

An 8-core, 64-thread, 64-bit, power efficient SPARC SoC (Niagara2)

Umesh Nawathe, Jim Ballard, Mahmudul
Hassan, Tim Johnson, Rob Mains, Paresh
Patel, Alan Smith

Sun Microsystems Inc., Sunnyvale, CA

Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

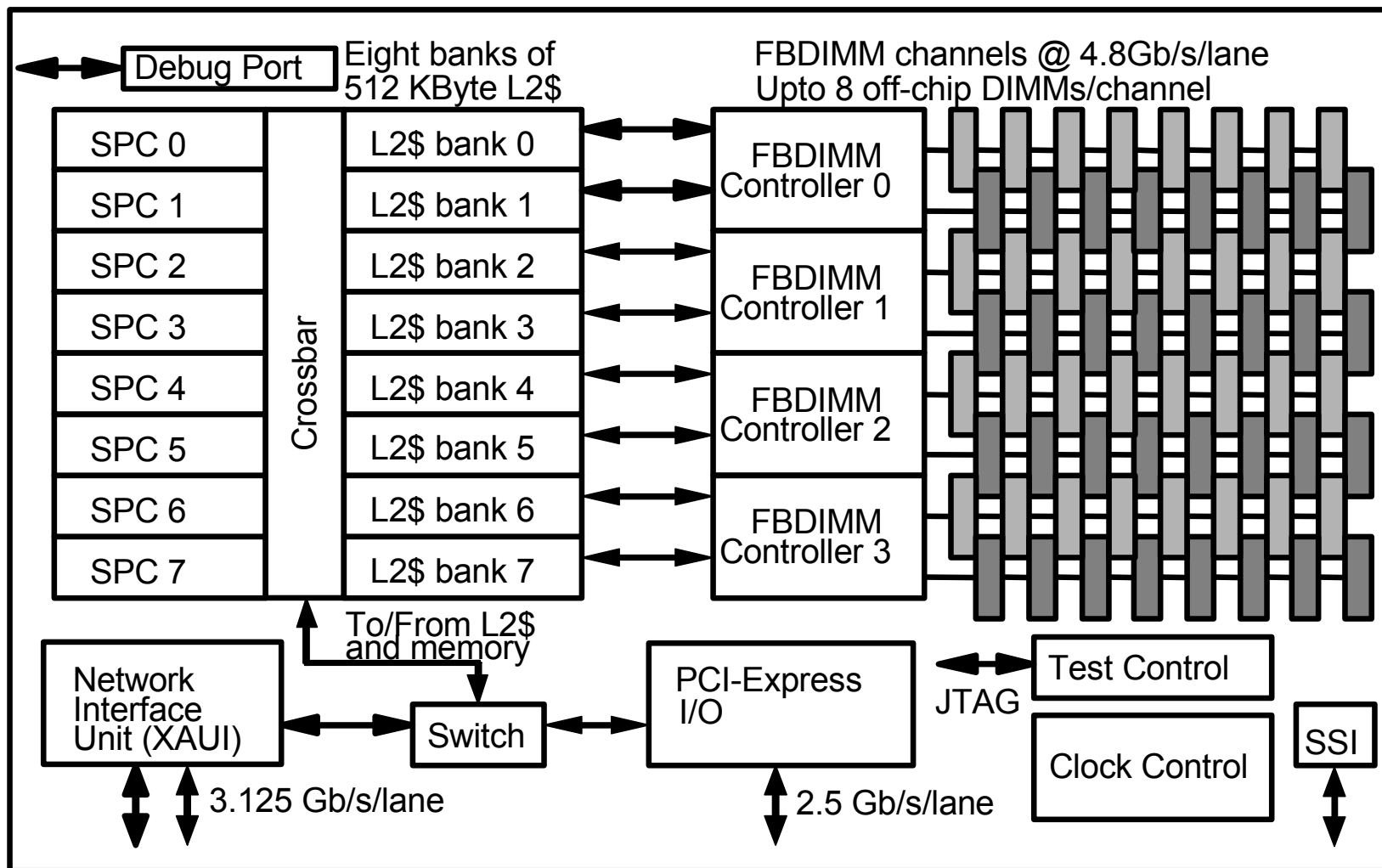
Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Niagara2's Key features

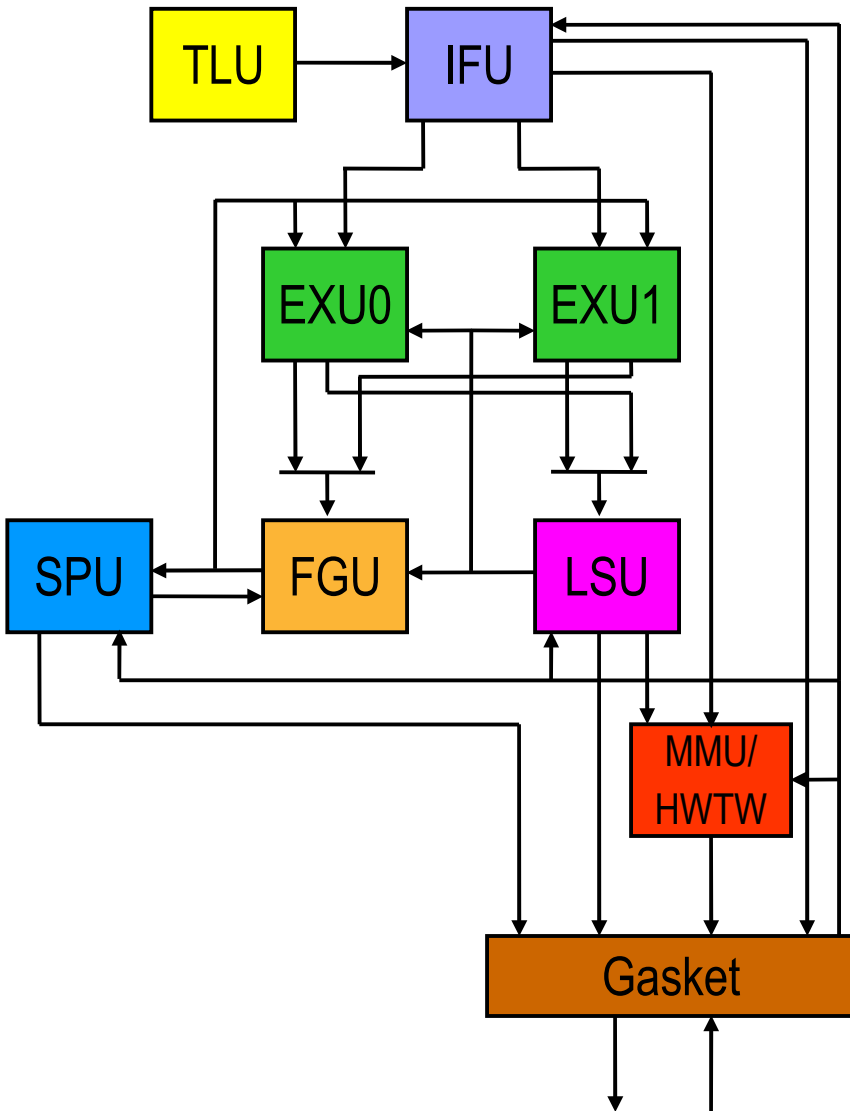
- 2nd generation CMT (Chip Multi-Threading) processor optimized for Space, Power, and Performance (SWaP).
- 8 Sparc Cores, 4MB shared L2 cache; Supports concurrent execution of 64 threads.
- >2x UltraSparc T1's throughput performance and performance/Watt.
- >10x improvement in Floating Point throughput performance.
- Integrates important SOC components on chip:
 - > Two 10G Ethernet (XAUI) ports on chip.
 - > Advanced Cryptographic support at wire speed.
- On-chip PCI-Express, Ethernet, and FBDIMM memory interfaces are SerDes based; pin BW > 1Tb/s.

Niagara2 Block Diagram



Key Point: System-on-a-Chip, CMT architecture => lower # of system components, reduced complexity/power => higher system reliability.

