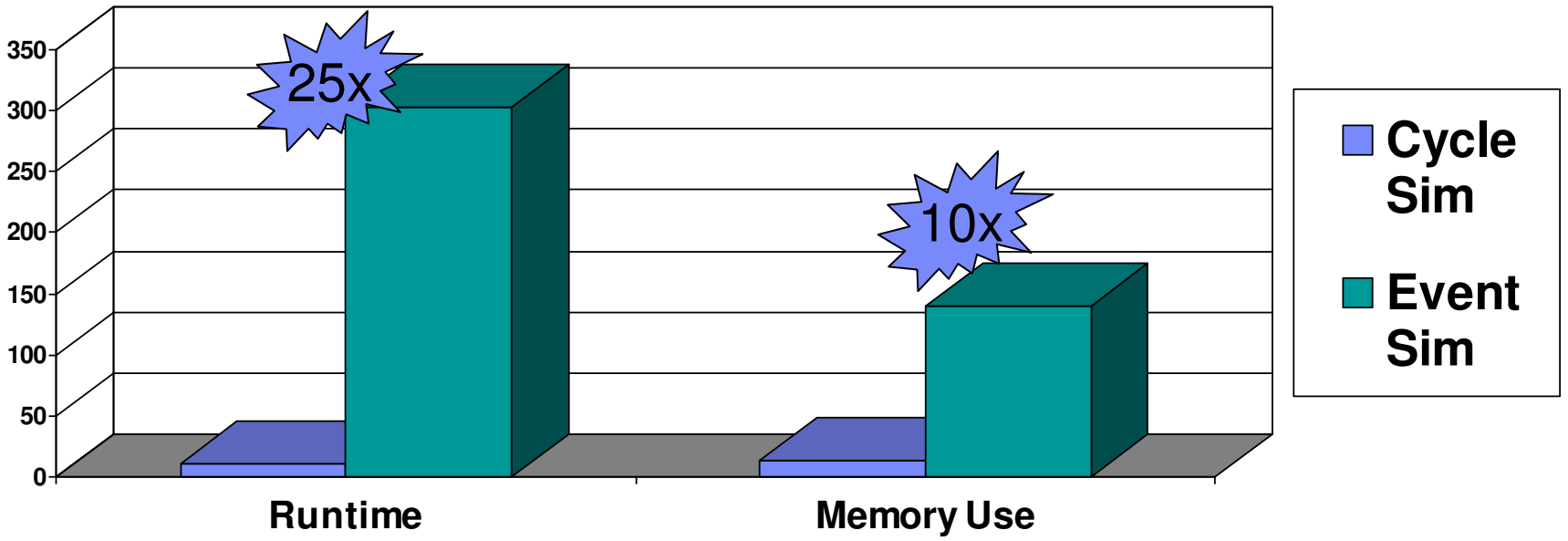


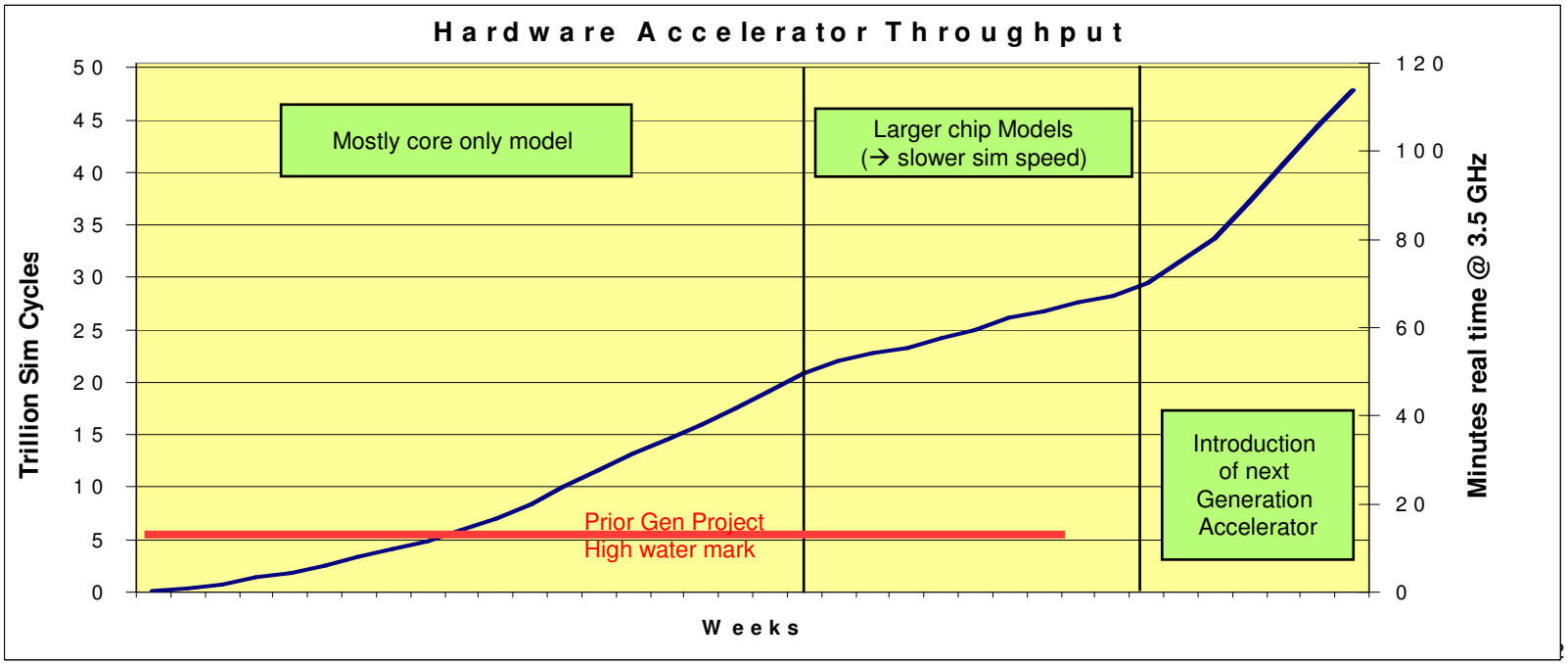
15% clock power reduction

Structured Clocking Construct

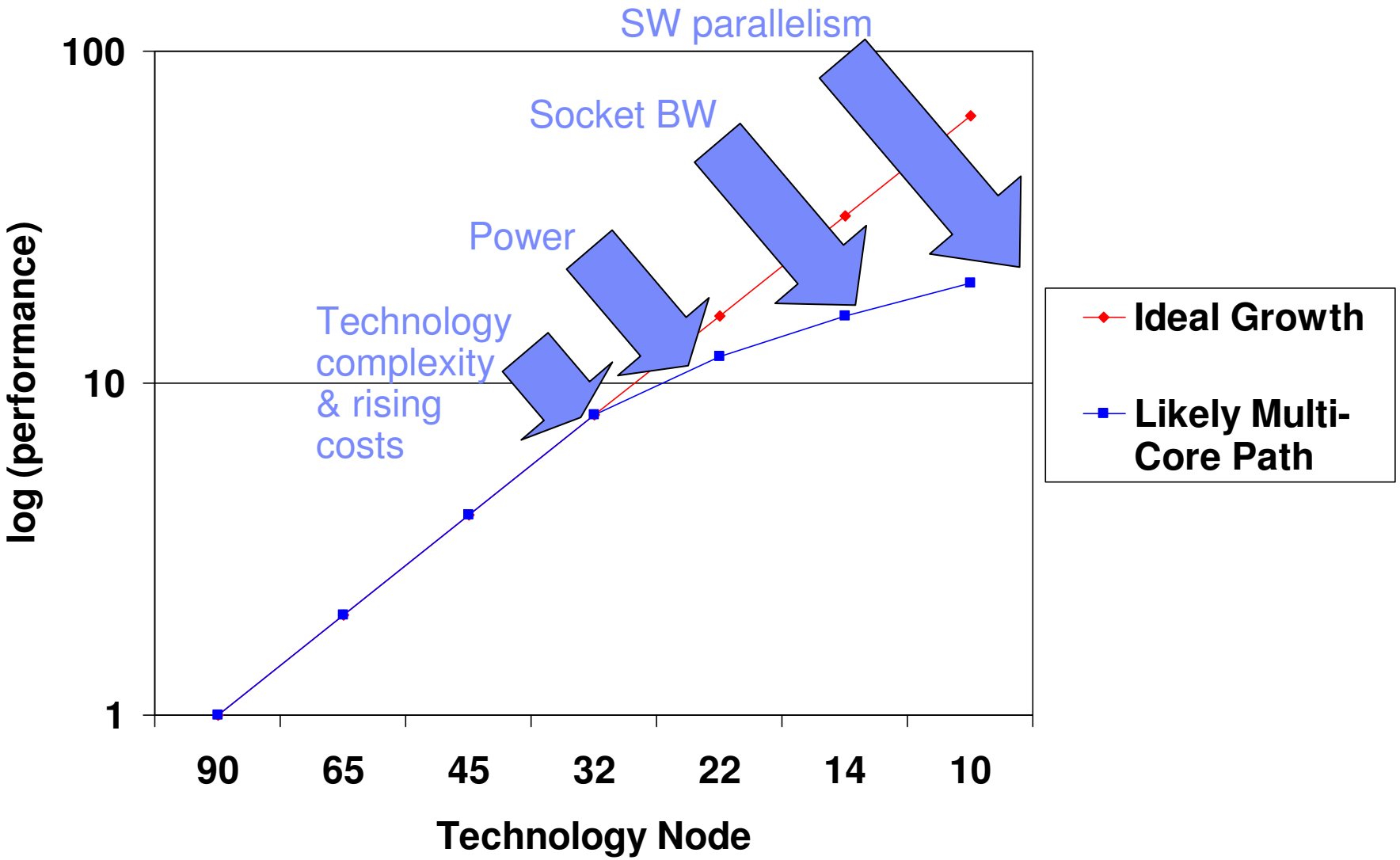


Multi-core Era EDA: Cycle Simulation

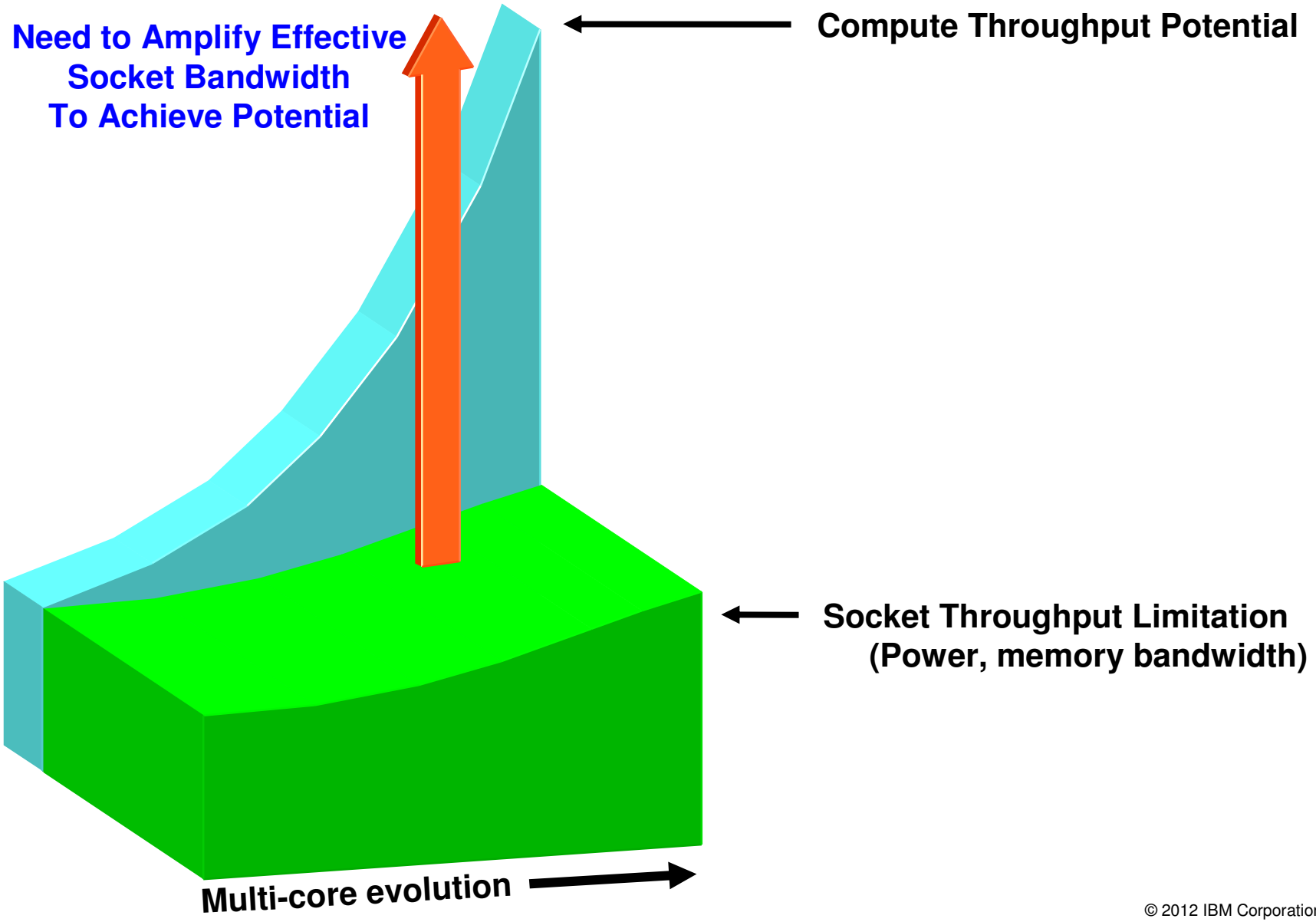




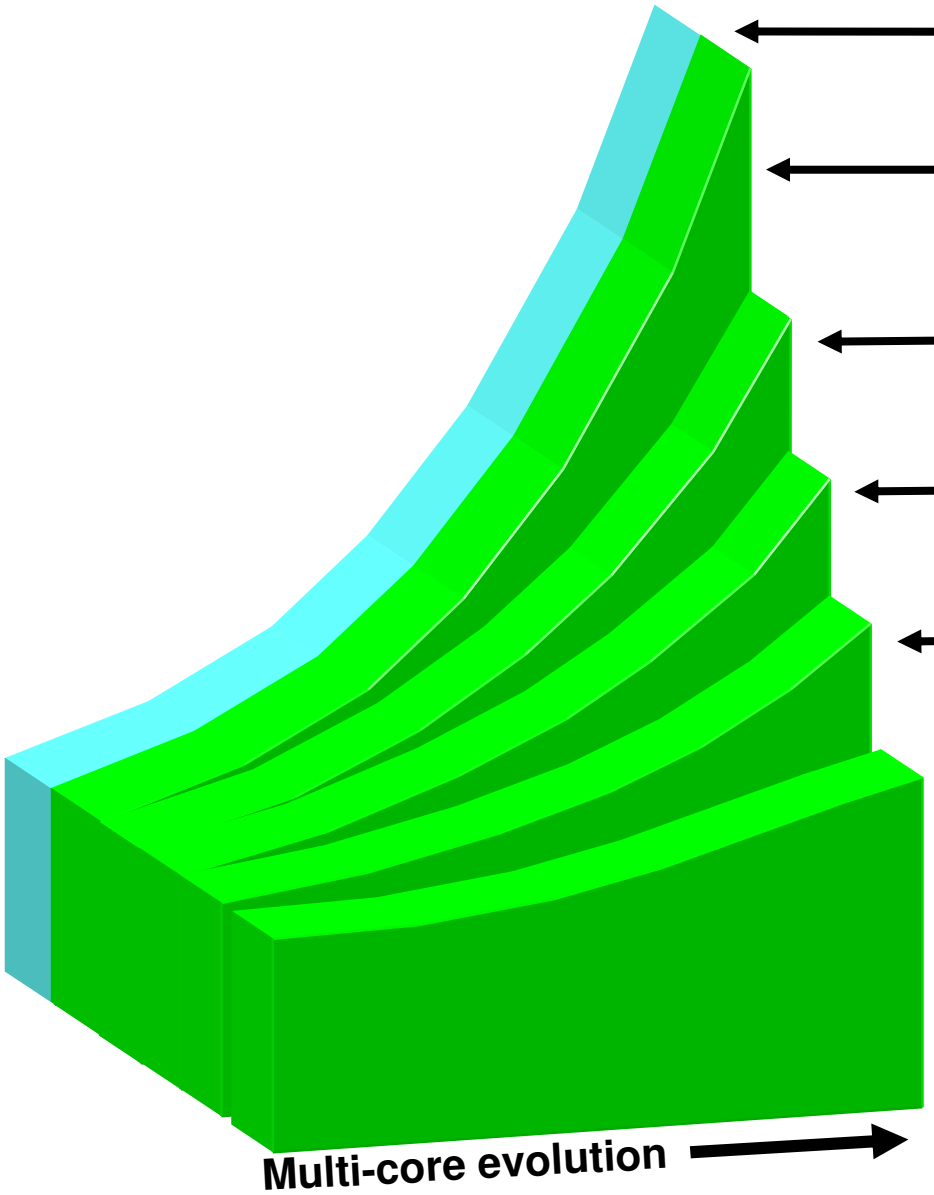
Multi-Core Era Limiters`



POWER's Multi-Core Advantage



