Design Technologies for Nanosystems on Chips

Giovanni De Micheli
Centre Systèmes Intégrés

Outline

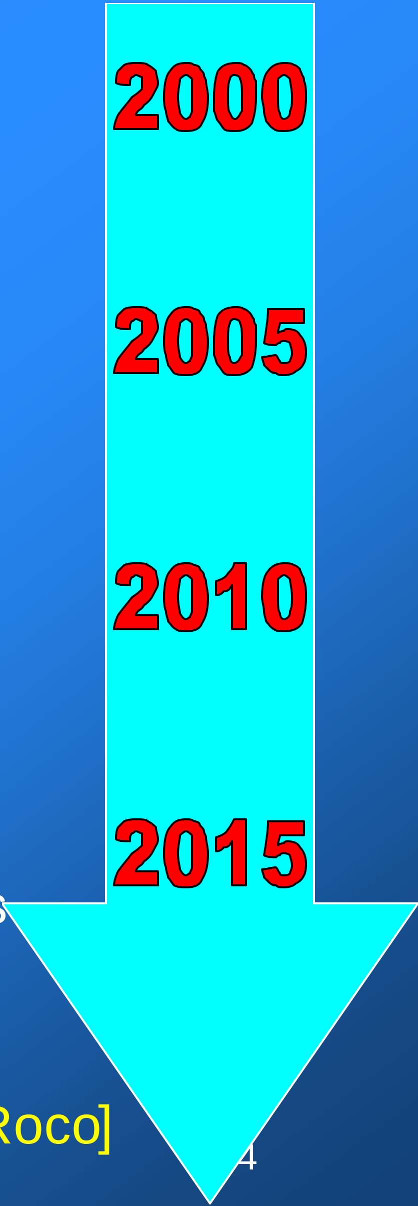
- Nanotechnology and nanoelectronics
- Application domains and requirements
- Design technology challenges
- Architectures for nanocircuits
- Summary and conclusions

Nanotechnology

- Materials with internal structures ranging from 1 to 100 molecular diameters
- Applications to all kinds of engineering
  - Including electronics, mechanical systems, medicine
- Large investments and hype
  - USD 4 Billion research investment in 2005
  - USD 1 Trillion aggregate revenue predicted for 2015
- Restructuring of manufacturing industry:
  - Job creation and reshaping

Evolution of nanotechnology

- Developing of passive nanostructures
  - Reinforcing fibers in new composites
  - Carbon nanotube wires in electronics
- Active nanostructures that change properties
  - Drug-delivery particles
  - Molecular electronic devices
- Nano-systems
  - Self-assembly of electronic systems
  - Tissue regeneration
- Heterogeneous nano-systems
  - Combination of molecular nanosystems, heterogeneous networks with molecules and supramolecular structures