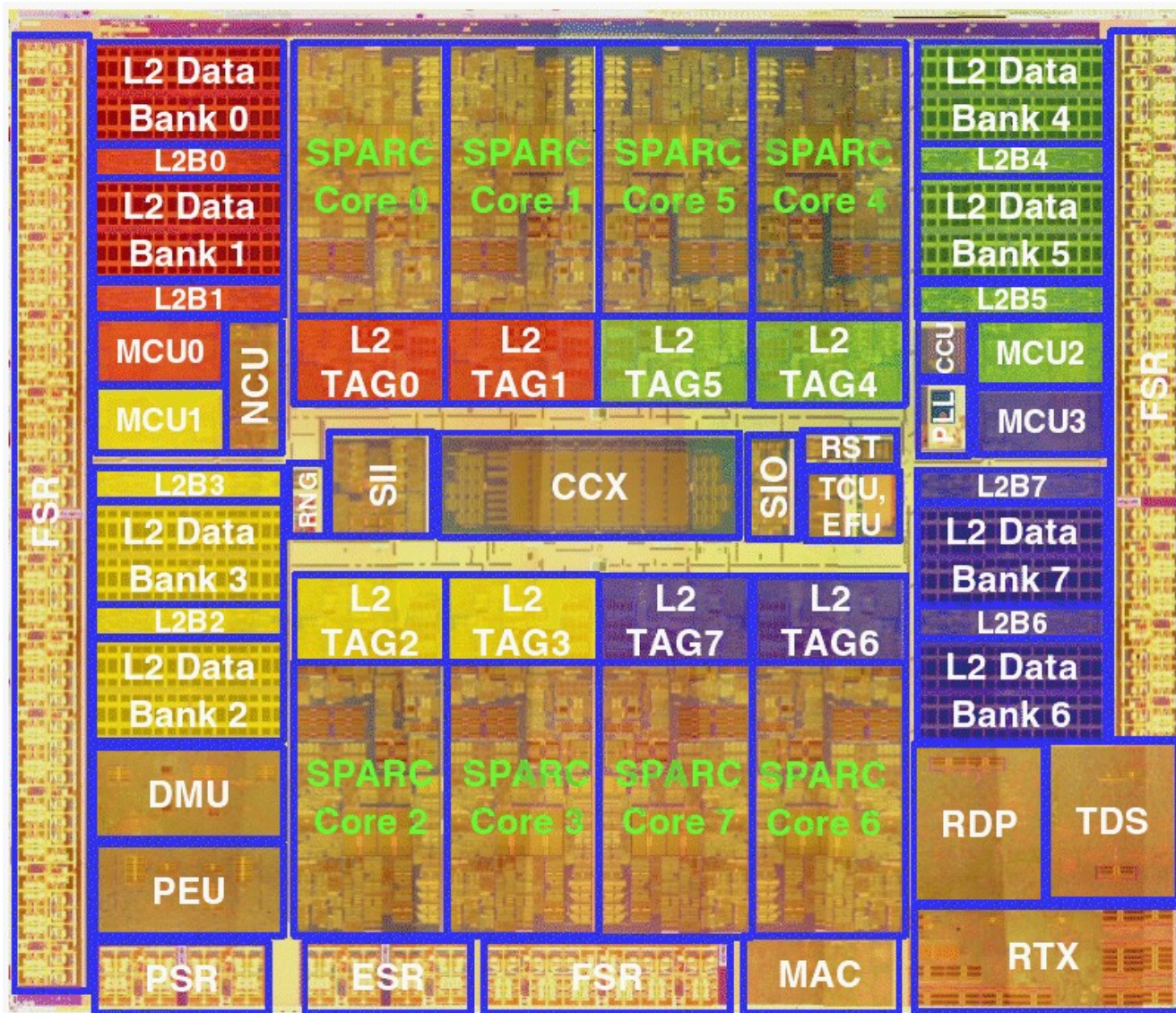
Sparc Core (SPC) Architecture Features



SPC Block Diagram

- Implementation of the 64-bit SPARC V9 instruction set.
- Each SPC has:
 - > Supports concurrent execution of 8 threads.
 - > 1 load/store, 2 Integer execution units.
 - > 1 Floating point and Graphics unit.
 - > 8-way, 16 KB I\$; 32 Byte line size.
 - > 4-way, 8 KB D\$; 16 Byte line size.
 - > 64-entry fully associative ITLB.
 - > 128-entry fully associative DTLB.
 - > MMU supports 8K, 64K, 4M, 256M page sizes; Hardware Tablewalk.
 - > Advanced Cryptographic unit.
- Combined BW of 8 Cryptographic Units is sufficient for running the 10 Gb ethernet ports encrypted.

Niagara2 Die Micrograph



- 8 SPARC cores, 8 threads/core.
- 4 MB L2, 8 banks, 16-way set associative.
- 16 KB I\$ per Core.
- 8 KB D\$ per Core.
- FP, Graphics, Crypto, units per Core.
- 4 dual-channel FBDIMM memory controllers @ 4.8 Gb/s.
- X8 PCI-Express @ 2.5 Gb/s.
- Two 10G Ethernet ports @ 3.125 Gb/s.

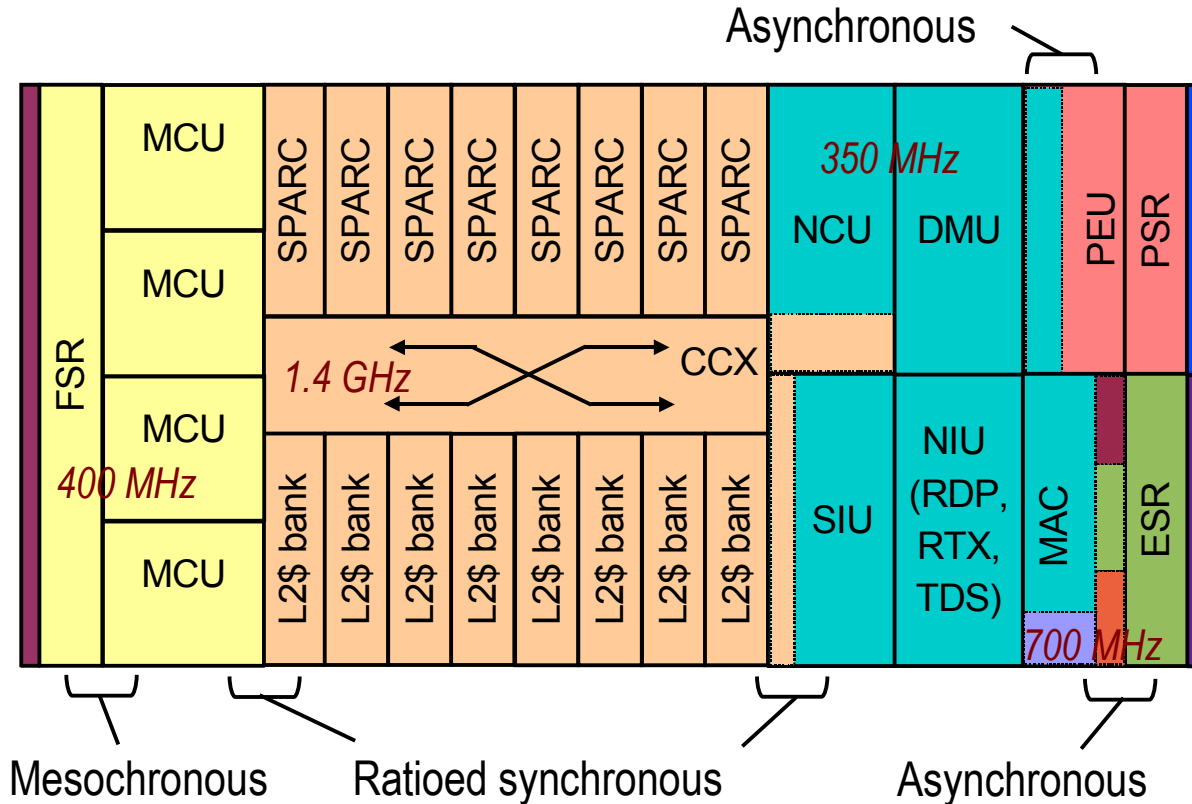
Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Physical Implementation Highlights

Technology	65 nm CMOS (from Texas Instruments)
Nominal Voltages	1.1 V (Core), 1.5V (Analog)
# of Metal Layers	11
Transistor types	3 (SVT, HVT, LVT)
Frequency	1.4 Ghz @ 1.1V
Power	84 W @ 1.1V
Die Size	342 mm ²
Transistor Count	503 Million
Package	Flip-Chip Glass Ceramic
# of pins	1831 total; 711 Signal I/O

