FPGA Challenges and Opportunities at 40 nm and Beyond

Vaughn Betz

Senior Director, Software Engineering

Altera

Scaling: Good for FPGAs

Scaling FPGAs: The Opportunity

- ~2X logic, RAM, etc. density
  - More processing on one chip
  - → Lower system cost & power
  - → Enable higher performance systems

- New features
  - Block RAM, DSP, serial interfaces, …
  - → Enable new systems *or* fewer chips

- (Maybe) Higher speed fabric
  - → Higher performance system *or*
    smaller & cheaper system (narrower datapath)

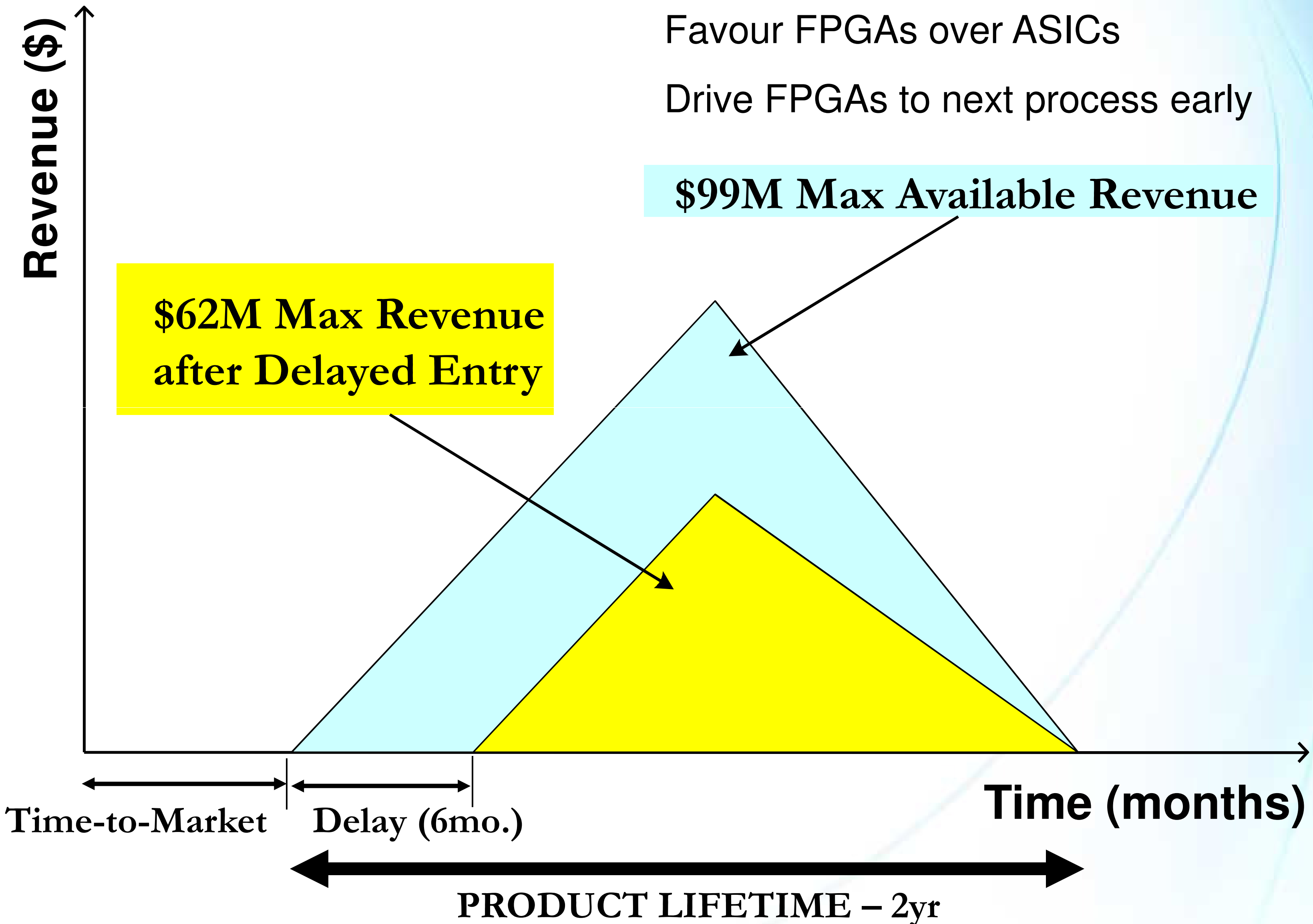
ASICs & Scaling

- Standard cell ASIC @ 40 nm
  - ~$4 M / mask set * 2 spins = $8 M
  - Test & product engineering ~$7 M
  - Design, verification, software ~$25 M
- Economics
  - $40 M development cost
  - 20% of revenue on R & D → need $200 M revenue
  - 10% market share → Need a $2 B market
- Result
  - Falling ASIC starts
  - Most still in 130 nm and above
  - Structured ASICs
  - ASICs increasing programmability → try to increase market size

Time-to-Market Economics

Favour FPGAs over ASICs

Drive FPGAs to next process early



ALTERA

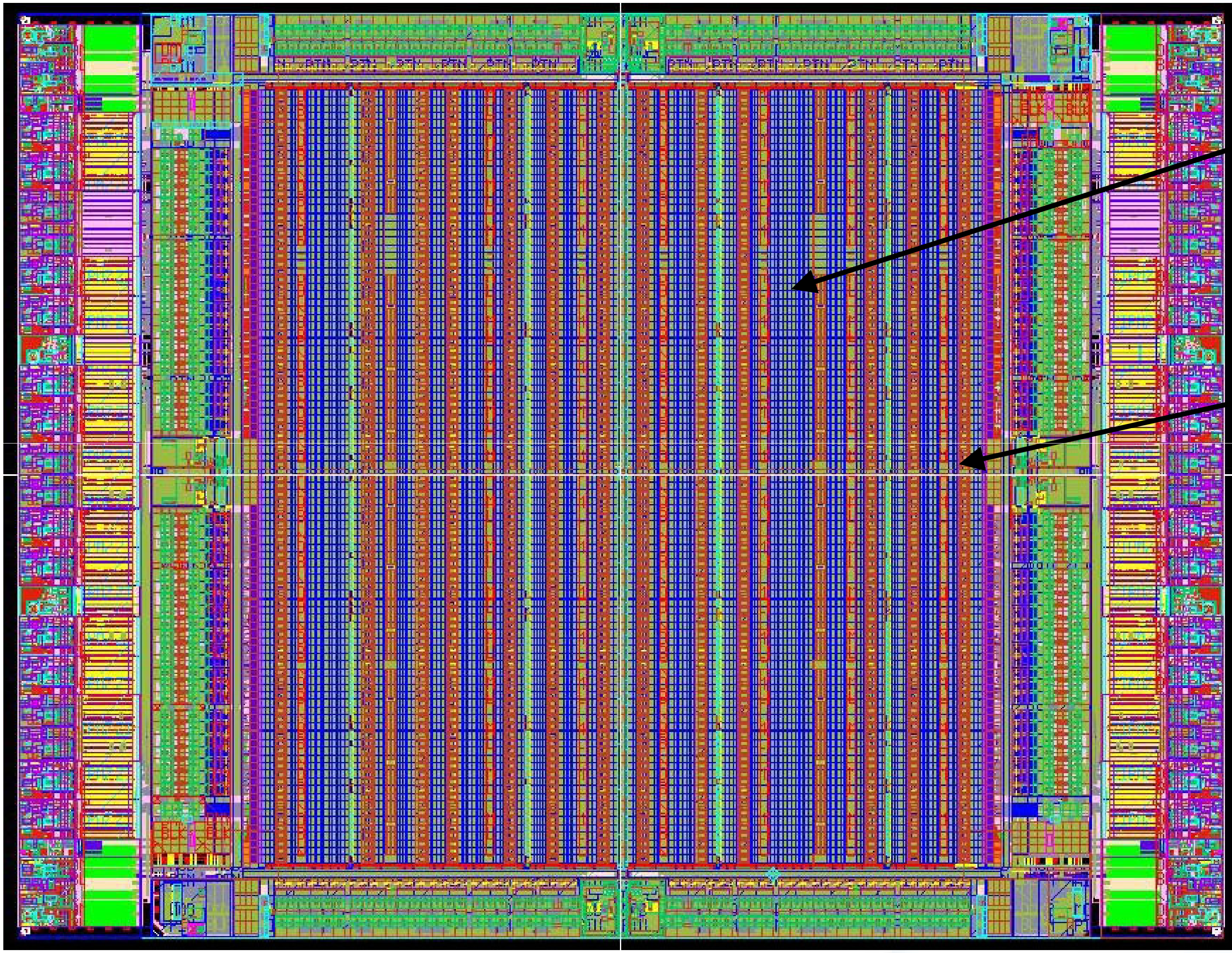
Stratix IV

FPGAs: Process Leaders

- Stratix IV
  - Shipped in 2008
  - First 40 nm FPGA & one of the first 40 nm chips
    - FPGA designed simultaneously with 40 nm process
  - 3 years & >$200 million to develop hardware + software + IP
  - Process driver: large & regular; contains logic & RAM
  - 40 nm allows integration of new hard (“ASIC”) functions

- Pipelined development
  - 28 nm underway for two years

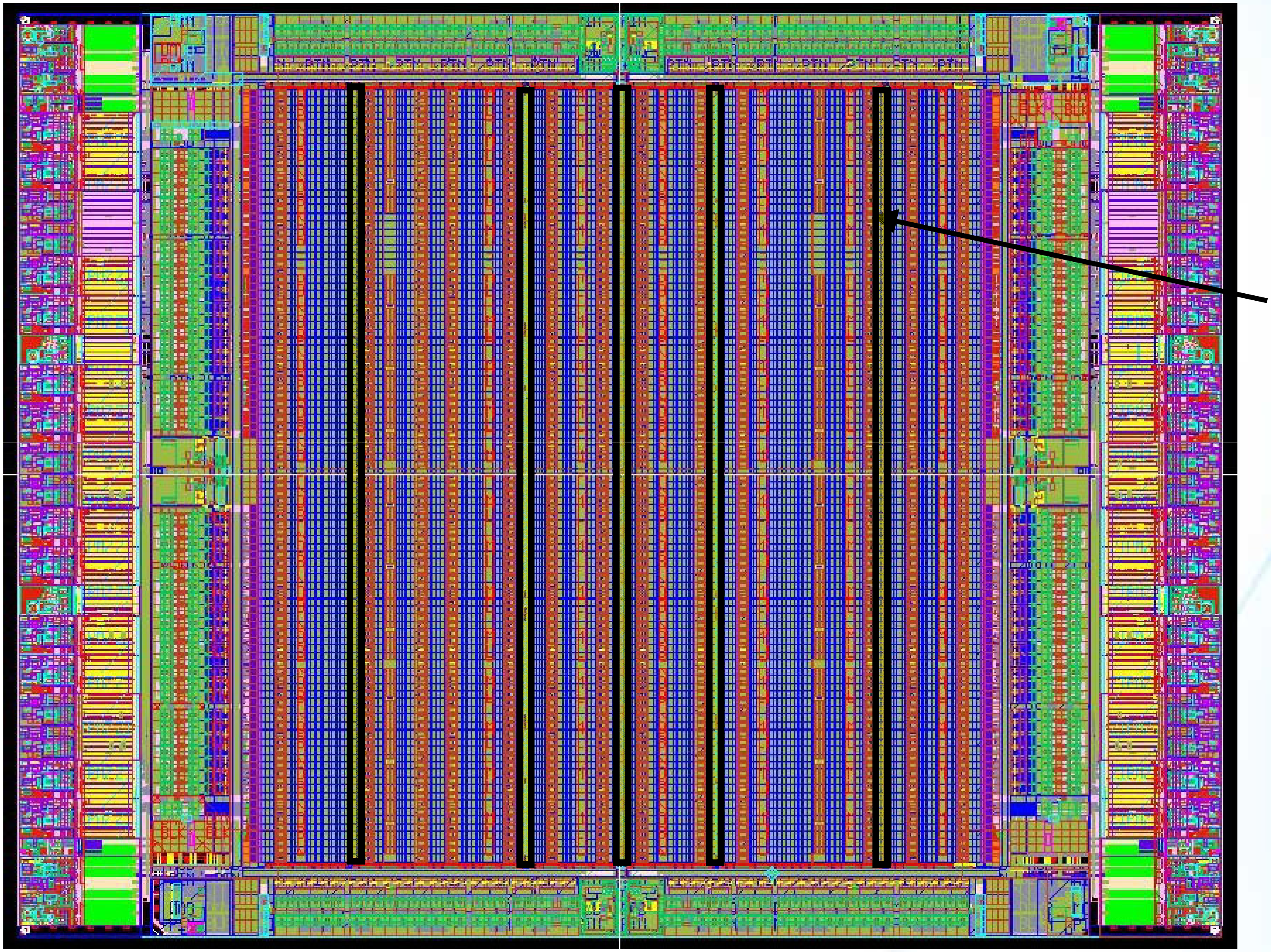
Stratix IVGX 230 (Mid-Size Device)