POWER Architecture & Technology Will Extend Multi-Core Gains



Compute Throughput Potential

Coherence Innovation to minimize socket-to-socket communication

Low-Power Off-Chip Signaling Technology

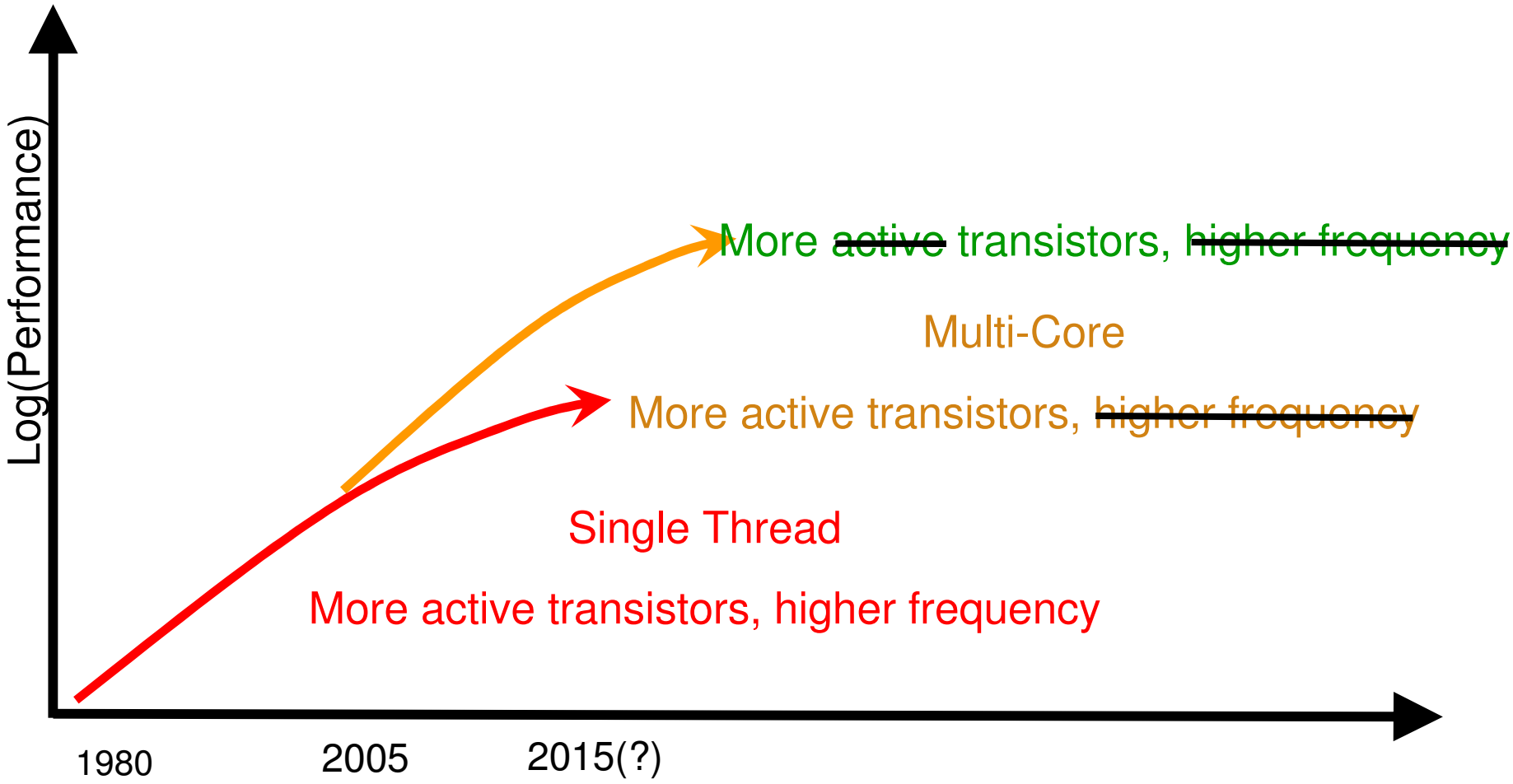
High bandwidth memory buffer

EDRAM = large, low power cache

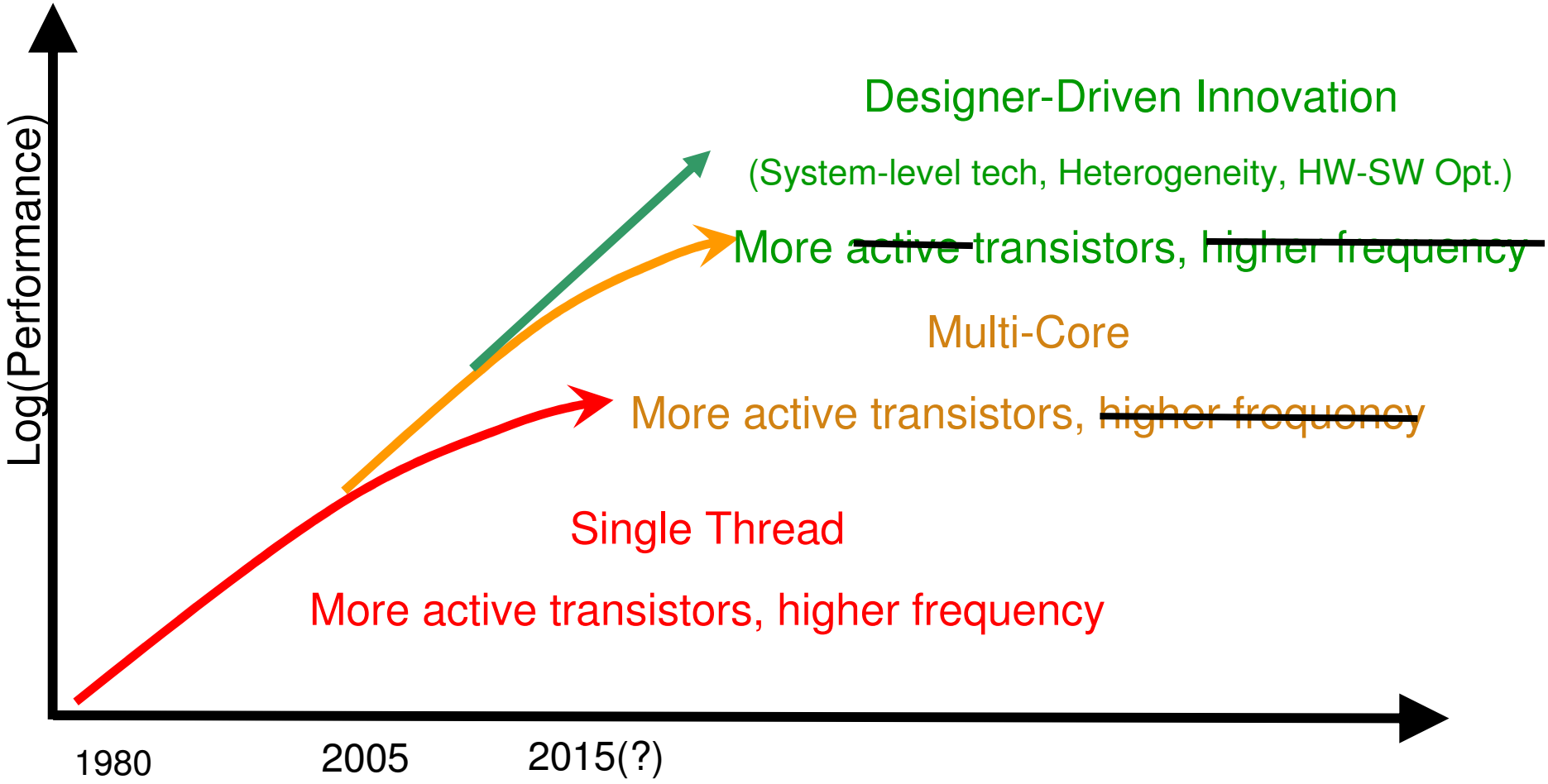
Socket Throughput Limitation (Power, memory bandwidth)

