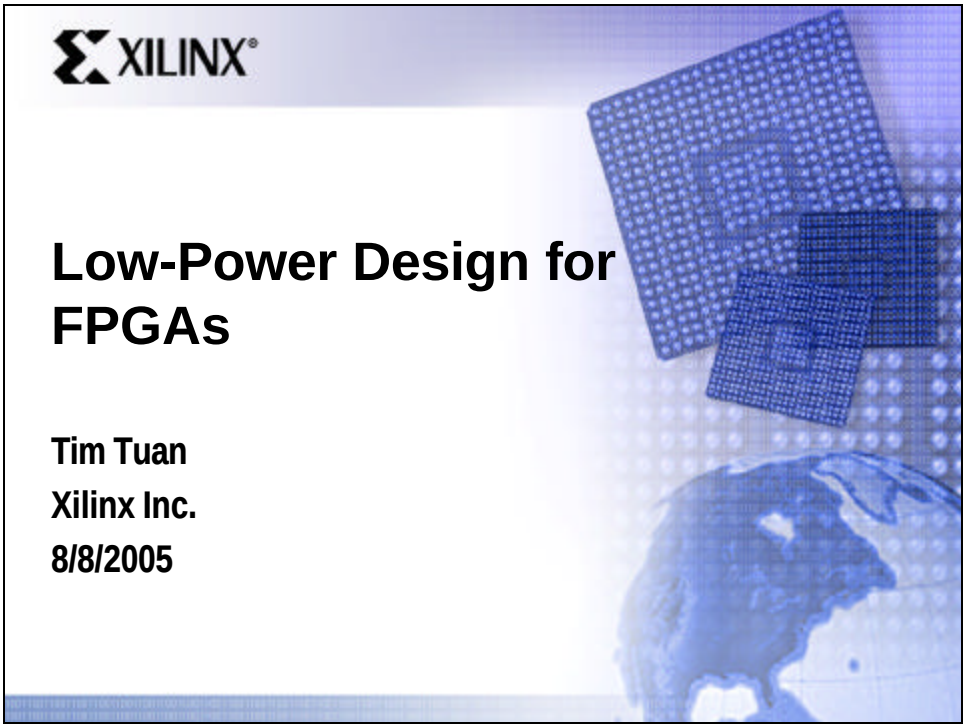
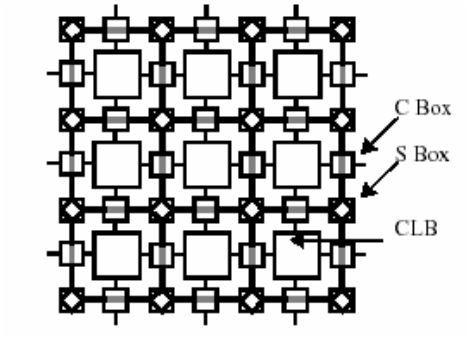


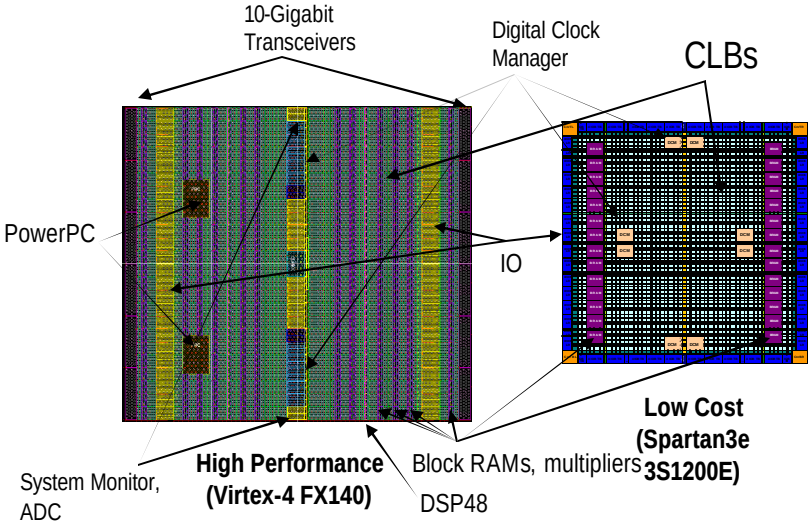
Outline

- **FPGA Power**
  - Power Trends
  - Power Models
  - Power Analysis
- **Architecture Innovations**
  - Dedicated Cores
  - Low Leakage Programmable Interconnect
  - Power Gating
  - Heterogeneous Architecture

