High Performance Microprocessor Design and Automation:

Overview, Challenges and Opportunities

Introduction

- About Myself

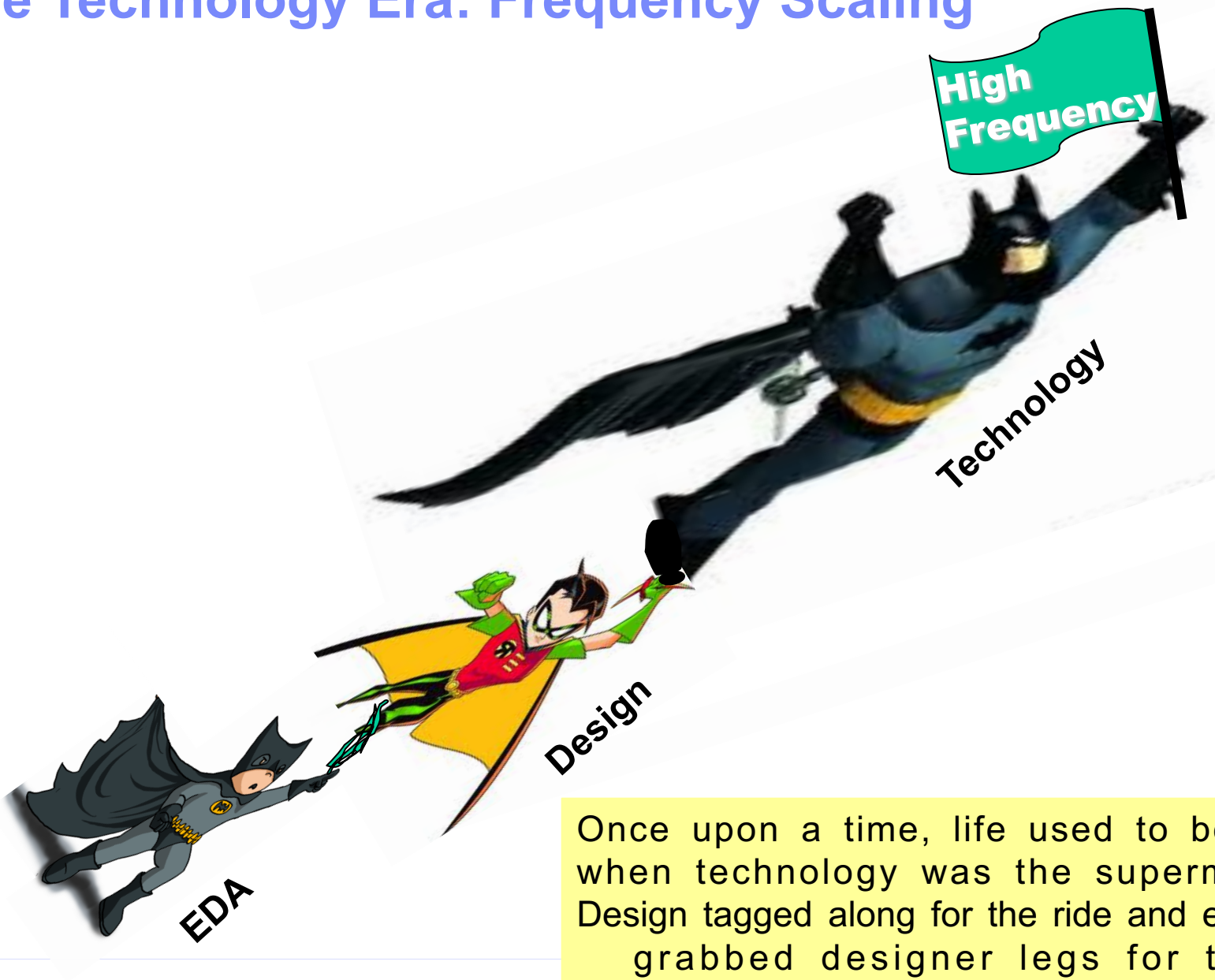
- What to expect out of this lecture
  - Understand the current trend in the IC Design
  - Challenges and opportunities

Agenda

- **Different Eras**
  - Technology Era
  - Multi core Era (Design Era)
  - Innovation Era (EDA Era)

- **Innovation**
  - Technology Innovation
  - Productivity Innovation

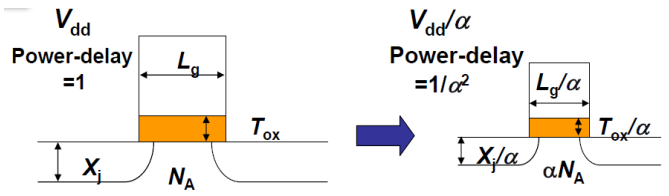
The Technology Era: Frequency Scaling



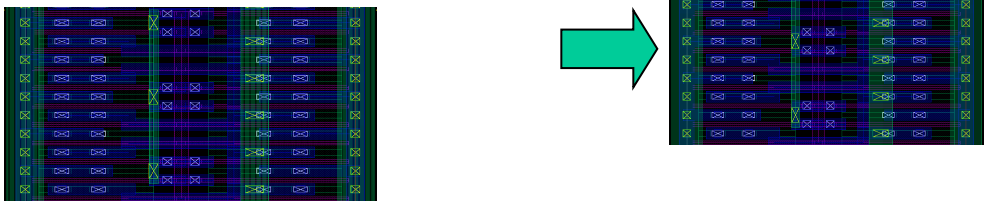
Once upon a time, life used to be Great, when technology was the superman and Design tagged along for the ride and even EDA grabbed designer legs for the fun!