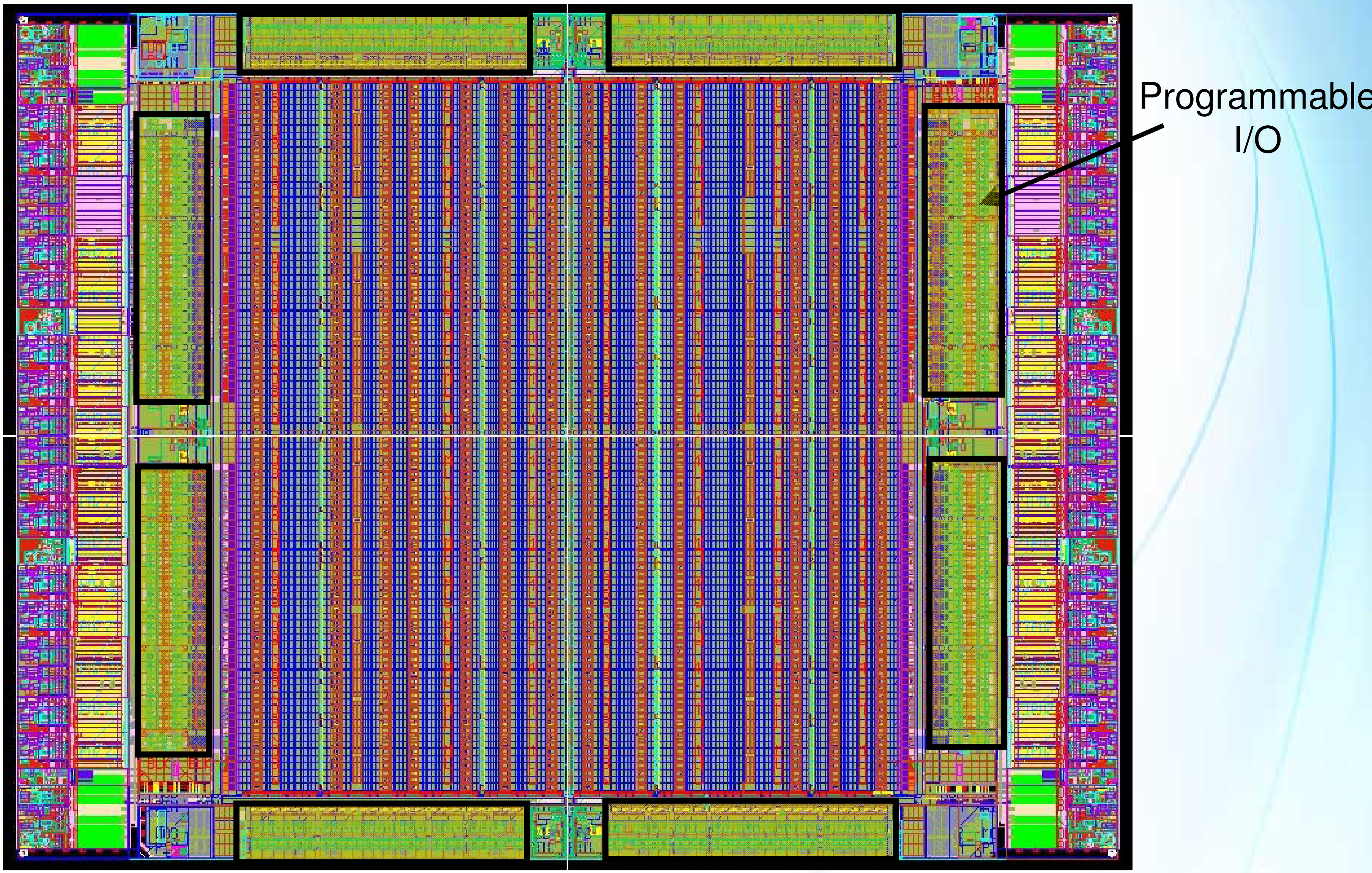
Adaptive Logic Modules

RAM Blocks (M9K & M144K)

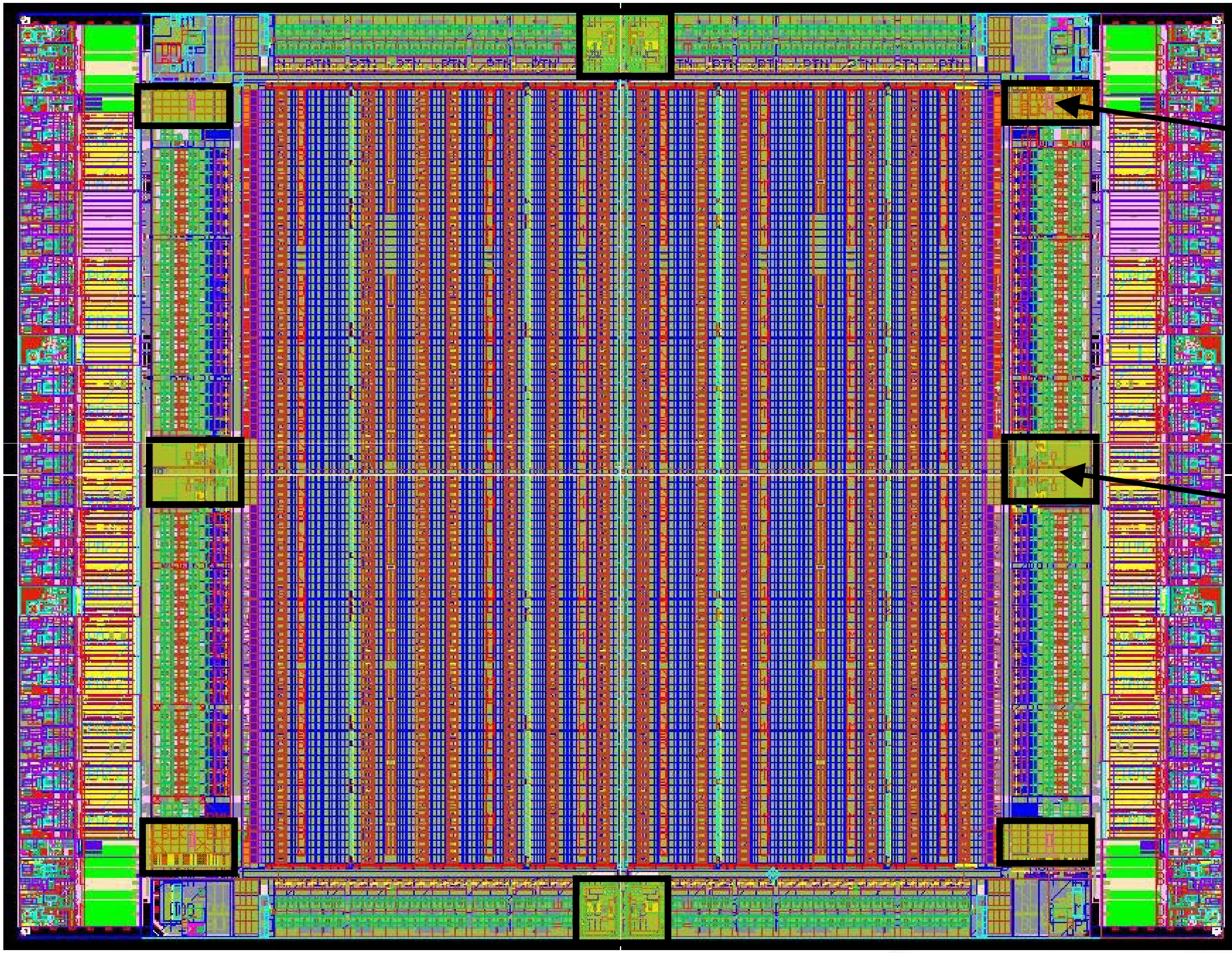
Digital Signal Processing



General I/O



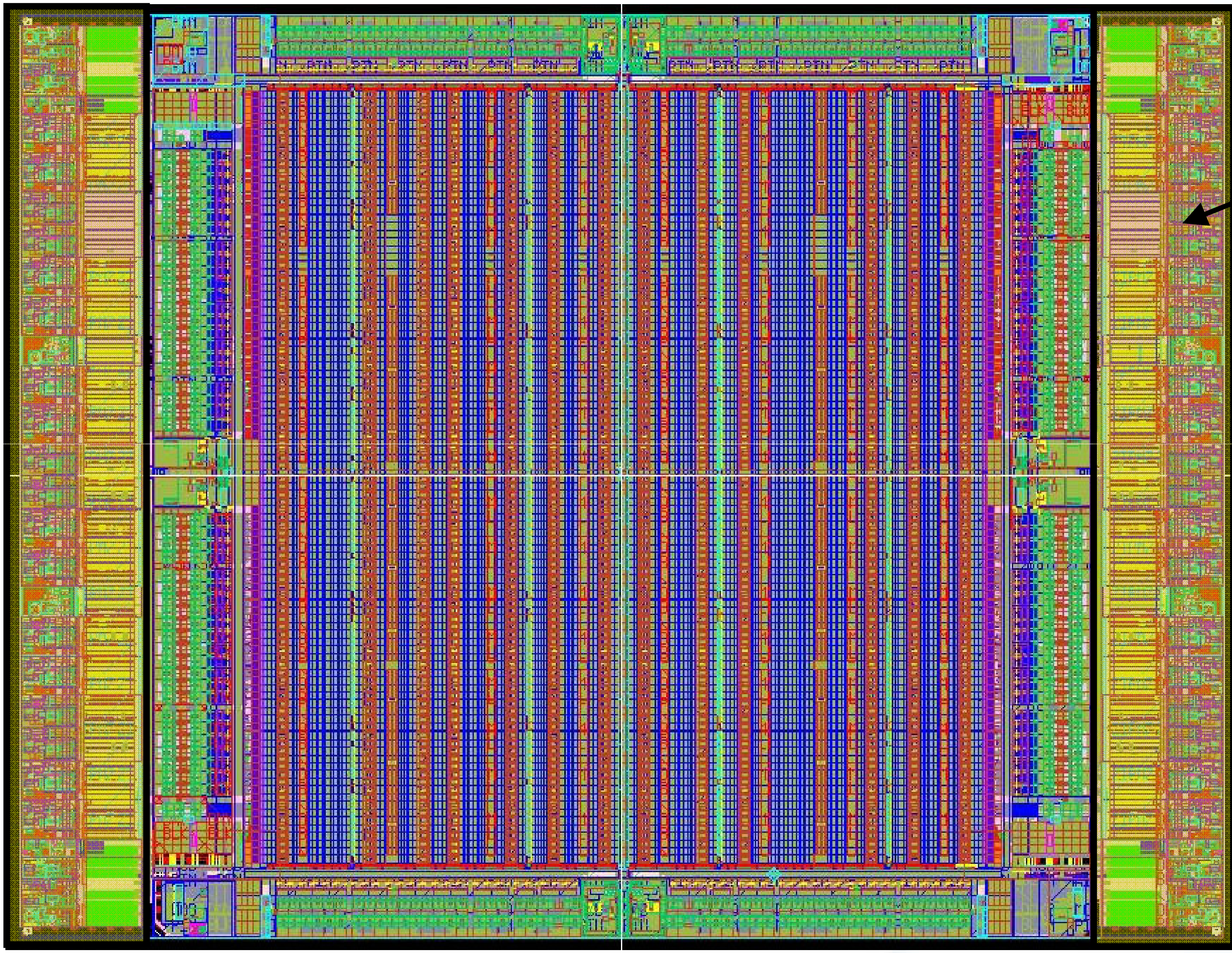
Clock Management



Delay Locked Loops (DLLs)

Phase Locked Loops (PLLs)

Serial Interfaces



Stratix IV Overview

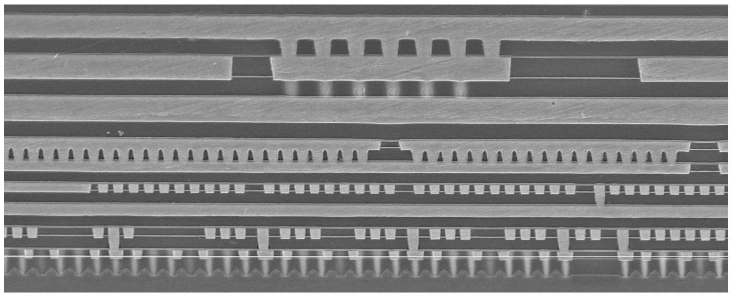
Feature	Stratix III (65 nm)	Stratix IV (40 nm)
Logic Elements	340k	820k
RAM bits	16 Mb + 4 Mb	33 Mb + 8.5 Mb
18x18 multipliers	768	1360
General I/O	1104	1104
High-speed serial links	0	48 transmit + 48 receive @ 11.3 Gb/s
Hard PCIe blocks	0	4
Clock generation	12 PLL(x10)	12 PLL(x10) + 32 serial recovered + + 24 serial transmit
Clock distribution	16 Global + 88 Quadrant + 132 PCLK	16 Global + 88 Quadrant + 132 PCLK

Scaling Challenges

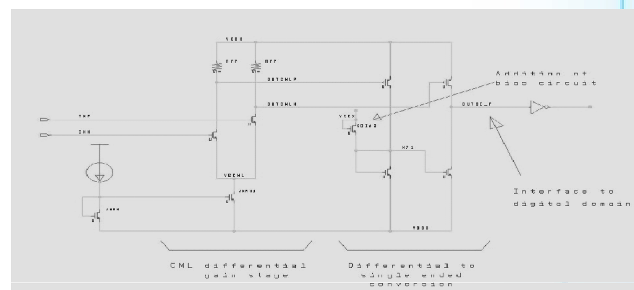
1. Designing & Optimizing the Fabric

FPGA Fabric: Converging Technologies

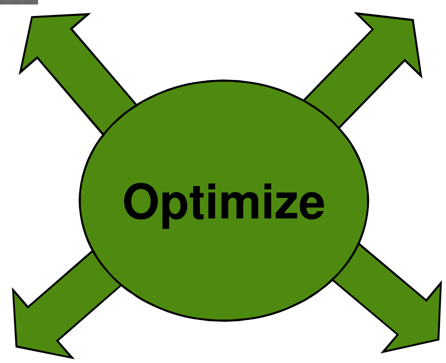
- Huge space → no one can optimize by instinct
- $250 M + 3 years to implement your ideas
  - Risk of over-conservatism

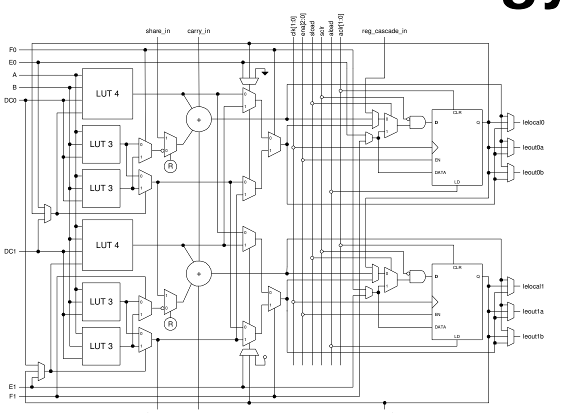


Process Technology

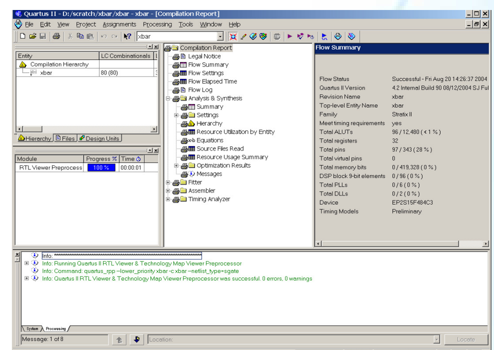


Circuit Design





FPGA Architecture



Software

Architect via Virtual Prototyping

FPGA Arch. Spec (150 pages)

```
# Logic architecture parameters
subblock_lut_size 2   # Using 2-LUT BLEs
subblocks_per_clb 1   # One 2-LUT per LAB

# Routing architecture parameters
switch_block_type subset
Fc_output 1
Fc_input 1

switch 0 buffered: yes  R: 1000.0  Cin: 1.0e-15 \
Cout: 1.0e-15

# All buffered length 1 wires
segment frequency: 1.0 length: 1 wire_switch: 0 \
opin_switch: 0 Frac_cb: 1 Frac_sb: 1 \
Rmetal: 100.0 Cmetal: 1.0e-14

# Process parameters
R_minW_nmos 5000
R_minW_pmos 10000
```

VPR++

FPGA Database (1 GB)