Clocking



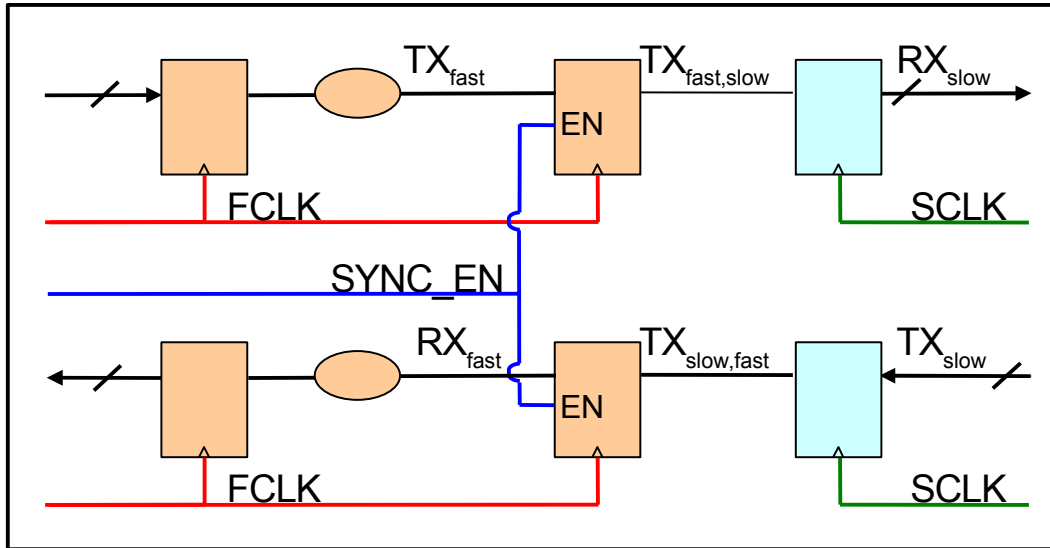
REF	133/167/200 MHz
CMP	1.4 GHz
IO	350 MHz
IO2X	700 MHz
FSR.refclk	133/167/200 MHz
FSR.bitclk	1.6/2.0/2.4 GHz
FSR.byteclk	267/333/400 MHz
DR	267/333/400 MHz
PSR.refclk	100/125/250 MHz
PSR.bitclk	1.25 GHz
PSR.byteclk	250 MHz
PCI-Ex	250 MHz
ESR.refclk	156 MHz
ESR.bitclk	1.56 GHz
ESR.byteclk	312.5 MHz
MAC.1	312.5 MHz
MAC.2	156 MHz
MAC.3	125/25/2.5 MHz

Key Point: Complex clocking; large # of clock domains; asynchronous domain crossings.

Clocking (Cont'd.)

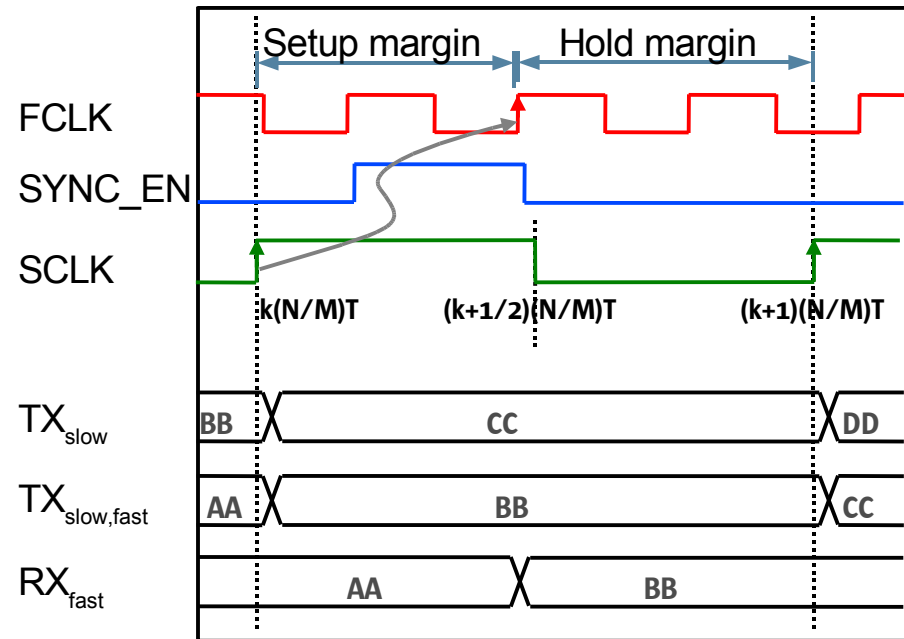
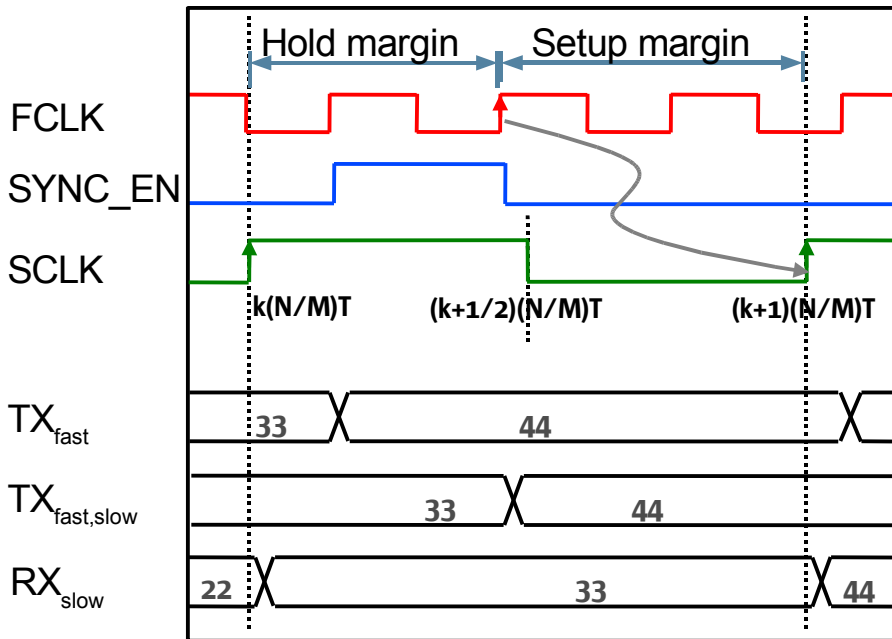
- On-chip PLL generates Ratioed Synchronous Clocks (RSCs); Supported fractional divide ratios: 2 to 5.25 in 0.25 increments.
- Balanced use of H-Trees and Grids for RSCs to reduce power and meet clock-skew budgets.
- Periodic relationship of RSCs exploited to perform high BW skew-tolerant domain crossings.
- Clock Tree Synthesis used for Asynchronous Clocks; domain crossings handled using FIFOs and meta-stability hardened flip-flops.
- Cluster/L1 Headers support clock gating to save clock power.

Clocking (RSC domain crossings)



- FCLK = Fast-Clock
SCLK = Slow-Clock
- Same 'Sync_en' signal for FCLK -> SCLK, and SCLK -> FCLK crossings.

Key Point: Equalizing setup and hold margins maximizes skew tolerance.

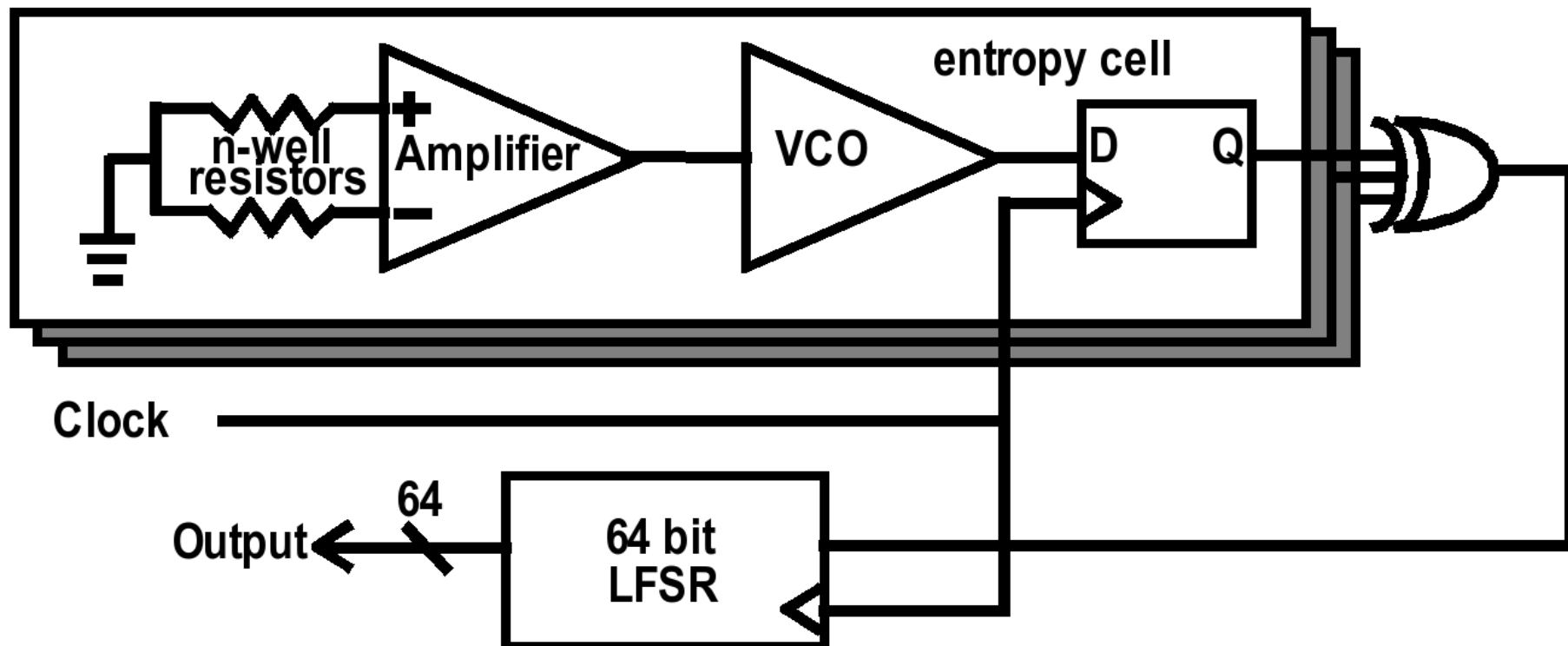


Niagara2's SerDes Interfaces

	FBDIMM	PCI-Express	Ethernet-XAUI
Signalling Reference	VSS	VDD	VDD
Link-rate (Gb/s)	4.8	2.5	3.125
# of North-bound (Rx) lanes	14 * 8	8	4 * 2
# of South-bound (Tx) lanes	10 * 8	8	4 * 2
Bandwidth (Gb/s)	921.6	40	50