Multi-core evolution

Microprocessor Trends

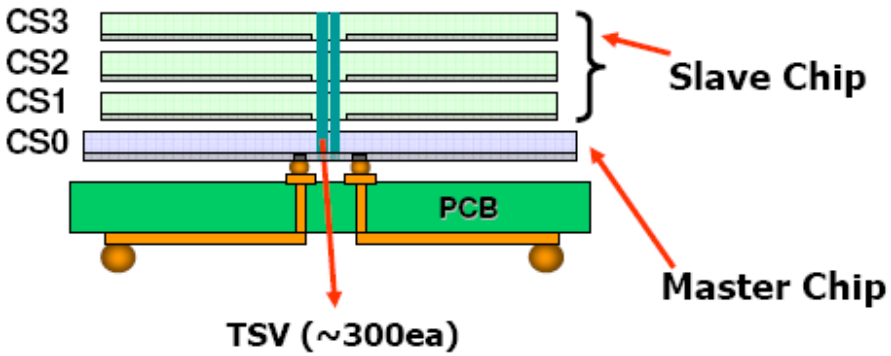


Microprocessor Trends: The Next Era

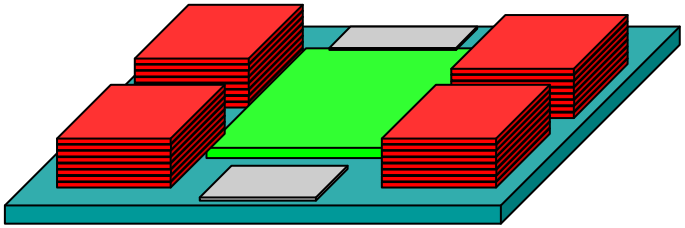


Innovation Era: System Level Technologies

3D Stacking with Through Silicon Vias

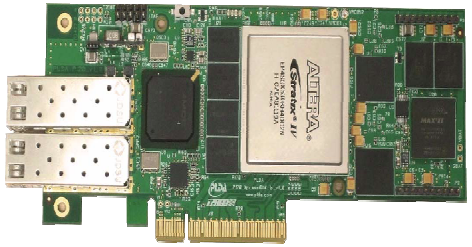


Single Processor–Memory Socket

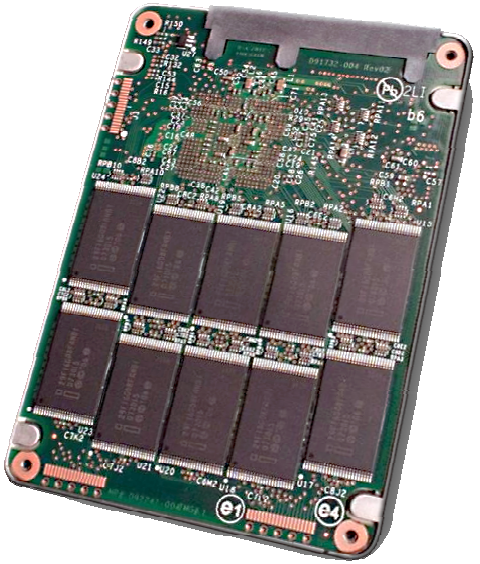


Silicon Photonics

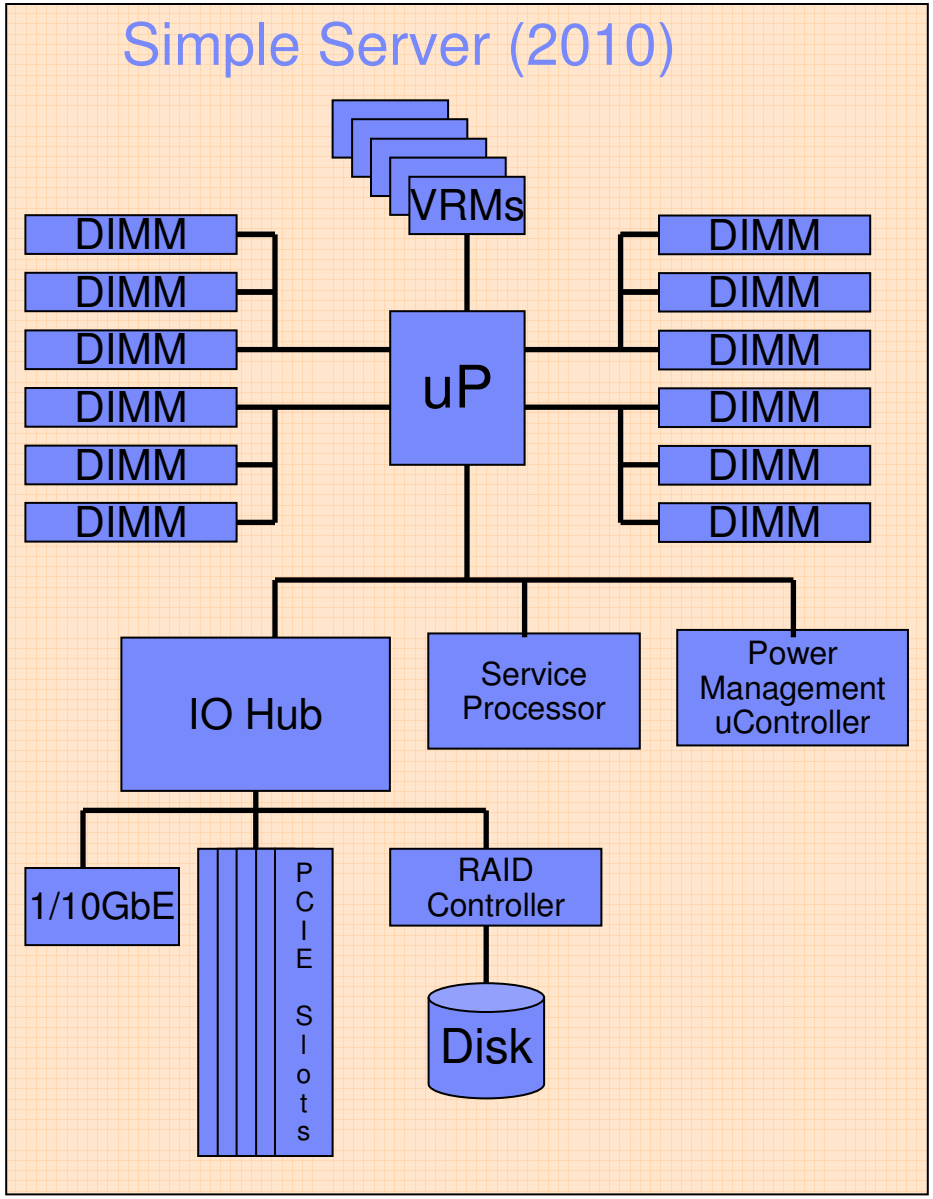
FPGA Accelerators