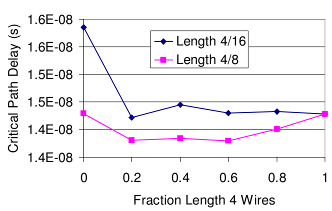
Customer Designs
IP, Reference Designs

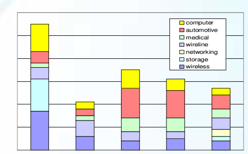
Parameterized Quartus Synthesis

VPR++ Place & Route

Timing, Area, Power Models

VPR++ Analysis: Speed, Area, Routability, Power

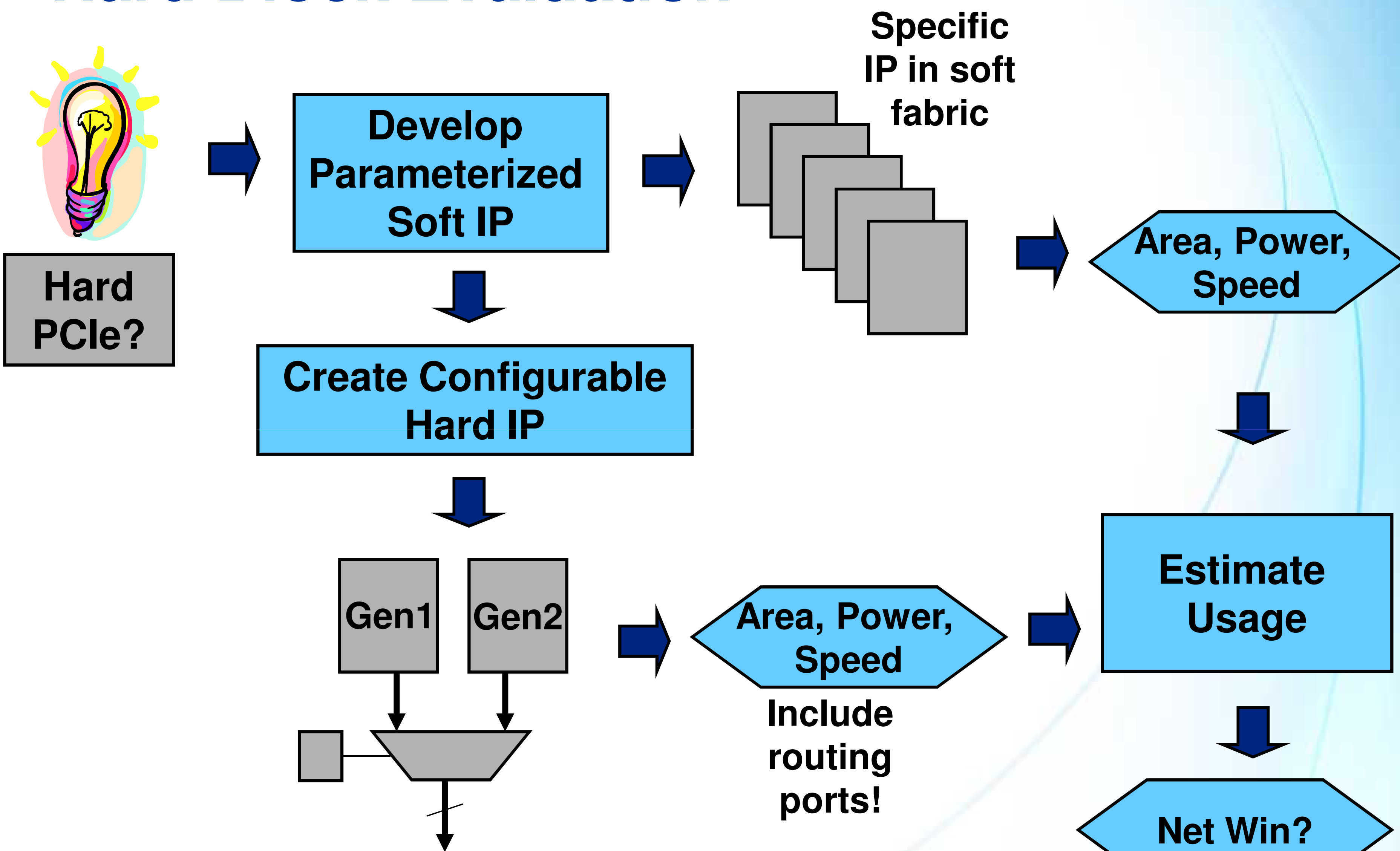




Major Innovations

Family	Major Innovation	Benefit
Stratix	100% Direct-drive, Optimized segmented routing	+40% Fmax, -40% area vs. APEX *(not including process adv.)* Redundant and repairable
Stratix II	Adaptive Logic Module	+26% Fmax, -7% area vs. Stratix *(not including process adv.)*
Stratix III	Programmable Power	Full speed of 65 nm process, ½ leakage of 90 nm

Hard Block Evaluation

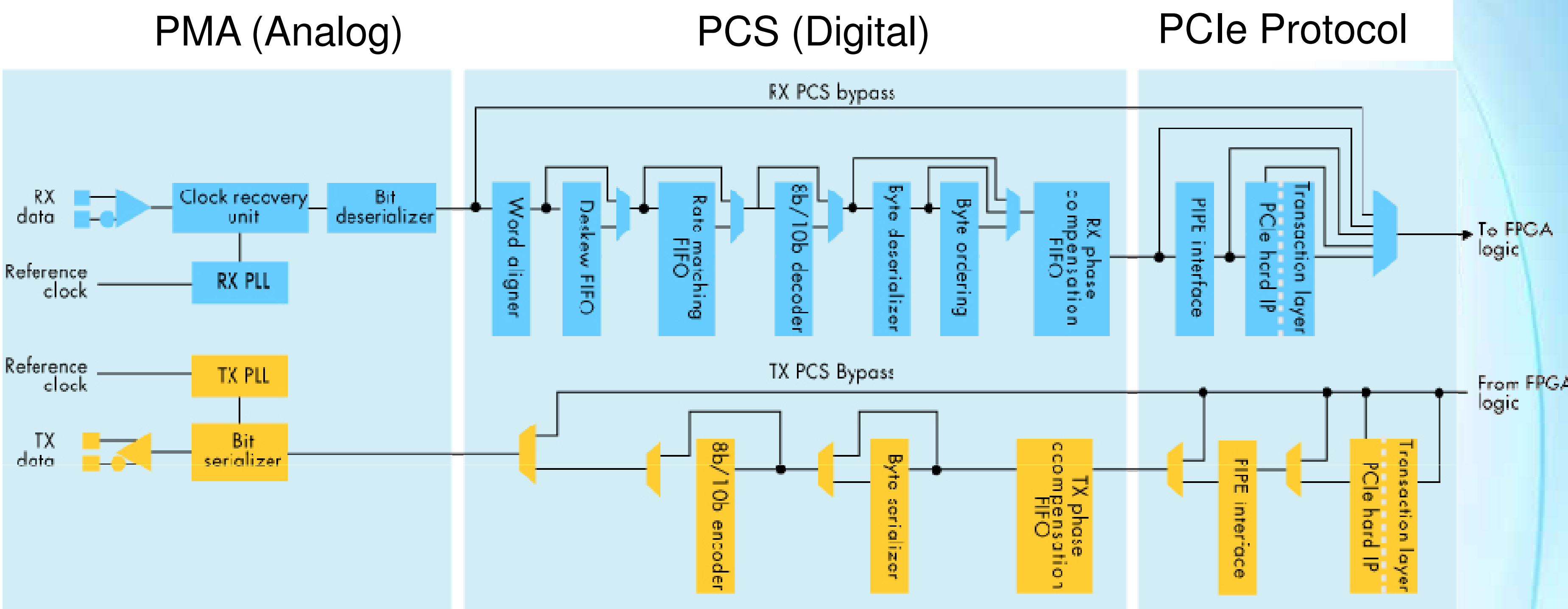


2. I/O Bandwidth

I/O Bandwidth

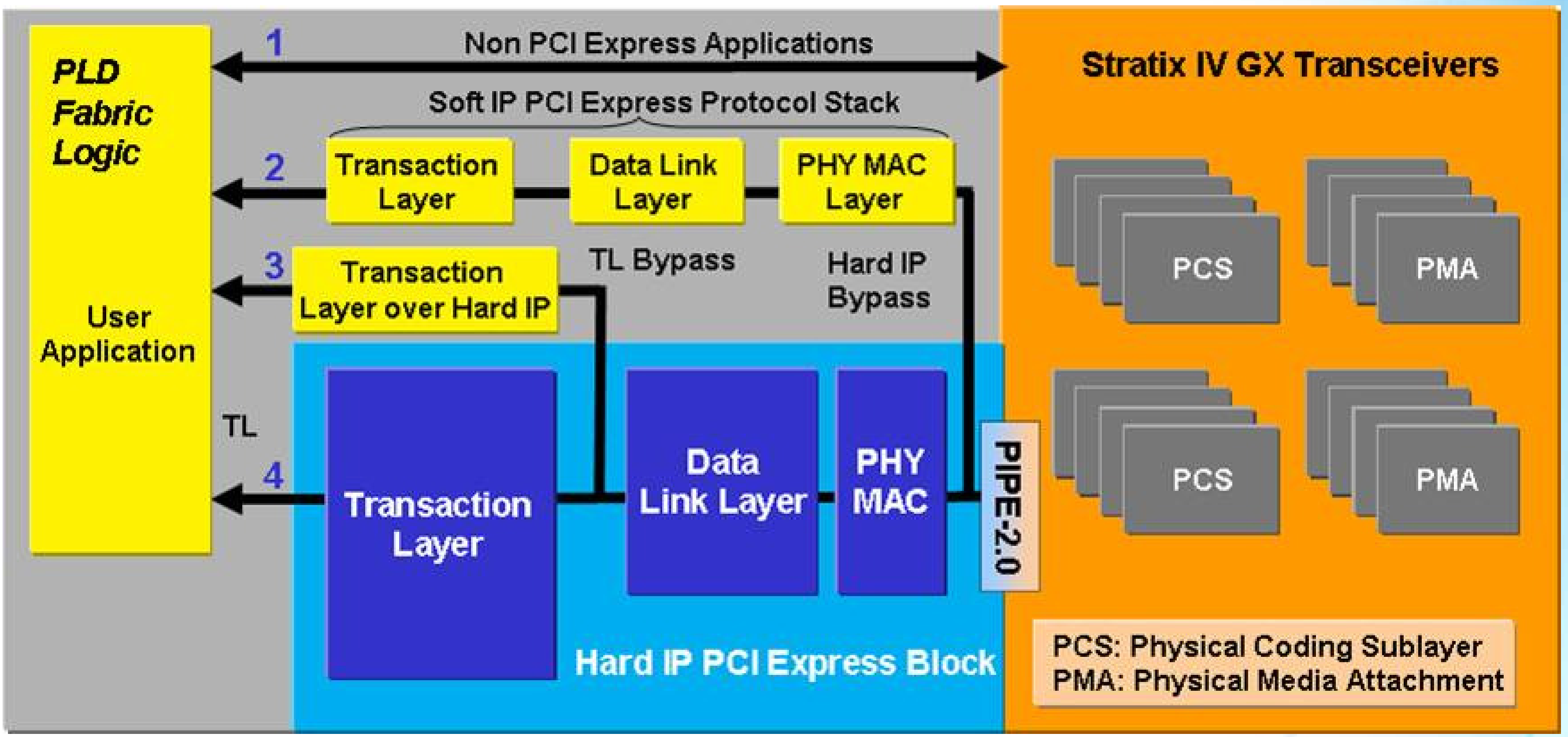
- Processing elements scale
  - ~2X more logic, RAM, DSP each generation
  - Stratix IV on-chip RAM bandwidth ~10 TB/s!
- I/O transistors, PCB traces, package balls don't scale
  - Roughly same number of I/Os per device
- Need: higher speed I/Os to keep datapath fed
  - 8.5 (SIVGX) to 11.3 (SIVGT) Gb/s serial transceivers
  - 1.067 Gb/s (533 MHz) memory interfaces
  - Total: ~150 GB/s bandwidth
- Challenges
  - Circuit speed & timing closure
  - Signal integrity

Stratix IV GX Embedded Transceivers



- Up to 48 receive + 48 transmit transceivers
  - 3 Gb/s to 11.3 Gb/s
  - Clock recovered from data stream
- Very high speed analog
- Many protocols → highly configurable analog & digital logic

Hard PCIe Block

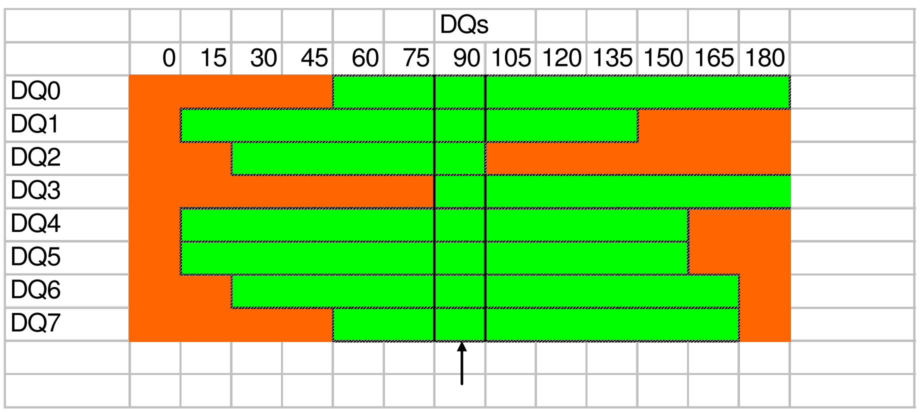


- Build hard PCIe?
  - ✓ Soft logic size: tens of thousands of LEs
  - ✓ Standard cells: smaller
  - ✓ Hard logic faster: can narrow datapath (smaller)
  - ✓ Protocol fixed and widely used
  - ✗ Add muxing and always build all options (Gen1, Gen2, …)

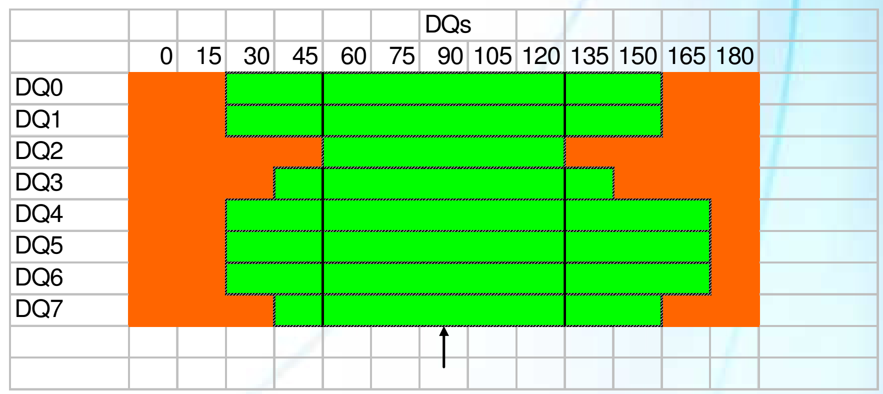
Memory Interfaces (DDR2, DDR3, …)