Traditional FPGA

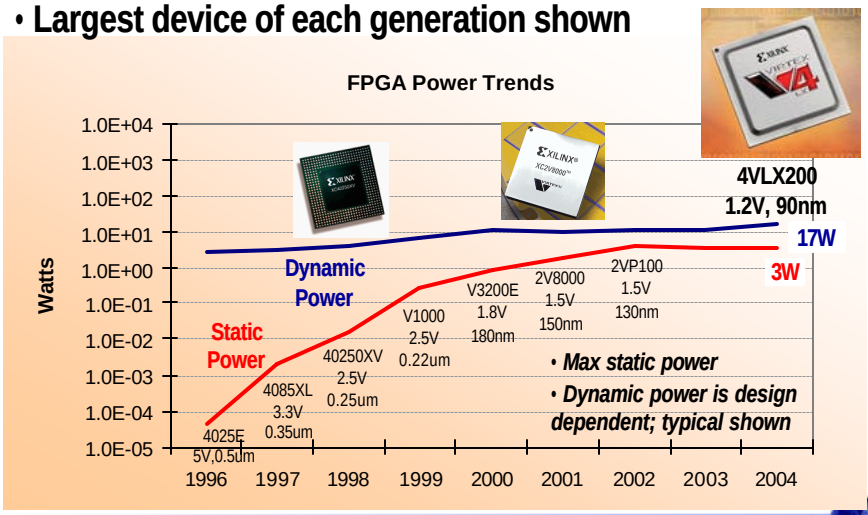


Modern FPGAs Architectures



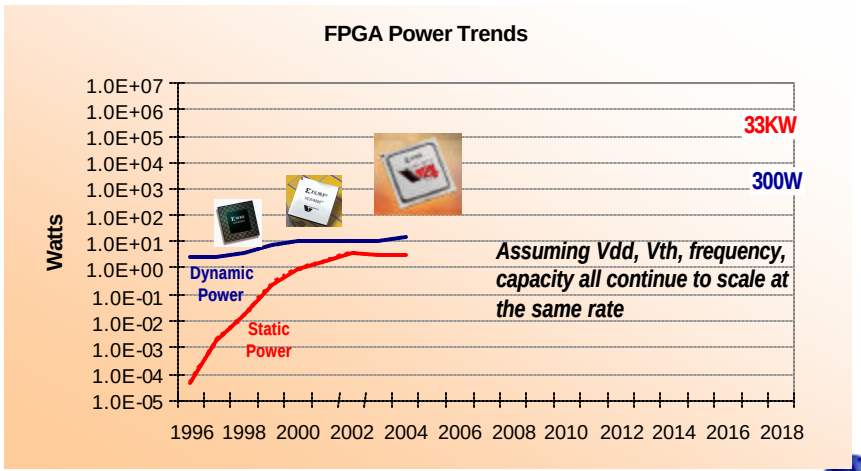
FPGA Power Trends

- Largest device of each generation shown



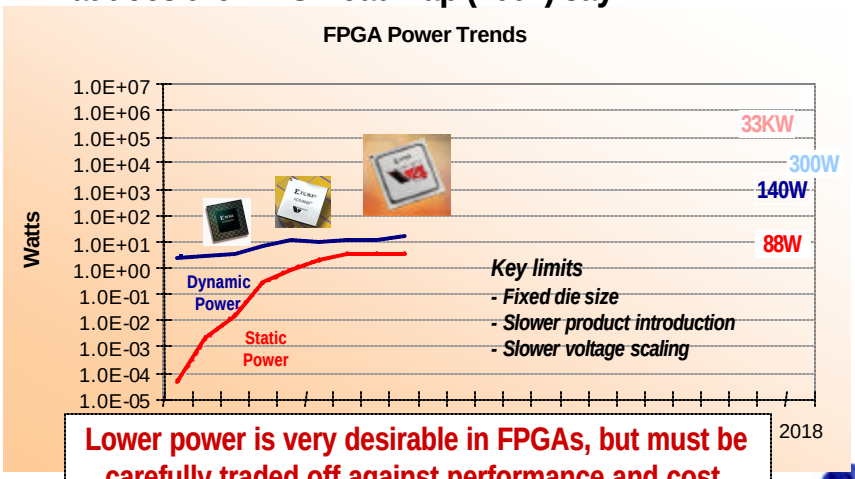
FPGA Power Trends