Characteristics of Single Thread Era

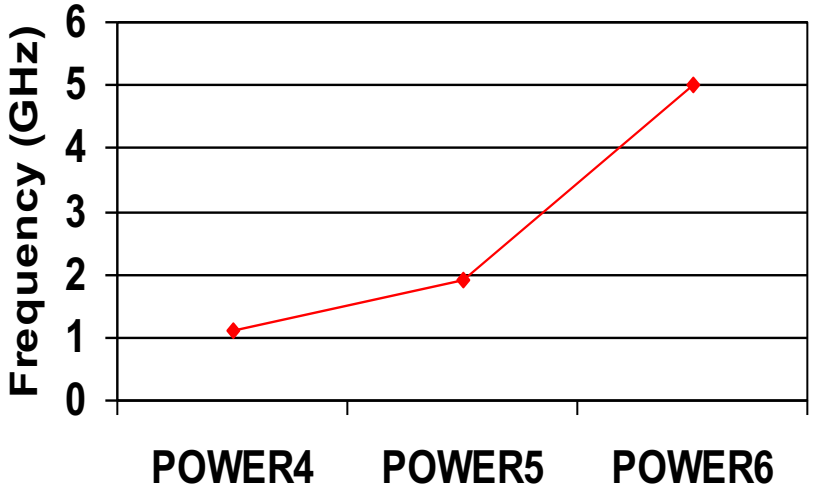
Dennard Scaling



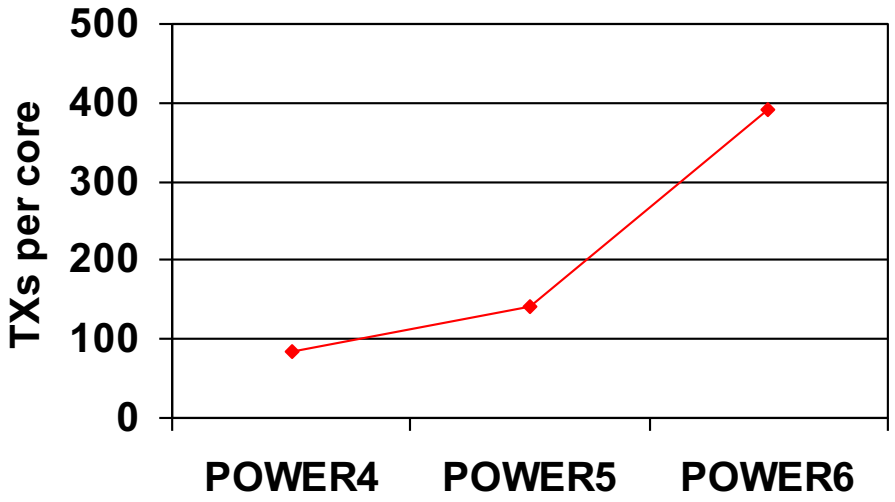
Optical Scaling / Node Migration



Exponential Frequency Growth

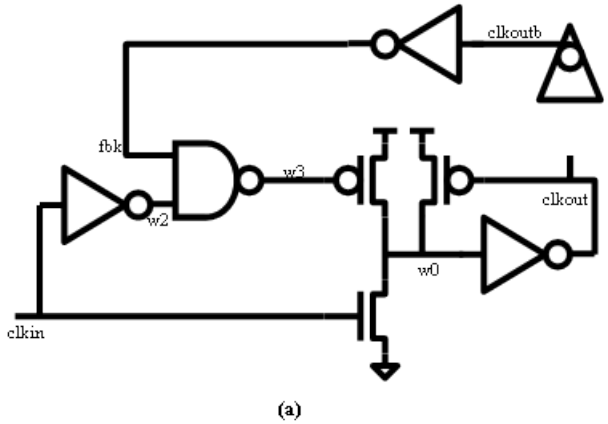


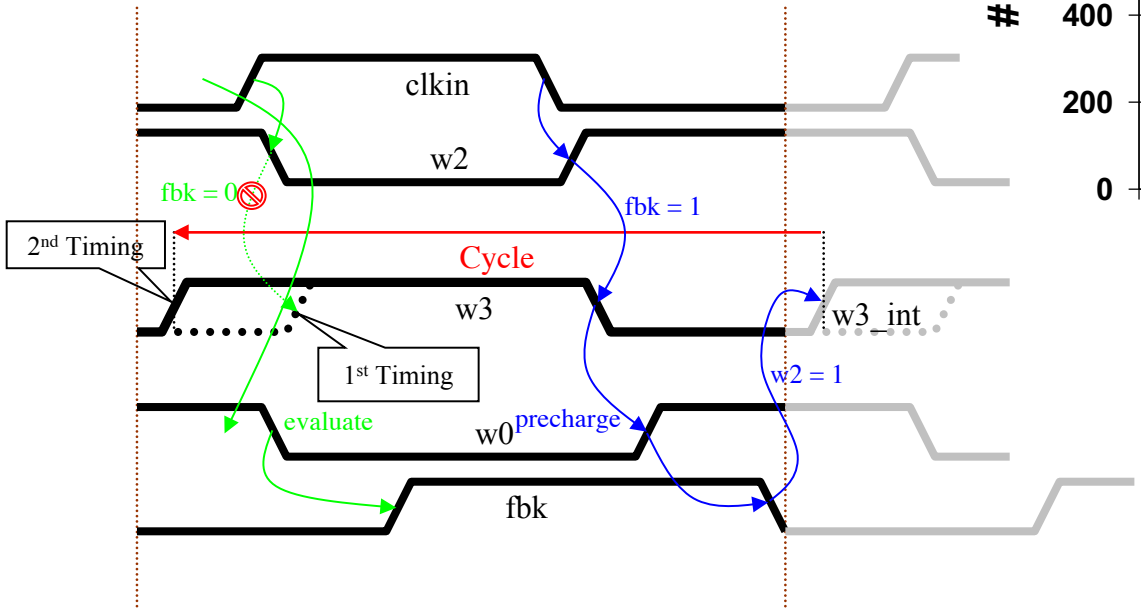
Expanding uArch Complexity



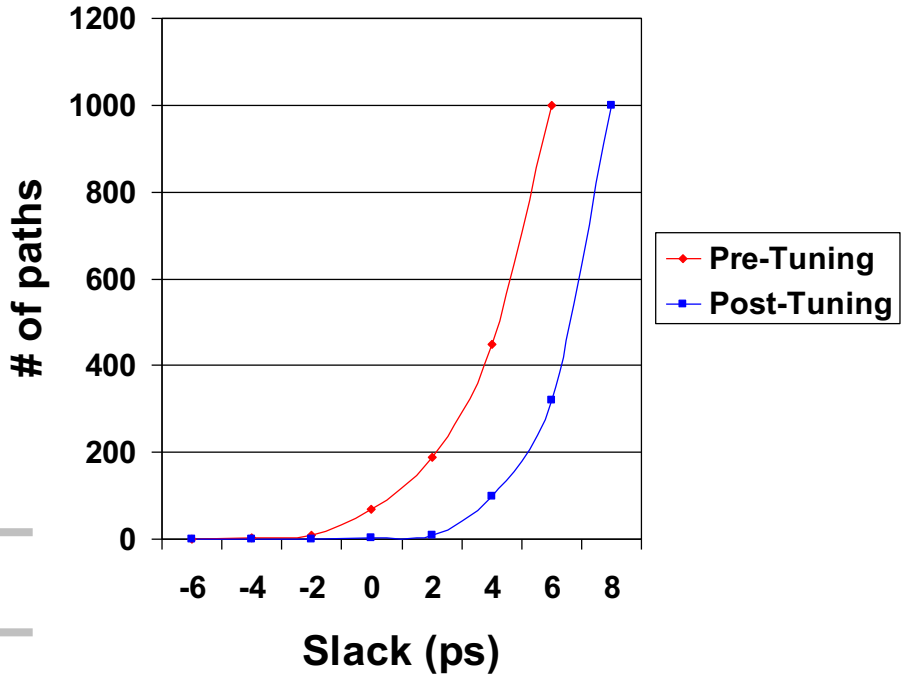
Single Thread Era EDA: Transistor Analysis & Optimization

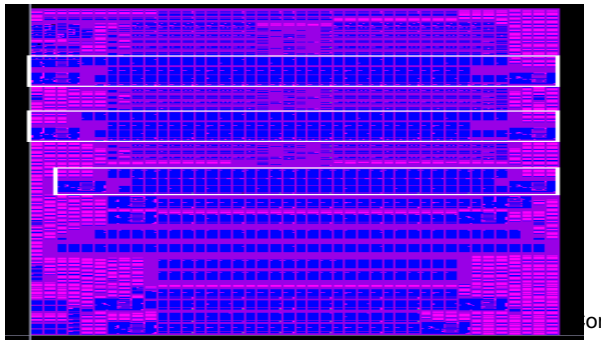
Static timing analysis of complex circuits