- Strobe (DQS) sent with several bits of data (DQ)
- Challenge: narrow data-valid window
- Solution:
  - Minimize jitter
  - Harden data capture logic, carefully match delays
  - **Calibrate** by modifying programmable delays for each DQ bit & DQS

Before de-skew -- small valid capture window

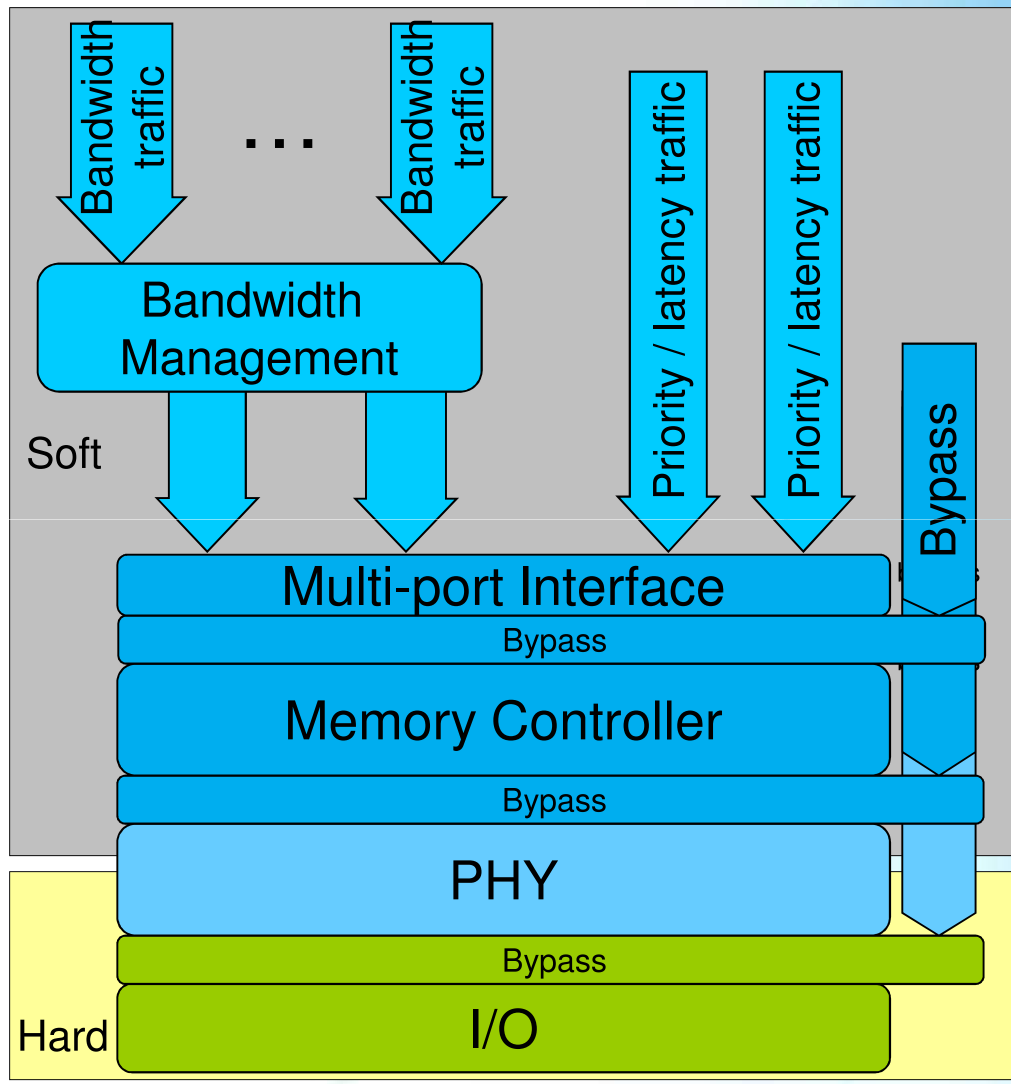


De-skew maximizes valid capture window



Memory Interface Stack: Hard vs. Soft

- Some PHY circuitry very high speed, needs carefully matched delays
  - → Hard circuitry
- PHY calibration: logic frequently changes
  - → Soft logic
- Memory controller, multi-port interface
  - – Relatively small (low thousands of LEs)
  - – Many algorithms and needs
  - – Soft logic



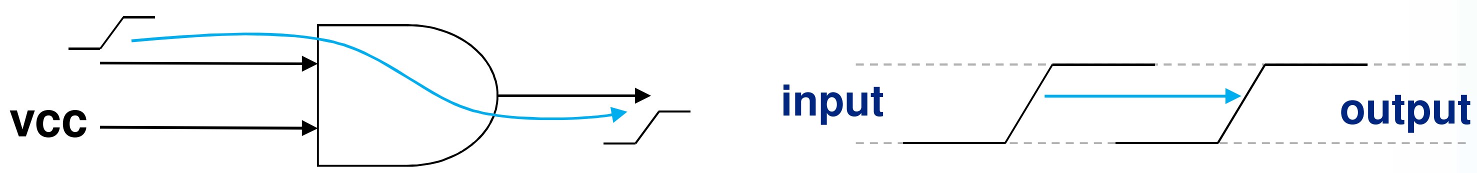
3. Device Modeling

Device Modeling Challenges

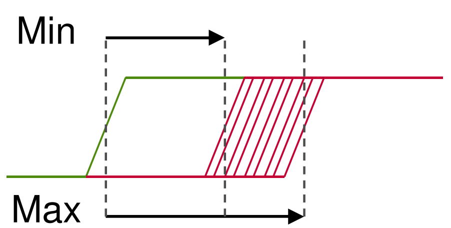
- Smaller transistors
  - → More process variation
- Lower operating voltage, with little V_{th} scaling
  - → Increased sensitivity to power supply noise
- 2nd-order transistor effects increasing
  - → More timing corners
- Faster clock speeds and edge rates
  - → Less ability to guardband
  - → Increased importance of jitter & signal integrity models
- Still need fast, easy-to-interpret analysis

ASIC Class Timing Analysis (Timequest)

- Model rise-rise, rise-fall, fall-rise, fall-fall delays
  - Propogate rise and fall delays through circuit
  - *unateness* aware – ignores impossible transitions



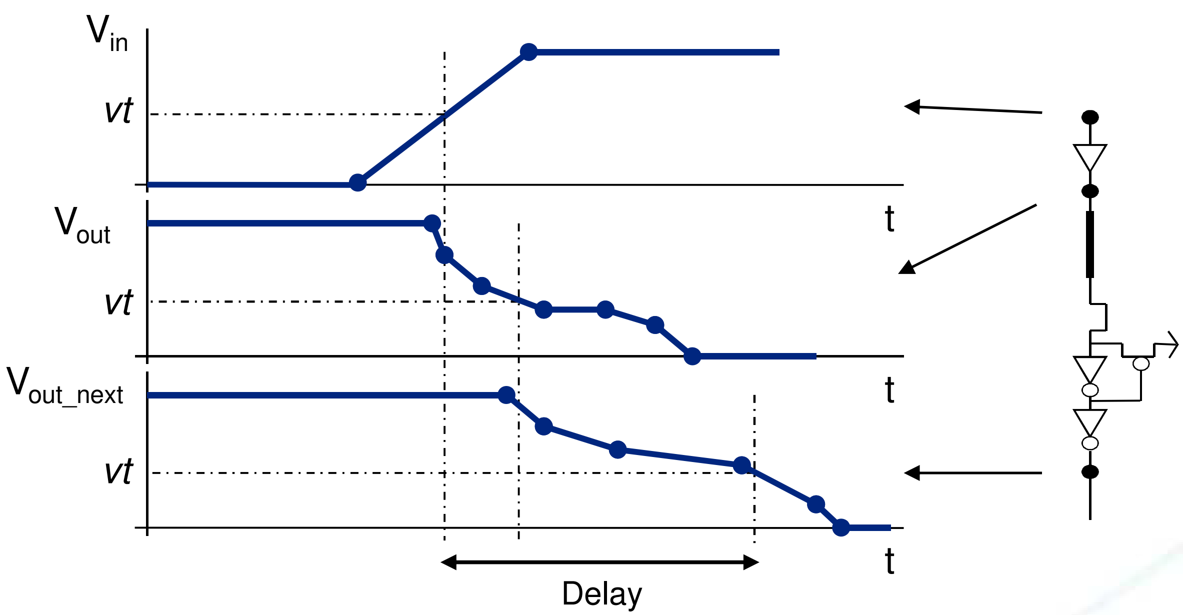
- Each delay is a min-max range
  - Covers on-die variation, transistor aging effects



- Analyze *and optimize* at 3 corners
  - A *corner* is a combination of process, temperature, voltage

Interconnect Timing: Leading ASICs

- ✘ FPGA routing: too many combinations for tables
- ✓ Quartus circuit simulates each route
  - Non-linear models for transistors
  - Extracted R & C for wires
  - Specialized to FPGA circuitry: 10,000x faster than HSPICE



Full waveform propagated, with all non-linear effects

Calibrated I/O Interfaces

- Calibrated interfaces (e.g. DDR3): timing is not static
- But can't simply assume calibration works
  - Need to analyze assuming worst-case devices, V, T, & P variation to ensure robust system
- Solution: extend TimeQuest timing analyzer
  - Calibration algorithm modeled
  - Ability to recapture margin explicitly included in timing analysis

Report

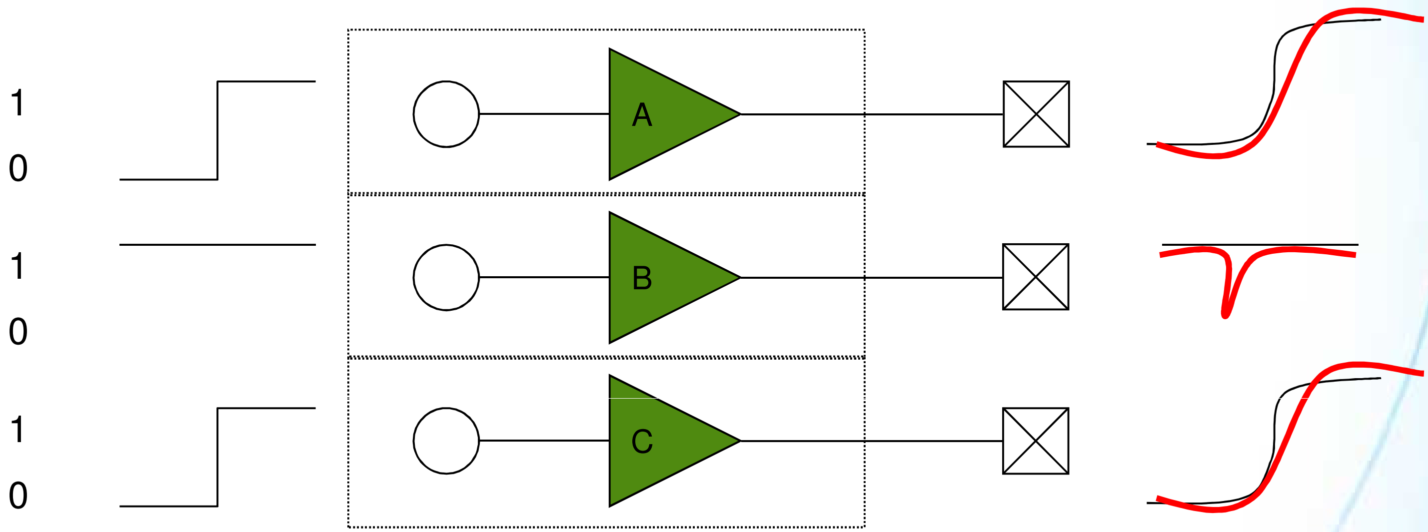
all_ddr3_10_phy Write
all_ddr3_10_phy Write (Before Calibration) (setup)
all_ddr3_10_phy Write (Before Calibration) (hold)
all_ddr3_10_phy Address Command (setup)
all_ddr3_10_phy Address Command (hold)

ddr3_hp_phy_timing_test_sweep_siv_all_ddr3_10_inst|ddr3_hp_phy_timing_te

	Operation	Setup Slack	Hold Slack
1	Before Calibration Write	-0.087	-0.048
2	Deskew Write and/or more clock pessimism removal	0.204	0.165
3	Quantization error	-0.025	-0.025
4	Calibration uncertainty	-0.016	-0.016
5	Duty cycle correction	0.04	0.04
6	Duty cycle correction quantization error	-0.05	-0.05
7	Duty cycle correction calibration uncertainity	-0.01	-0.01
8	After Calibration Write	0.056	0.056

Simultaneous switching noise (SSN)

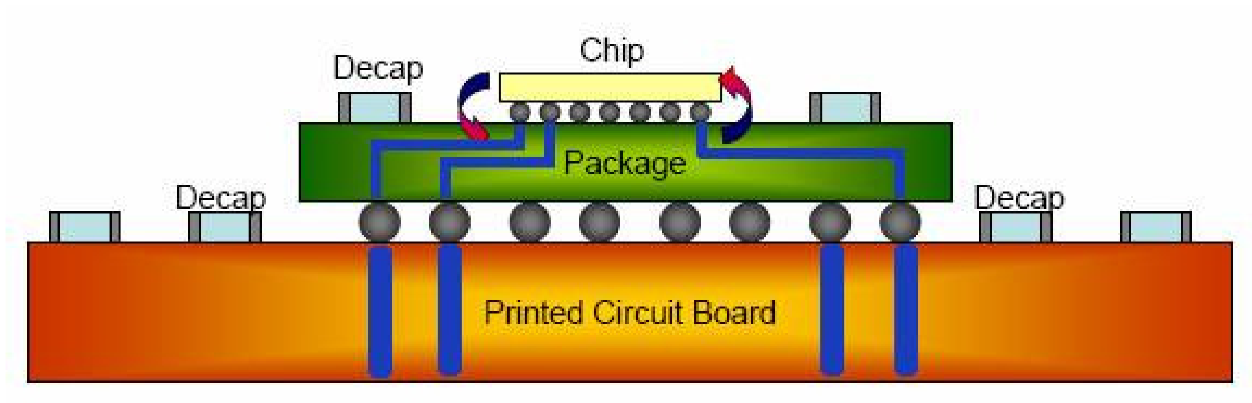
- Noise induced on *victim* I/O due to switching of other *aggressor* I/Os