- All SerDes share a common micro-architecture.
- Level-shifters enable extensive circuit reuse across the three SerDes designs.
- Total raw pin BW in excess of 1Tb/s.
- Choice of FBDIMM (vs DDR2) memory architecture provides ~2x the memory BW at <0.5x the pin count.

Niagara2's True Random Number Generator



- Consists of 3 entropy cells.
- Amplified n-well resistor thermal noise modulates VCO frequency; VCO o/p sampled by on-chip clock.
- LFSR accumulates entropy over a pre-set accumulation time.
 - > Privileged software programs a timer with desired entropy accumulation time.
 - > Timer blocks loads from LFSR before entropy accumulation time has elapsed.

Outline

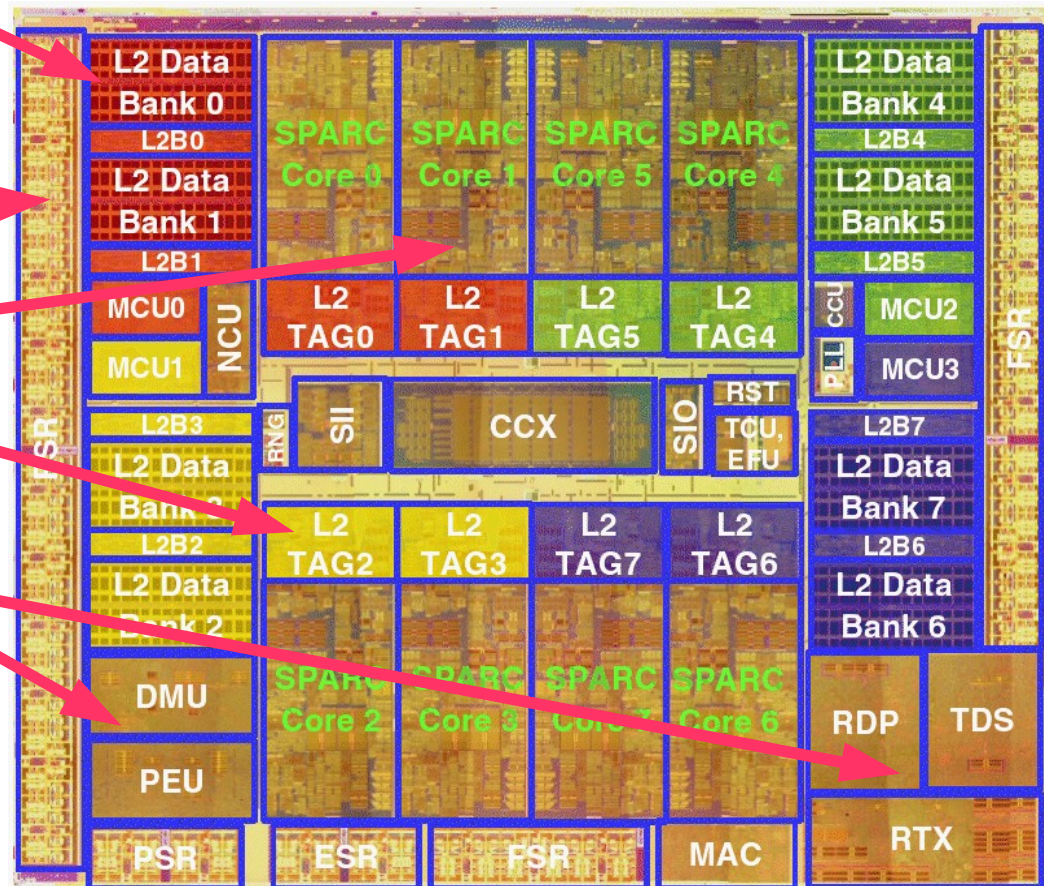
- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Niagara2's System on Chip Methodology

- Chip comprised of many subsystems with different design styles and methodologies:

Key Point: Chip Design Methodologies had to comprehend blocks with different design styles and levels of abstraction.

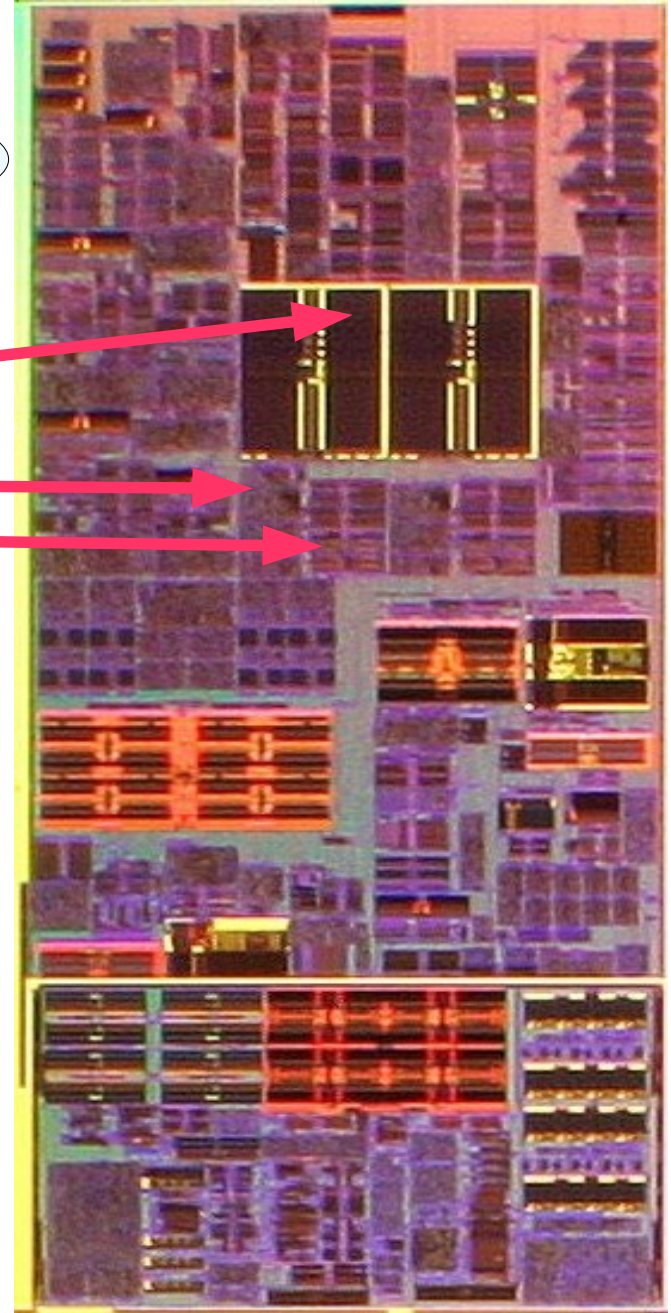
- > Custom Memories & Analog Macros:
 - > Full custom design and verification.
 - 40% compiled memories.
 - > Schematic/manual layout based.
- > External IP:
 - > SerDes full custom IP Macros.
- > Complex Clusters:
 - > DP/Control/Memory Macro.
 - > Higher speed designs.
- > ASIC designs:
 - > PCI-Express and NIC functions.
- > CPU:
 - > Integration of component abstracts.
 - > Custom pre-routes and autoroute solution.
 - > Proprietary RC analysis and buffer insertion methodology.



Complex Design Flow

Key Point: Design Flow different for different design phases.

- Architectural pipeline reflected closely in the floorplanning of:
 - > Memory Macros.
 - > Control Regions.
 - > Datapath Regions.
- Early Design Phase:
 - > Fully integrated SUN toolset allows fast turnaround.
 - > Less accurate, but fast - allows quick iterations to identify timing fixes involving RTL/floorplan changes.
 - > Allows reaching route stability.
- Stable Design Phase:
 - > More accurate, but not as fast, allows timing fixes involving logic and physical changes; Allows logic to freeze.
- Final Design Phase:
 - > More accurate, but longer time to complete; More focus on physical closure then logic.
- Freeze and ECO Design Phases:
 - > Allows preserving large portion of design from one iteration to next.