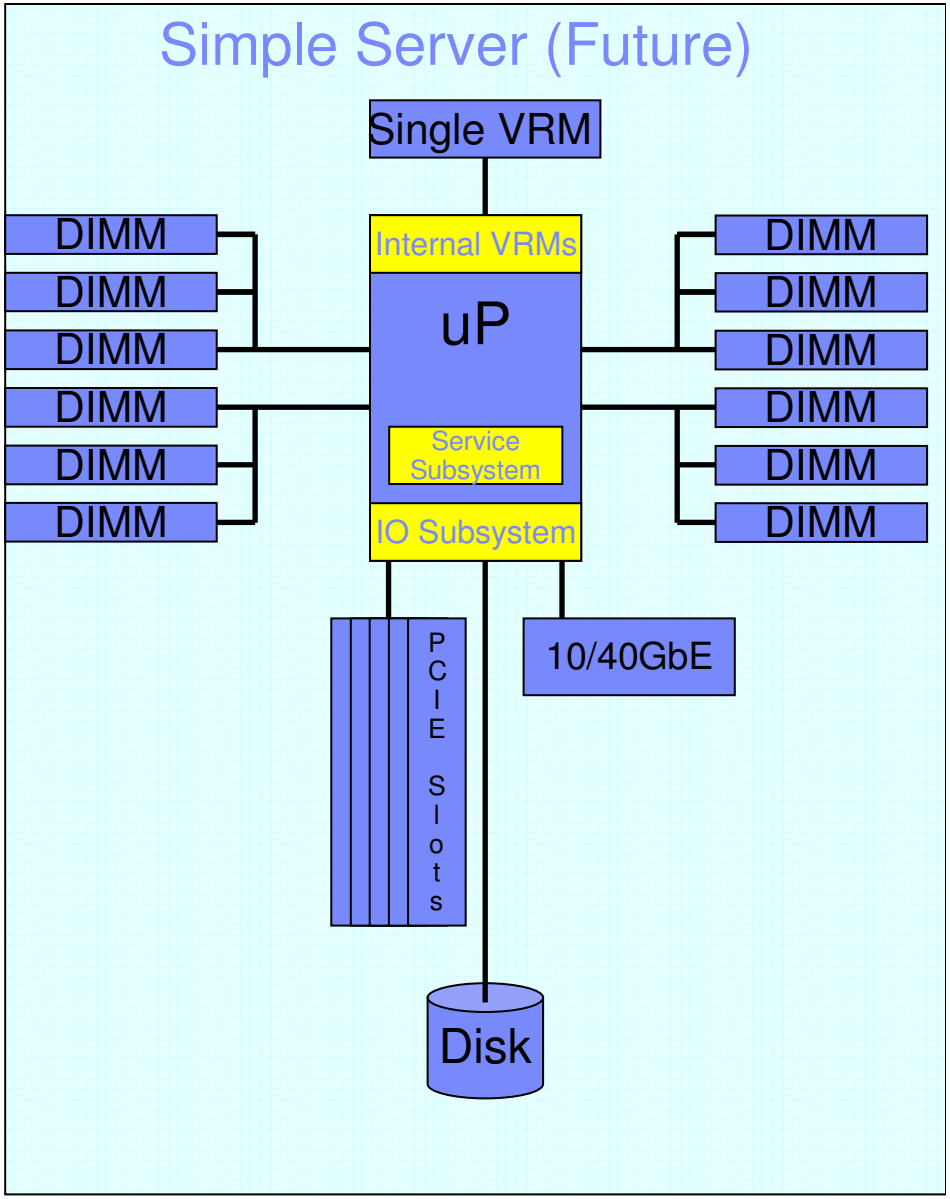


Flash Memory / SSD

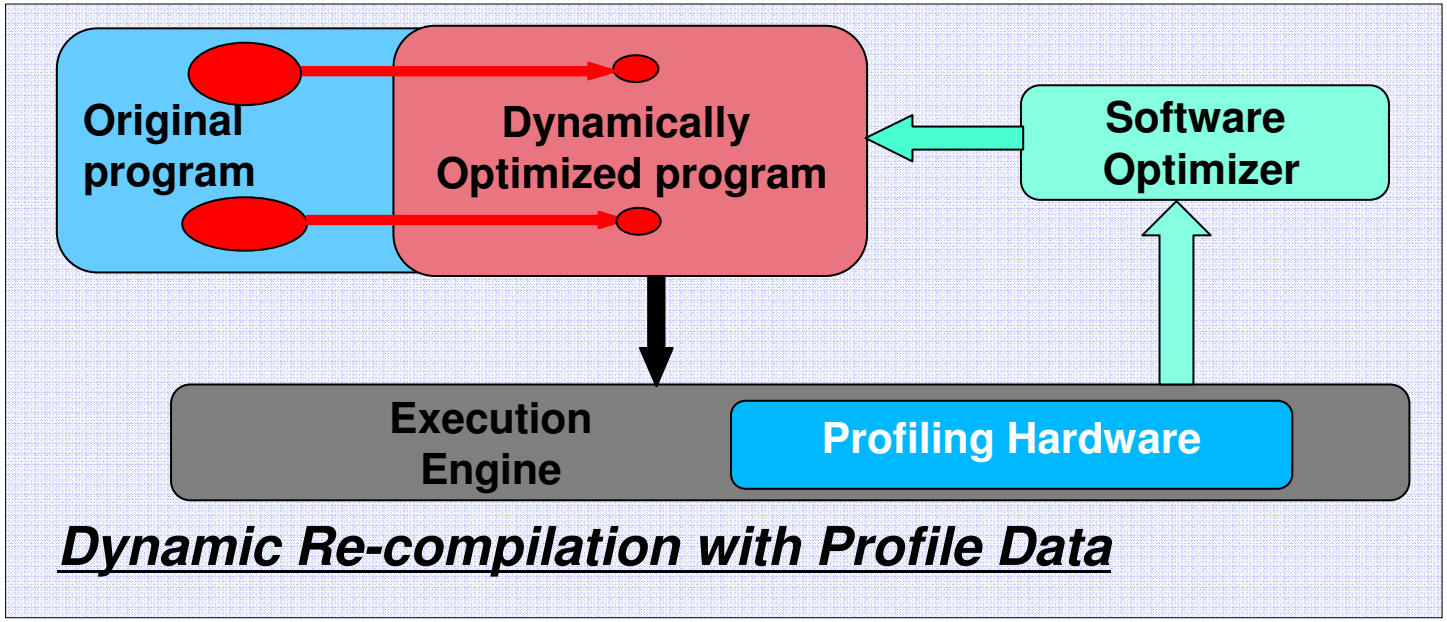


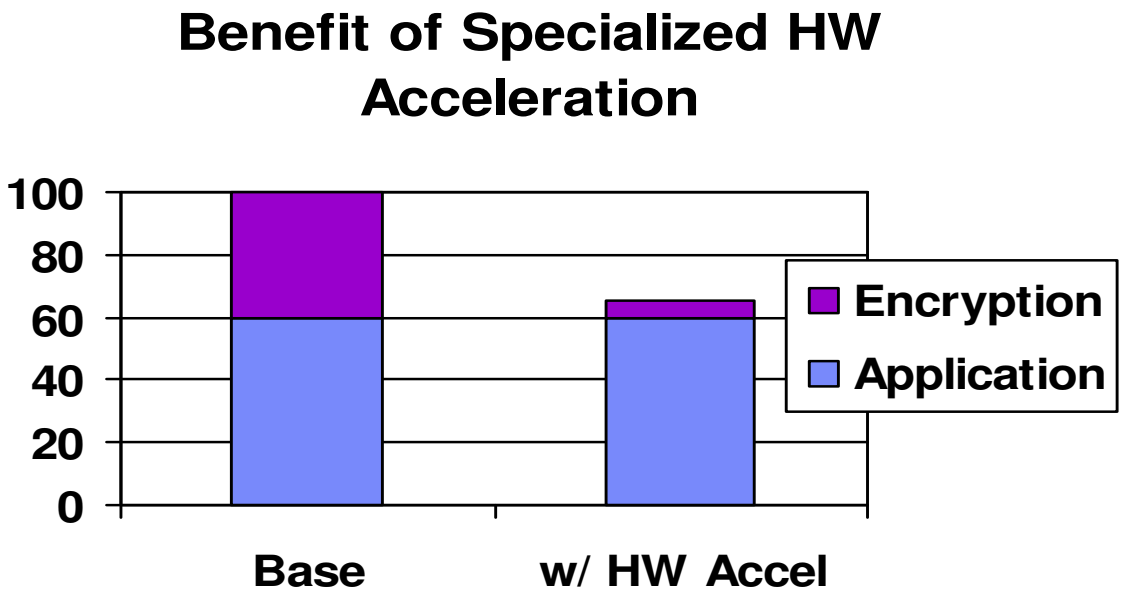
Innovation Era: Heterogeneous Integration





Innovation Era: Hardware-Software Co-Optimization



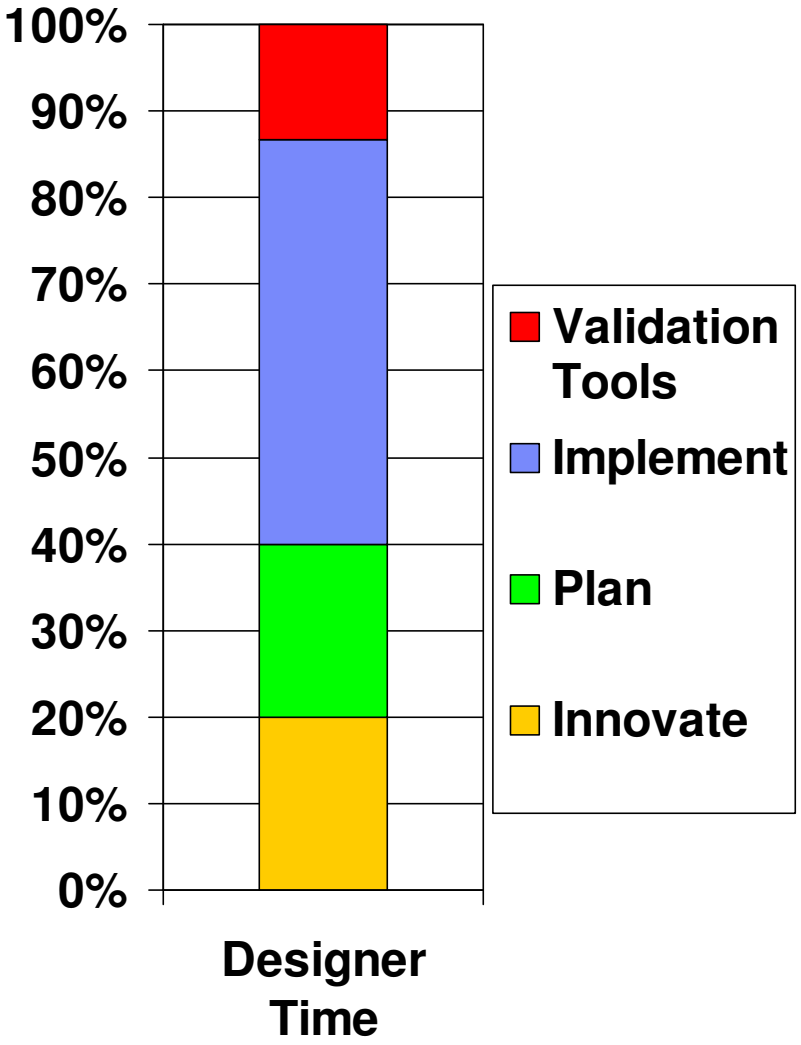


Specialized HW acceleration viable in many areas

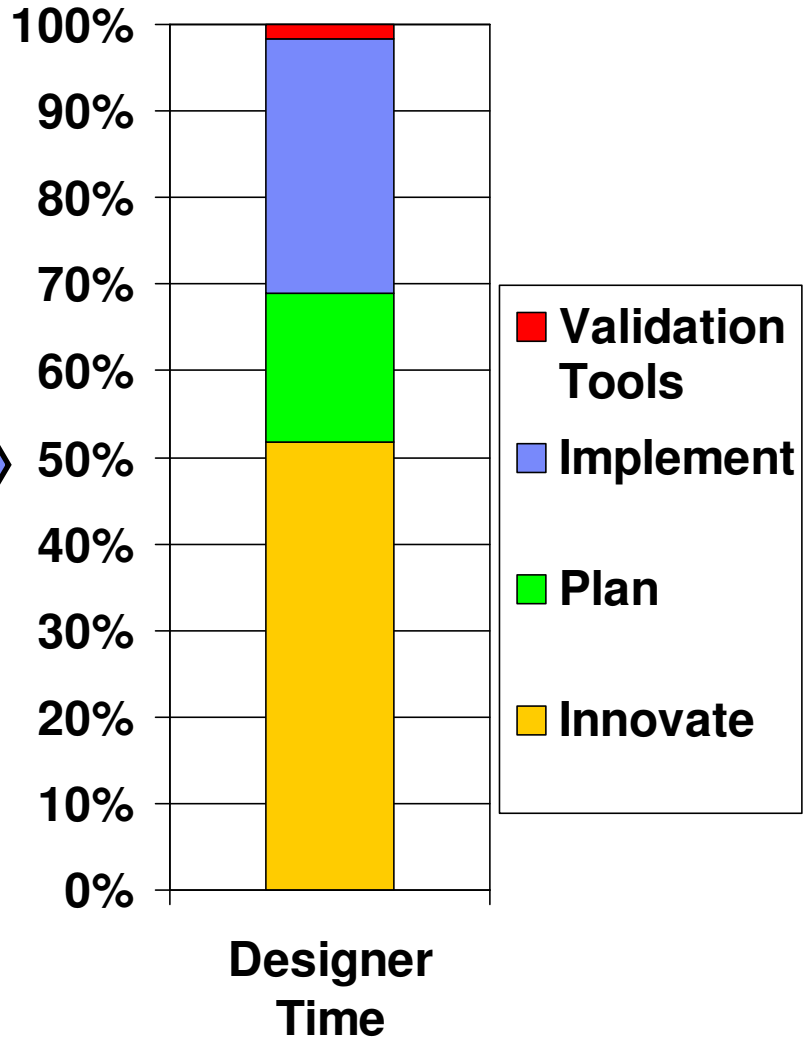
- Graphics
- Compression
- Cryptography
- More….