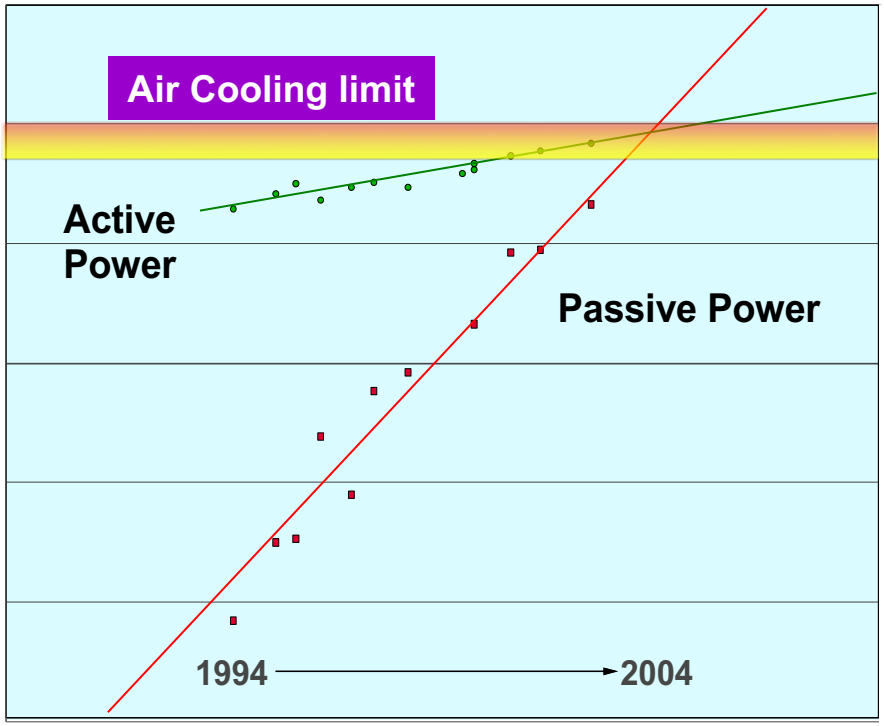


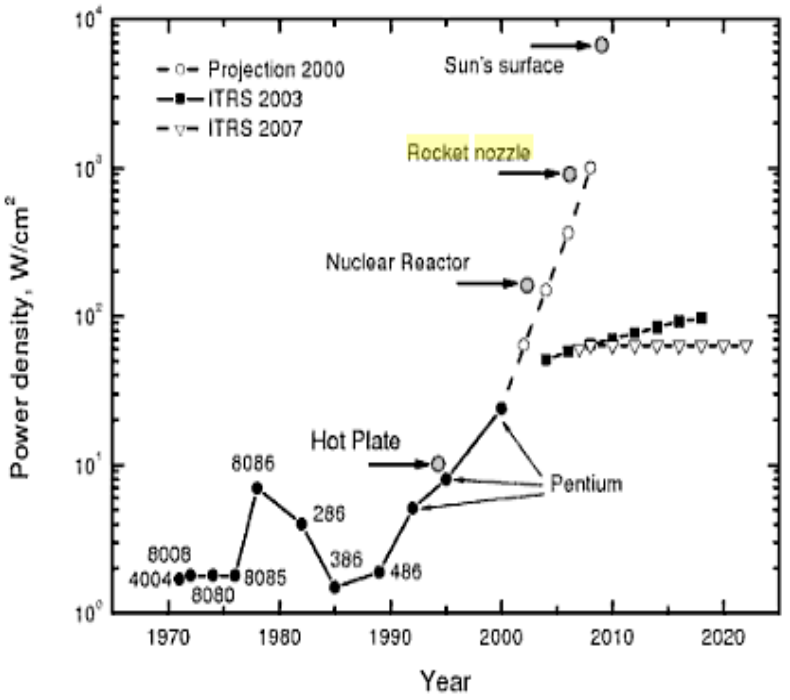
Transistor Level timing optimization





End of Frequency Scaling : The Power Wall

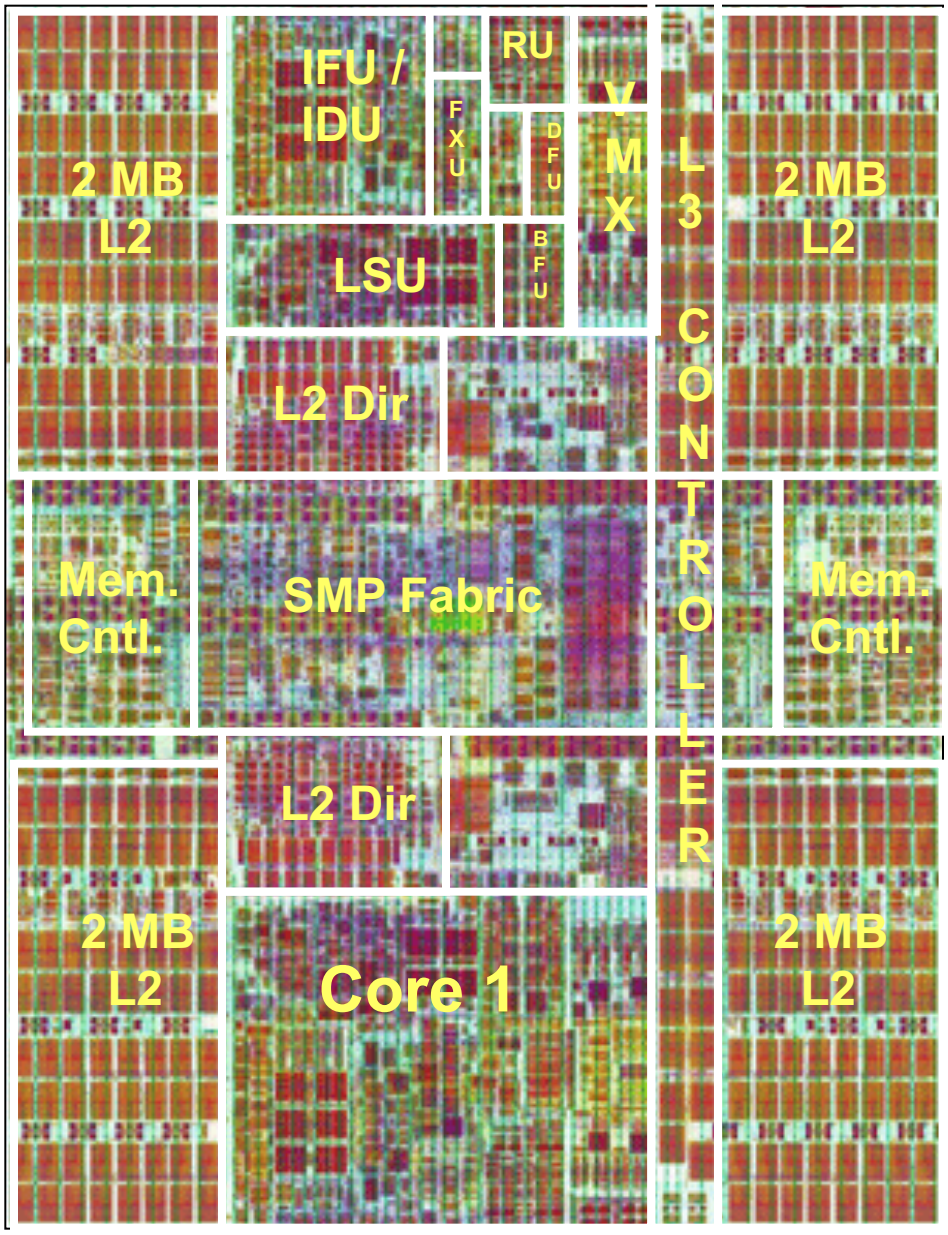




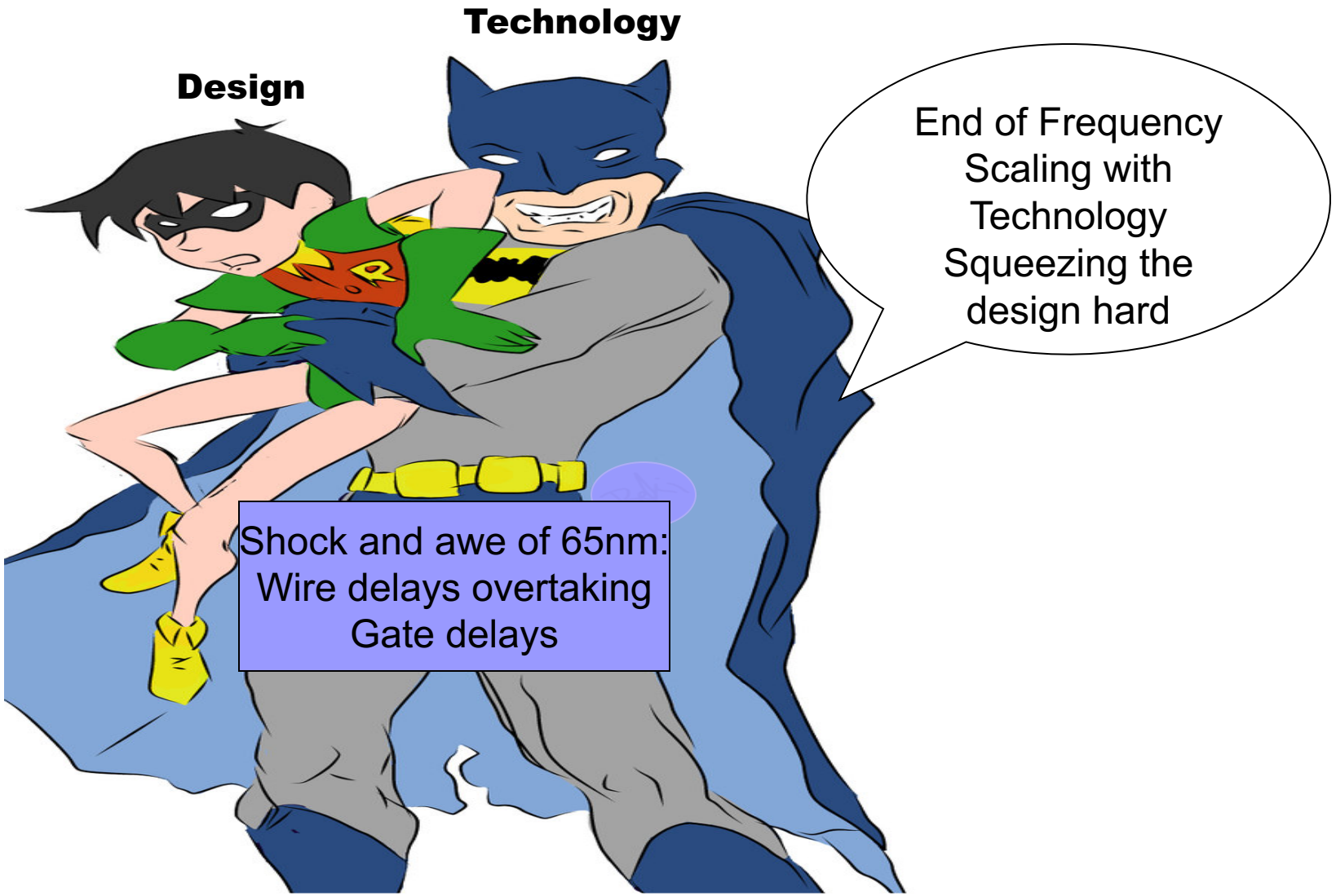
Inability to scale Oxide thickness & lower voltage resulted in a power wall for single thread performance

Frequency Scaling : POWER6 (65nm, 2007)

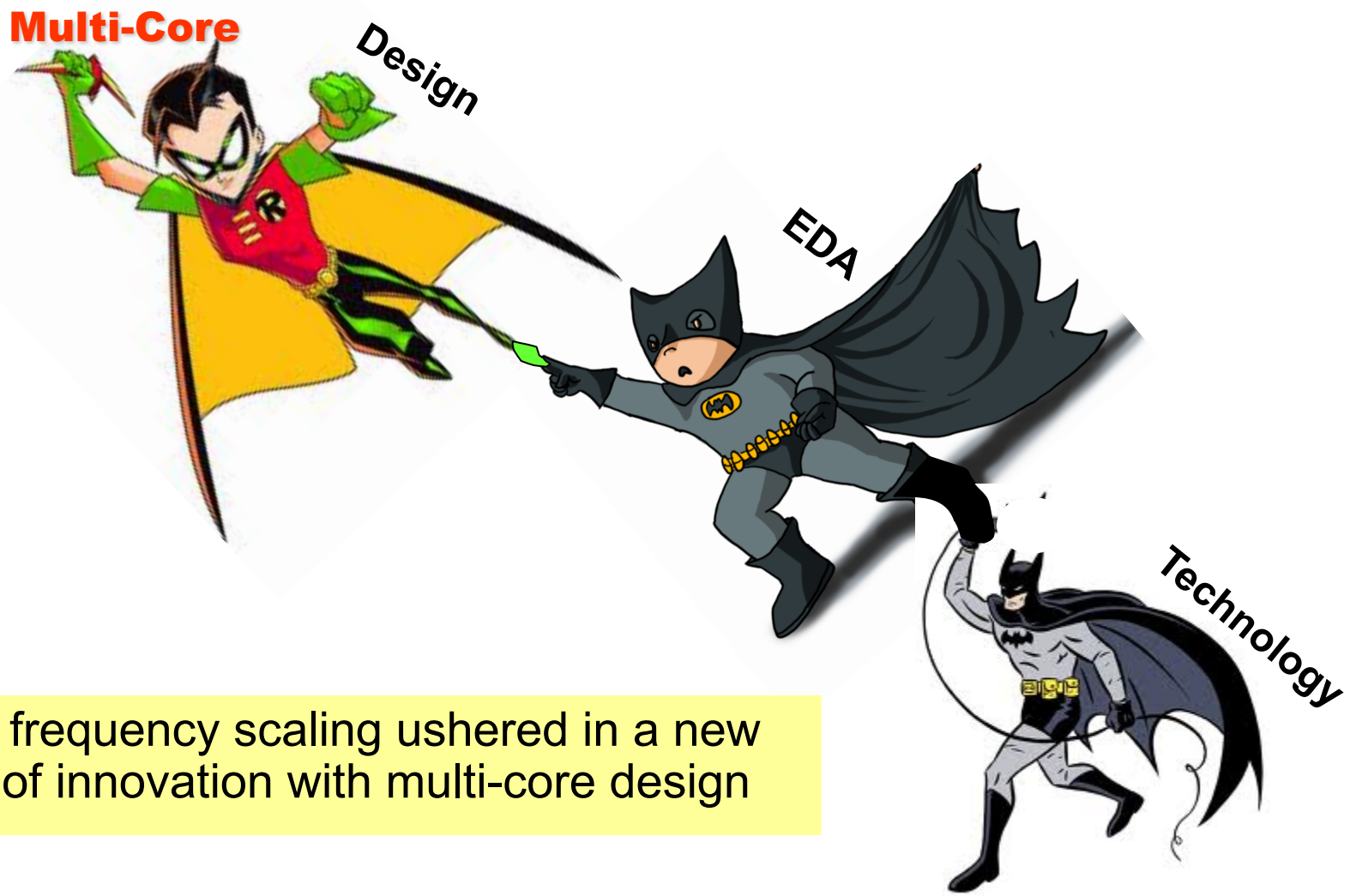
- 5+ GHz operation, >790M transistors, 341mm^2 die
- 65nm SOI with 10 levels of Cu interconnect
- Same pipeline depth & power @ 2x frequency versus POWER5



Technology Tantrums

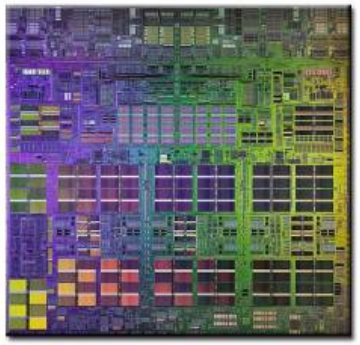


Multi-Core Era

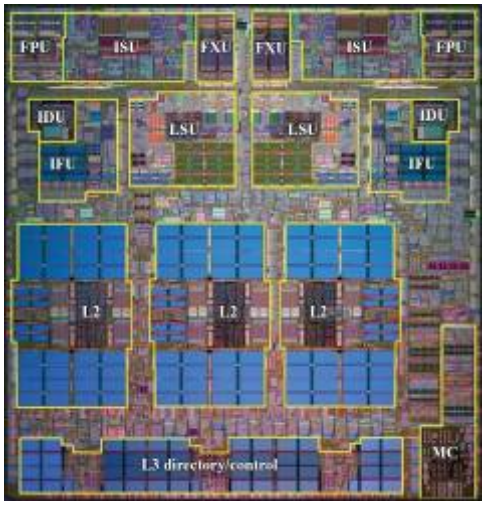


End of frequency scaling ushered in a new era of innovation with multi-core design

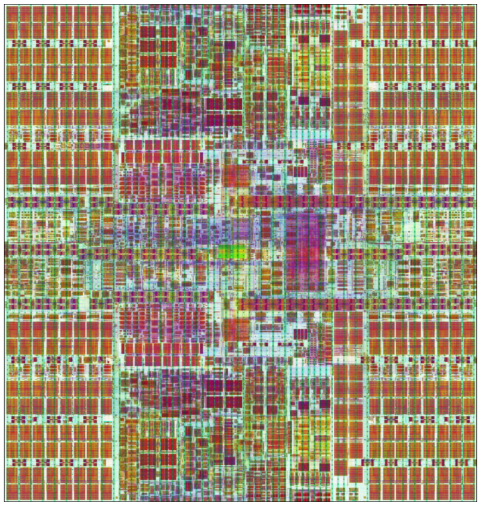
POWER Processors Began the Multi-Core / Multi-Thread Era