Key Cluster Methodology Features

(Floorplanning, Synthesis, Placement)

- Cluster Floorplan partitioned into cell areas or regions:
 - > Types - Datapaths, Control Blocks, Custom Macros, “top” level.
 - > All blocks are relatively placed.
 - > Datapaths and Control Blocks placements flattened; Logical hierarchy != physical.
- Cluster pins driven top-down from fullchip level with bottom-up negotiation.
- Routing is done flat at cluster level for better route optimization.
- Datapaths:
 - > Pseudo-verilog inferred datapath rows (macros).
 - > Embedded flop headers and mux-selects.
 - > Rows relatively placed within the DP regions.
 - > Minimum sized cells – will be sized after global route.
- Control Blocks:
 - > Synthesis and placement of each Control Block done stand-alone.
 - > Bounding box for placement obtained from assigned region in parent cluster.
 - > 'Virtual' pin-locations for placement derived from previous iteration of global route.
 - > Placement (def) converted to flat relative placement in parent cluster.
 - > Pseudo-verilog for flop instantiation.

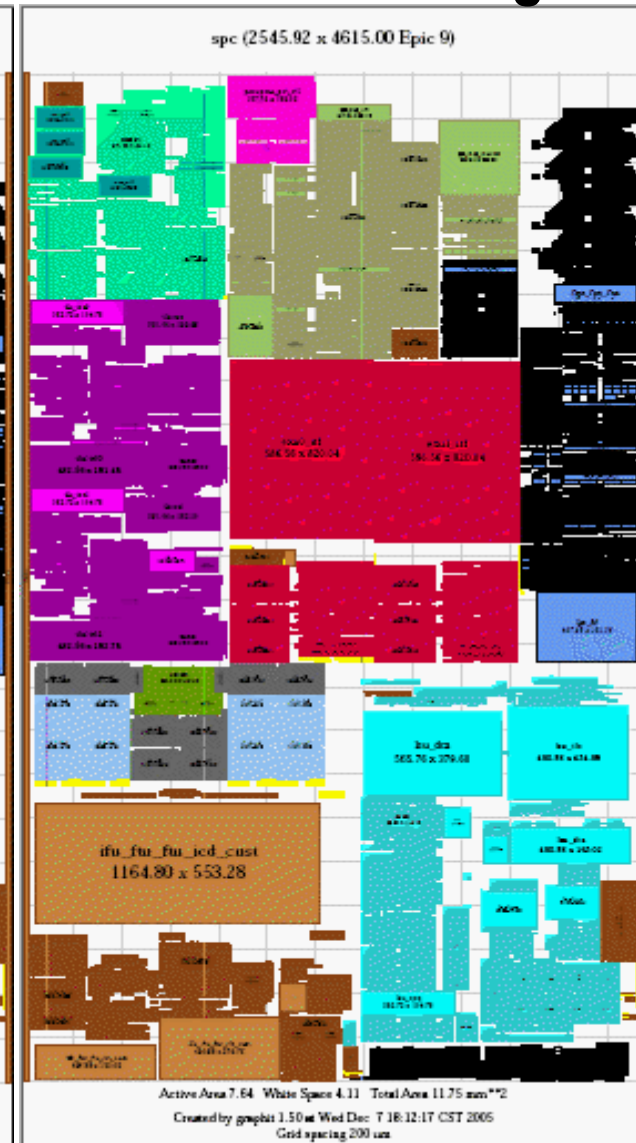
Key Cluster Methodology Features (Gate Sizing)

- Flat Global route -> 2-D symbolic extract -> Timing Analysis -> Gate Sizing.
- 1st Phase -> upsize to meet nominal slew spec.
- Load and Delay 'Stamping' of long wires.
- 2nd Phase -> 'Over-size' to meet delay spec for critical paths.
- Cluster re-relatively placed using new gate sizes.
- Prior iterations help set 'pre-gate-size' floorplan for best 'post-gate-size' results.

Before Gate-Sizing



After Gate-Sizing



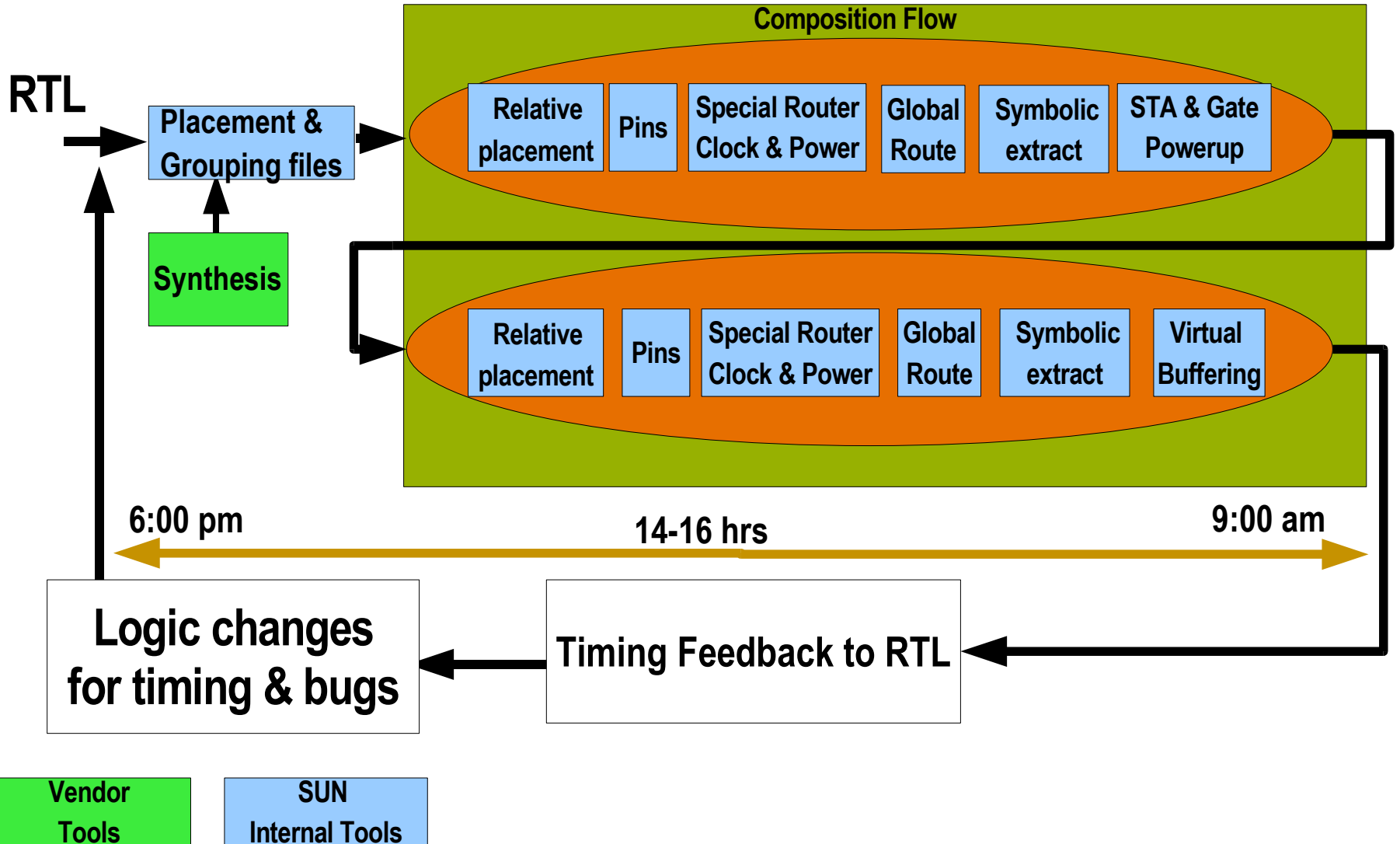
Key Cluster Methodology Features (Clock and Power Grid insertion)

- Clock Grid Insertion:
 - > Auto-constructed from rules governing usage metal, metal width/space, shielding requirements.
 - > Level2 Clock Grid placed on M7/M8:
 - > Anchored by grid drivers placed manually along opposite edges of clusters.
 - > Skew reduced using low-R unloaded metal straps from grid drivers to specific points within grids.
 - > Up to 3 Level2 clock grids per cluster – multiple clock domains.
- Power Grid insertion:
 - > Auto-constructed from rules governing usage metal, metal width/space.
 - > M2 - M4 power constructed uniquely for each cluster region; Cluster regions then stitched together:
 - > Cells having specific M3 needs contain M3 power pins to guide power insertion – example clock shielding.
 - > M5 - M8 power constructed flat across the whole cluster.