Designer Time: Technology-Driven Eras vs. Innovation Era

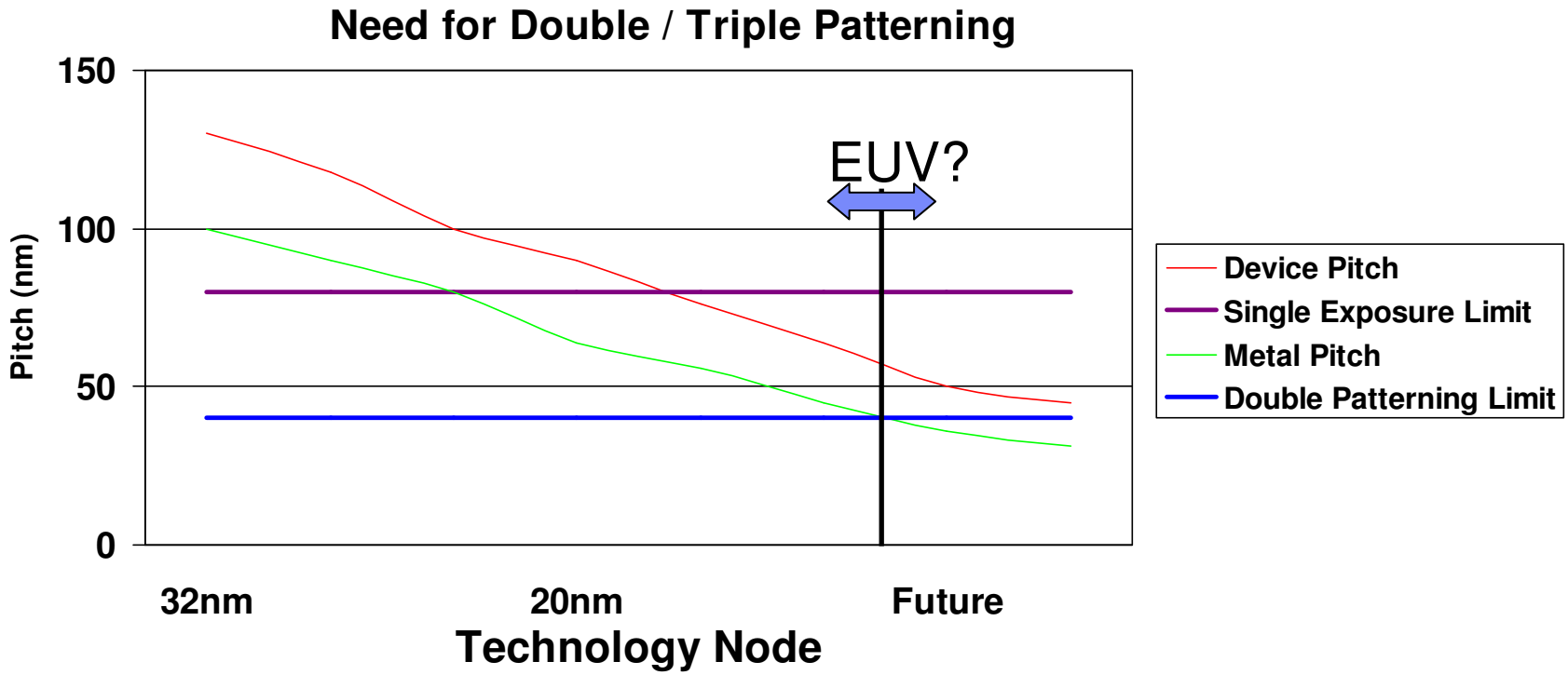
Technology Focused

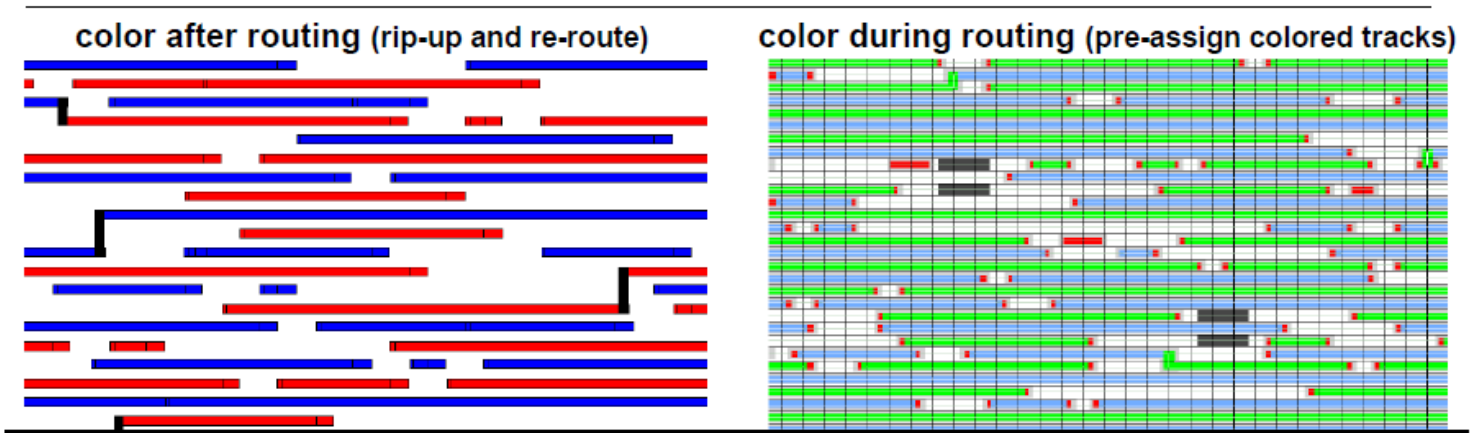


Innovation Focused

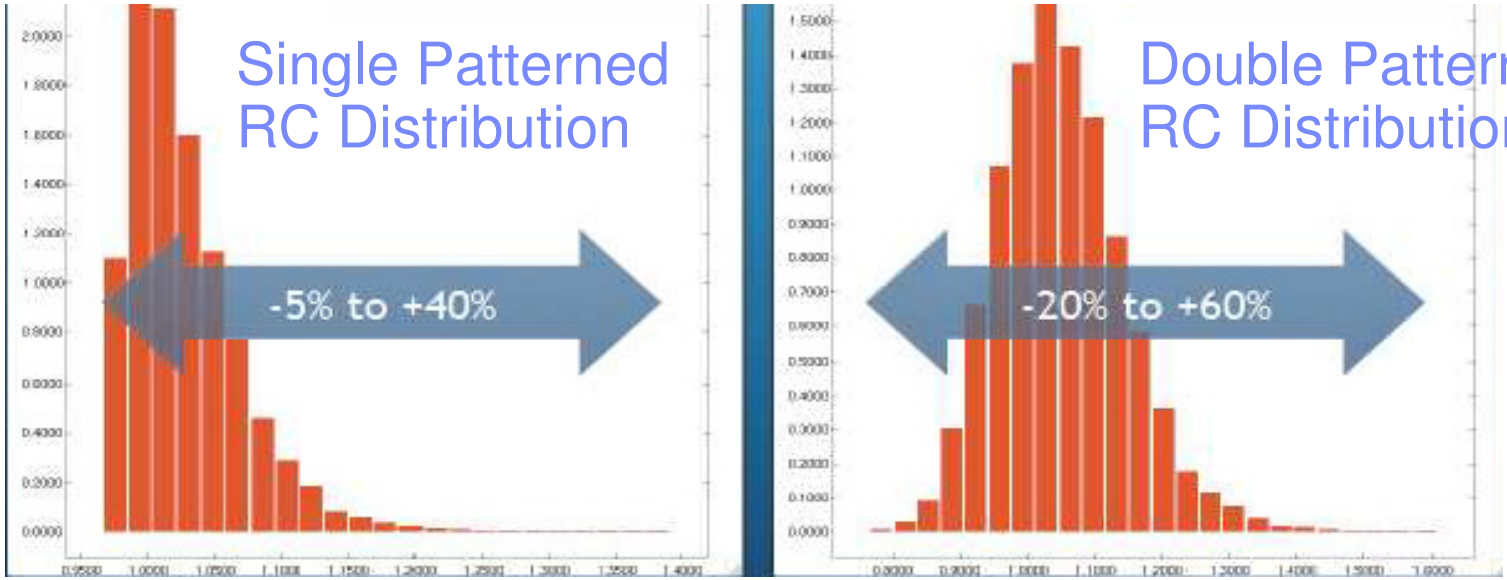


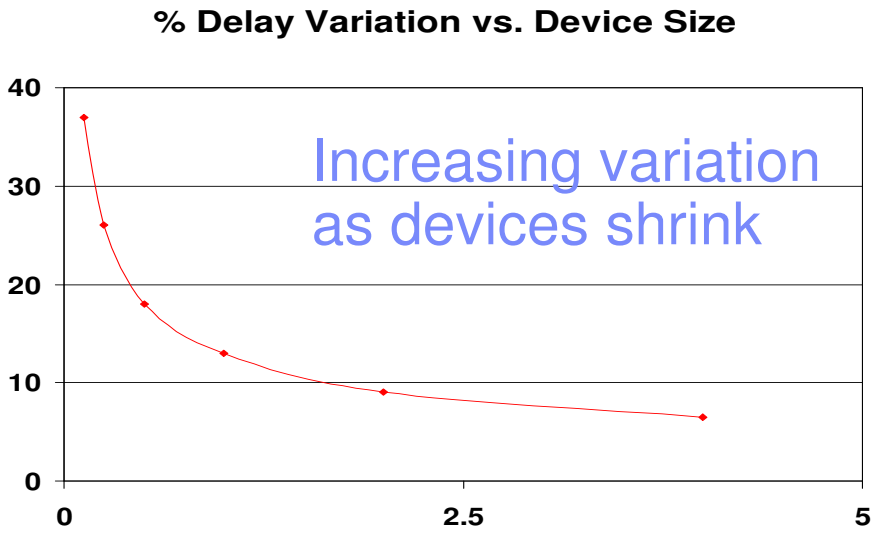
Manage Technology Complexity: Double Patterning