- If past trends continue …



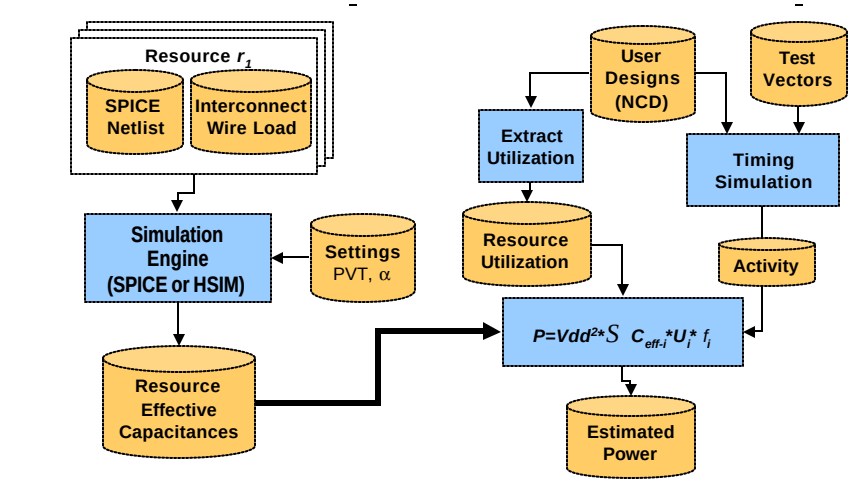
FPGA Power Trends

- What does the ITRS Roadmap (2004) say?

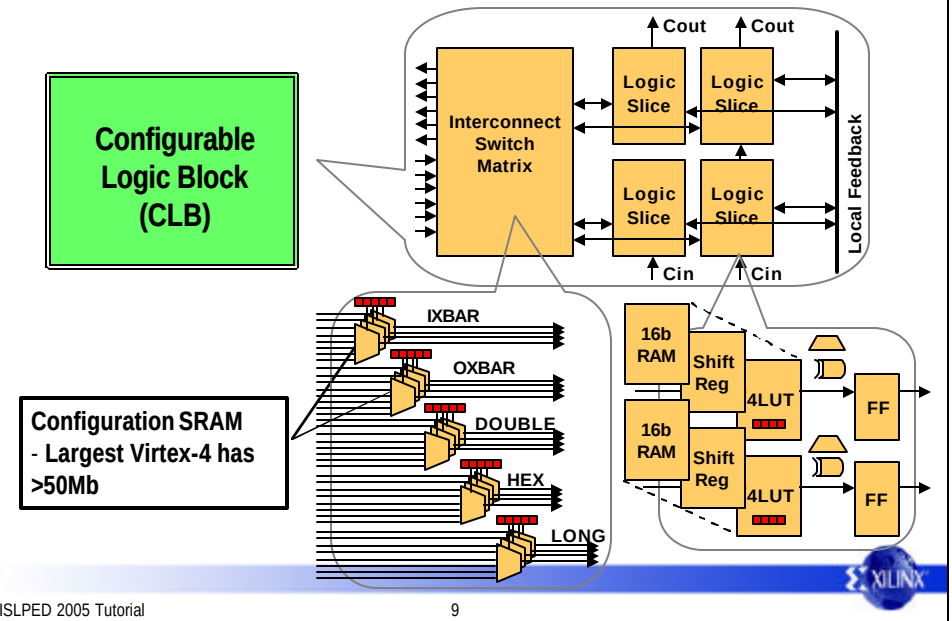


Lower power is very desirable in FPGAs, but must be carefully traded off against performance and cost.