Early Design Phase

Ideal clocks
Global Routing
Virtual buffering

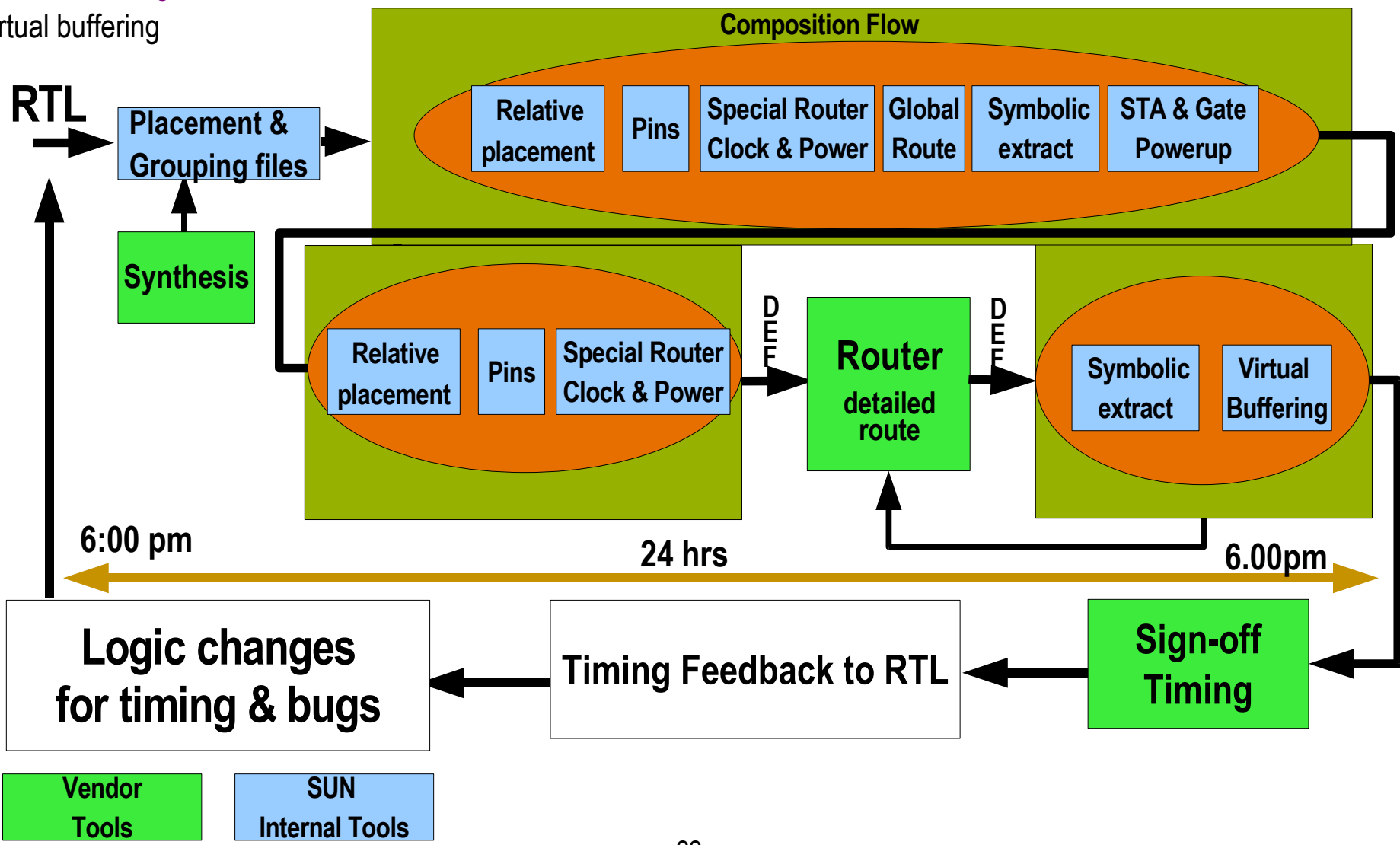
Identify timing paths involving logic changes and floorplan changes



Stable Design Phase

Identify timing paths having impact on logic due to detail route

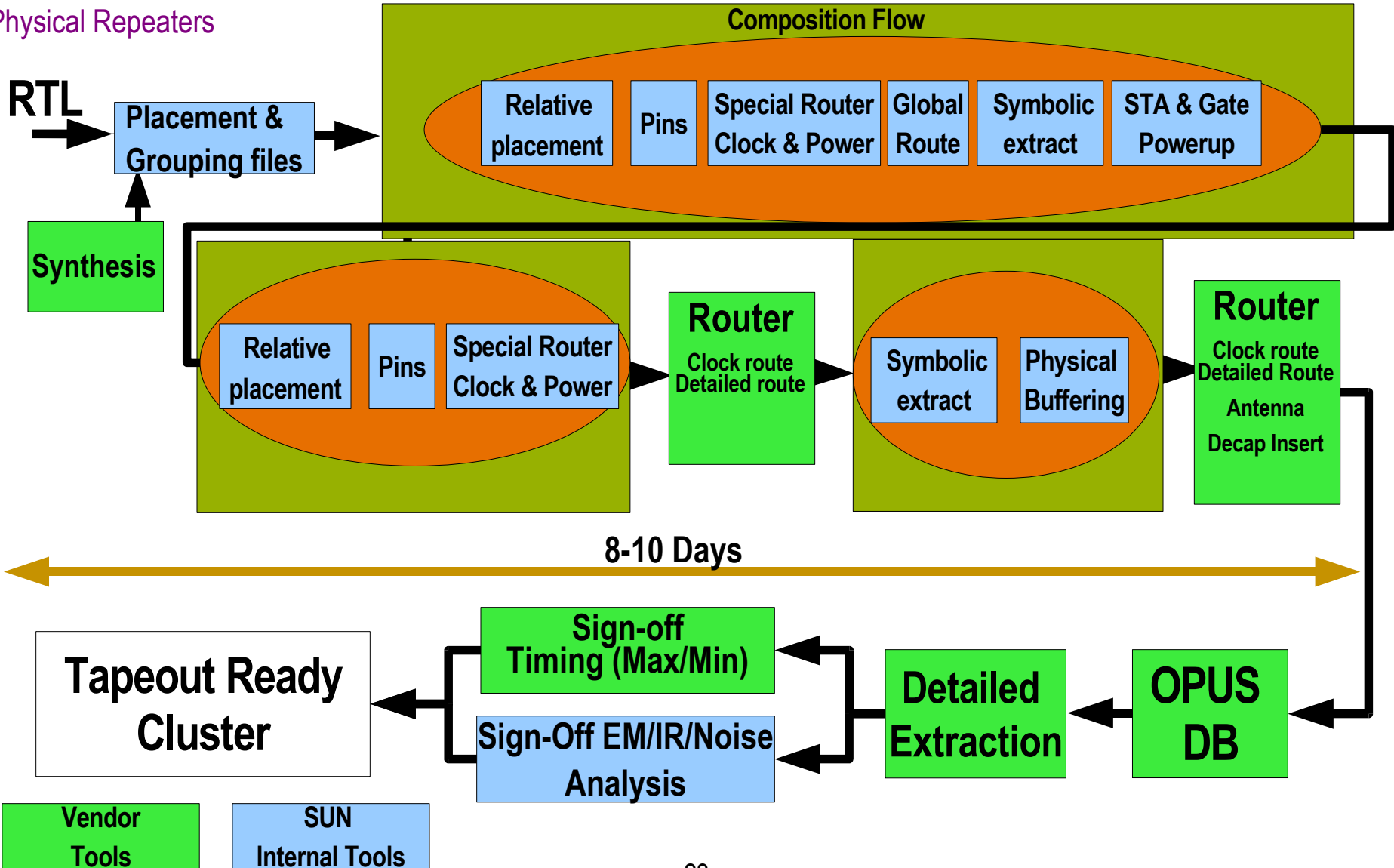
Ideal clocks
Detailed Routing
Virtual buffering



