CE653 – Asynchronous Circuit Design

Instructor: C. Sotiriou

http://inf-server.inf.uth.gr/courses/CE653/

Course Focus: Asynchronous VLSI Design

- General Goal of Asynchronous VLSI Design
  - Explore and take advantage of more general circuit structures that do not have a single global clock.
- CE653 Goal:
  - Prepare engineers to
    - Understand different async design styles and their tradeoffs
    - Critically analyze their applicability
    - Use existing asynchronous CAD tools
    - Provide background for future study
- Goal of this lecture
  - Why is asynchronous VLSI interesting?
  - Put EE552 in context of USC VLSI & CAD Curriculum

What is VLSI?

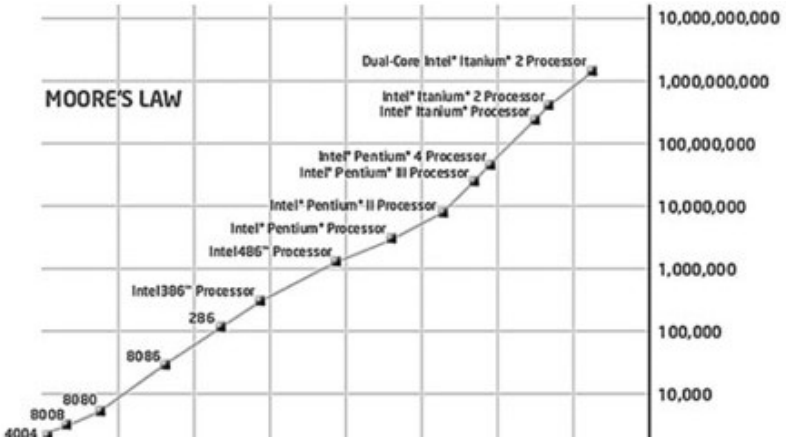
- Fundamental goal: to implement algorithms in hardware
- Many possible technologies
  - TTL, NMOS, ECL, CMOS
  - Ge, GaAs
  - Currents instead of voltages
  - Analog Computation
  - Synchronous vs. Asynchronous
  - Quantum computing
- How do we judge a technology?
  - Correctness
  - Performance: Latency, Throughput, Energy
  - Cost…

Complexity

If the semiconductor industry [had] a million-transistor technology like VLSI, I'm not so sure it would know what to do with it.

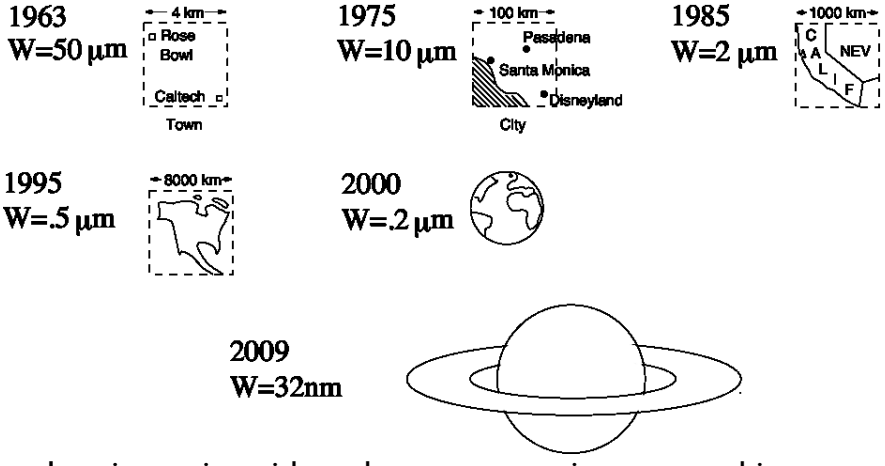
- January 1979, Gordon Moore's Keynote Address, First VLSI Conference

Moore's Law



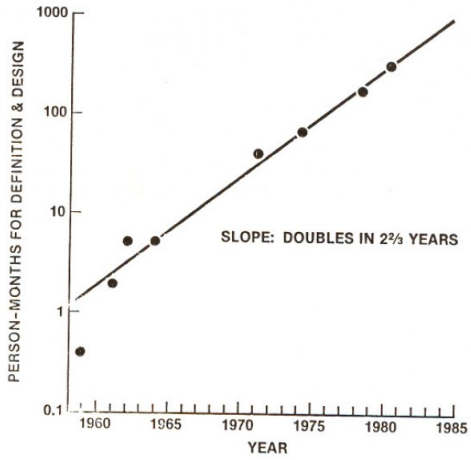
- The number of transistors doubles every 18mo or so...
- This is not a statement about performance

Complexity



- Imagine a city with each street as a wire on our chip, 200m between blocks. (Seitz and Mead, 1979.)