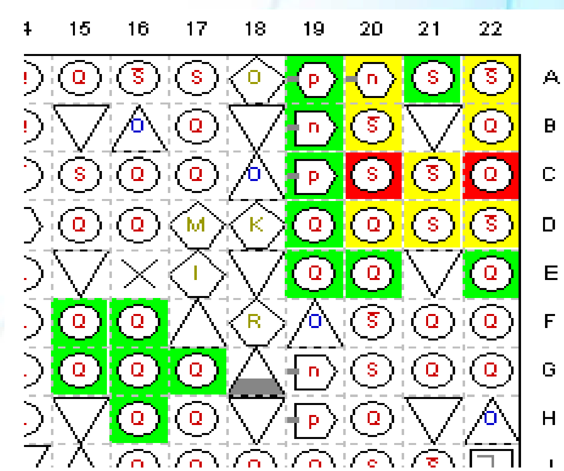
- Faster edge rates & higher I/O density worsen
- Reduce with FPGA & package design
- But cannot eliminate failures for all designs, on all boards

Quartus II SSN Analyzer

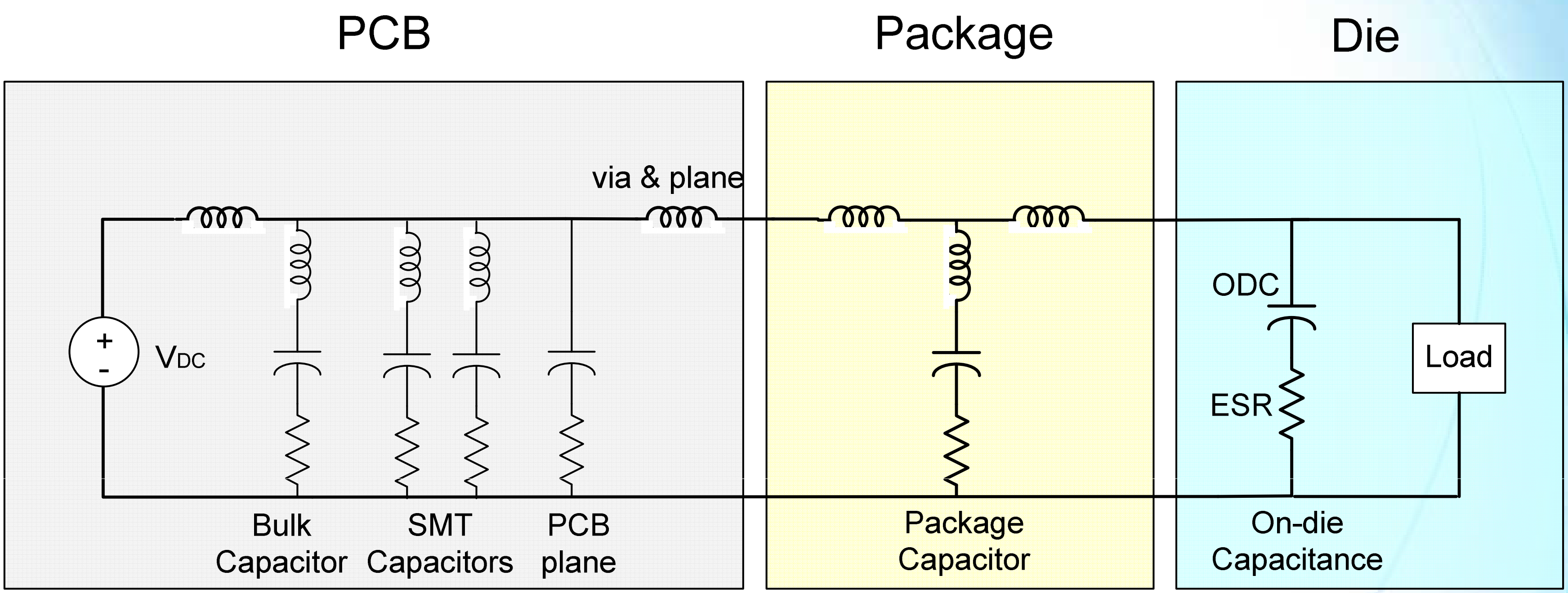
- Models FPGA, package and board
  - Signal paths and power-distribution network



- HSPICE (full design): 1 week
- SSN Analyzer: 30 minutes
- Displays signal margin & problem pins
- Enables analysis of mitigation techniques



Power Integrity



- 0.9 V operating voltage – can only afford tens of mV drop
- Hard & expensive: low-impedence for all freq.
- But design could go from 2 A – 12 A in one cycle!
- Over-engineer *or* analyze and forbid?

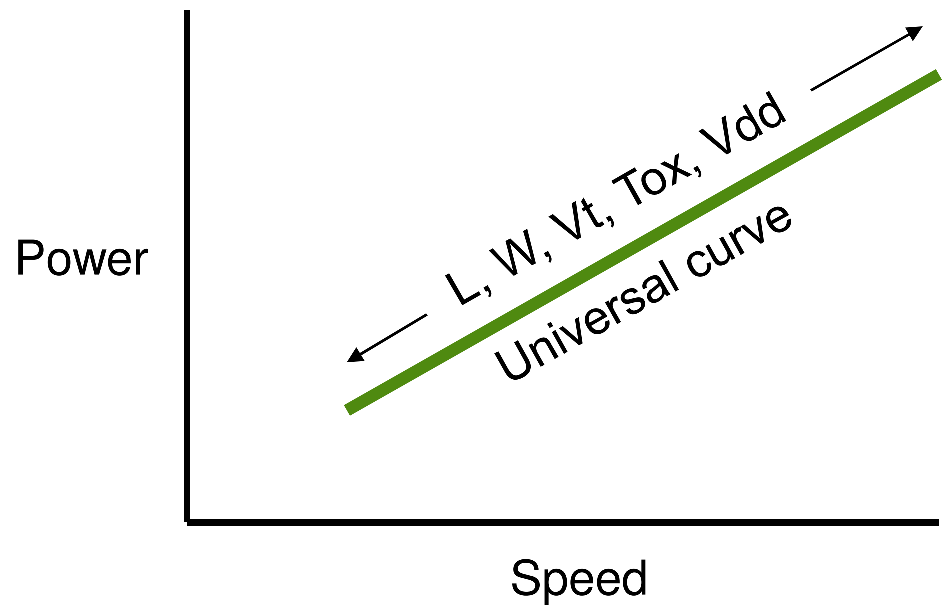
4. Power

Power

- Twice as many transistors
- Naturally more leaky
- But power budget per device fixed
  - About 2 – 20 W for high-end FPGAs

→ Innovate to control power
while sacrificing minimum performance

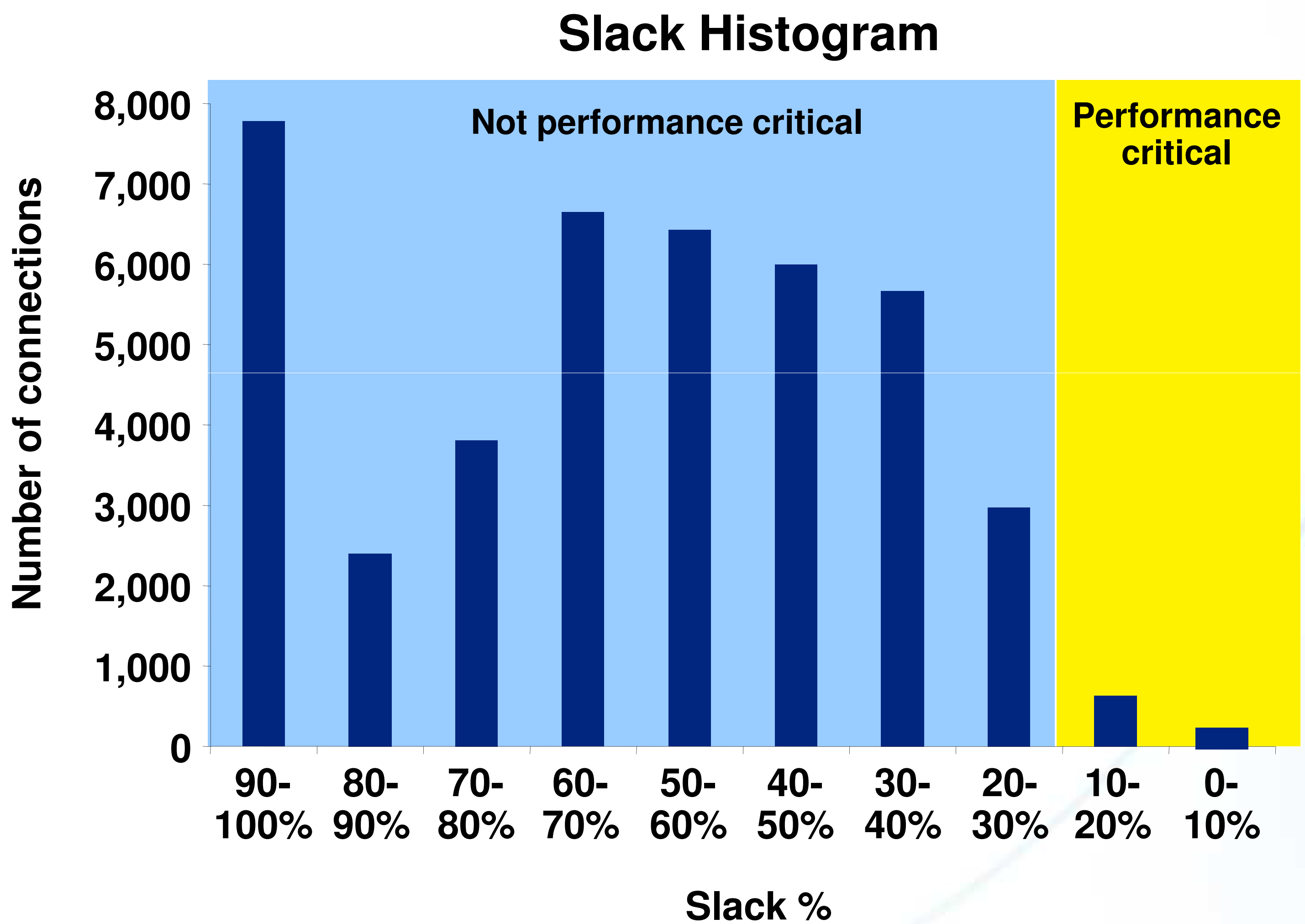
Can We Do Better than the 'Universal Curve'?



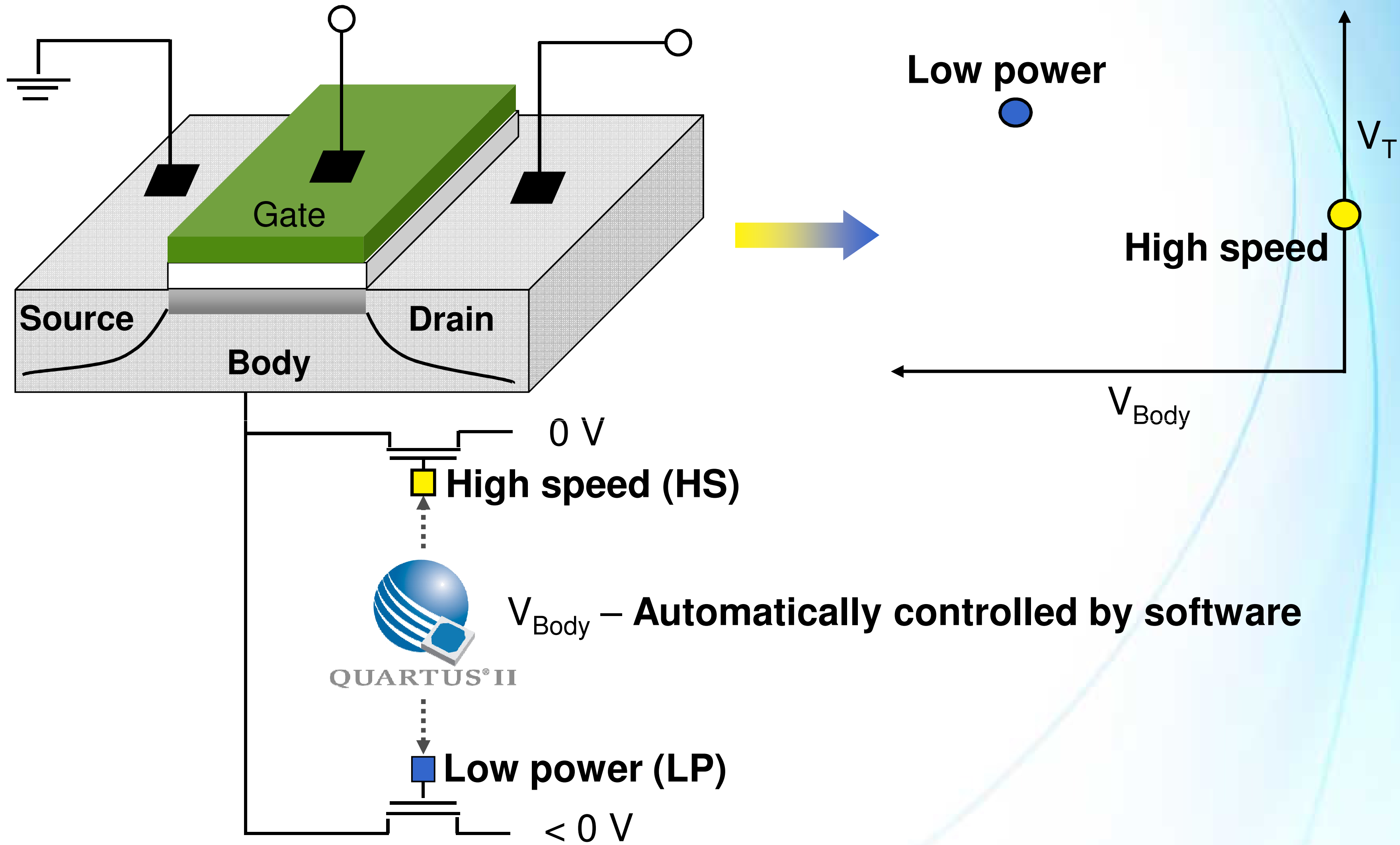
- Optimizing along the curve gives <u>fixed</u> choices
- Choose different values for each transistor
  - Helps for some transistors which are never speed critical (e.g. CRAM)
  - But most transistors will need to compromise between speed & power