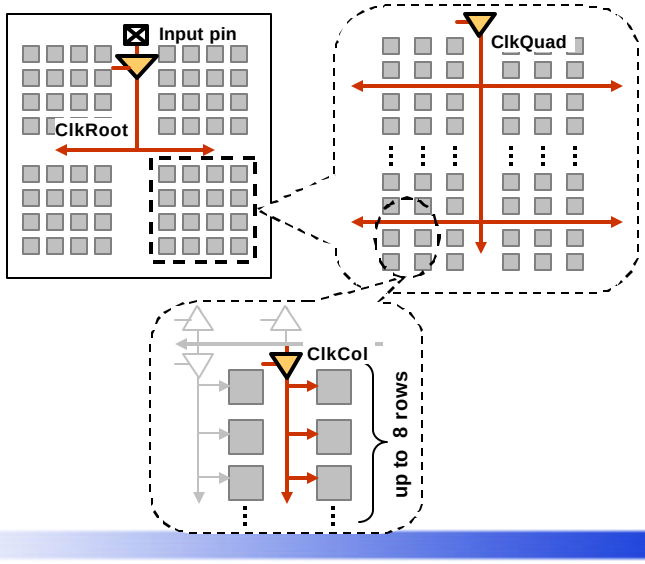
Power Analysis Flow



CLB Architecture Model



Clock Tree Model

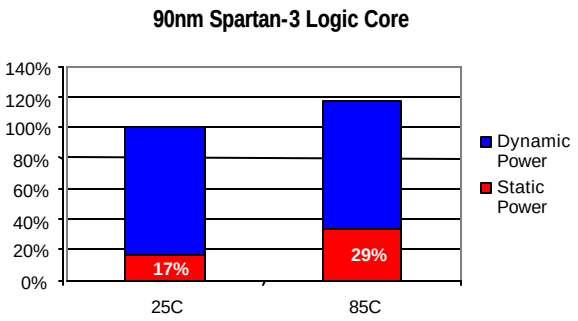


Power Breakdown

- **Based on typical resource utilization, 12.5% switching activity, f_{CLK}=100MHz**

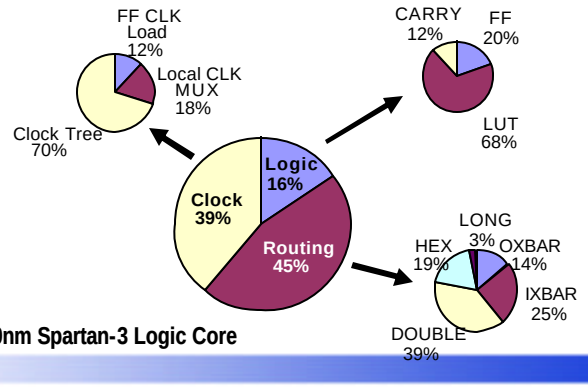
Block	% of Available
Logic Slice	67%
LUT	47%
FF	26%
Carry	4%
IXBAR	40%
Direct	27%
Double	19%
Hex	5%
Long	1%
ClkRoot	16%
ClkQuad	16%
ClkCol	4%
IXBAR_Clk	36%

Typical resource utilization



Dynamic Power Breakdown

- **Based on typical resource utilization, 12.5% switching activity, f_{CLK}=100MHz**