Power 4
2001
Introduced First Dual core

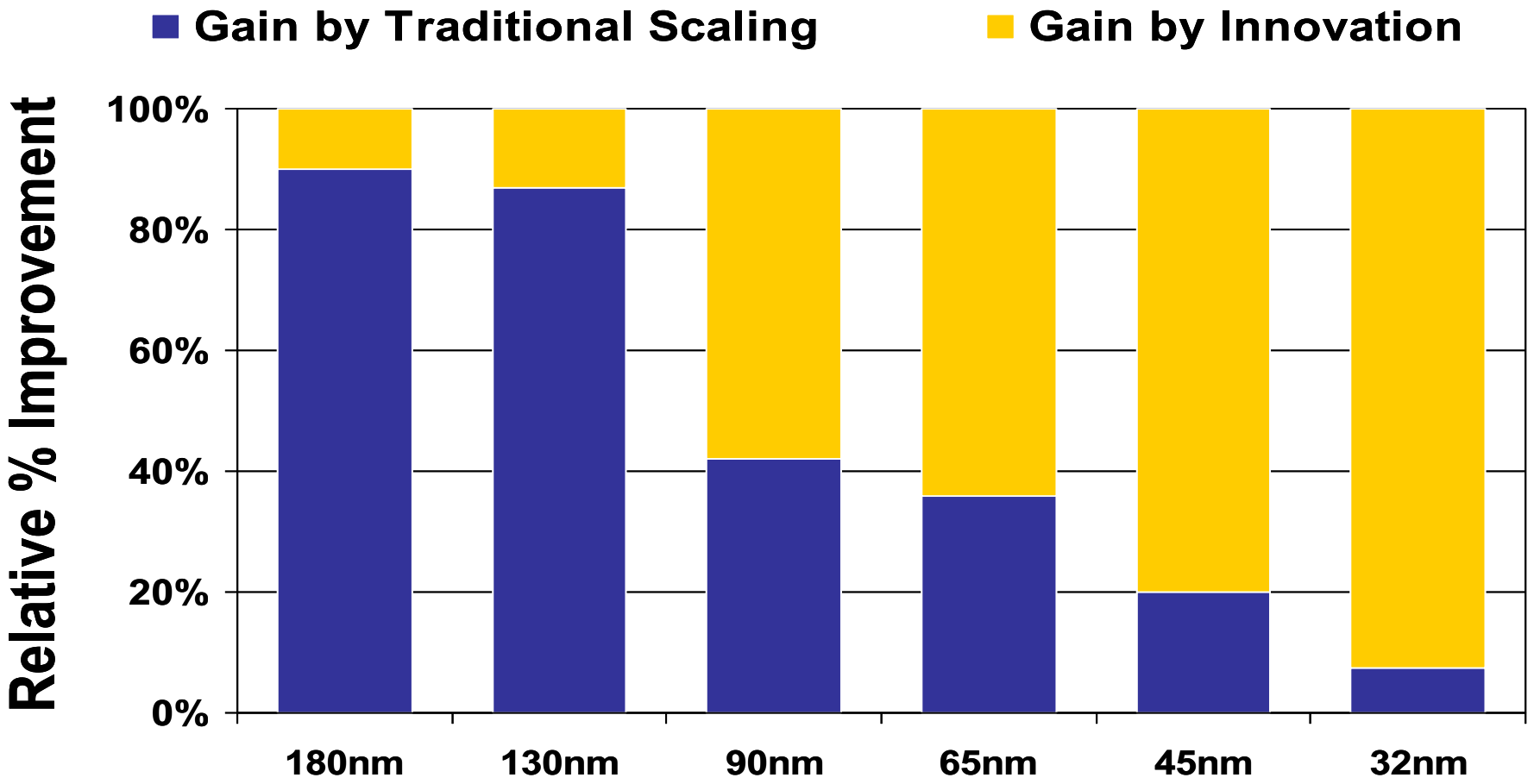


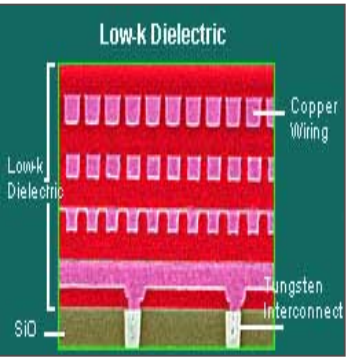
Power 5
2004
Dual Core
Introduces SMT (4 threads)

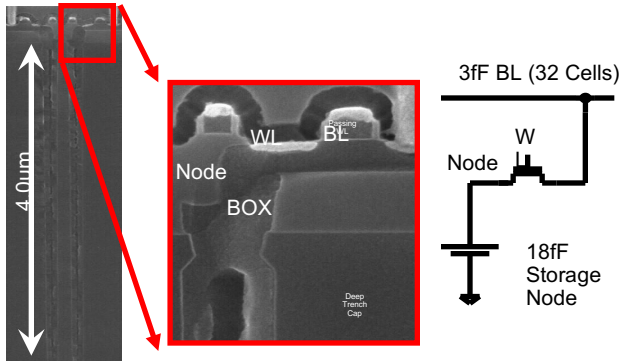


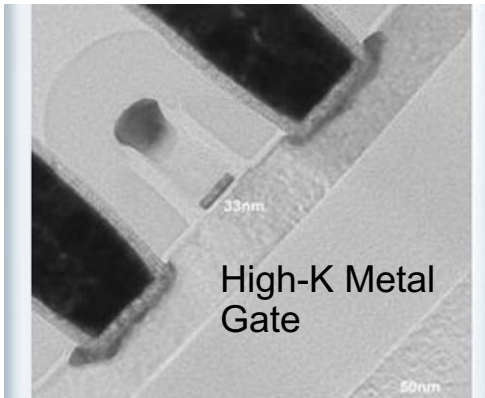
Power 6
2007
Dual Core – 4 threads
Enhances SMT Efficiency