Complexity

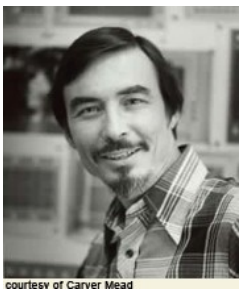


- Gordon Moore's other law... presented in 1979

Complexity

VLSI is a statement about system complexity... Many fundamental ideas [pertaining to large system design] have yet to be discovered. The architecture and algorithms for highly concurrent systems is even less well developed.

- January 1979, Carver Mead's Keynote Address to the First VLSI Conference

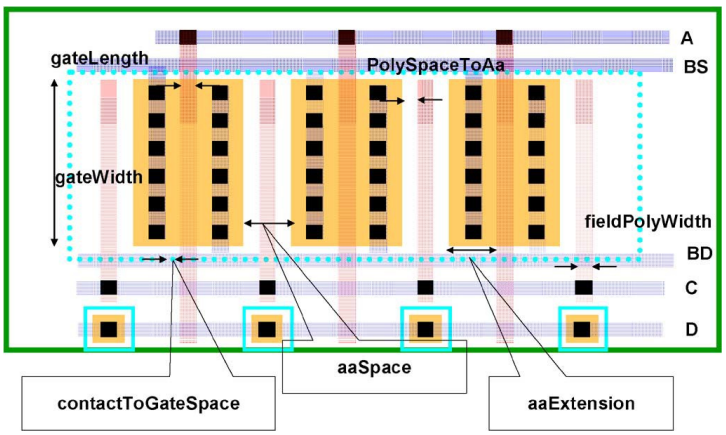


courtesy of Carver Mead

Performance--Constant E-Field

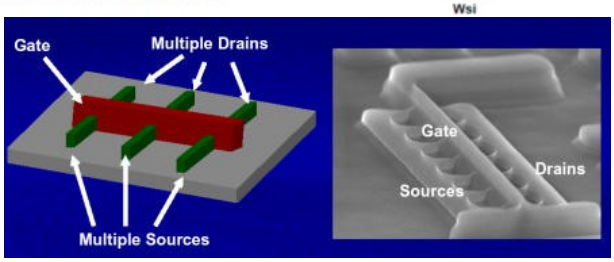
Parameter	Was	Now	What happened
scale factor	**1**	**1/a**	
linear size	**x**	**x/a**	**Dimensions shrink**
voltage	**V**	**V/a**	**Reduce voltage**
E-field	**E**	**E**	**Constant E-Field**
current	**i**	**i/a**	**Reduce current**
power	**vi**	**vi/a^2**	**Per-device**
delay	**T**	**T/a**	**Reduce delay**
energy	**viT**	**viT/a^3**	