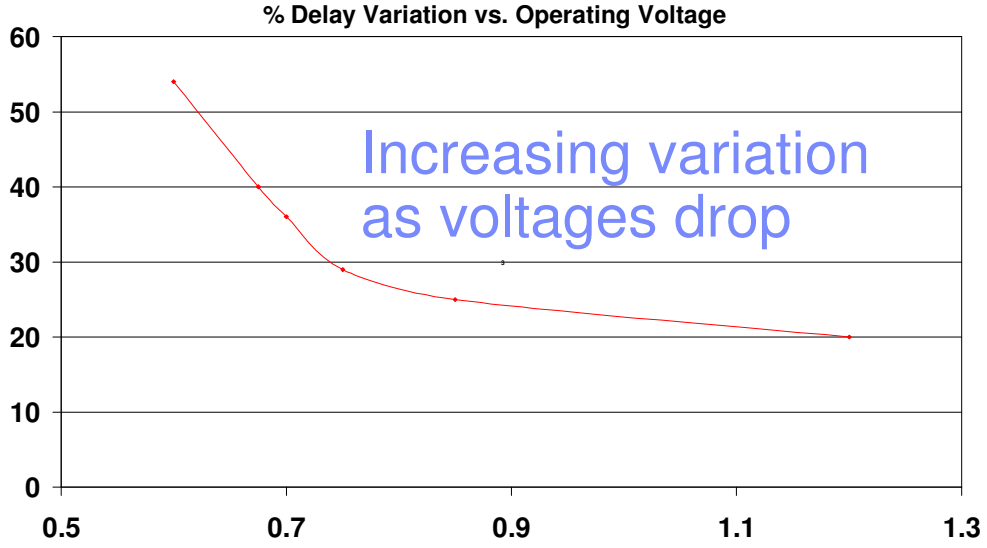




Manage Technology Complexity: Extreme Variability







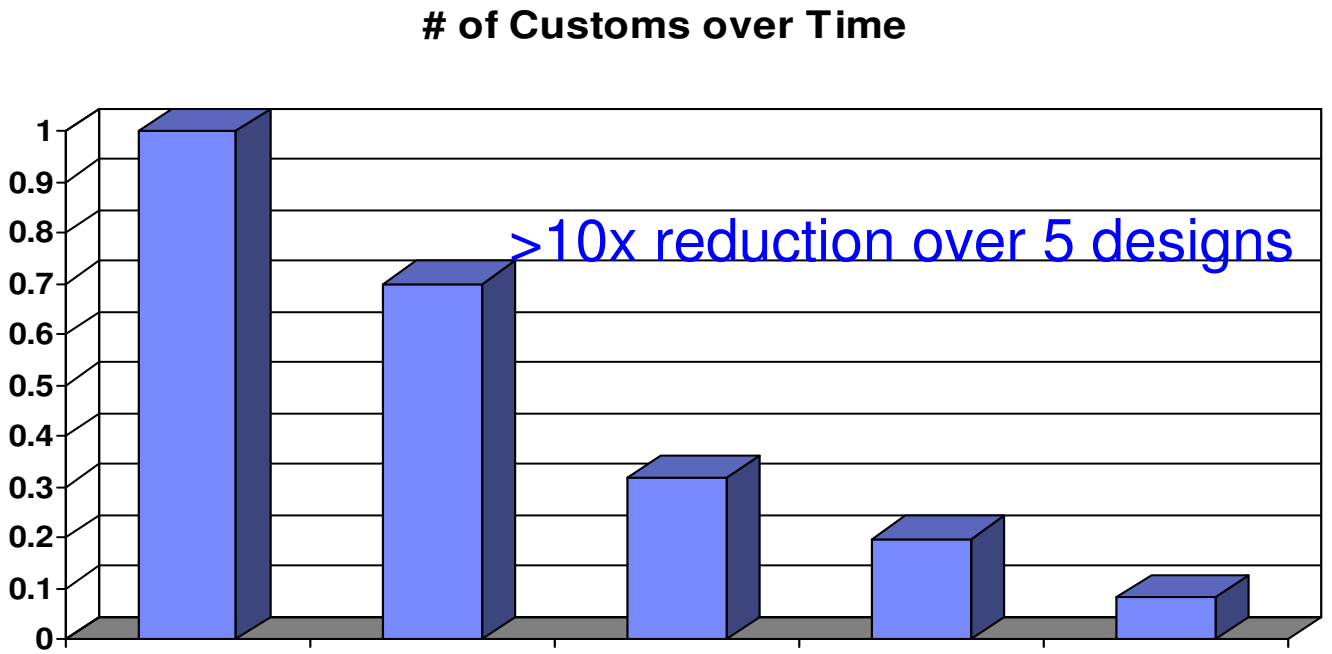
Variability grows rapidly w/ each technology generation.

Productivity & TAT: Merge ASIC Productivity with Custom Performance

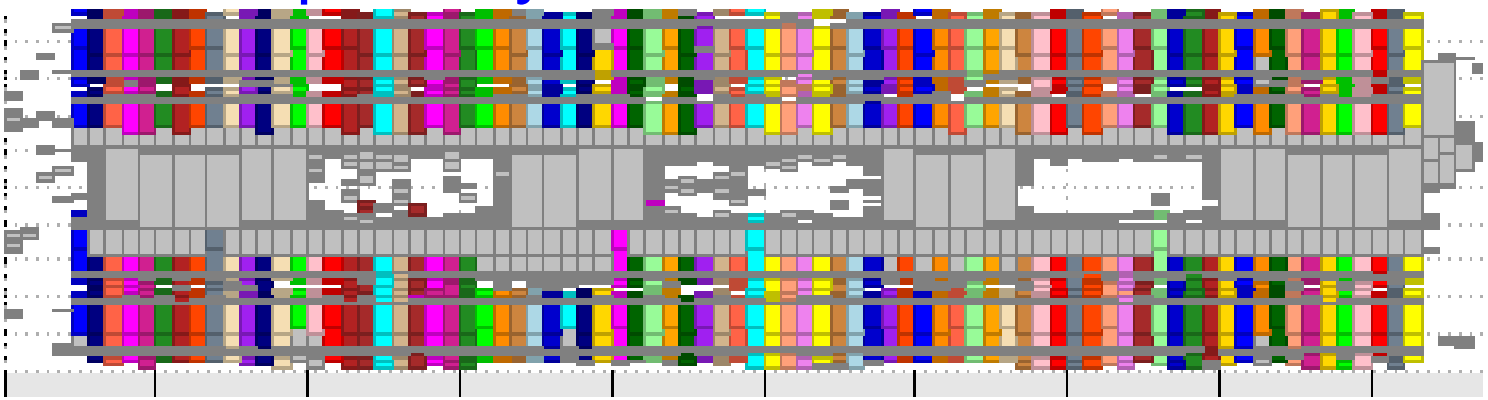
- **Custom Processor Methodology**
  - Manual, custom design
  - Tool support for Array / analog
  - Transistor level analysis
  - Clock mesh (low skew)
  - Able to engineer wires & placement
  - Small macros
- **ASIC Methodology**
  - Synthesis Dominated
  - Blackbox array / analog
  - Gate level analysis
  - Clock trees (high skew)
  - Auto-place / auto-route only
  - Large blocks

Custom performance w/ ASIC productivity

Productivity: Reduce Custom Design



Datapath Synthesis Results

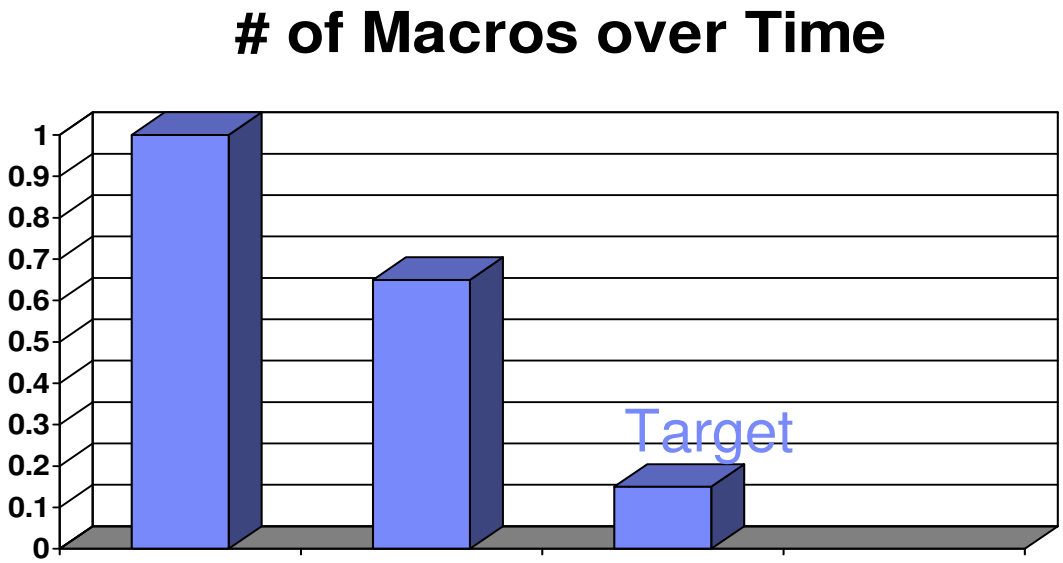


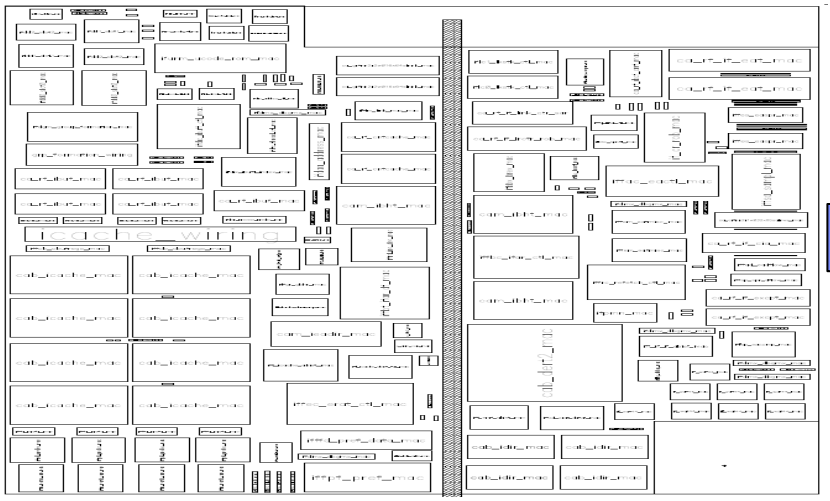
custom-like data flow alignment

91% density

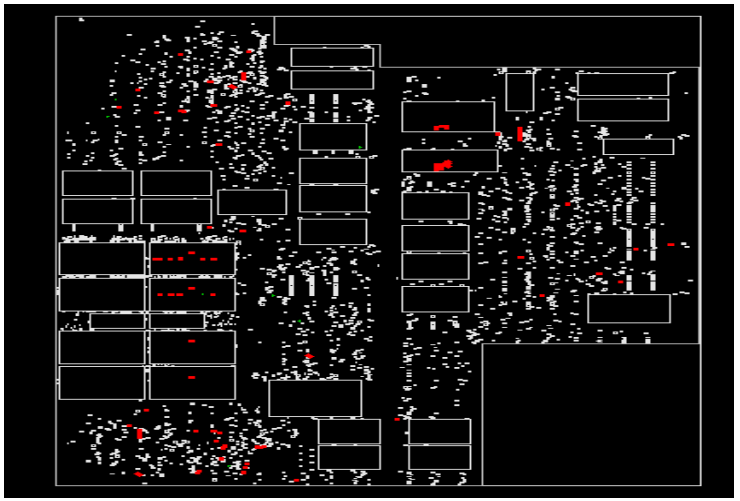
Each color indicates cells associated w/ a different logical bit.

Productivity: Reduce # of Design Partitions



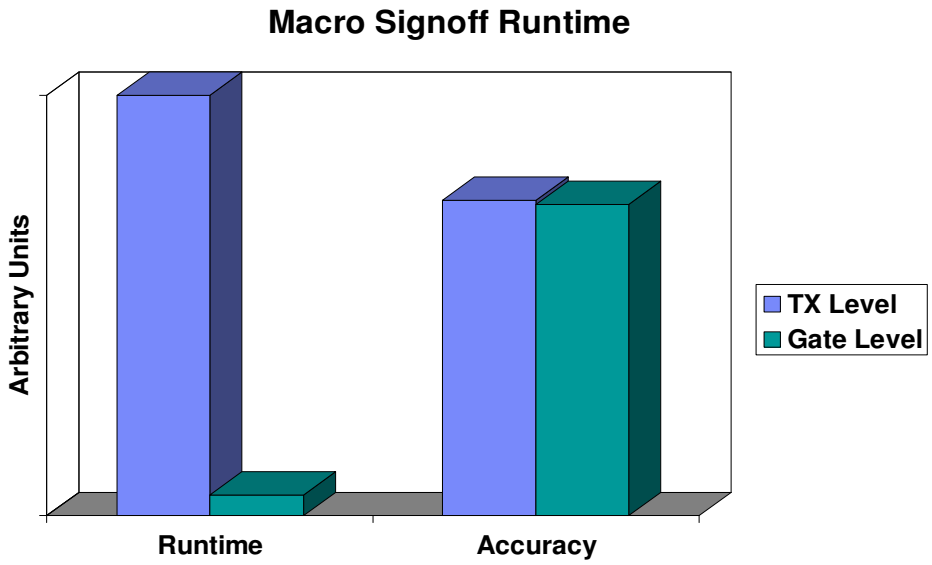


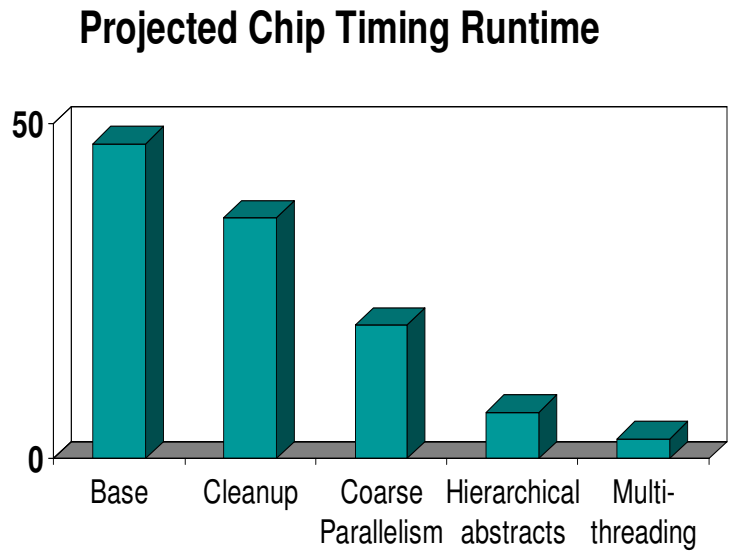
60 logic macros, 25 customs, 14 unique arrays/RFs