Design-Specific Power Optimization

- Only a small fraction of logic is performance critical



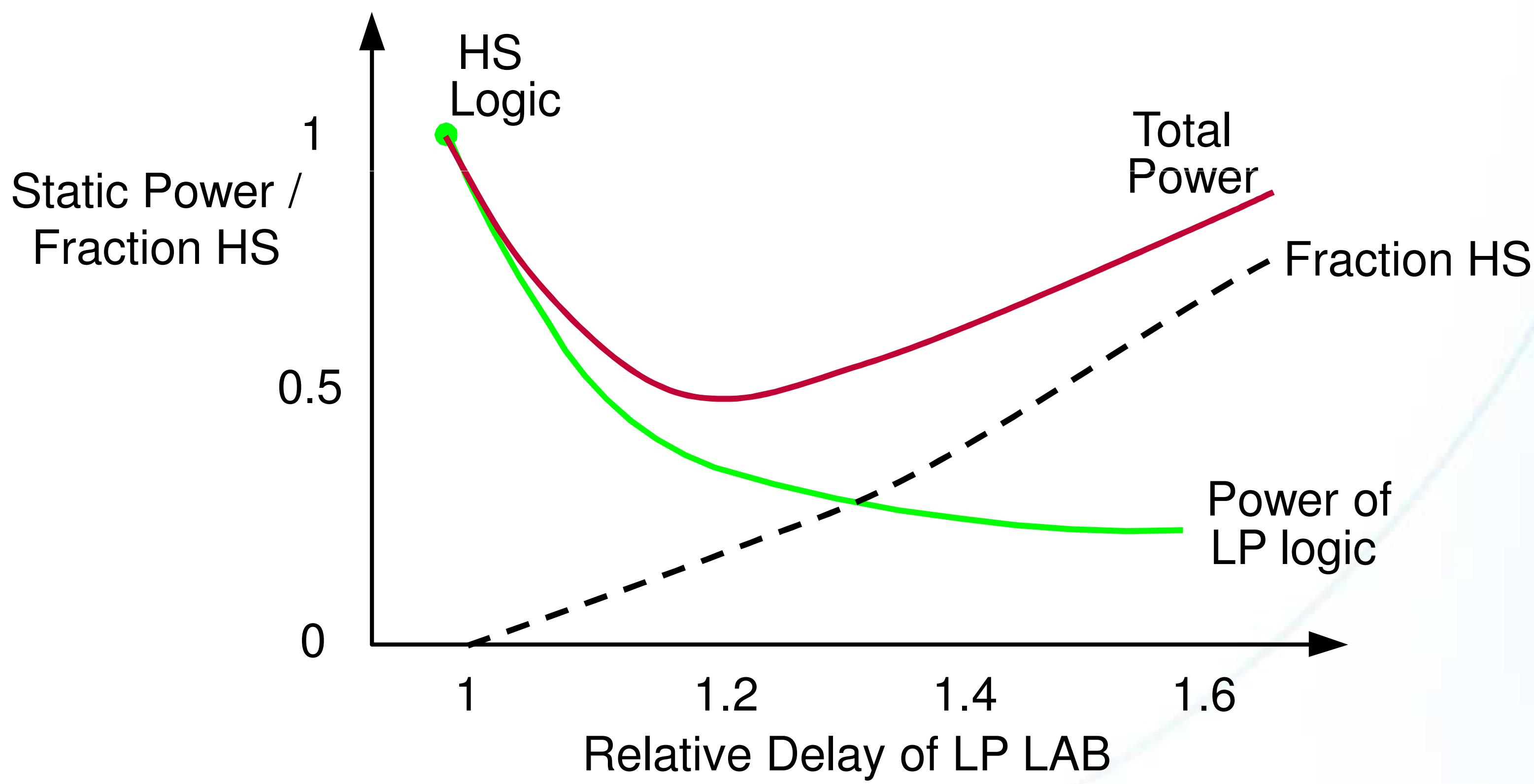
Programmable Speed vs. Leakage



Note: A simple "model" showing Programmable Power Technology. Actual implementation varies and is patented.

Tradeoff 1: How Much Back-Bias?

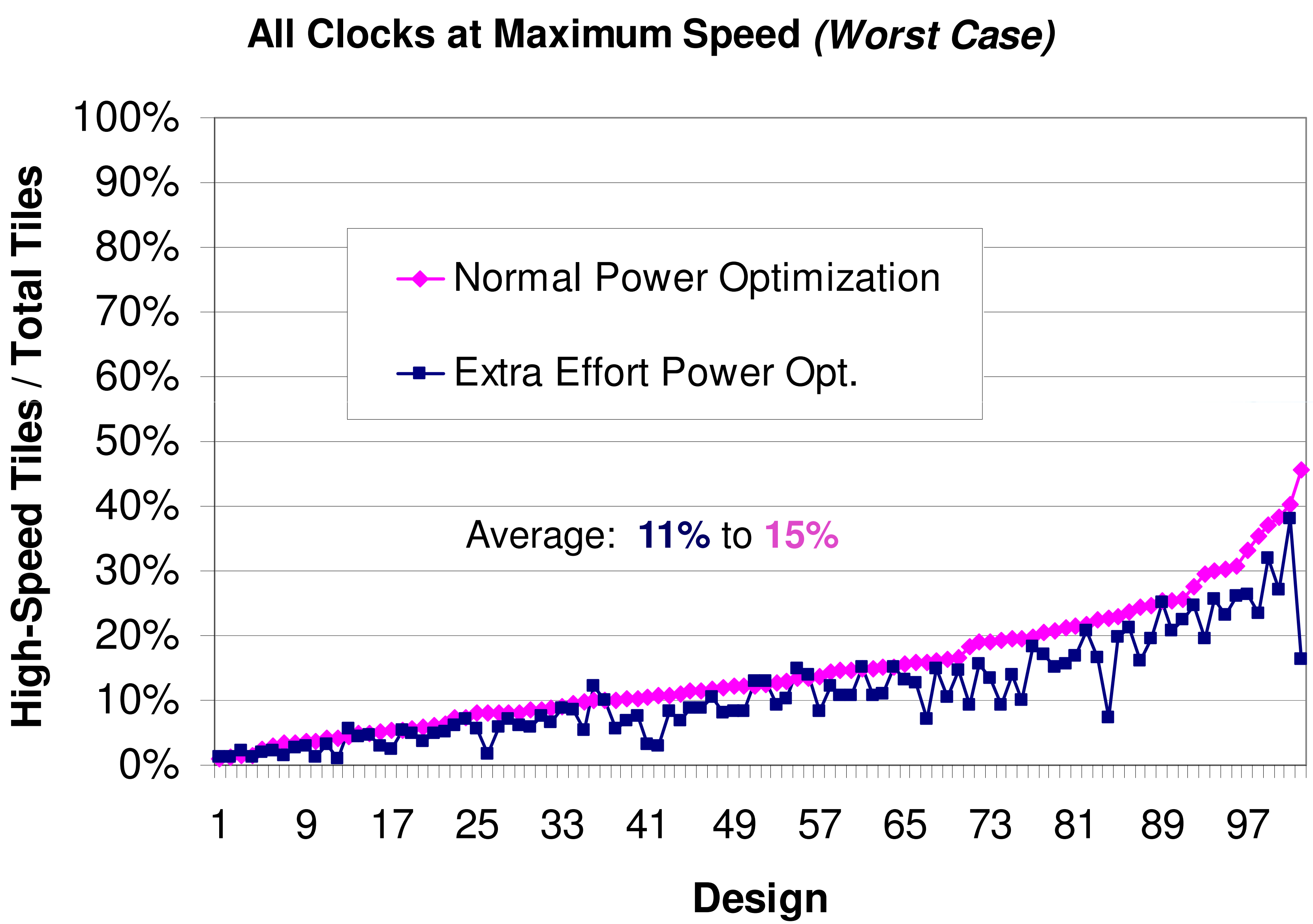
- Power minimum at LP ~20% slower than HS
  - Low Power logic has 40% the static power of HS at that point



Tradeoff 2: Control Granularity

- Finer grain control
  - Can set more transistors to low-power state
  - But costs area (well spacing, CRAM, pass gate)
- Critical paths naturally cluster
  - OK for many transistors to be grouped into a *tile*
- Controlling logic and routing with a single setting
  - Small increase in power, larger area reduction
- Controlling pairs of LABs together: also good
- Stratix IV tile: LAB pair, DSP or RAM block

Most Tiles Are Low Power

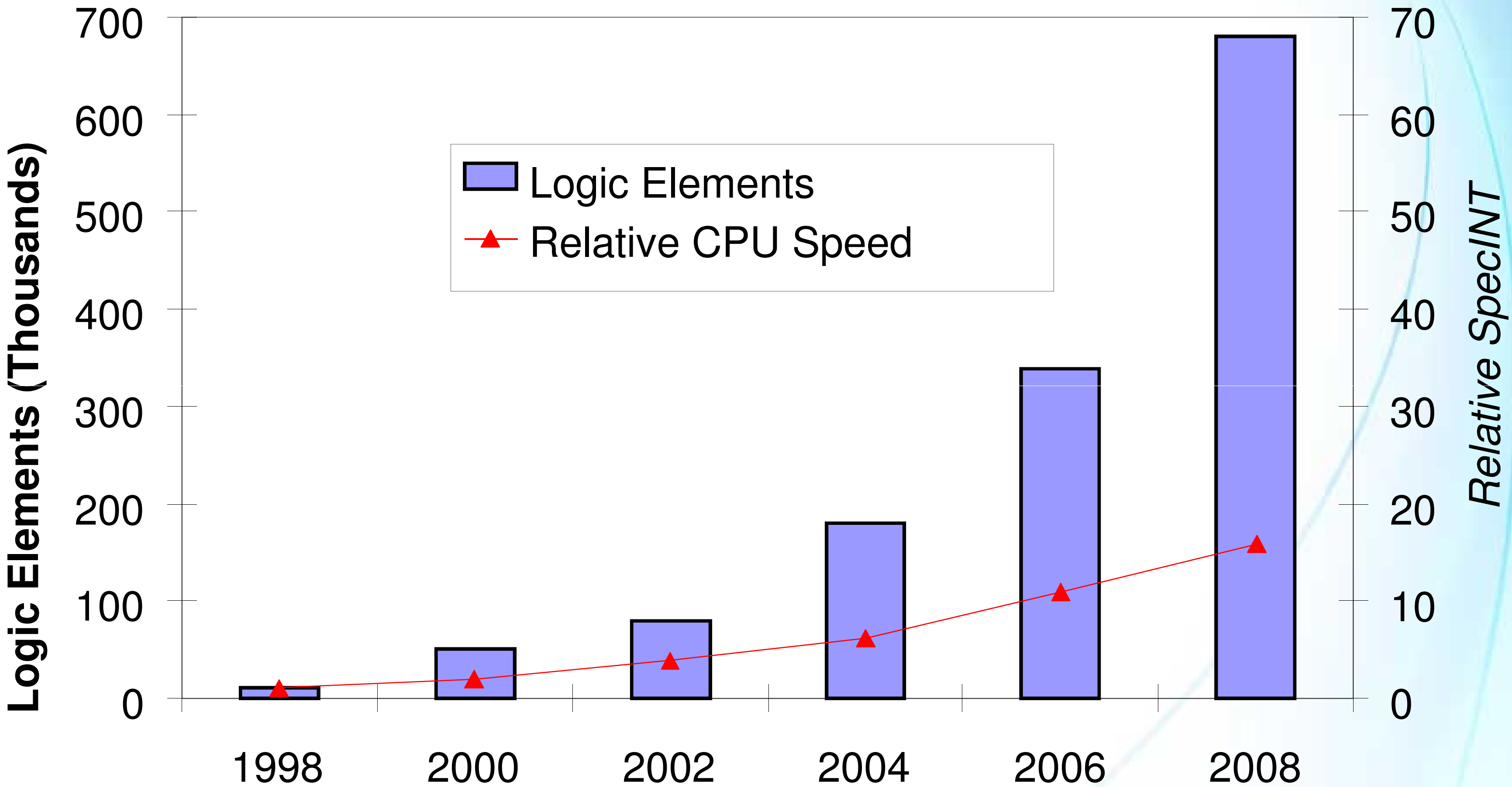


5. Designer Productivity

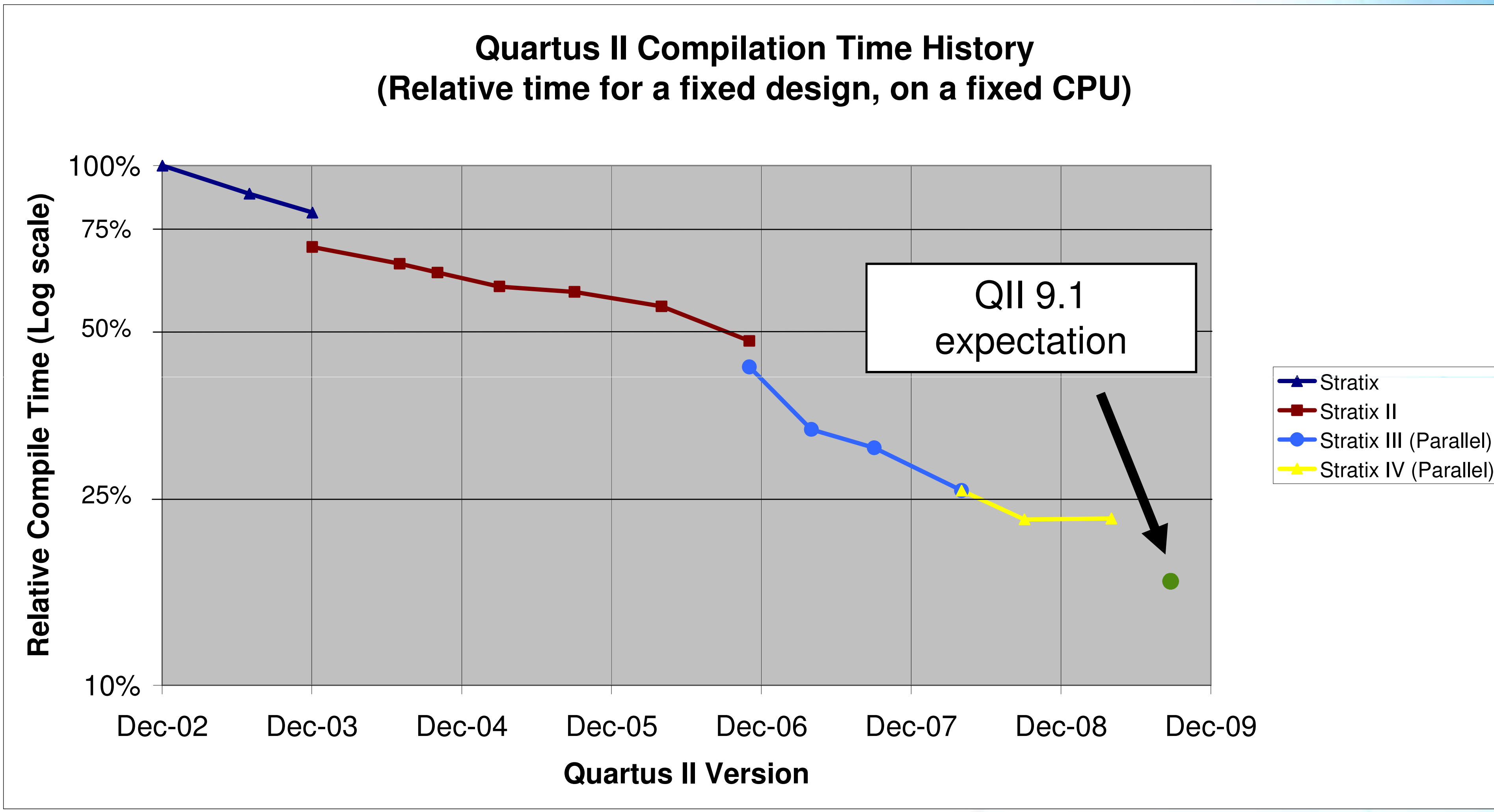
Designer Productivity

- FPGA density doubles
  - → CAD problem size doubles
  - → Designers need to create 2X the logic in the same time
- But CPU speed increase << 2X
  - Faster algorithms
  - Parallel CAD
  - Incremental compile
- Designer typing, debug speed increase << 2X
  - High quality CAD → reduce designer intervention
  - Higher levels of design abstraction