Gates



3D Tri-Gates / FinFets

Source: Globalfoundries

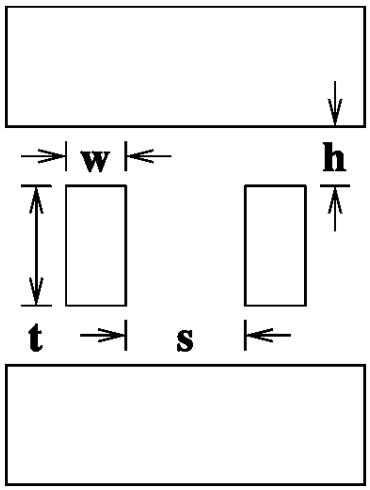


Source: Intel

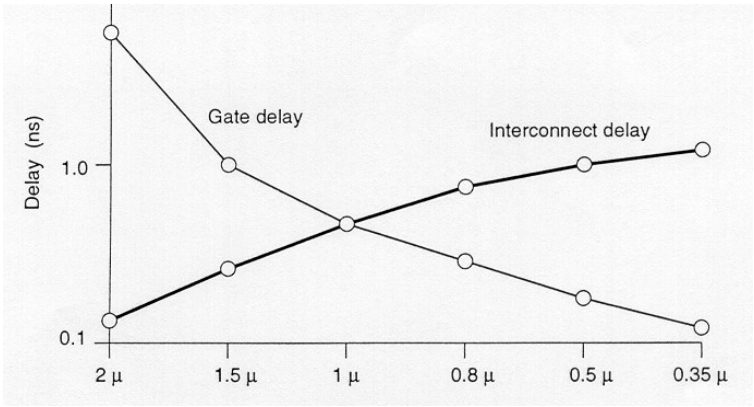
- Higher I_{dsat} when on -> Higher performance, Less leakage when off -> Lower power
- Faster changing between on and off

Wires

- Pitch = w + s
- Aspect ratio = t/w
- Old processes had AR<<1
- Modern processes AR~2
- Under scaling:
  - Narrower and thinner
  - Wiring layers are closer
  - Wires are drawn closer



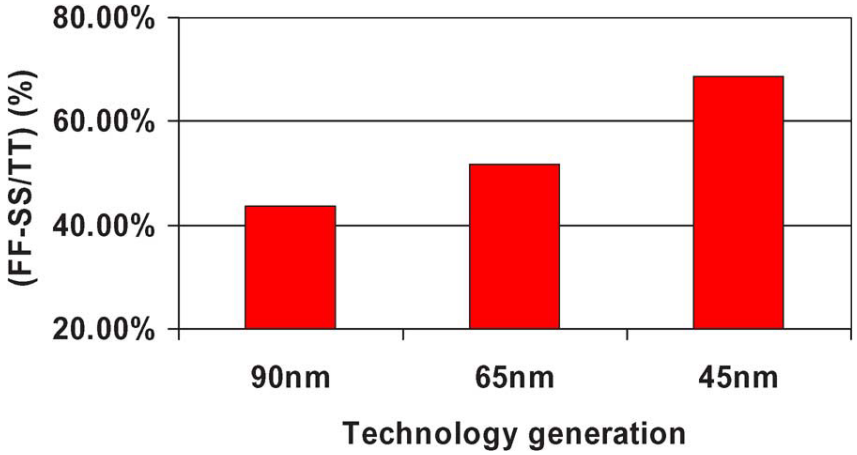
Wire Delays vs Gate Delay



Gate parasitics are decreasing but not interconnect parasitics

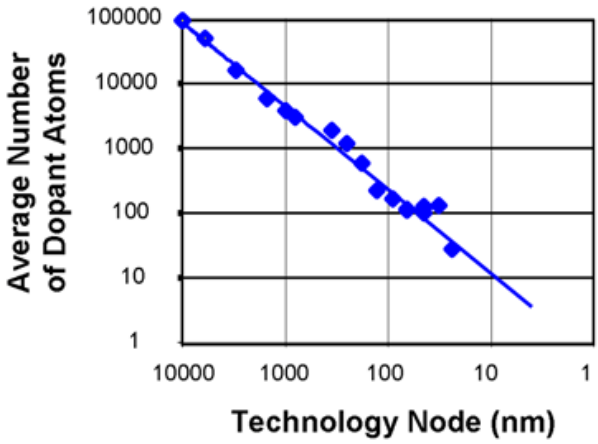
- gate delay is decreasing
- but distance is becoming relatively more costly

Delay & Transistor Variation



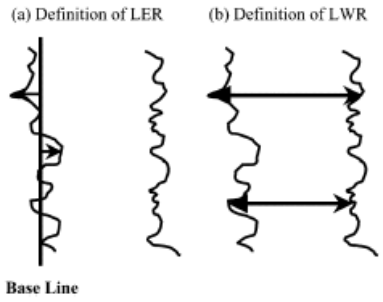
S. Saxena, et al., "Variation in Transistor Performance and Leakage in Nanometer Scale Technologies," *IEEE Trans. Electron Devices,* vol. 55, no. 1, Jan 2008