Life starts to become interesting: Technology ride very bumpy

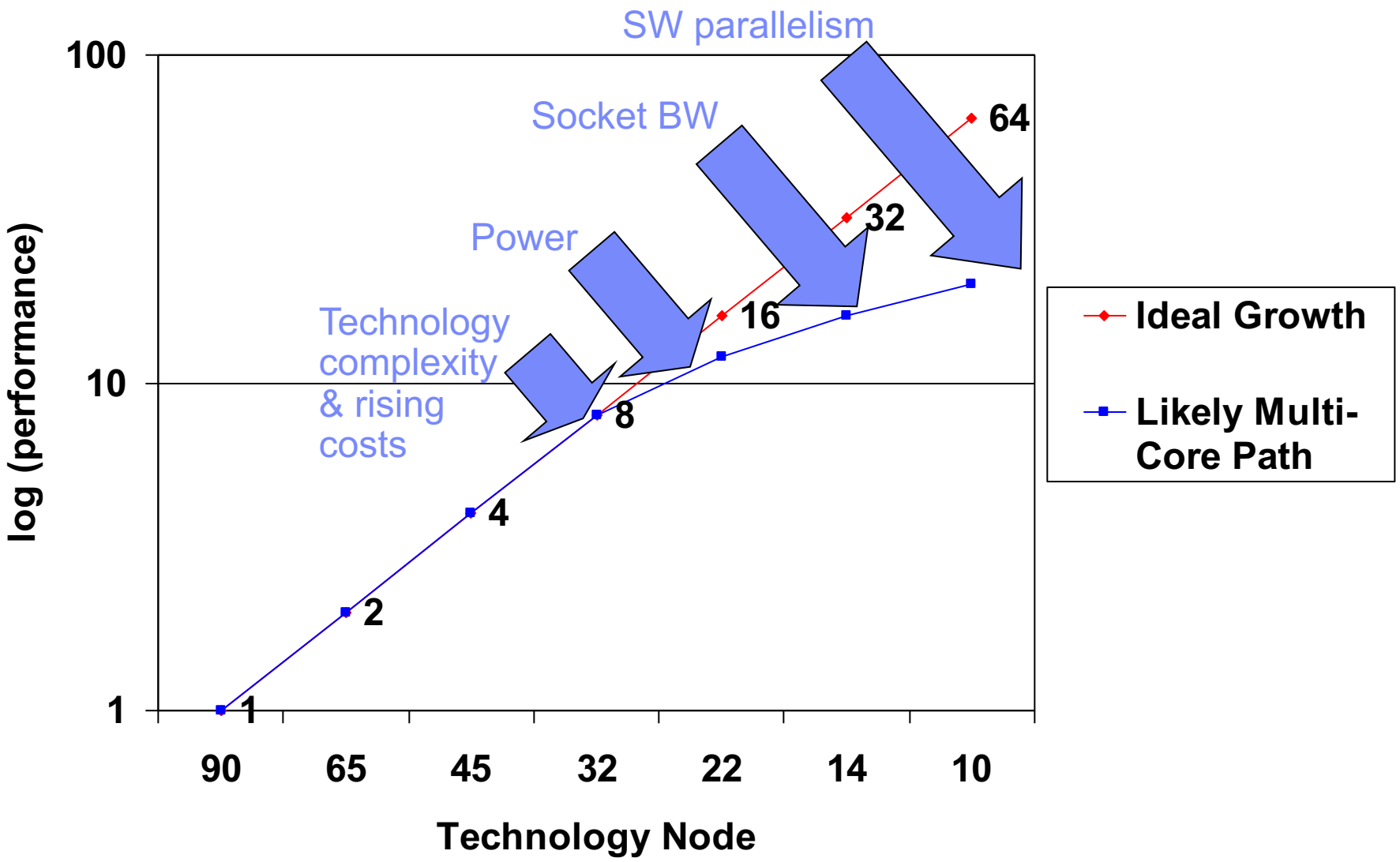




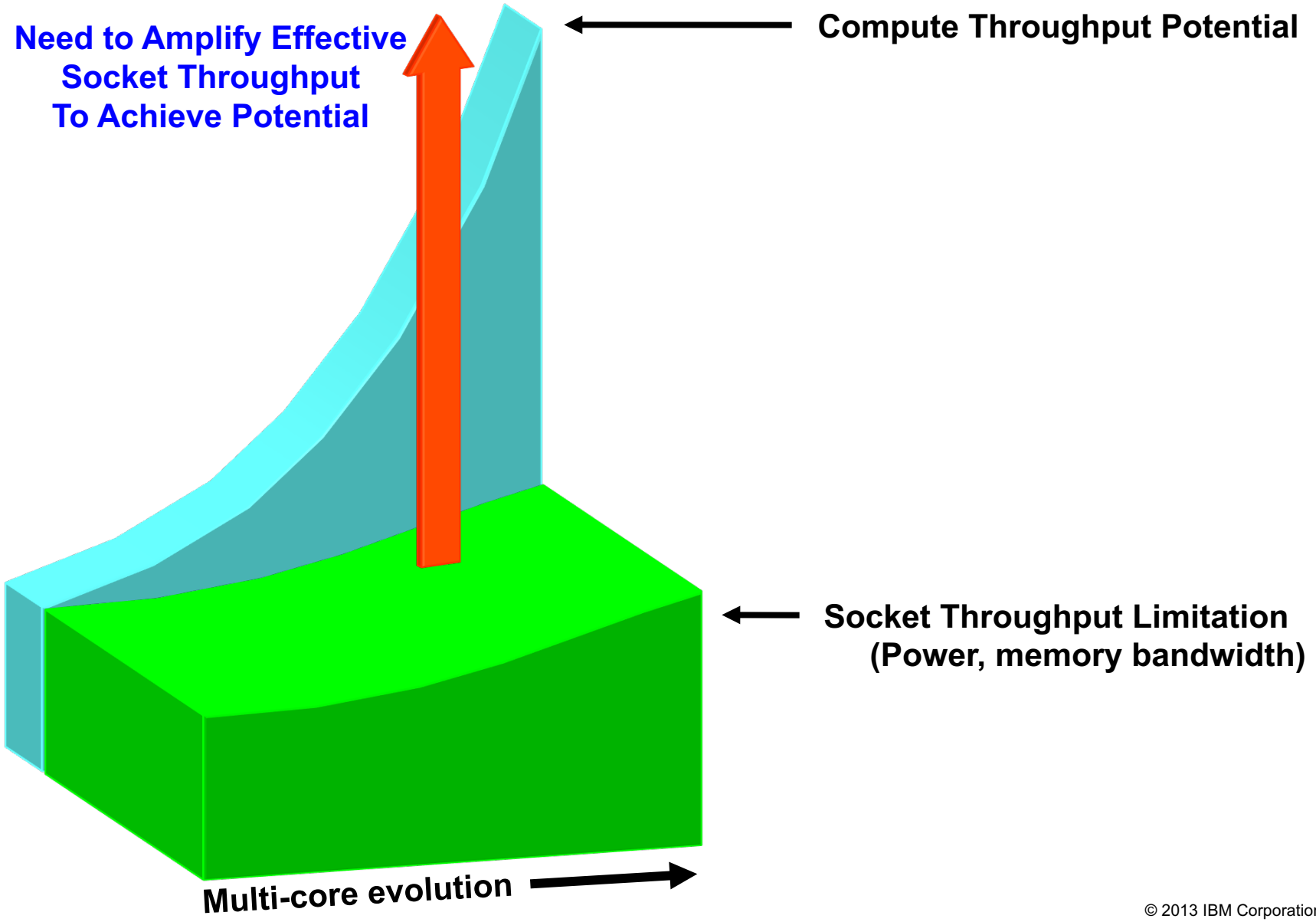




Multi-Core Era Limiters



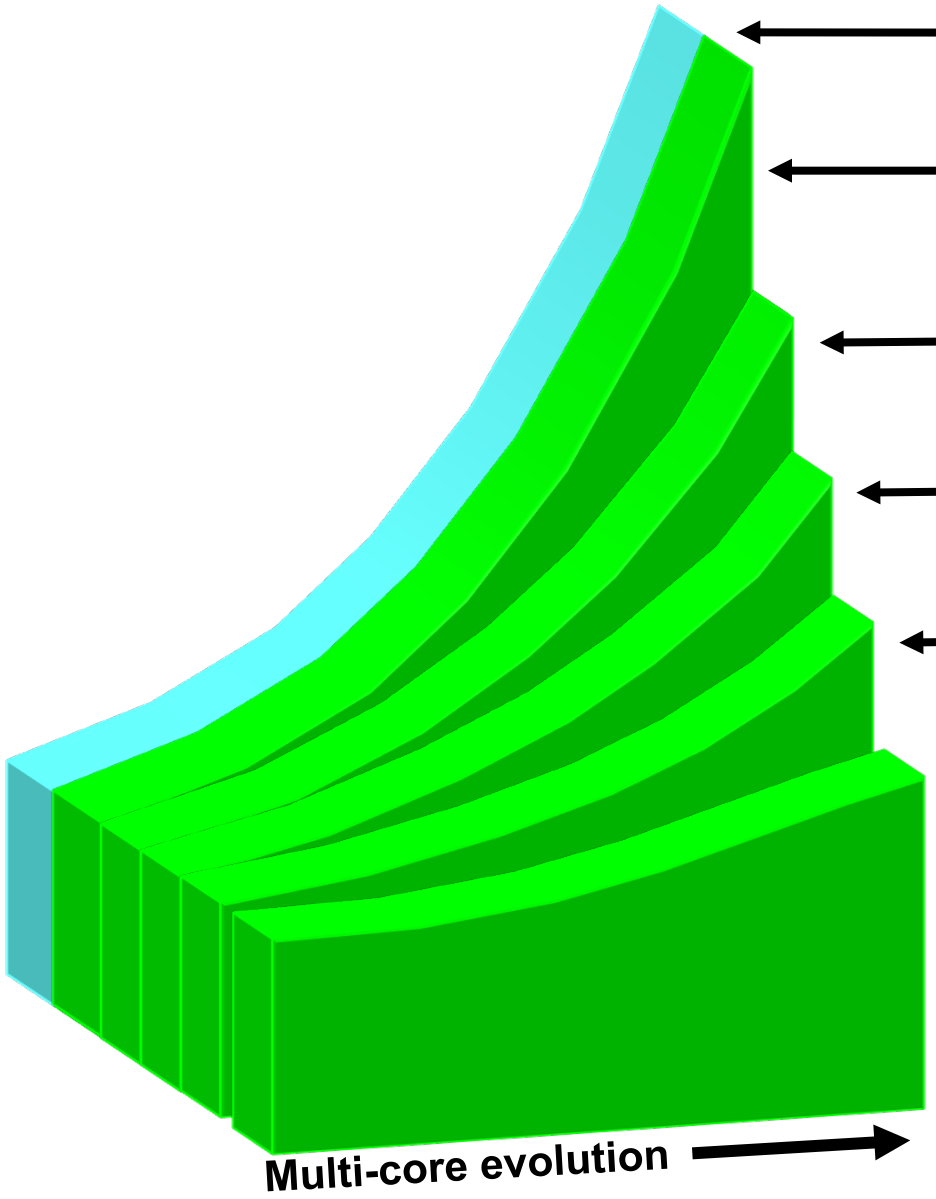
Multi-Core Advantage



Innovation Drive

Architecture & Productivity Innovation

High performance uP Designs:
Extending Multi-Core Gains (Power processor)



Compute Throughput Potential

Coherence Innovation to minimize socket-to-socket communication

Low-Power Off-Chip Signaling Technology

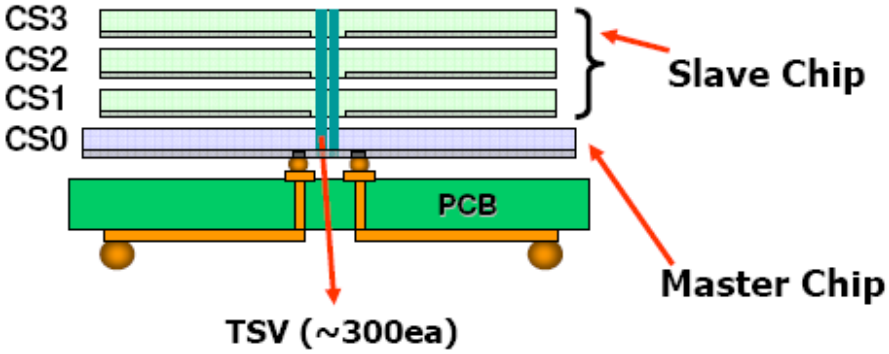
High bandwidth memory buffer

EDRAM = large, low power cache

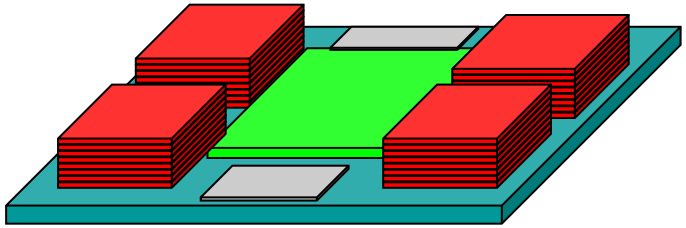
Socket Throughput Limitation (Power, memory bandwidth)

Innovation Drive : System Level Technologies

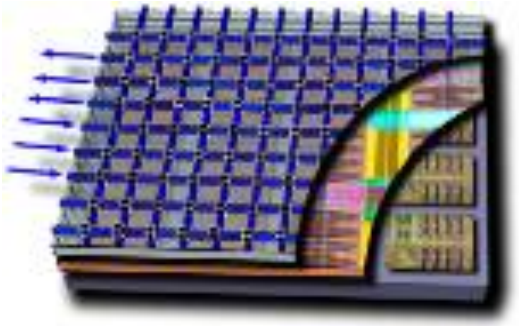
3D Stacking with Through Silicon Vias



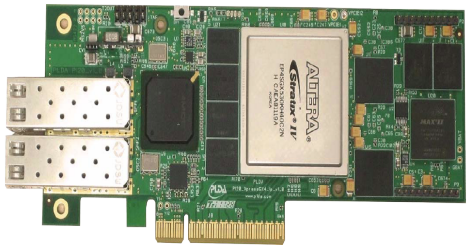
Single Processor–Memory Socket



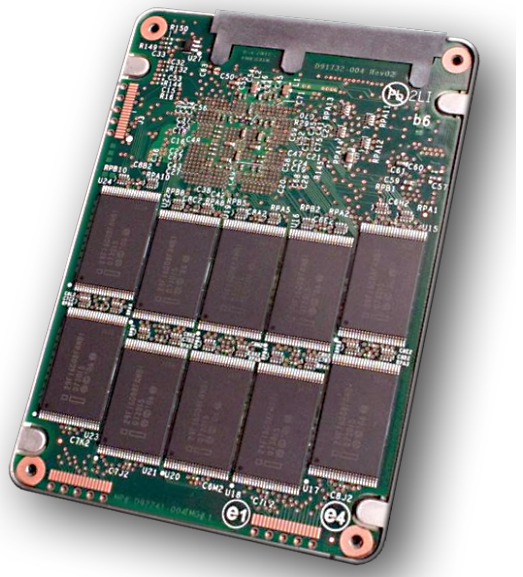
Silicon Photonics



FPGA Accelerators

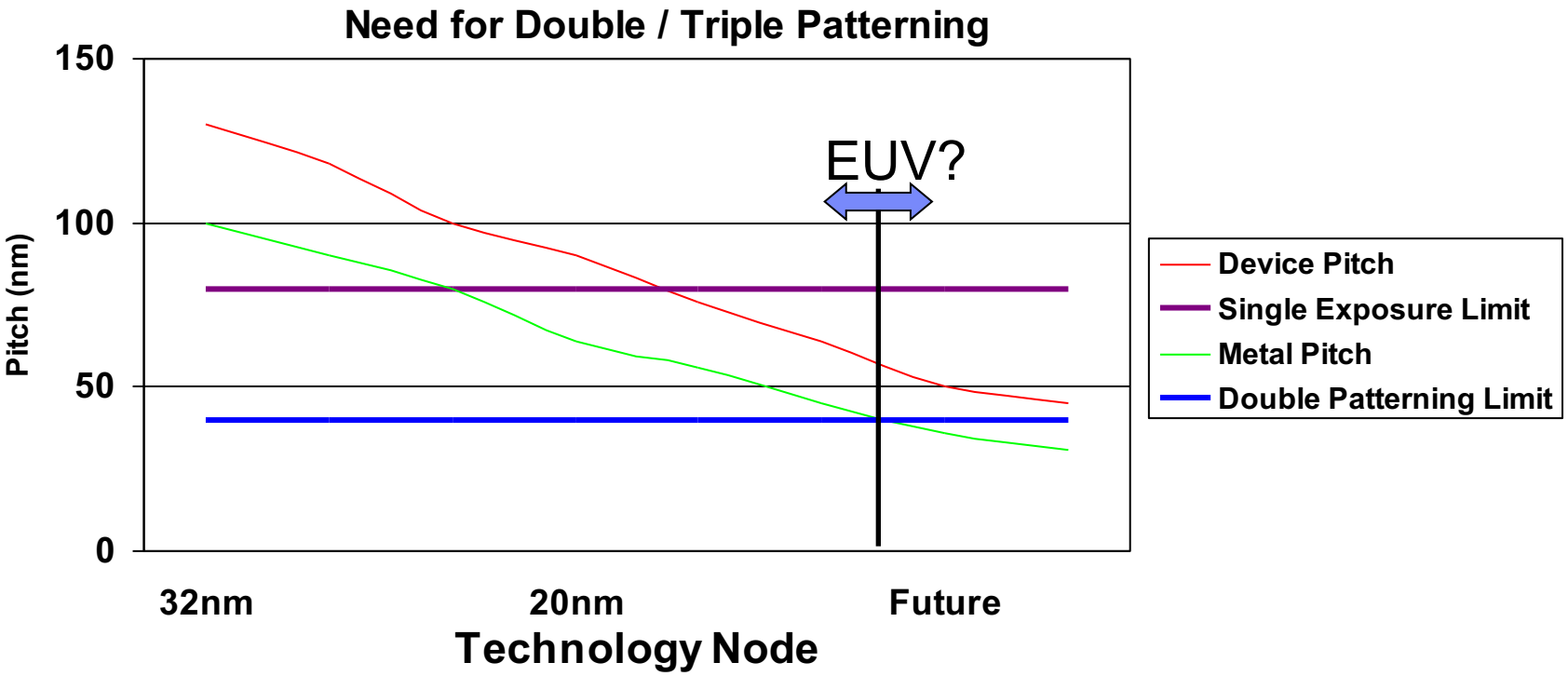


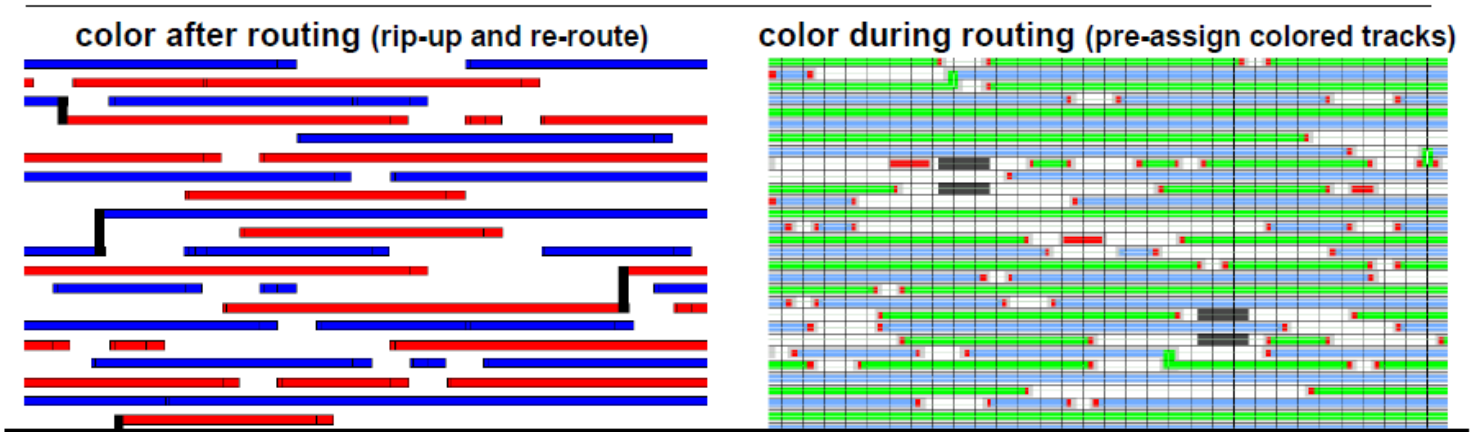
Flash Memory / SSD



Heterogeneous systems on Chip
Specialized functions
Specialized cores:
- Single thread focused
- Throughput focused