The Compile Time Challenge

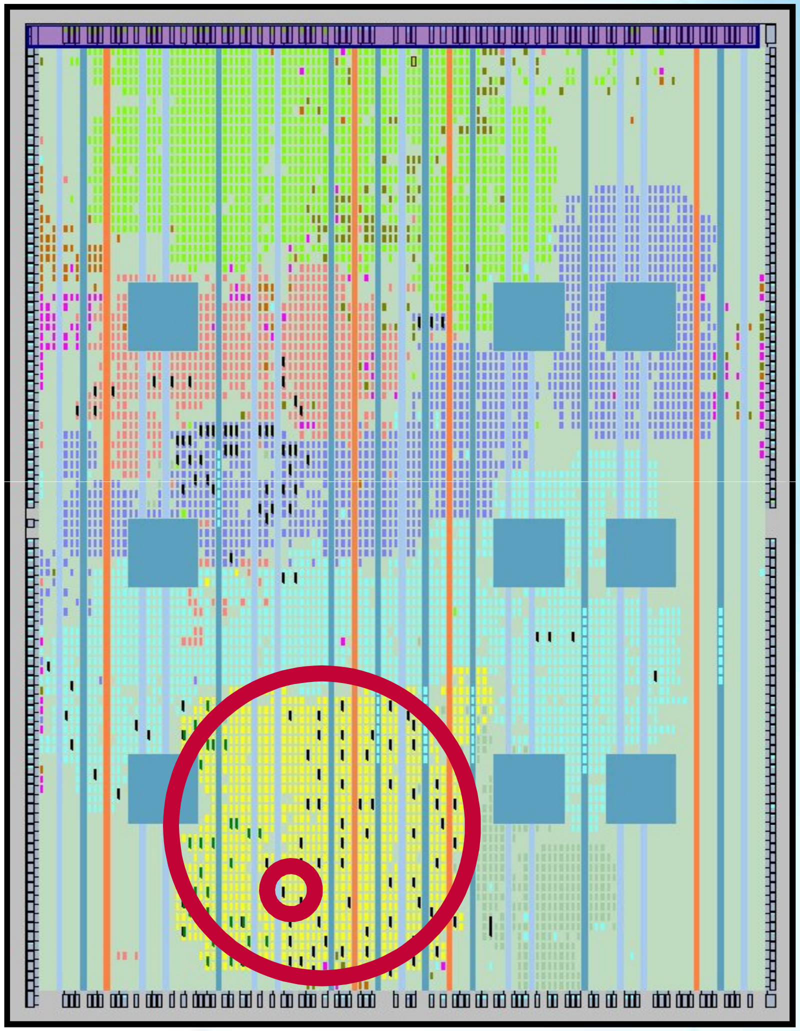


Improvements More than Bridge the Gap

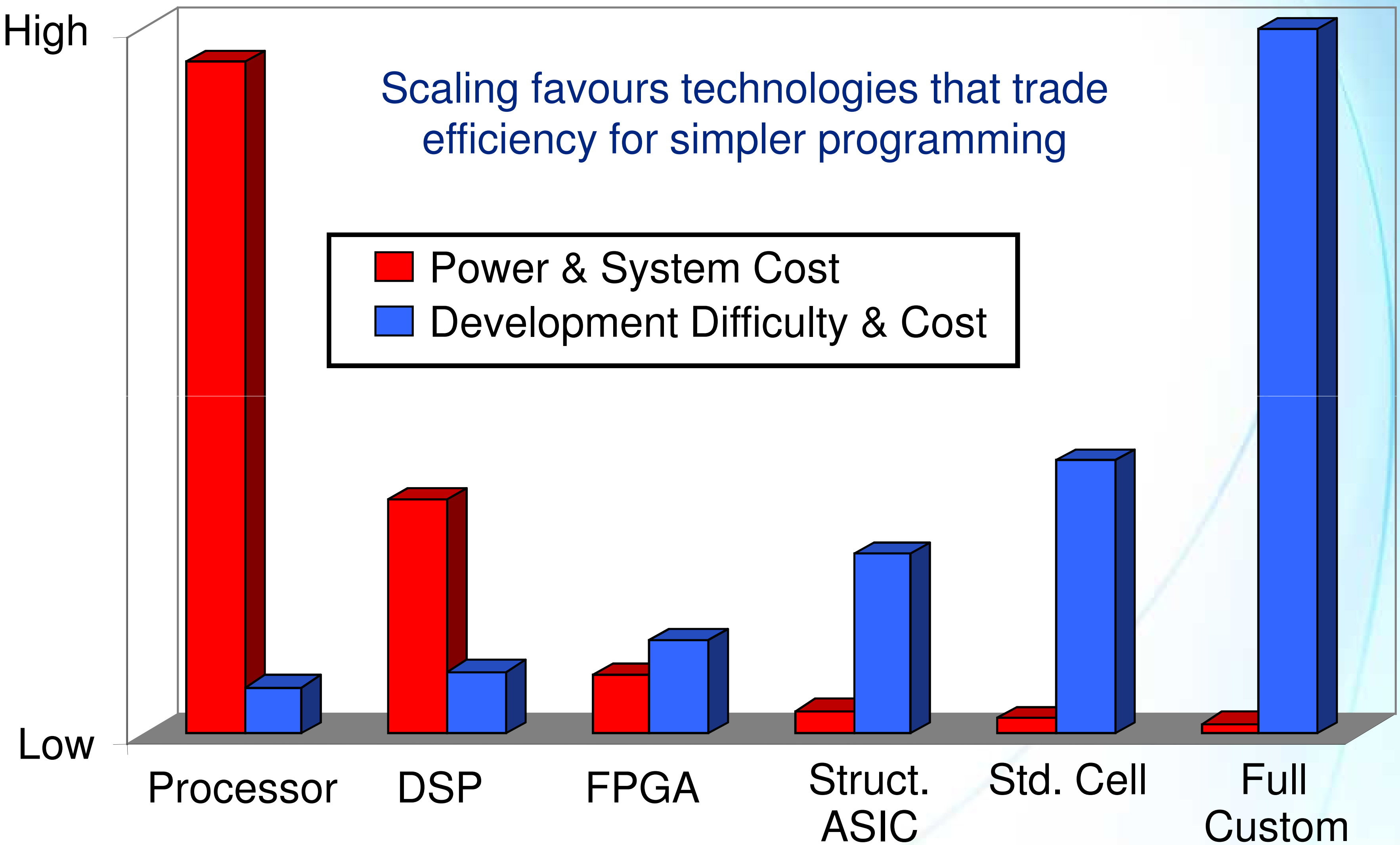


Incremental Compilation

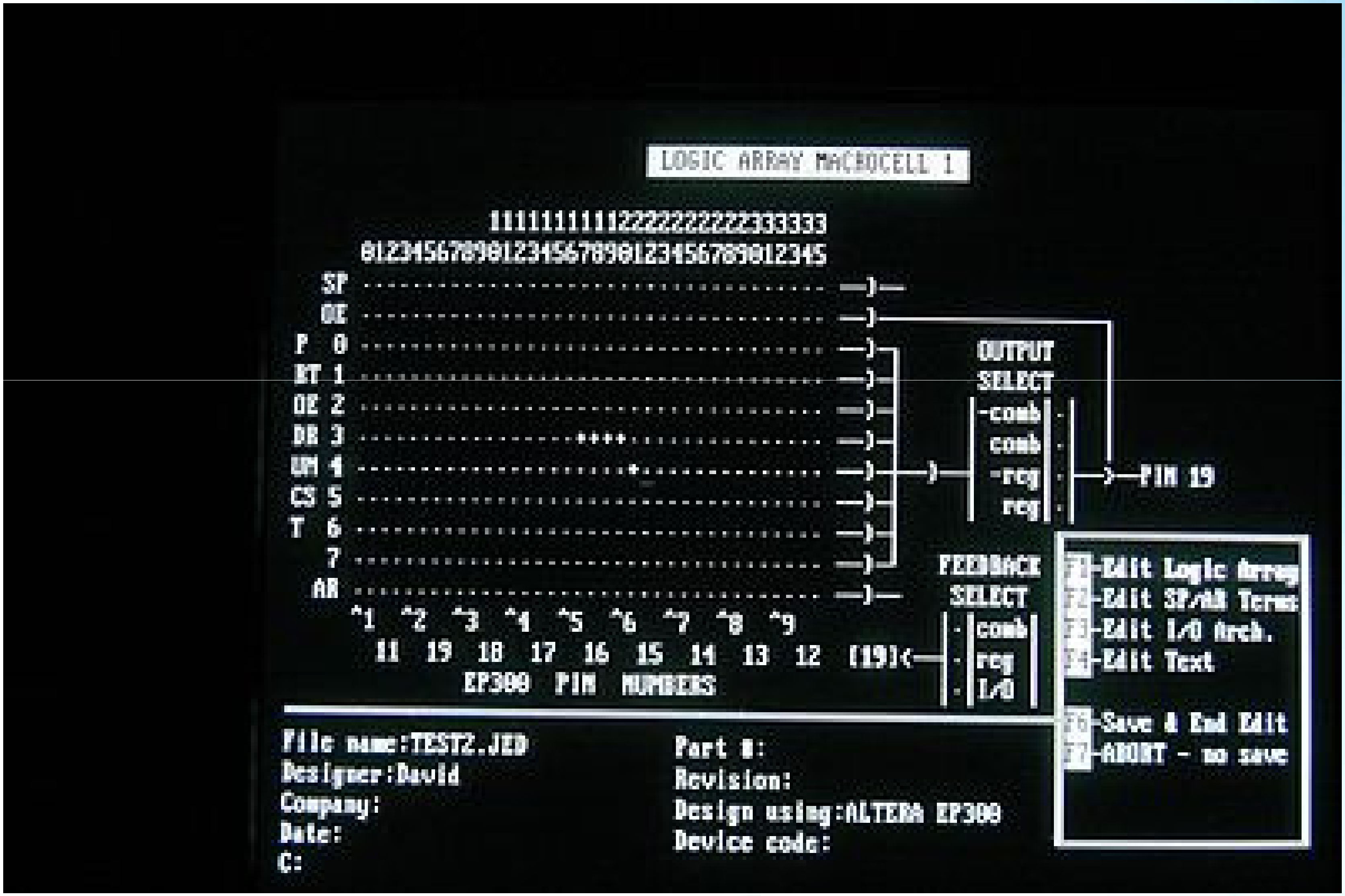
- Define partitions
  - CAD will not optimize across partitions
  - Can re-synthesize, place and route one partition alone
  - Faster compile time
  - Fewer iterations because other logic unchanged
- Work in progress
  - Incremental compile without the designer identifying partitions?
  - Challenge: global optimizations