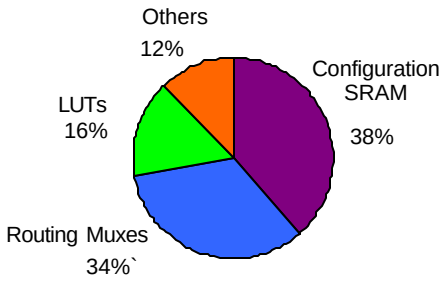


90nm Spartan-3 Logic Core

[Degalahal ASPDAC 2005]

Leakage Power Breakdown

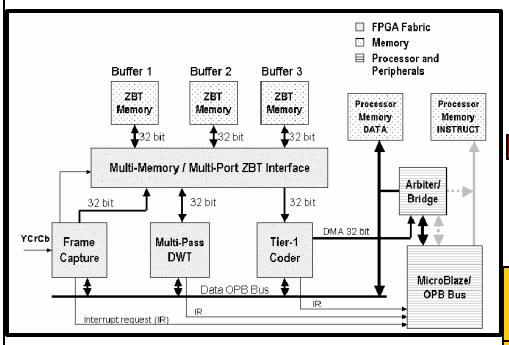
- **A few basic circuits consume the majority of total leakage power**

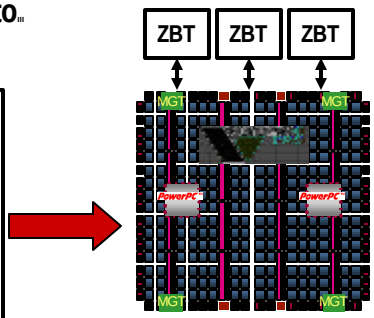


[Tuan CICC 2003]

Are FPGA Power-Inefficient?

- **Depends on what you compare to…**
- **Case study: JPEG2000 encoder**



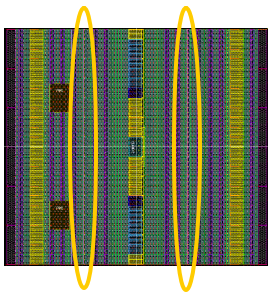


[Schumacher, SPIE 2003]

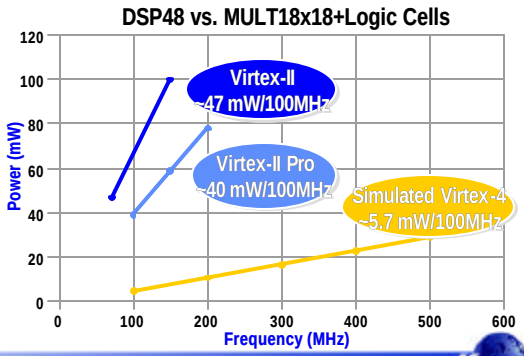
	Power Dissipation (in Watts)			Relative MOPS/mW
	Device	Memory	Total	
FPGA	3.5	3.5	7.0	8.0
DSP	1.7	2.4	4.1	1.0
ASIC	2.0	0.2	2.2	28.6

Dedicated Blocks

- **Customs logic is more energy-efficient than programmable logic**
- **Introduce custom logic cores for common functions**
  - Virtex-4 optimizes for DSP applications with DSP48 custom cores