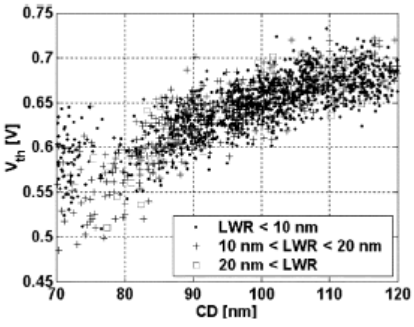
Delay & Transistor Variation



- Fewer dopant atoms => greater change in Isat from single atom fluctuations

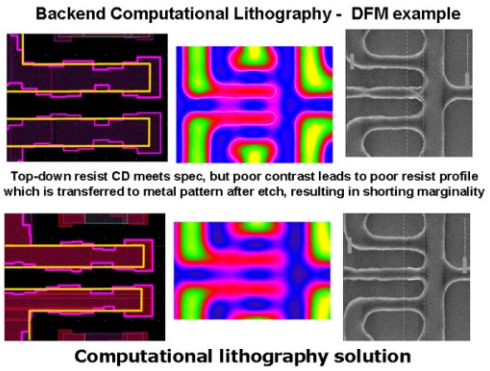
Line Roughness





H.W. Kim, et al, "Experimental Investigation of the Impact of LWR on Sub-100-nm Device Performance", IEEE Trans. Electron Device, vol 51, no 12, 2004

Delay & Wire Variation

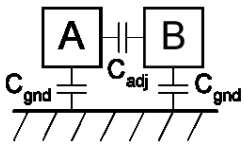


- Wire pitch is well controlled in modern processes
- Pitch is determined by mask precision itself
- Width and space individually is not
- Width and space depend on etching and photoresist

K Kuhn, C. Kenyon, et al. "*Managing Process Variation in Intel's 45 nm. CMOS Technology*." Intel Technology Journal, vol. 12, no. 2, *Jun. 2008*

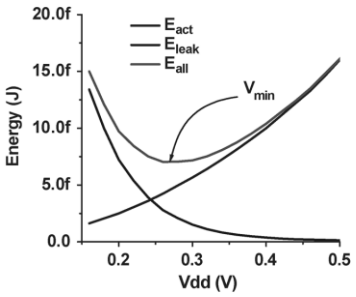
Delay & Wire Cap-Coupling

- If we assume layers above and below are quiet
  - Can estimate Cabove and Cbelow as lumped
  - Can assume to be GND
- Effective capacitance depends on switching of neighbors
- Capacitances changes in time: glitches due to charge-conservation (Q=CV)
- Wire-cap often now exceeds gate-cap therefore capacitance variation is significant, therefore delay variation is significant

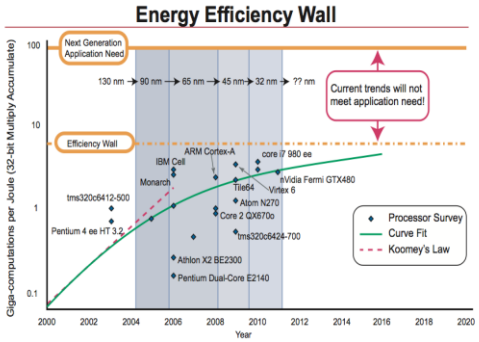


B	deltaV	Ceff(A)	MCF
Constant	**Vdd**	**Cgnd + Cadj**	**1**
Switching with A	**0**	**Cgnd**	**0**
Switching opposite A	**2Vdd**	**Cgnd+2adj**	**2**

Power and Energy



Zhai et al. 2009



Marr 2010

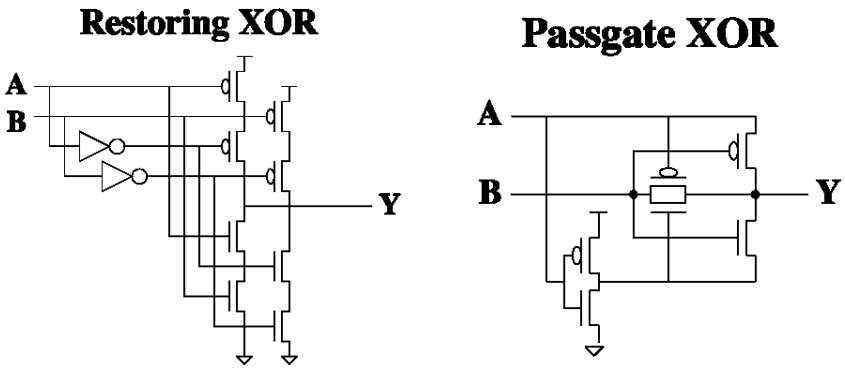
- Minimum energy design achieved when Vdd is near or sub threshold
  - Process variations at this Vdd are very high
- Synchronous techniques hitting an energy efficiency wall
  - Opens options for alternative techniques