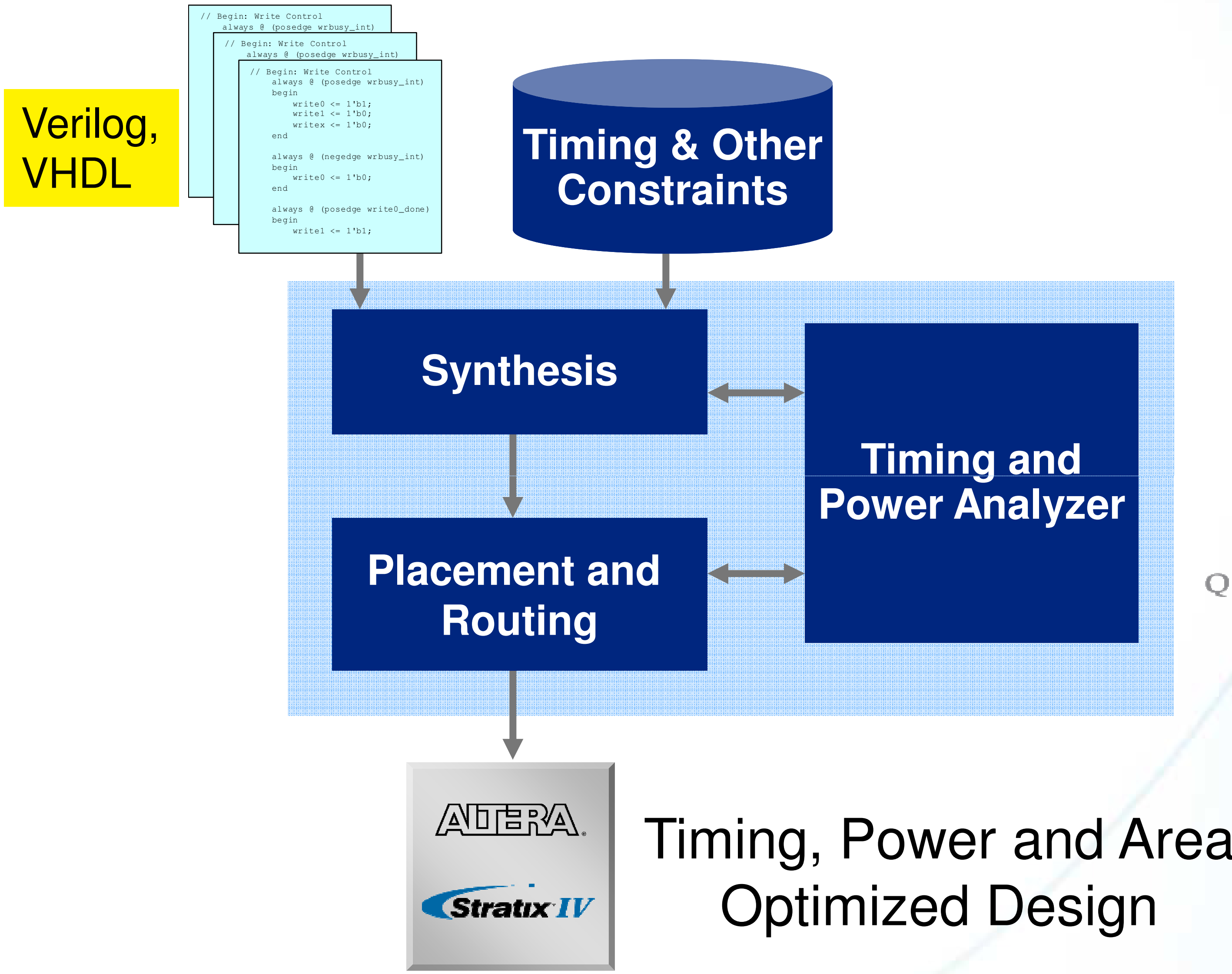
Efficiency & Programming Ease



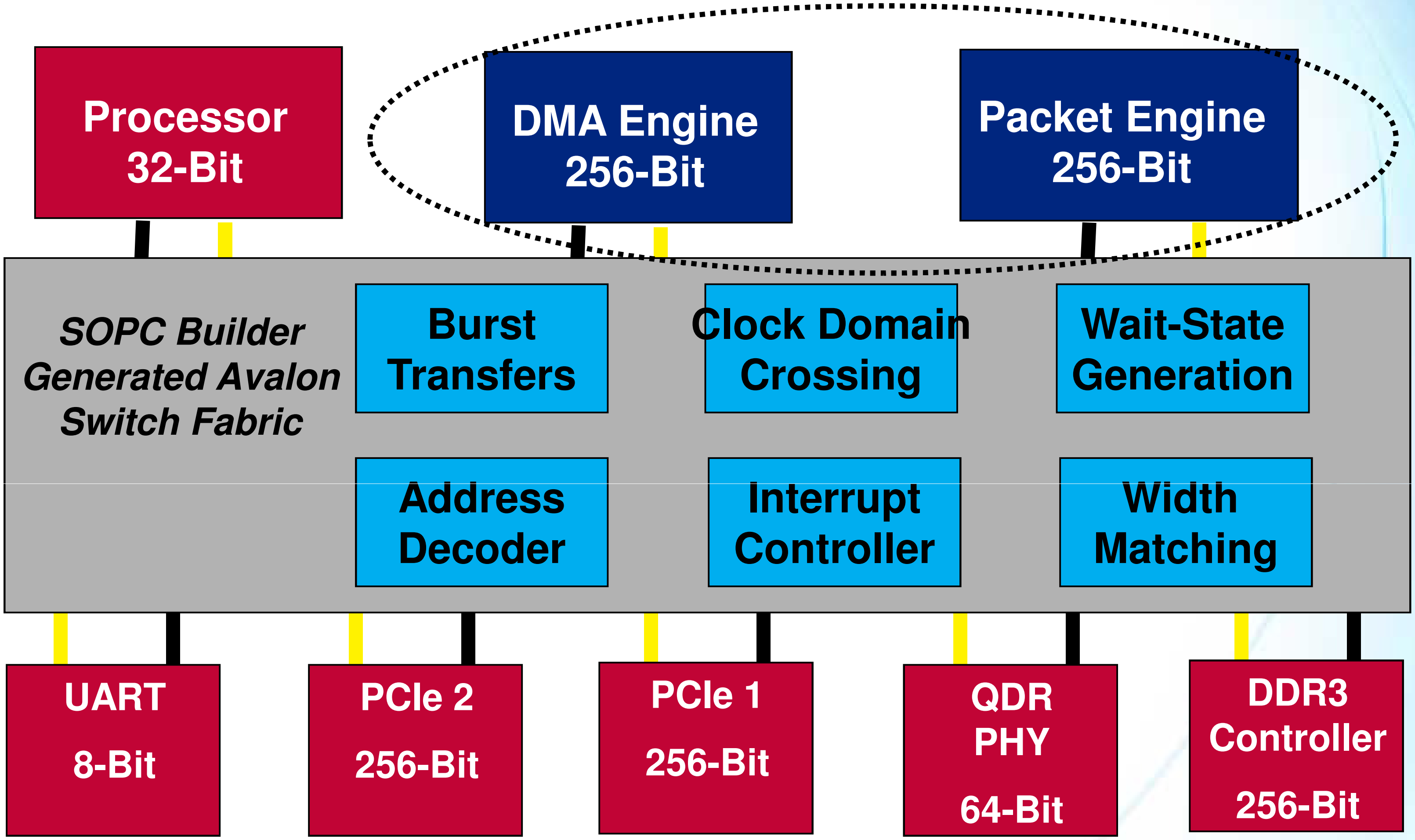
The Past (1984): Editing Switches



The Present: HDL Design Flow

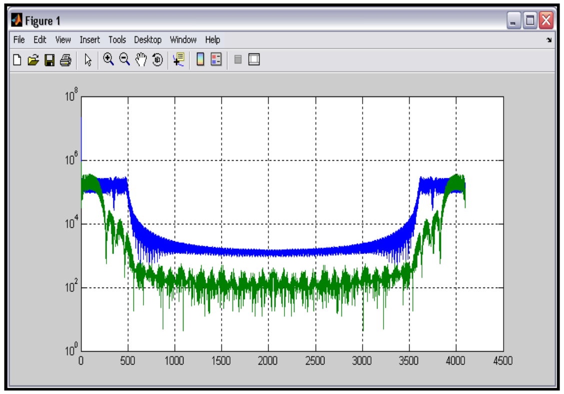


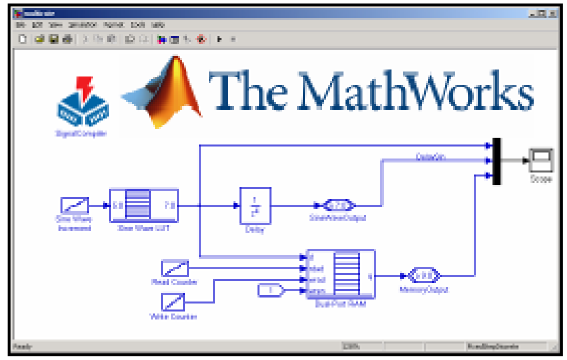
The Future: SOPC Builder Switch Fabric

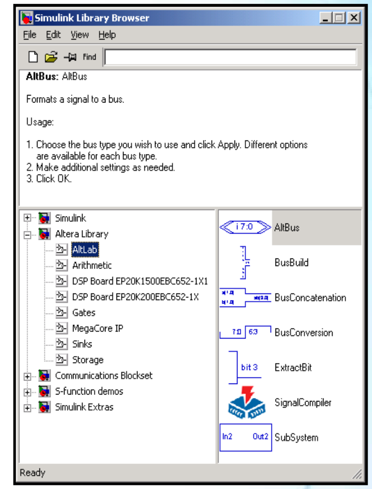


- Focus on your unique functions
- Re-use IP and let SOPC Builder integrate the system

The Future: DSP Builder Design Flow

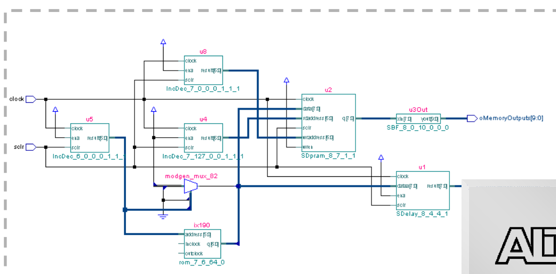


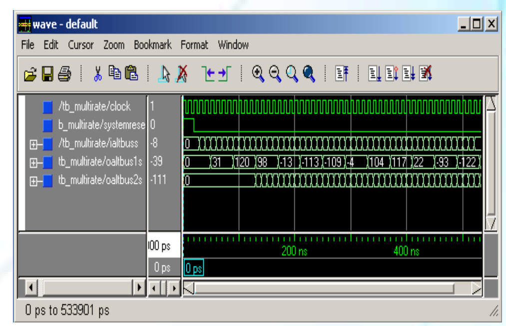




Matlab/Simulink domain
(System simulation and verification)

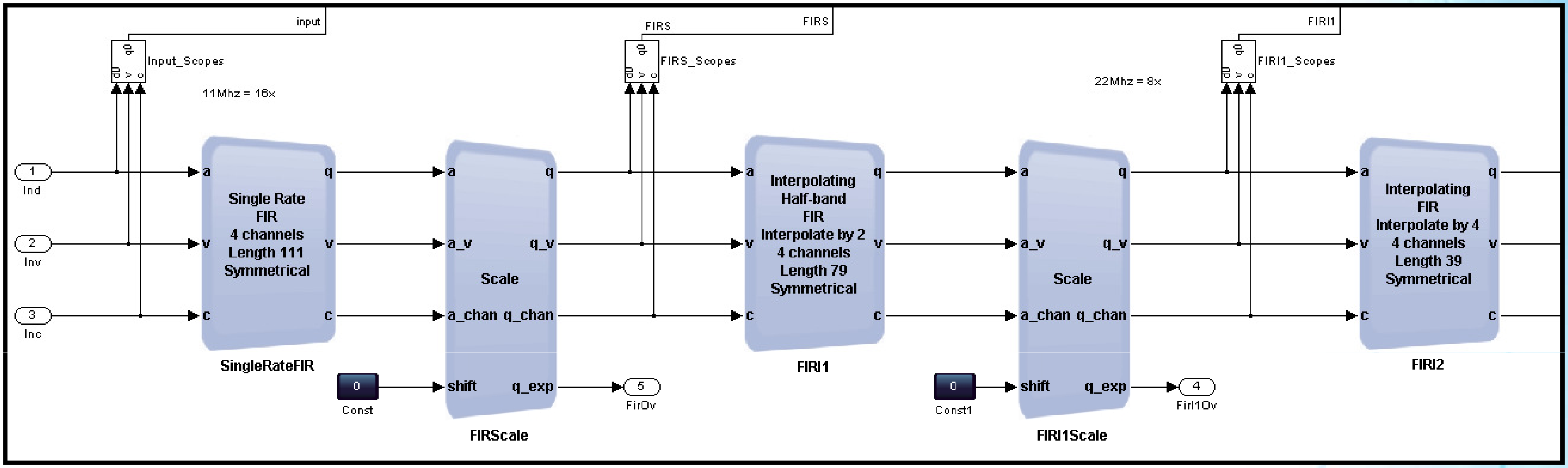
HDL/hardware domain
(Hardware implementation/RTL simulation)





DSP Builder Advanced Blockset

2 Antenna WiMax Digital Up Converter



- Fixed frequency design for high throughput
  - > 400 MHz for huge designs
- *Automatic* datapath widen/narrow to match data rate
- *Automatic* register insertion & pipeline balancing
- *Automatic* time-domain multiplexing of hardware

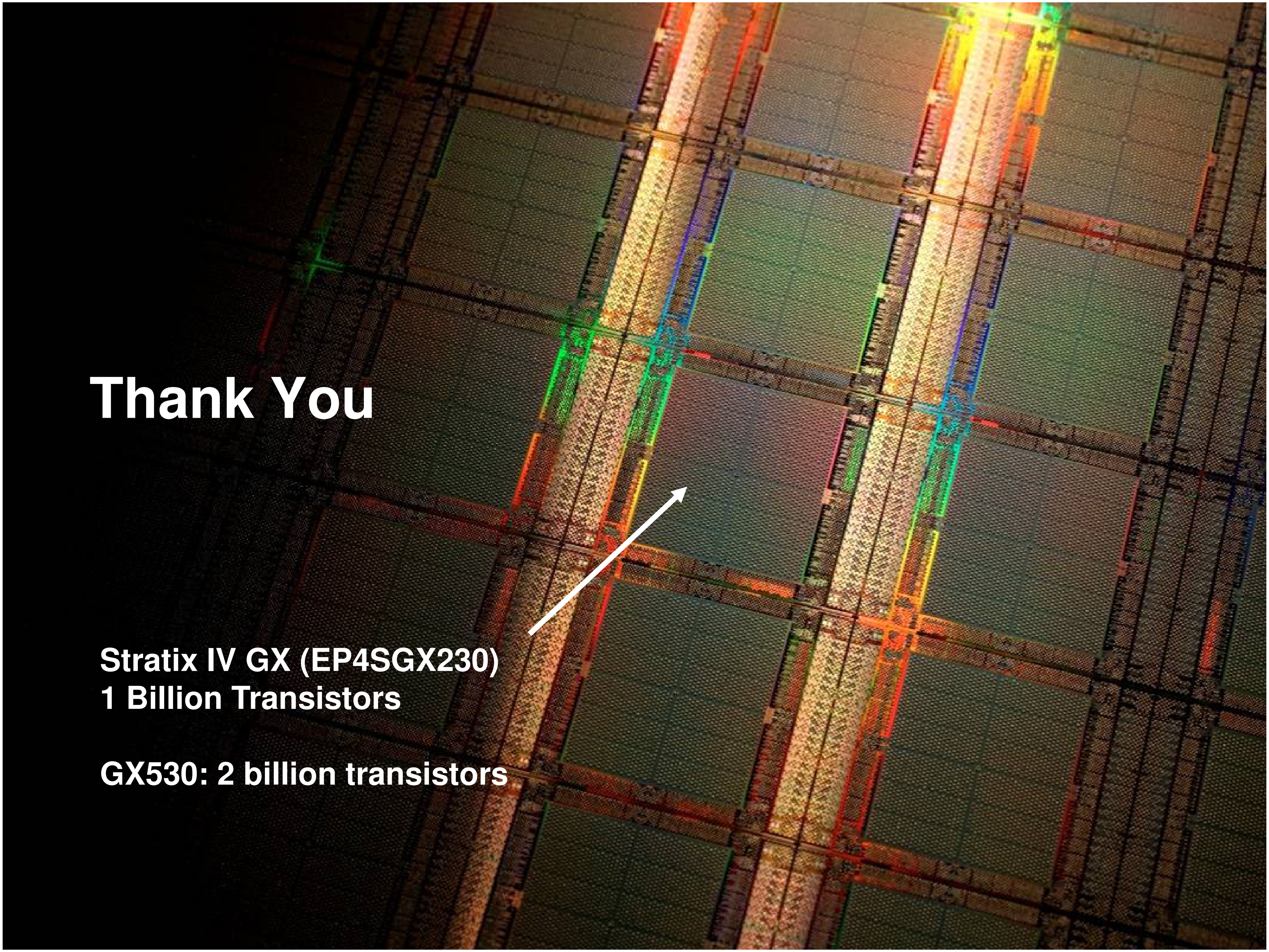
The Future: High-Level Languages to HW

- Incremental
  - SystemVerilog
- Bigger gains
  - Catapult C, C2H, ImpulseC, AutoPilot
  - RapidMind for FPGAs?
  - Others?
- Co-develop target applications and tool
- Plus novel debug tools

Summary

Summary

- Scaling favours the programmable
  - FPGAs
  - Processors
  - Can ASICs embed enough programmability?
- Challenges
  1. Architecture: fabric and configurable hard blocks
  2. I/O bandwidth
  3. Device modeling
  4. Power & power integrity
  5. Keeping designers productive: compile time, new design & debug tools


Thank You
Stratix IV GX (EP4SGX230)
1 Billion Transistors
GX530: 2 billion transistors