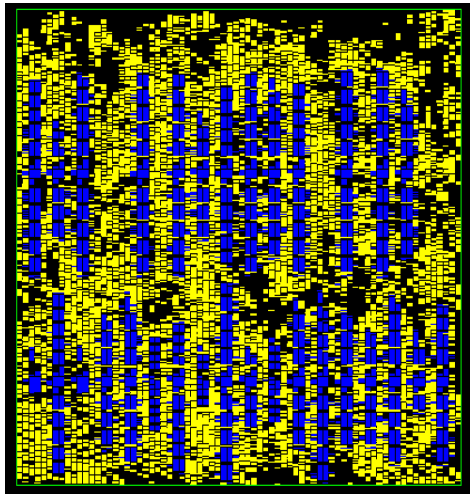
Innovation at Technology, Design Interface: Double/Triple Patterning



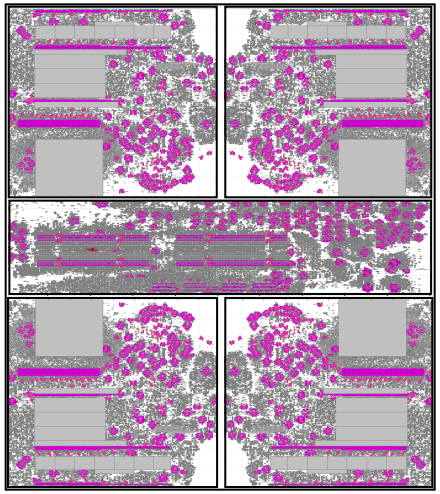


Productivity Innovation: Structured Synthesis and Large Block Synthesis

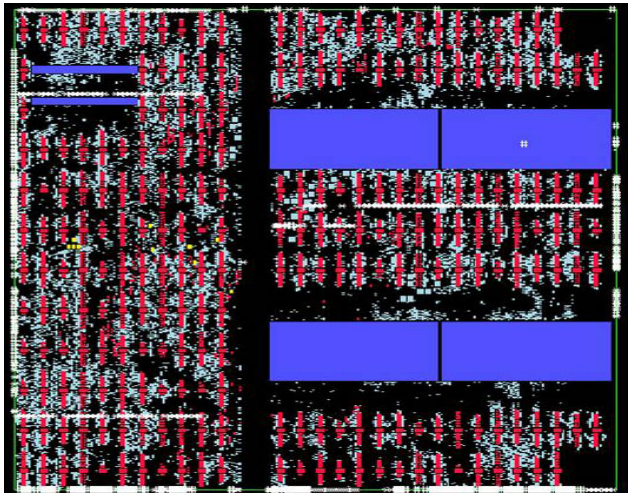
- Customs take large amount of resources and productivity is key
- Merge the domain of customs and Synthesis targeting design productivity and improved quality
  - through merging of custom and synthesis hierarchy with structure in synthesis (not random logic any more)
  - Global Optimization view; Targeted structured data paths and synthesis
  - A methodology with numerous algorithmic and practical innovations spanning from incremental logic design processing, to data paths to structured clocking to custom synthesis merged techniques.



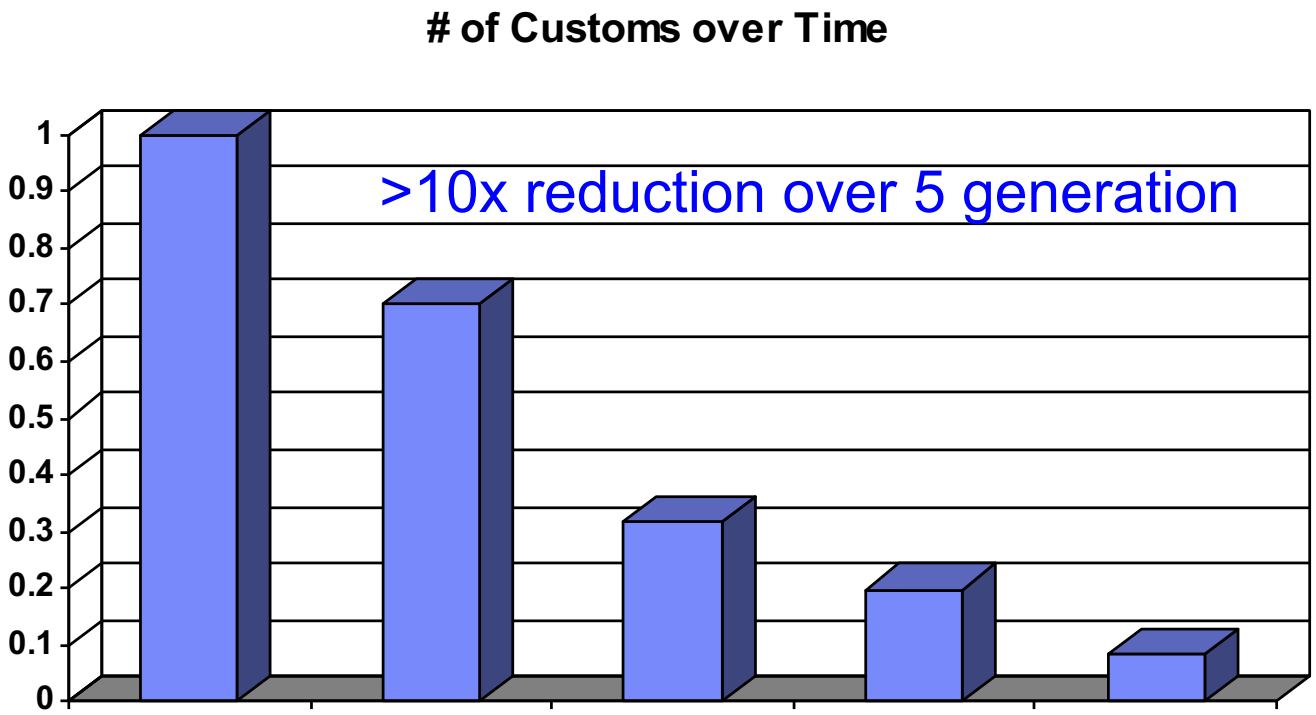
P/Z server Macro

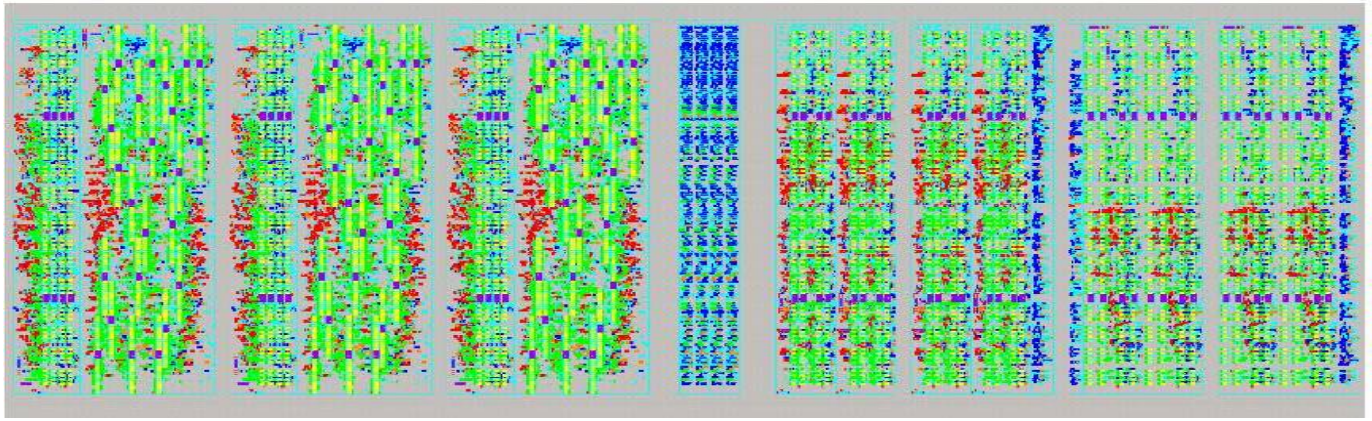


Quad FPU



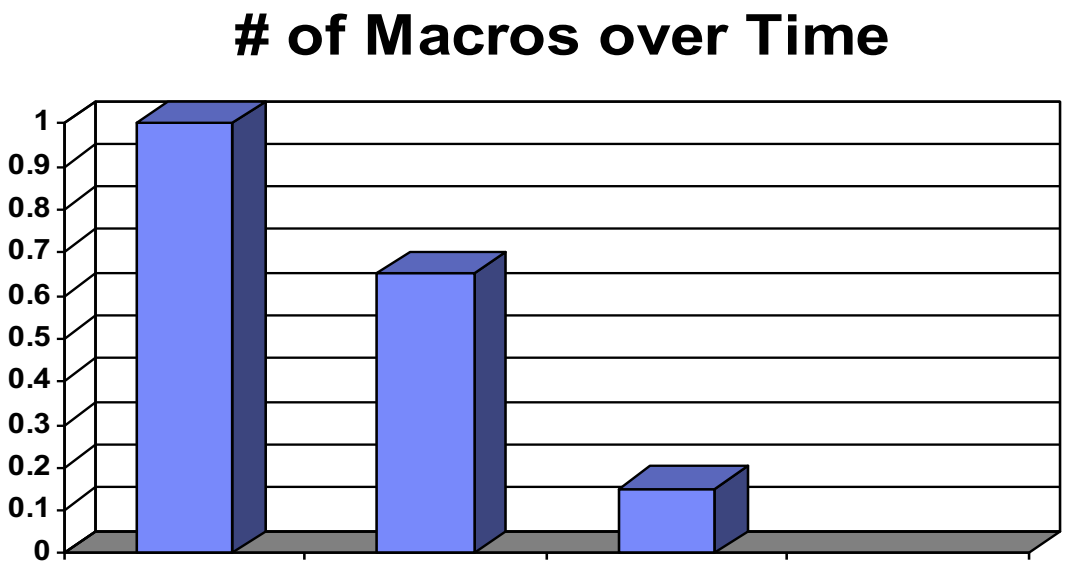
Productivity Innovation : Reduce Custom Design (Structured Synthesis)