Abstractions

Use design patterns (discipline) to abstract away details of semiconductor devices and parasitics

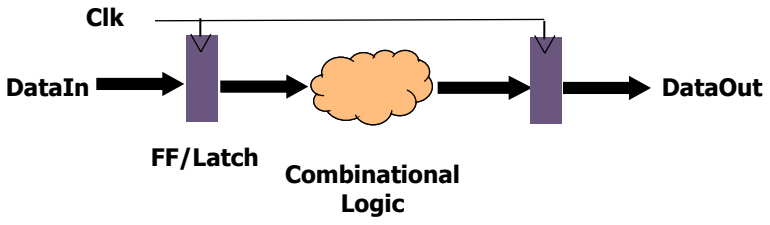
- Combinational Logic Abstraction
  - Treat transistor networks as digital switches to implement Boolean functions
    - Pass-gate logic
    - Restoring combinational logic
    - Domino logic (DEC Alpha)
- Synchronous design abstraction
  - Abstract time into ticks of a clock
    - 2-phase non-overlapping clocks w/ latches (Mead+Conway, 1979)
    - Pulse-mode clocking with latches
    - Single global clock with edge-triggered flip-flops
  - Separates performance considerations from function (kind-of)

Switching Networks



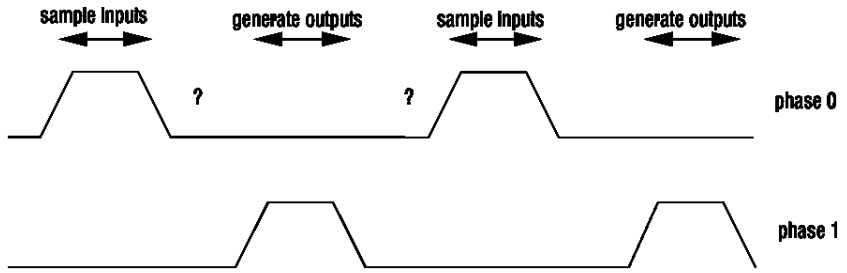
Pass gate version is much simpler...

The De-facto Standard: Synchronous Design Abstraction



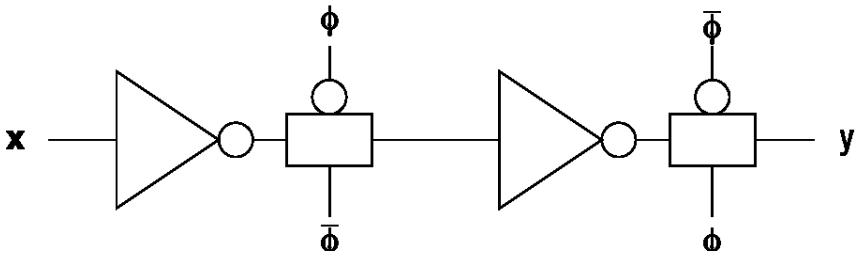
- Flip-flops (aka registers/latches)
  - Memory elements that store "state" of system
- Combinational Logic
  - Performs logical functions on data (e.g., add, mult, etc…)
- Clock
  - Periodic square wave that controls update of memory elements
  - Assume data is stable upon latching of data

Clocking: Two-Phase



- Setup and hold can be met post-silicon by manipulating clock-frequency---one-sided timing constraint. Safe but slow.