1 macro, 0 customs, 9 unique arrays / RFs
Reduced area & power; equal cycle time

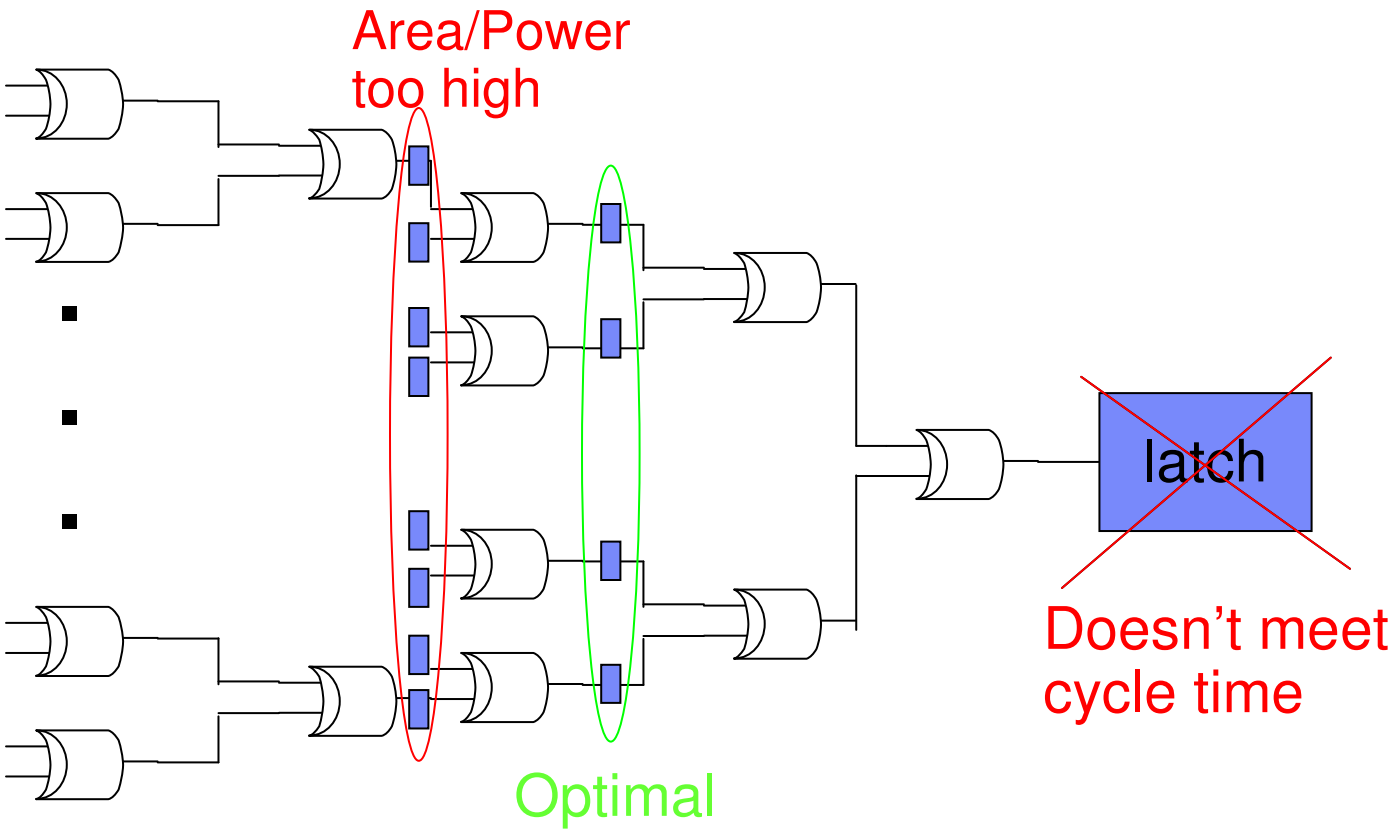
TAT: Gate-Level Analysis, Hierarchical Abstraction & Multi-Threading





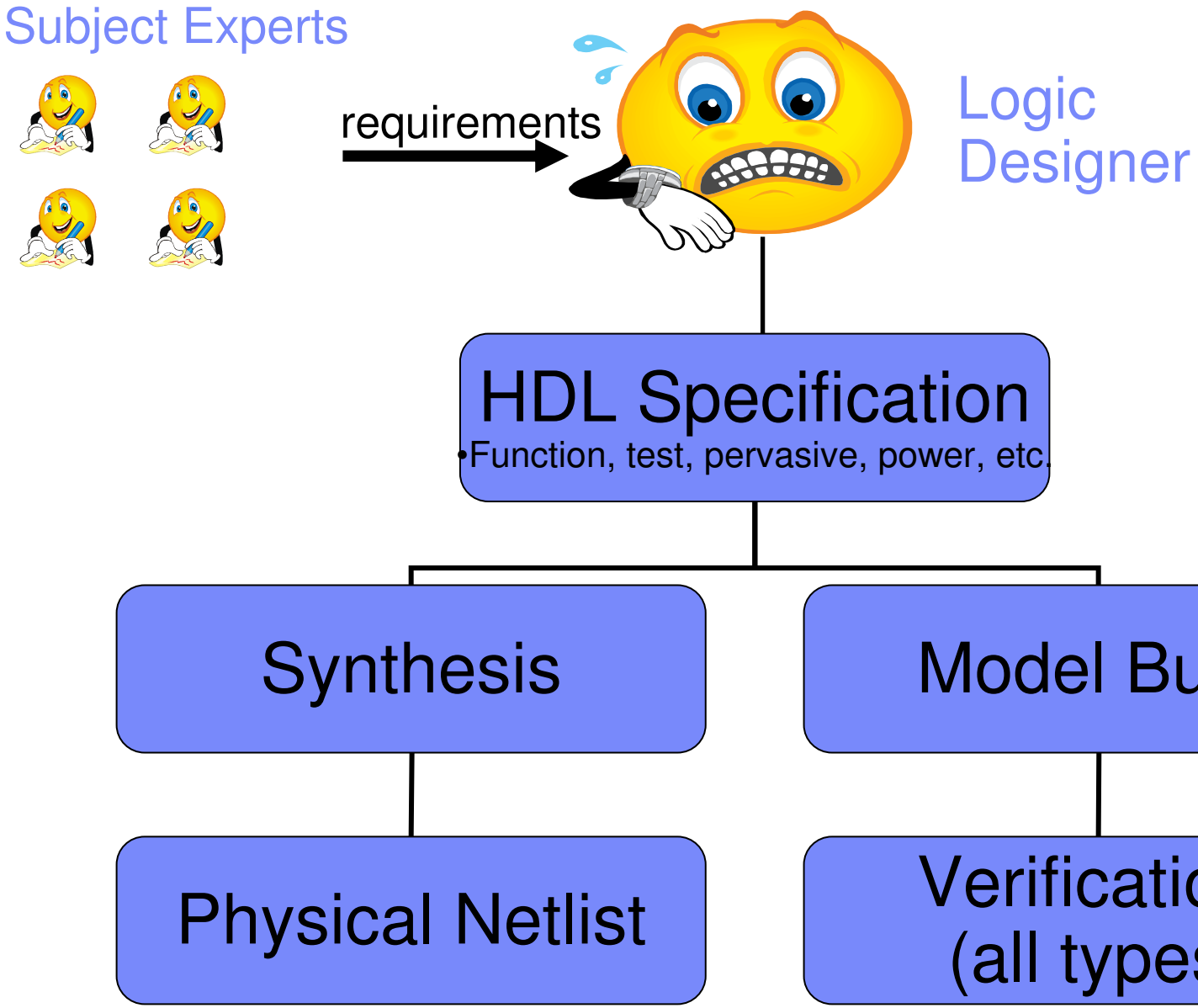
- Fast macro and global analysis tools allow designers to iterate more quickly resulting in reduced implementation time
- Gate level analysis provides orders-of-magnitude speedups over transistor level analysis
- Hierarchical abstraction & multi-threading can provide 5-10x gains in chip level analysis

Productivity: Retiming

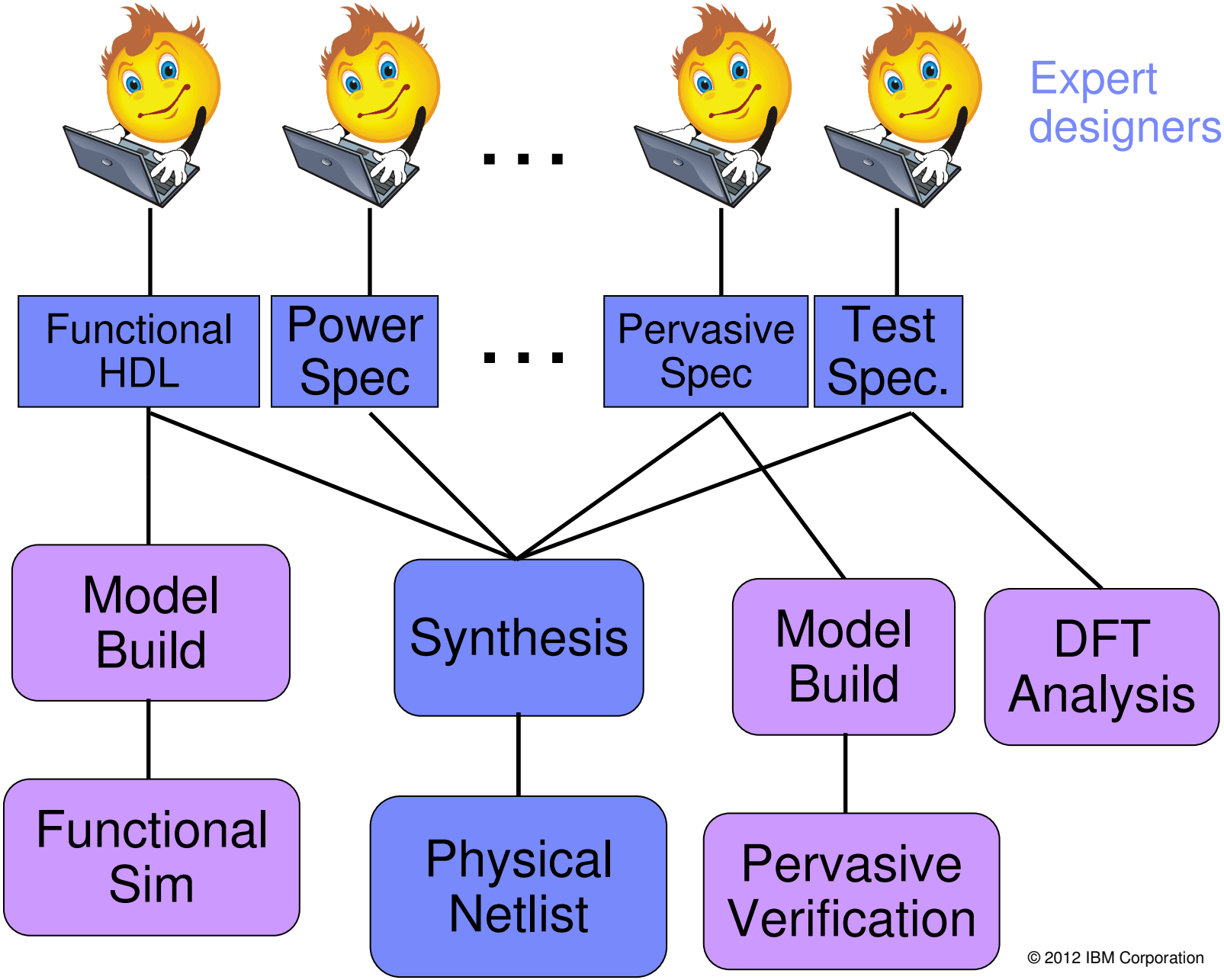


- Significant designer effort spent in optimizing cycle boundaries
- Retiming enables synthesis to optimally locate latches to balance timing/area/power
- Invention is required to seamlessly handle verification implications