Final Design Phase

Focus on logic freeze and physical closure

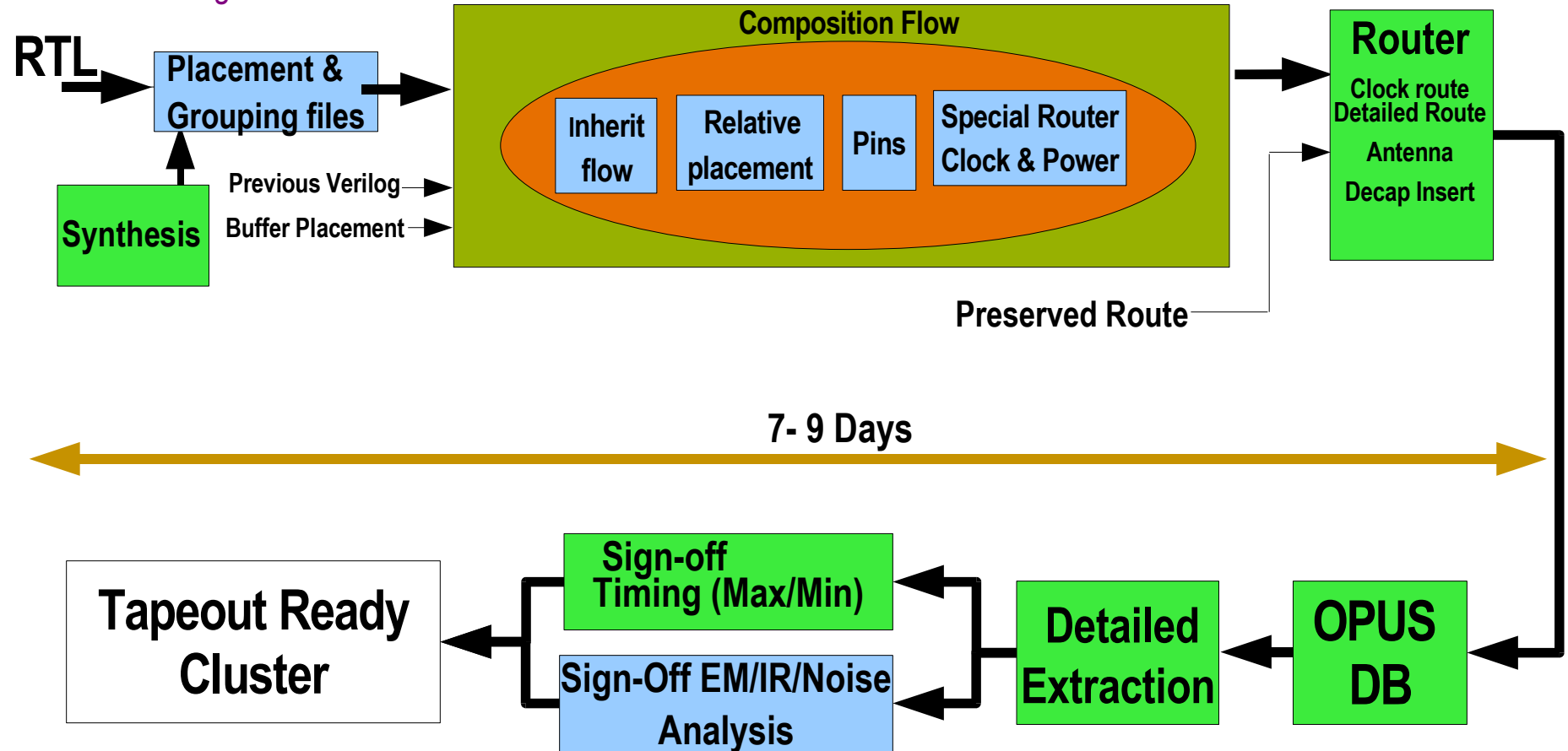
Detailed Clocks
Detailed Routing
Physical Repeaters



Freeze Design Phase

Focus on logic & physical freeze (Preserve most of the design)

- Preserve Cell Size
- Preserve Physical Buffers
- Preserve Critical routes
- Detailed Clocks
- Detailed Routing- non critical

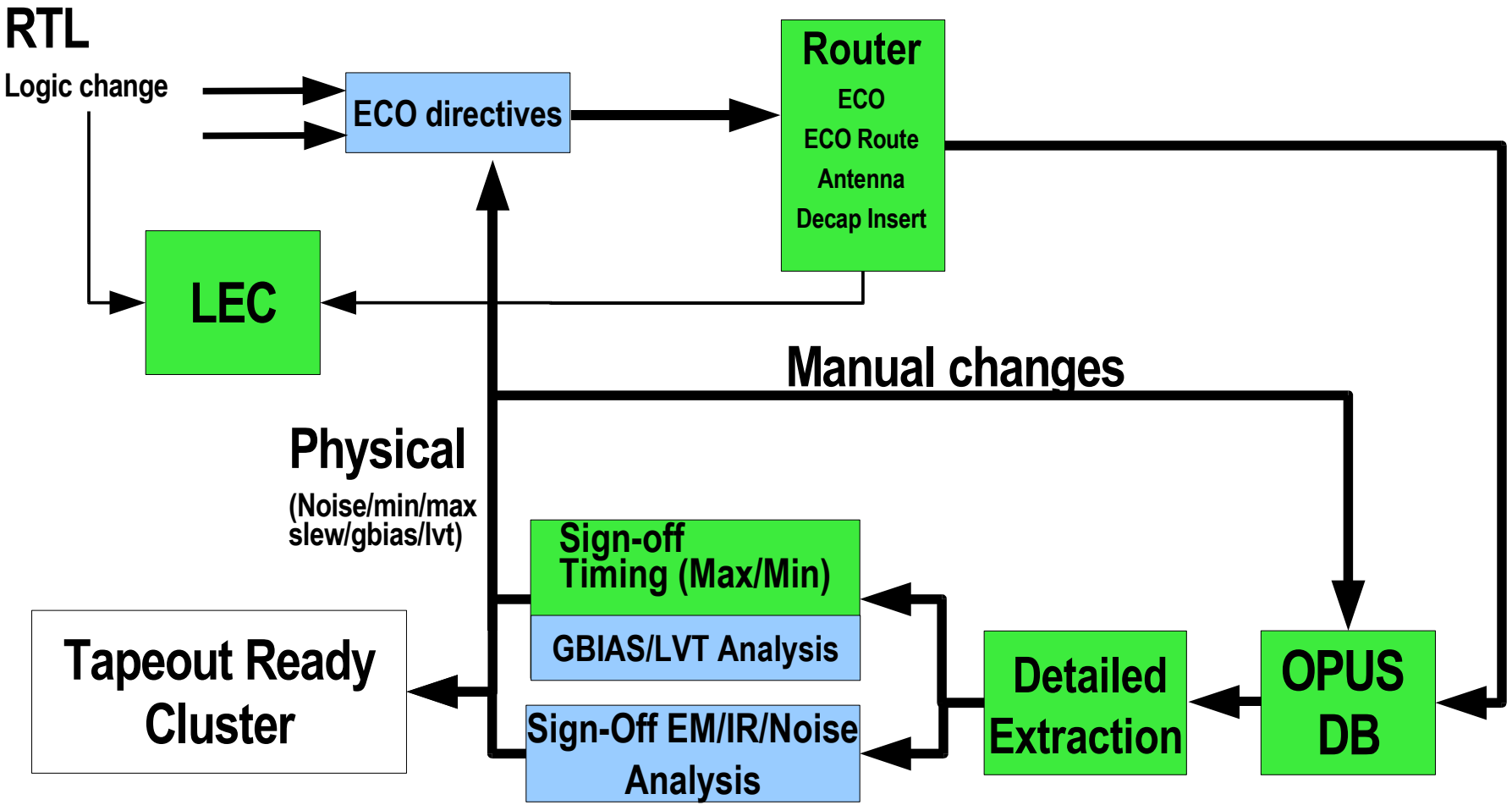


Vendor Tools

SUN Internal Tools

ECO Design Phase

ECO gatelevel netlist
ECO Route



Vendor Tools (Green box) **SUN Internal Tools** (Blue box)

Niagara2's Timing Methodology

- Timing flows built on Industry Standard STA tools.
- Sun-internal char engine used for Standard cell char.
- Design converged on aggressive miller assumptions.
 - > Timing windowing used close to Tapeout for final tuning.
- Formal RTL checks used to verify correctness of multicycle path statements used for Sync_En-ed paths.
- Hierarchy:
 - > Flat gate-level netlist for Cluster analysis.
 - > ILM-based analysis at full-chip level.
 - > Fullchip gates used only for dft/reset timing verification.
- Automated substitution flow to swap in Low-Vt cells for critical path optimization.
 - > What-if analysis enables quick min/max convergence.

Niagara2's Timing Methodology: Mintime

- Extensive cell library support for fixing hold-time violations:
 - > Footprint compatible flop variants with higher intrinsic Clock->Q delay.
 - > Footprint compatible buffer variants with larger delay.
 - > Option of inserting regular or 'min-time' buffers as well.
- Automatic slack-based 'fix' flow generated P&R ECO control directives.
- Clock skew spec derived from spice simulations to account for worst case PVT variations.
 - > Extensive Common Path Pessimism Relaxation used for Global clock tree and CTS trees.
 - > L2 grid skew modeled directly into STA tool; allowed distance-based skew relaxation.