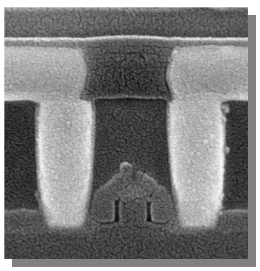


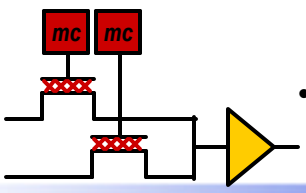
DSP48 blocks



Low Leakage Interconnect

Triple-Oxide

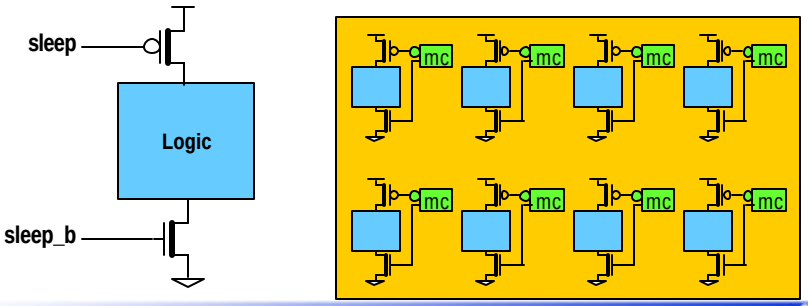




- **Use low-speed, low-power configuration SRAM**
  - Not timing critical
  - Also used in interconnect passgates with higher gate boosting
- **High-Vt is not enough in 90nm due to substantial gate leakage**
  - Virtex-4 uses a mid-oxide device
  - 40% leakage reduction, no leakage increase from 130nm (Virtex2-Pro)
- **Reduces much of the overhead of reconfigurability with no performance penalty**

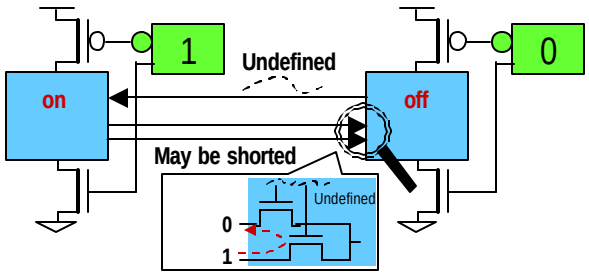
Power Gating

- **In MPUs, use a high-Vt device to cut off power to inactive blocks**
  - 20-2000X leakage reduction during sleep for <10% delay and area penalty
- **In FPGAs, applicable to inactive and *unused* blocks**
  - Each PG block may be one or more CLBs, or sub-CLB resources (e.g. LUT)



Design Issues

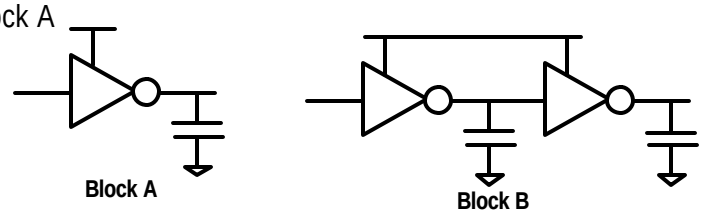
- **Inputs, outputs often require interfacing logic**
  - Relatively less overhead for larger blocks



Coarser-grain power gating requires less overhead

Design Issues

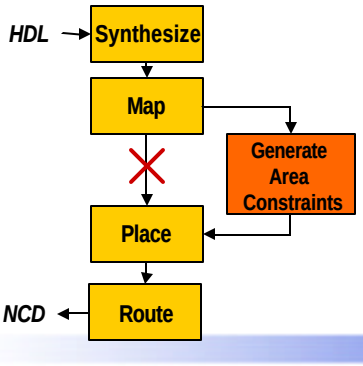
- **Power gates must be sized for peak current**
  - Larger blocks have less power gate overhead, as they typically have lower peak current relative to block size
  - E.g.: Block B has roughly the same power gate requirement as Block A

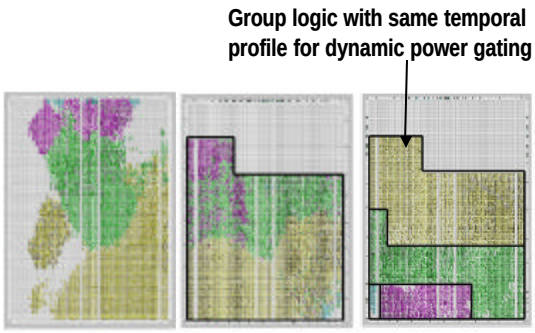


Coarser-grain power gating requires less overhead

CAD for Power Gating

- **Power saving may be limited by coarse-grain power gating**
- **Region-constrained placement (RCP) circumvents that limitation**
  - Use placement constraints to group "used logic"