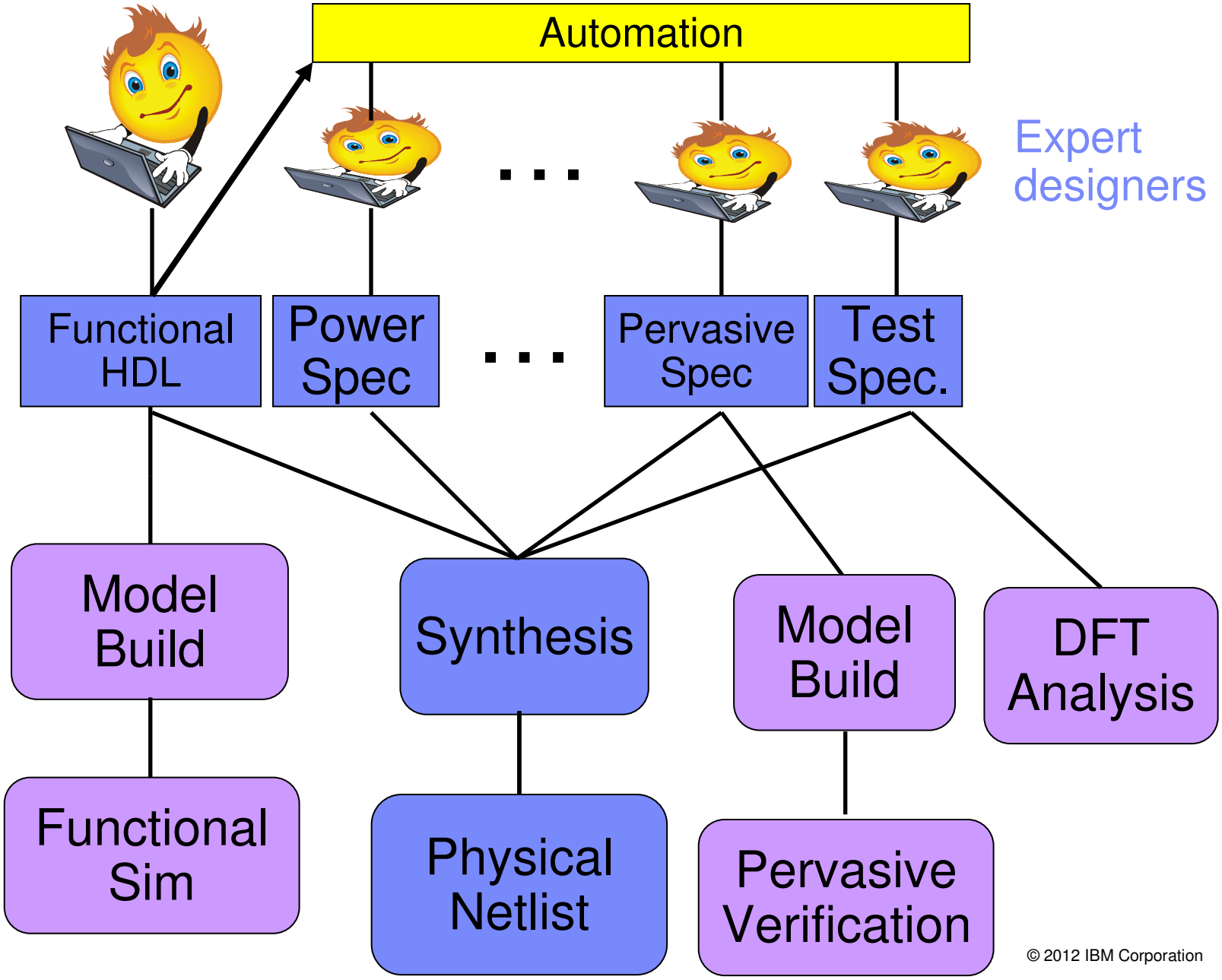
Productivity: Traditional Design Flow



Productivity: Aspect-Oriented Design Flow



Productivity: Aspect-Oriented Design Flow

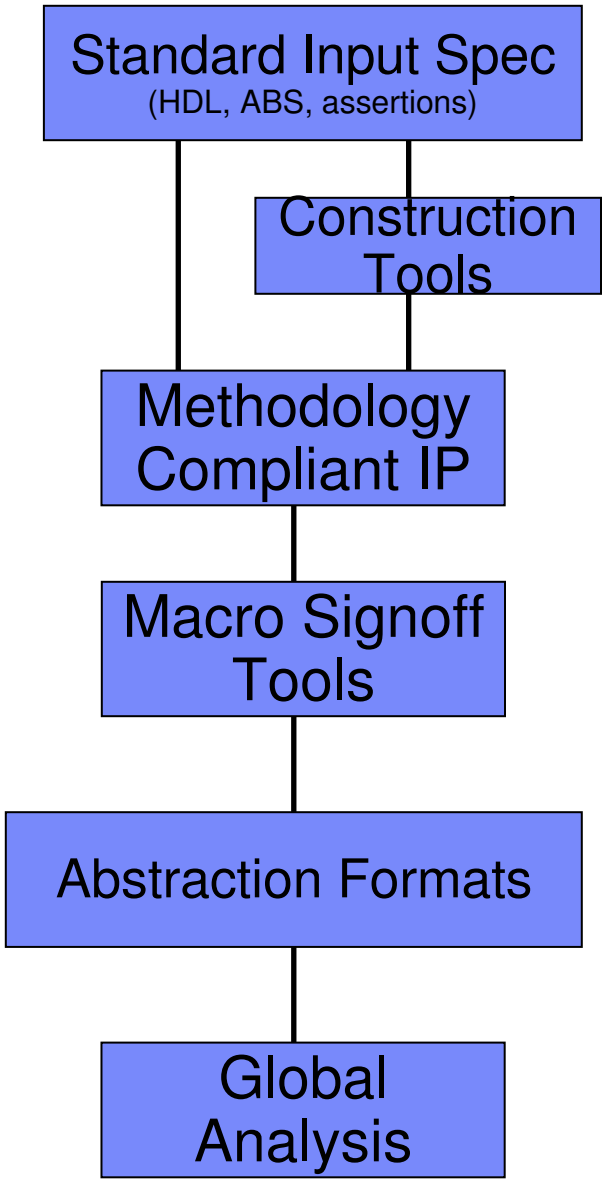


Multi-Core Era Integration: Homogeneous IP

Homogenous Design

Limited # of development groups

Strictly enforced design styles / rules with full methodology support



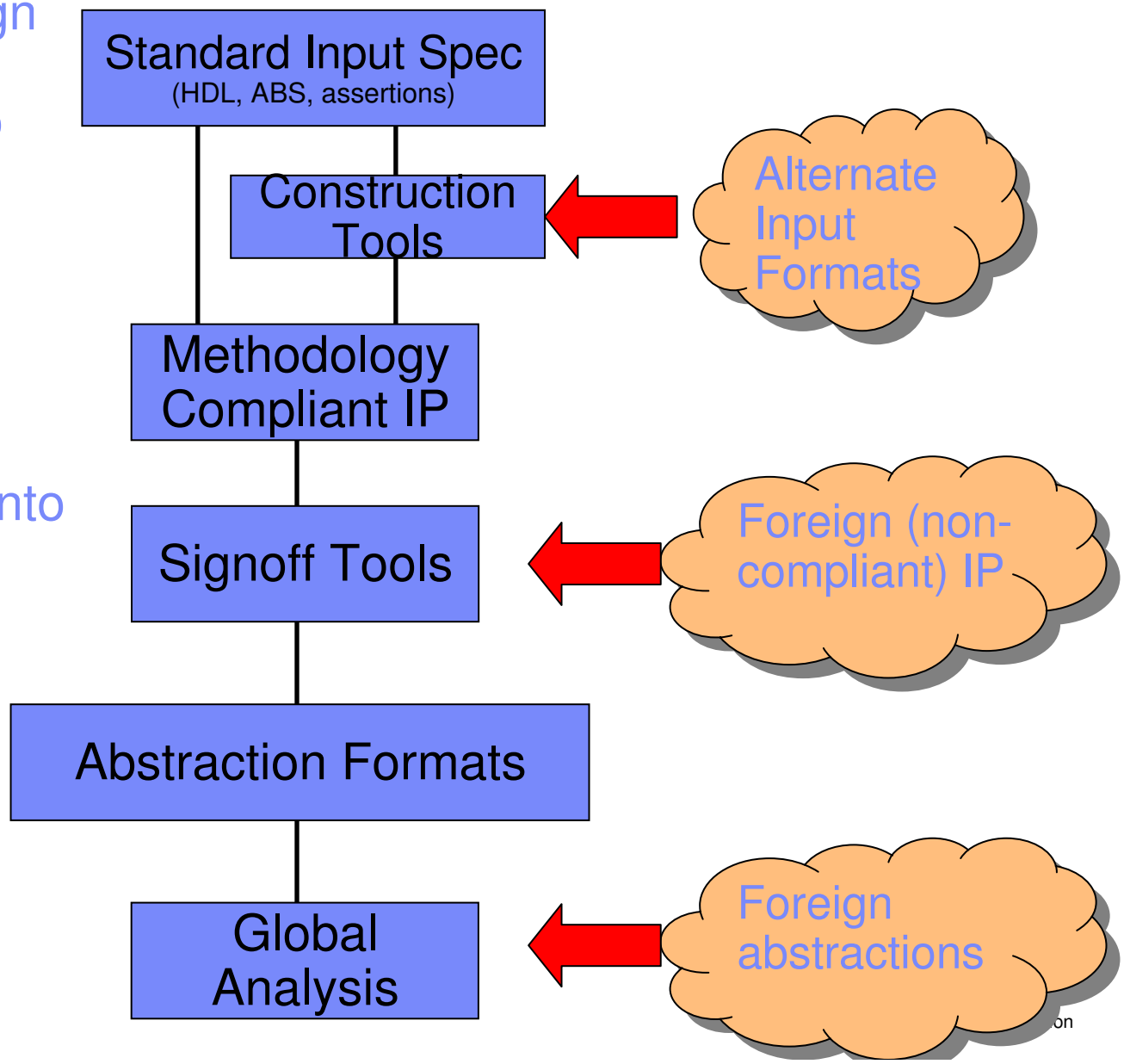
Innovation Era Integration: Heterogeneous IP

Heterogeneous Design

Multiple IP sources to obtain best-of-breed component designs

Limited control on design styles

Multiple entry points into design flow



Summary

- Moore's law continues…..
- However, we will need **MORE INNOVATION**, not just more cores to leverage it
- EDA tools need to provide **MORE PRODUCTIVITY** to enable designer innovation