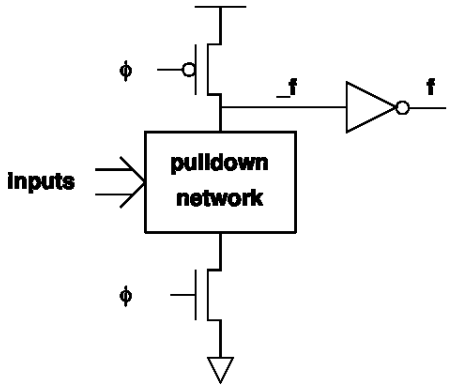
Clocking: One-Phase



- Inputs latch closes at the "same moment" that output latch opens
- Inputs must be held (hold time) until input latch is definitely closed

Domino Logic

- Combines (Latch + Logic function) --less area
- Eliminates P-Network, PMOS is usually 2-3x larger than NMOS
  - Equivalent functions have less load capacitance
  - Equivalent functions are faster
  - Allows for much more logic complexity with less latency (and energy)



Time-Quality Tradeoff

- Full-custom design
  - Hand-crafted, highly optimized
    - Advanced circuit styles and structures
  - Less automation
    - Larger design time
  - Appropriate for stringent design requirements
- Semi-custom design
  - Semi-automated, generally not as optimal design
    - Constrained circuit styles and flows
    - Makes CAD tool design easier
  - Full automation
    - Lower design time
  - Appropriate for less stringent design requirements

Baseline Analysis
Semi-Custom Synchronous

- Positives
  - + Good performance/power with 12-month design times
  - + Supported by mature CAD tools
    - + Characterized cell library
    - + Automated synthesis from RTL
    - + Mature physical design flows
- Negatives
  - – Use constrained circuits and methodology
    - – Static CMOS standard gates
    - – Limited clocking and gated-clocking methodologies
    - – Limited flip-flops with large D-Q overheads
  - – Variation in deep-submicron
    - – Timing closure problems causing schedule slips
    - – Variability causes large margins in performance and power
  - – High electro-magnetic interference

Baseline Analysis
Full-Custom Synchronous

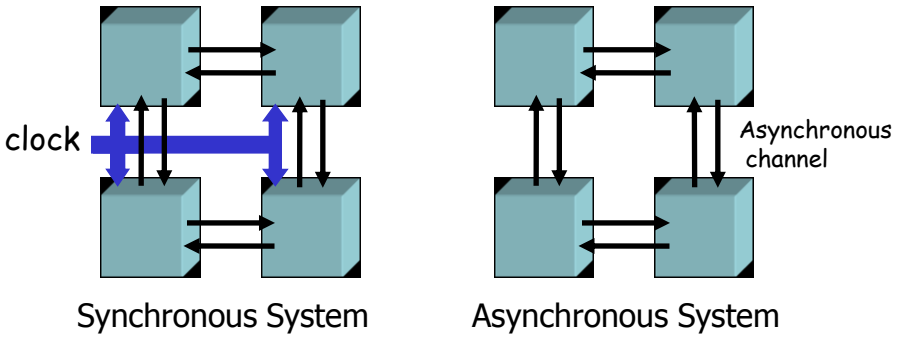
- Positives
  - + 2X improvement in performance and power
    - + Carefully designed macro cells
    - + Dynamic logic (e.g., self-resetting domino)
    - + High-speed flip-flops and latches
    - + Low-voltage design
- Negatives
  - – Increased design time (~36 months)
    - – Charge sharing problem with dynamic logic
    - – Aggressive two-sided timing assumptions
    - – Extensive analog verification pre and post-layout
  - – Still may have high electro-magnetic interference

- **Full-custom versus semi-custom --**
- **Basic tradeoff between productivity (design-time) and quality**

Synchronous Design - Challenges

- Design Complexity is Increasing
  - High frequencies depend on careful pipelining
  - Pipelining has algorithmic implications
    - Clock-stage misalignments are a significant source of design error
    - Reusing sub-circuits depends on accommodating pipeline depth
    - Mismatches may not manifest under testing if critical data happens to hold constant cycle-to-cycle
- Variation demands more margin
  - ~30-40% lost to process-corner (wafer) variation
  - ~5-10% lost to in-die variation
  - ~10-20% lost to signal integrity
  - Standard discipline does not accommodate multiple clocks easily
  - Multiple clock-domains challenging due to metastability issues
  - Global clock-period must enclose worse-case path
    - ***Cost of doing business as usual is increases*** →
      - ***Interest in alternatives increases!***