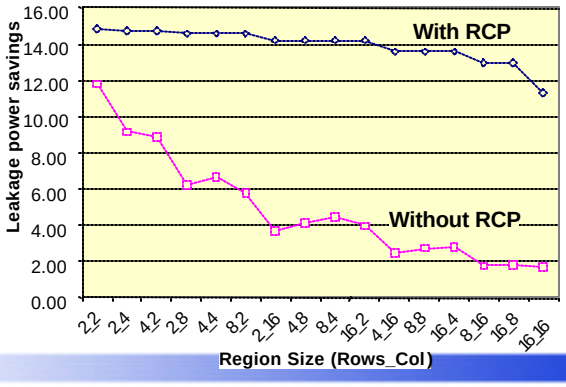




[Gayasen, FPGA '04]

CAD for Power Gating

- **RCP enables high power savings in coarse grain power gating architectures**

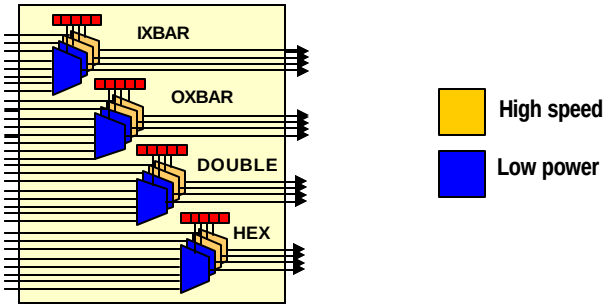


Heterogeneous Architectures

- **Heterogeneity addresses the over-design inherent in homogeneous FPGAs**
  - ASICs use slack timing: Use fast cells for critical paths
  - In FPGAs, critical paths are unknown when designing silicon, hence everything is designed for high performance
- **Heterogeneous FPGA**
  - Mix fast and slow (but low power) resources. Use CAD tools to map timing critical logic to fast resources

Heterogeneous Interconnect

- Mixed high-speed and low-power interconnect resources
- Use router to perform slack timing



- Also can be applied to logic blocks

[Rahman CICC 2005]

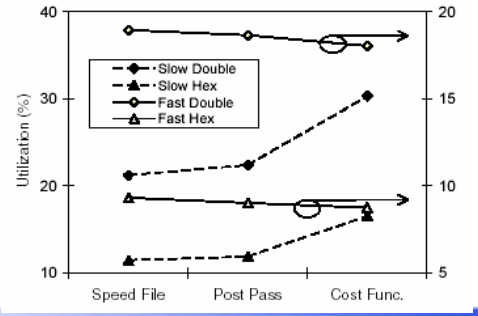
CAD for Heterogeneous Arch

- Resize some interconnect muxes to consume less power
- Use standard timing-driven place and route tools
  - Critical paths should naturally be mapped to fast resources
  - 40% leakage reduction in the interconnect

Architecture: Low power resources	Power	Performance
all imux, longs, 25% double, hex, 50% omux	-30.1%	-4.0%
all long, 75% imux, 25% double, hex, 50% omux	-26.0%	-3.5%
all imux, long, 75% double, hex, 50% omux	-41.5%	-6.8%
all imux, long, hex, 50% double, omux	-40.8%	-11.6%

CAD for Heterogeneous Arch

- **Potential CAD improvements**
  - Post-routing optimization: After PAR, re-assign non-critical nets to low power resources, if possible
  - Modify cost functions to include timing **and** power



Summary

- **FPGAs trend towards higher power, but power reduction must be carefully weighed against performance and cost**
- **In 90nm Spartan-3 FPGA, dynamic power is still dominant (70%-84% of total core power)**
- **Dedicated blocks improve power efficiency at the cost of flexibility**
- **Triple-oxide process offers an effective, one-time solutions**
- **Power gating is effective for reducing static and standby power, especially with CAD support**
- **Heterogeneous architectures allow slack timing for power reduction with proper CAD support**