Design For Manufacturability (DFM)

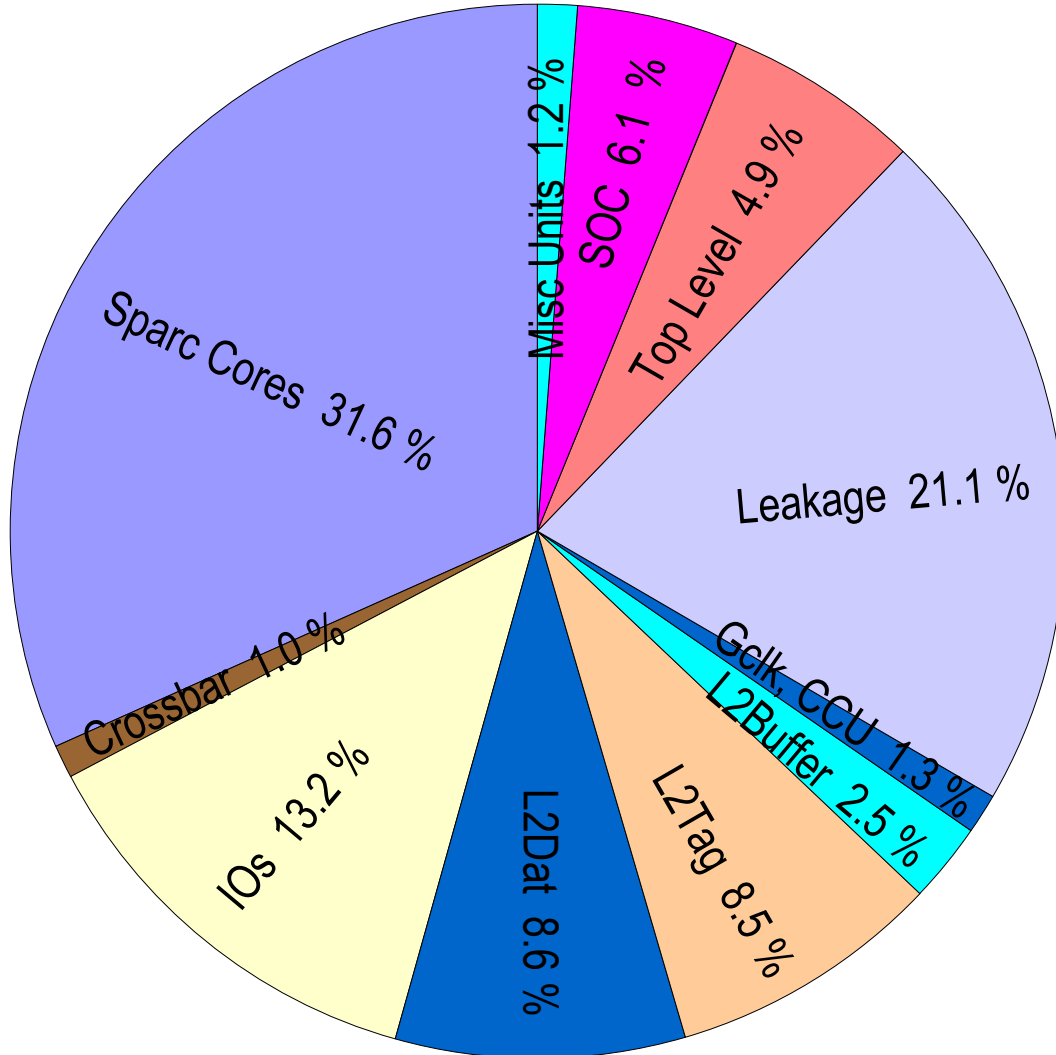
- Single poly orientation (except I/O blocks).
- Larger-than-minimum design rules:
 - > To minimize impact of poly/diffusion flaring.
 - > Near stress-prone topologies to reduce chances of dislocations in Si-lattice.
 - > Larger Metal overlap of via/contact where possible.
- Improved gate-CD control:
 - > Dummy polys used for gate shielding.
 - > Limited gate-poly pitches used; OPC algorithm tuned for them.
- OPC simulations of critical cell layouts to ensure sufficient manufacturing process margin.
- Extensive use of statistical simulations:
 - > Reduces unnecessary design margin that could result from designing to FAB-supplied corner models that often are non-physical.
- Redundant vias placed without area increase.
- All custom ckts proven on testchips prior to 1st Si.

Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Power

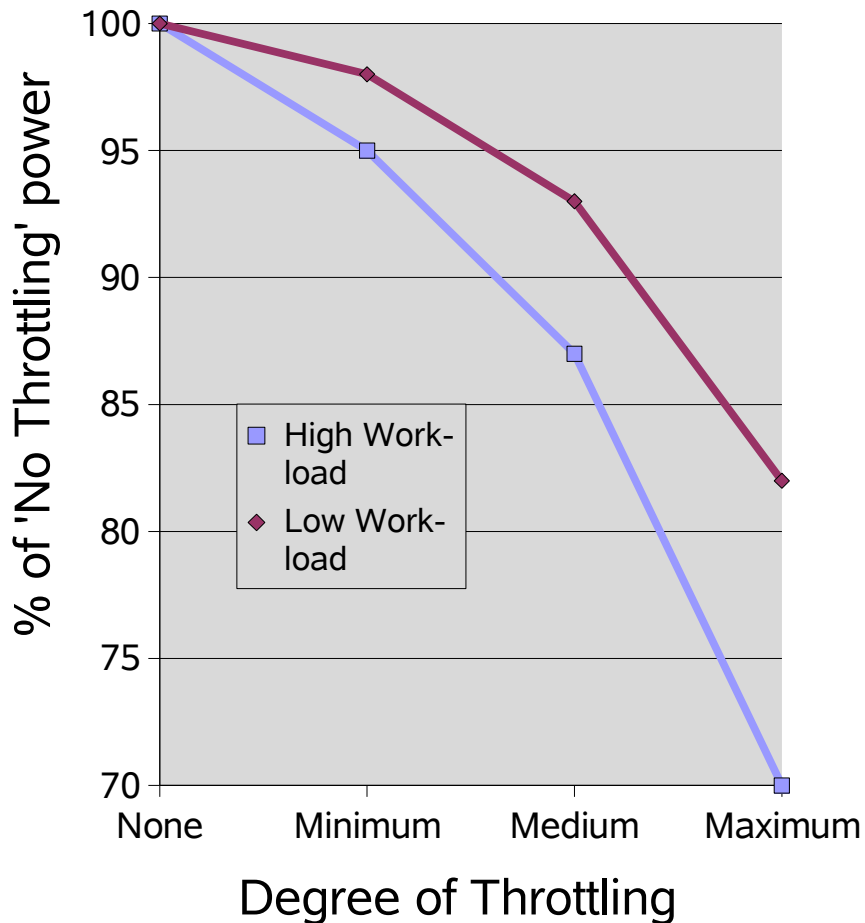
Niagara2 Worst Case Power =
84 W @ 1.1V, 1.4 GHz



- CMT approach used to optimize the design for performance/watt.
- Clock gating used at cluster and local clock-header level.
- 'GATE-BIAS' cells used to reduce leakage.
 - > ~10 % increase in channel length gives ~40 % leakage reduction.
- Interconnect W/S combinations optimized for power-delay product to reduce interconnect power.

Power management

Effect of Throttling on Dynamic Power



- Software can turn threads on/off.
- 'Power Throttling' mode controls instruction issue rates to manage power consumption.
- On-chip thermal diodes monitor die temperature.
 - > Helps ensure reliable operation in case of cooling system failure.
- Memory Controllers enable DRAM power-down modes and/or control DRAM access rates to control memory power.

Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > Clocking Scheme
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Design for Testability

- Deterministic Test Mode (DTM) used to test core by eliminating uncertainty of asynchronous domain crossings.
- Dedicated 'Debug Port' observes on-chip signals.
- 32 scan chains cover >99 % flops; enable ATPG/Scan testing.
- All RAM/CAM arrays testable using MBIST and Macrotest.
 - > Direct Memory Observe (DMO) using Macrotest enables fast bit-mapping required for array repair.
- Path Delay/Transition Test technique enables speed testing of targeted critical paths.
- SerDes designs incorporate loopback capabilities for testing.
- Architecture design enables use of <8 SPCs/L2 banks.
 - > Shortened debug cycle by making partially functional die usable.
 - > Will increase overall yield by enabling partial-core products.

Outline

- Key Features and Architecture Overview
- Physical Implementation
 - > Key Statistics
 - > On-chip L2 Caches
 - > SerDes interfaces
 - > Cryptography Support
 - > Physical Design Methodology
- Power and Power Management
- DFT Features
- Conclusions

Conclusions

- Sun's 2nd generation 8-core, 64-thread, CMT SPARC processor optimized for Space, Power, and Performance (SWaP) integrates all major system functions on chip.
- Doubles the throughput and throughput/watt compared to UltraSparcT1.
- Provides an order of magnitude improvement in floating point throughput compared to UltraSparcT1.
- Enables secure applications with advanced cryptographic support at wire speed.
- Enables new generation of power-efficient, fully-secure datacenters.

Acknowledgements

- Niagara2 design team and other teams inside SUN for the development of Niagara2.
- Texas Instruments for manufacturing Niagara2.

Thank You !