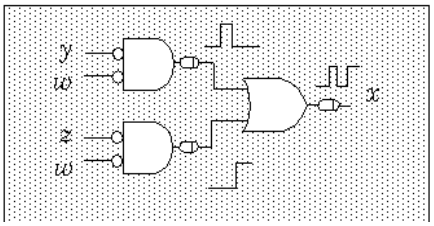
The Asynchronous Alternative



Synchronization and communication between blocks implemented with asynchronous channels that send and receive tokens

Logic Hazards (Glitches)

- Synchronous Circuits
  - Glitches tolerated because outputs sampled only after signals settle
  - Clocking constraints
    - Clock edge occurs only after data settles
    - Limits clock frequency
- Asynchronous Circuits
  - Control circuits
    - Avoided completely
    - Hazard-free logic synthesis techniques
  - Datapath (Either)
    - Outputs sampled after signal settles OR
    - Avoided completely



Power Consumption Comparison

- Synchronous Circuits
  - Clock: up to 50% of chip power
    - Particularly for high-performance designs
  - Hazards: up to 70% of computation power
    - For very unbalanced logic; typically much smaller
- Datapath
  - Elements expend energy in every clock
    - Clock gating can help here
- Asynchronous Circuits
  - \+ No global clock
  - \+ Only expend energy in datapath element when used
    - \+ Perfect gated clocking
    - \+ Can turn data slices off
  - \- Handshaking circuitry has power overhead
  - \- Some types of designs have high switching activity

Performance Comparison

- Synchronous Design
  - Clock frequency set to worst-case conditions
  - Semi-automated
    - Limited circuit styles supported by automated tools
    - Limited clocking styles and Flip-Flops/Latches
    - Process variations lead to large over-design and limited clock frequency
  - Full-custom
    - Advanced circuit styles available
    - Advanced clocking styles and latches
    - Robustness versus performance tradeoffs
- Asynchronous Design
  - Some design styles are very robust to process variations
  - Close to full-custom performance with ASIC design times possible

Other Asynchronous Advantages

- Reduced Electromagnetic Noise
  - No global clock
  - Frequency spread out much more
  - Co-locate with sensitive analog
    - Less noise - does not effect analog circuitry
  - Others argue not critical
    - Analog circuitry can be designed to be insensitive to clock noise
- Ease of modular composition
  - Supports GALS design for multi-frequency disparate SoC designs
  - Handshaking offers plug-and-play IP
- Adjusts to operating conditions/process variations
  - Supports easy voltage scaling (no need to pair voltage/frequency)
- Immediate start-up (no need to wait for clock to stabilize)
  - Go quiet
  - Run fast
  - Go quiet again

Asynchronous Challenges

- Lack of CAD tools
  - Limited support from major EDA companies
  - Asynchronous EDA start-up environment challenging
- High-performance asynchronous design
  - Some blocks can be 2-5x larger due to dual-rail design
  - May consume more peak power than desired
- Low-power asynchronous design
  - Can be slower than desired if control overhead not managed
- Debug
  - Circuit can't be slowed via clock to aid in debugging
- Asynchronous test
  - Testers geared toward synchronous
  - Standards do not exist and test methodologies still evolving
  - Automatic test pattern generation in infancy

Asynchronous Commercialization Efforts

- Fulcrum Microsystems (www.fulcrummicro.com)
  - Fabless semi-conductor company
  - High-performance computing and networking markets
  - Uses high-performance async design as secret sauce
  - Founded out of Caltech in 2000; Bought by Intel in 2011
- Achronix (www.achronix.com)
  - High-performance async FPGA core with synchronous interfaces
  - Founded out of Cornell research in 2006
- TimeLess Design Automation
  - ASIC Flow for Asynchronous Design
  - First target – high-performance – GHz+ silicon in 65nm
  - Founded out of USC in 2008
  - Sold to Fulcrum Microsystems in 2010
- Tiempo (www.tiempo-ic.com)
  - IP Cores and ASIC Flow (Power/Performance Tradeoff)
  - Lower performance than TimeLess Design Automation
- Numerous failed start-ups
  - Handshake Solutions, Silistix, Elastix, Nanochronous