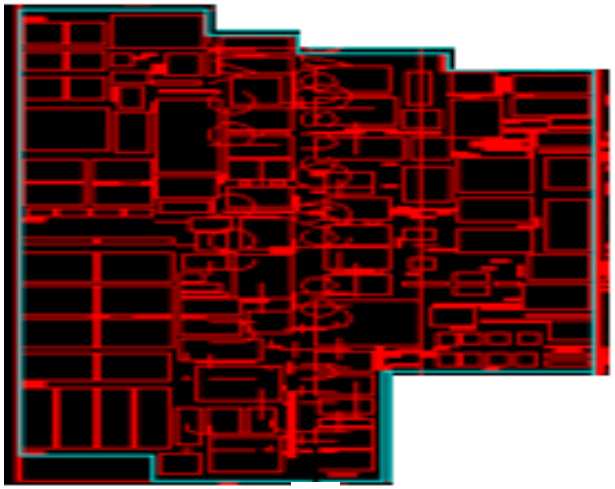


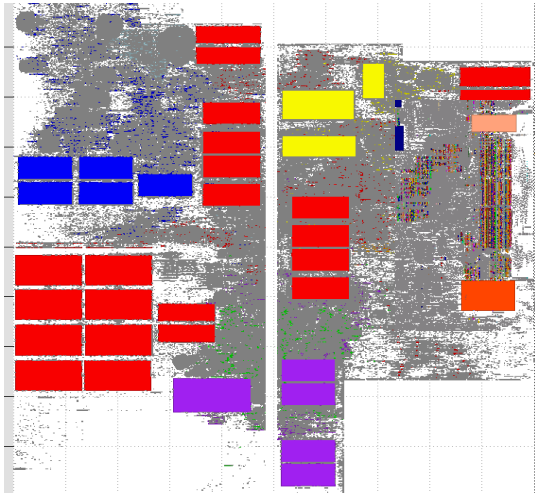


Synthesis results w/ custom-like data flow alignment.

Productivity Innovation: Reduced # of Design Partitions (Large Block Synthesis)



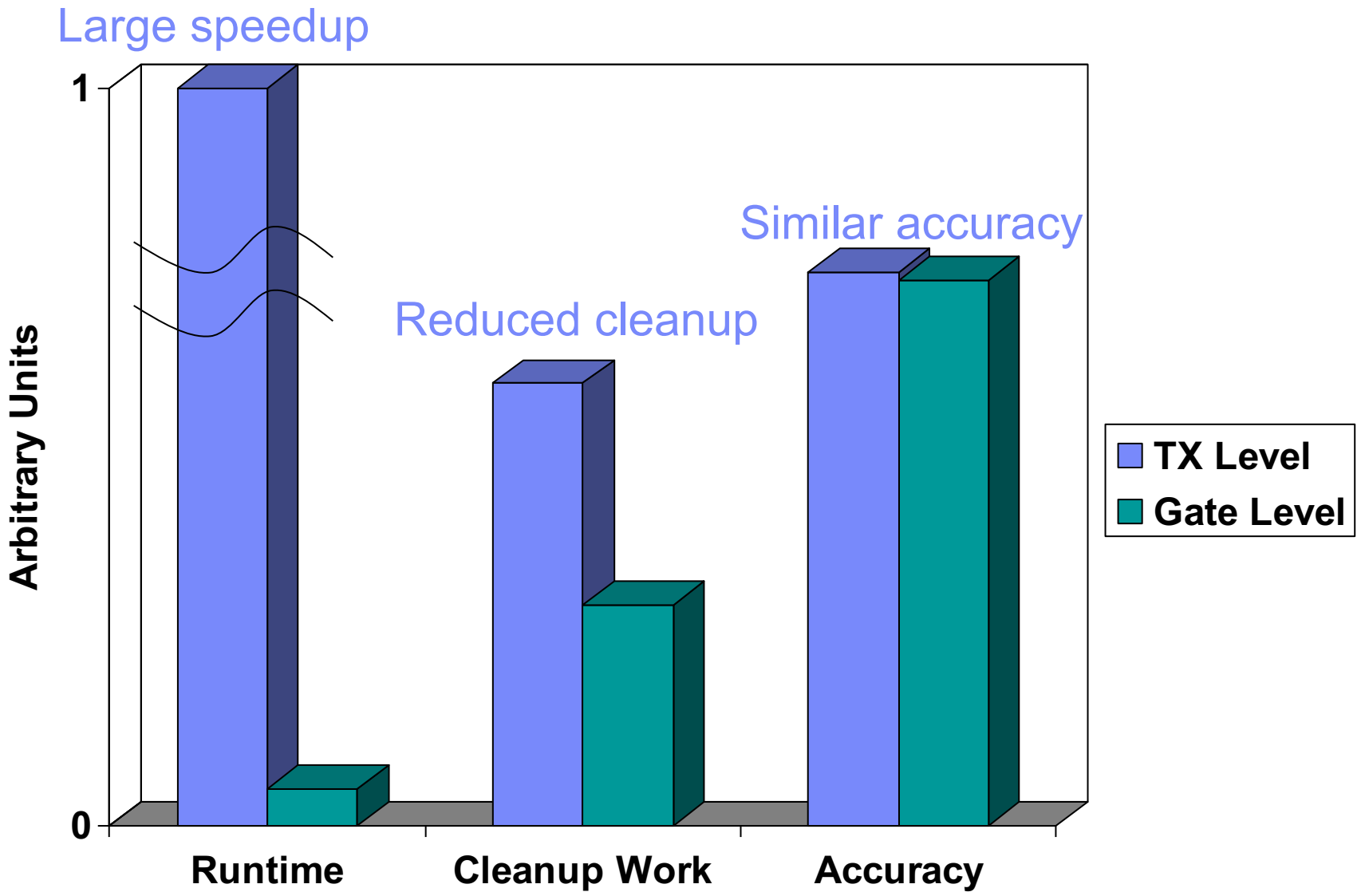




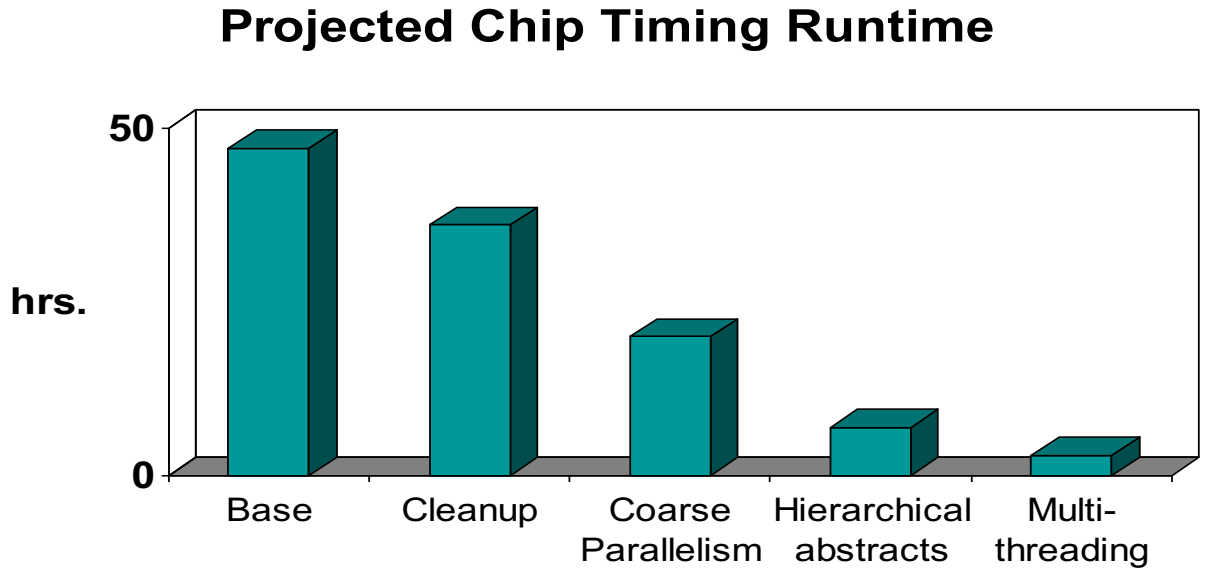
60 logic macros, 25 customs, 14 unique arrays/RFs

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

Productivity & TAT Innovation: Gate Level Analysis & Signoff

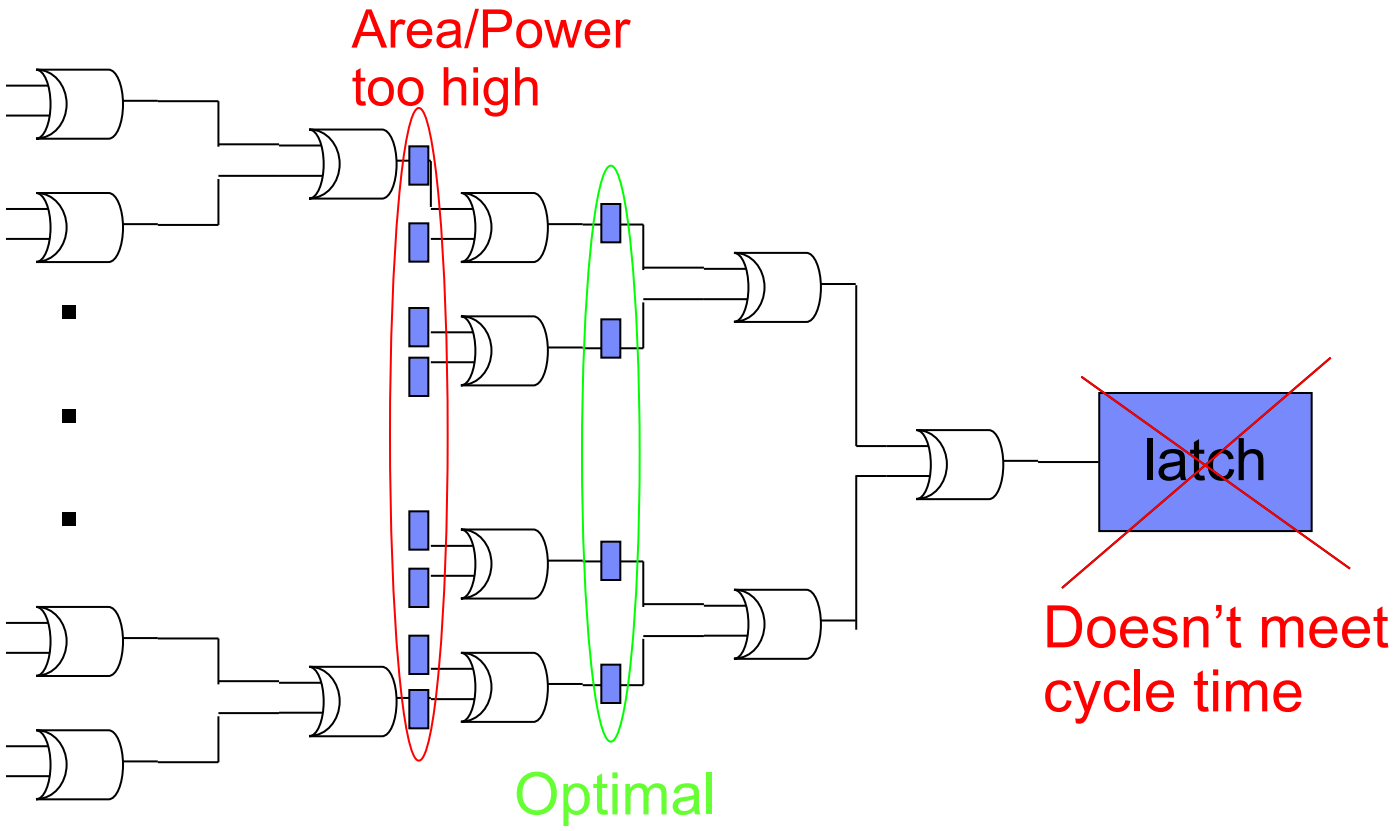


Productivity & TAT Innovation: Hierarchical Abstraction & Multi-Threading



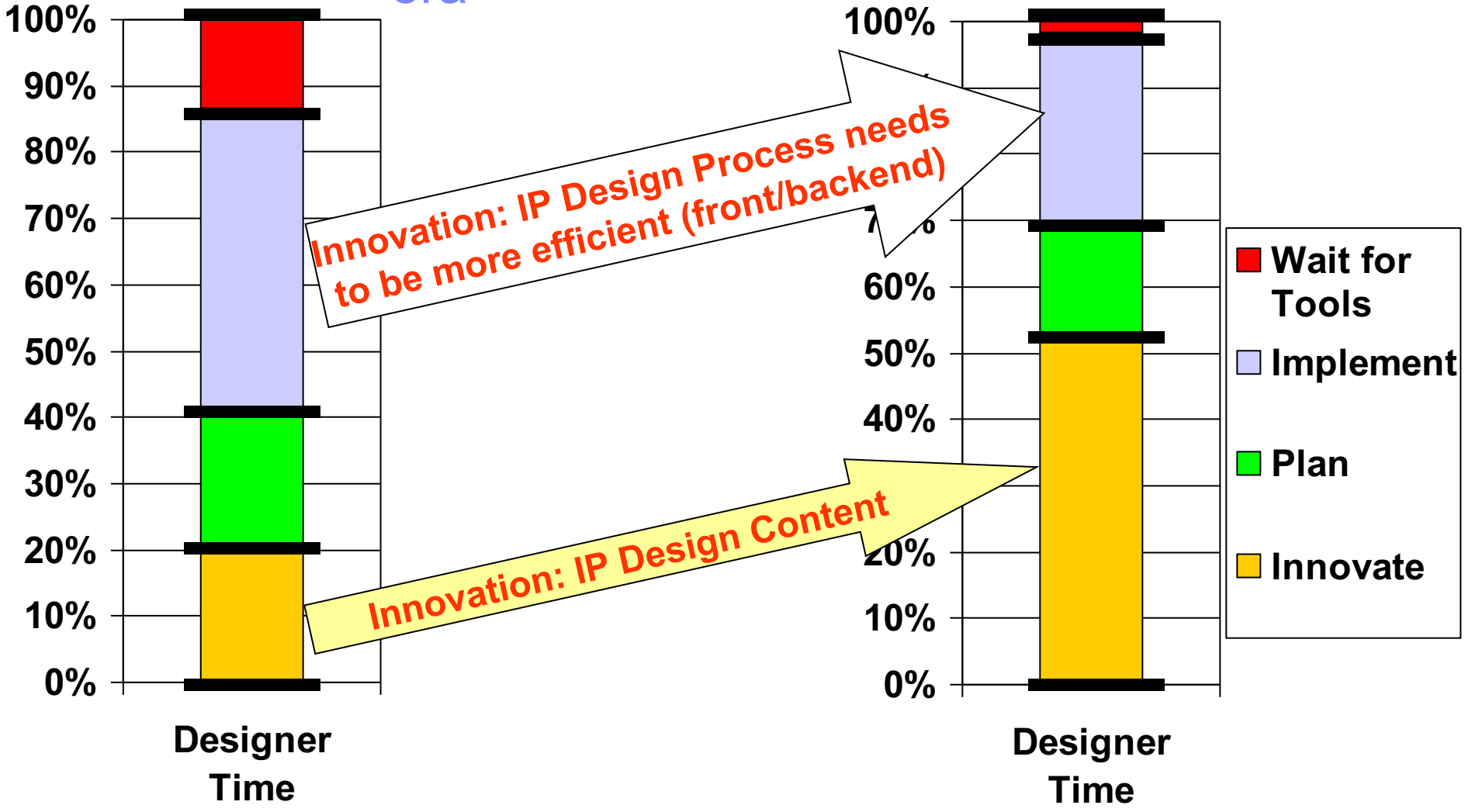
- Fast global analysis tools allow designers to iterate more often resulting in improved final designs.
- Hierarchical abstraction & multi-threading are the most promising ways to minimize TAT.
  - –Applies to all disciplines (timing, verification, etc)

Productivity Innovation: Retiming

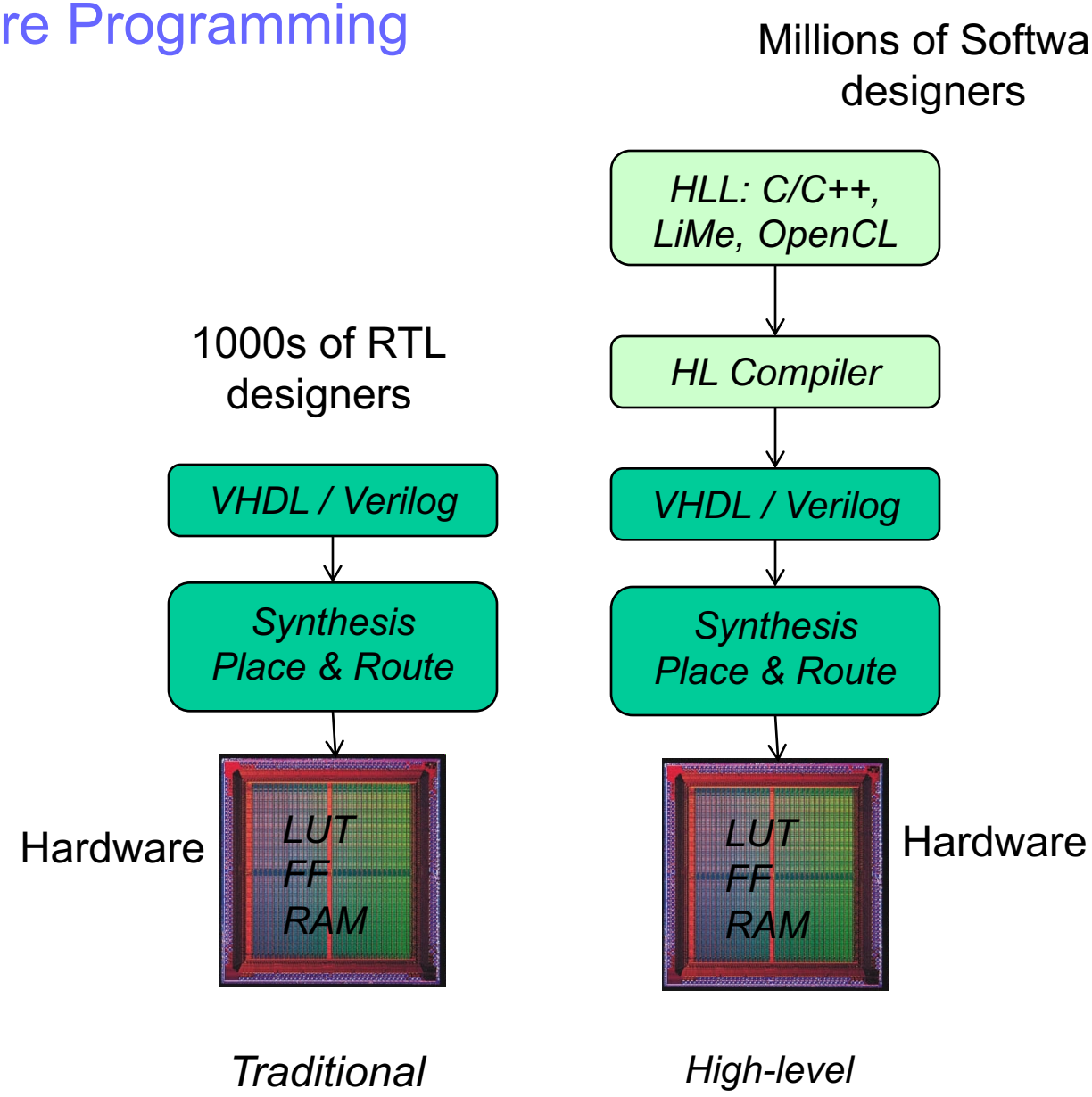


- Significant fraction of logic designer effort spent in optimizing cycle boundaries
- Retiming enables physical synthesis to optimally place latches in logic cones to balance timing/area/power
- Invention is required to seamlessly handle divergence between functional RTL (Verilog/VHDL) and physical implementation throughout methodology.

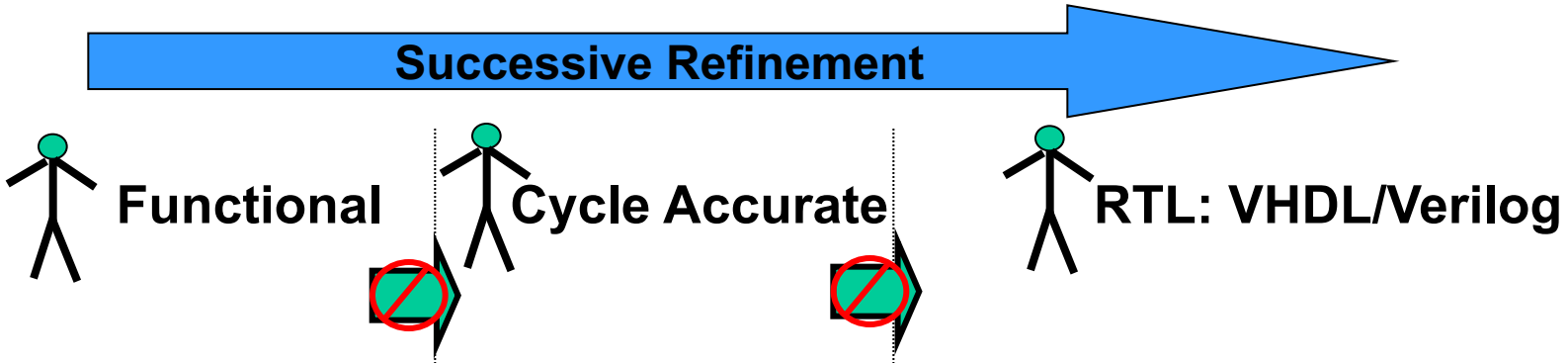
Innovation: The sweet spot in this new era



Hardware Programming



Architectural Synthesis



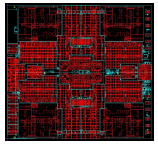
C/C++ Model

Metric: Cache Miss rate etc.

C/C++ Model

Metrics: Performance Models, CPI etc

Back End Design Implementation and Analysis



Metrics: Electrical, Timing, Area, Noise etc.

Summary

- Information technology landscape is changing dramatically
  - Value is in innovating across the entire stack and increasingly higher up in the stack
  - Key problems remain to be solved in technology, design and automation as technology continues to scale
  - Significant emerging opportunities in new ways to solve system bottlenecks at every levels: Logic, Architecture, Memory.
  - In last several years, life became very challenging but also very interesting as the ride has gotten a lot choppier…
  - With challenges and opportunities abound, organizations that grab these challenge and innovate their way out of the current dilemmas will be the winners.