2000

2005

2010

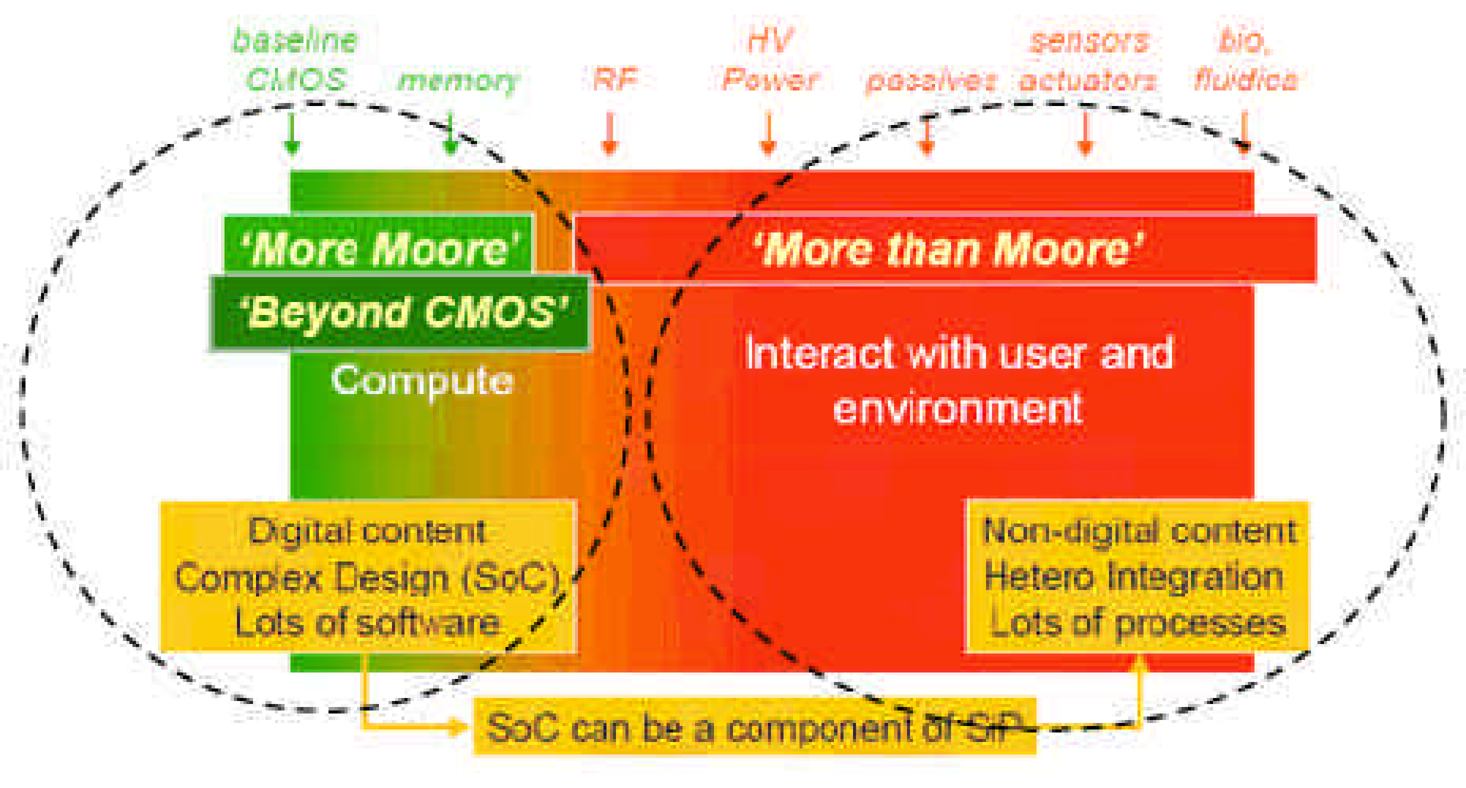
2015

[Source: M. Roco]

Nanoelectronis and integrated nanosystems

- Nanoelectronics:
  - Deeply-scaled standard semiconductor technologies
  - Novel technologies with devices in the [5 - 50] nanometer range
  - Molecular electronics with single-molecule switches and/or storage nodes
- Nanosystems:
  - Combination of nanoelectronics with nanomechanical and nanofluidic components

Where are we heading ?



baseline CMOS

memory

RF

HV Power

passives

sensors actuators

bio, fluidics

'More Moore'

'Beyond CMOS'

Compute

'More than Moore'

Interact with user and environment

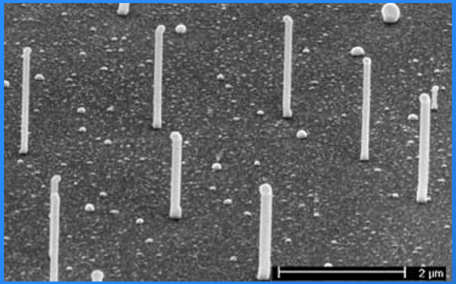
Digital content
Complex Design (SoC)
Lots of software

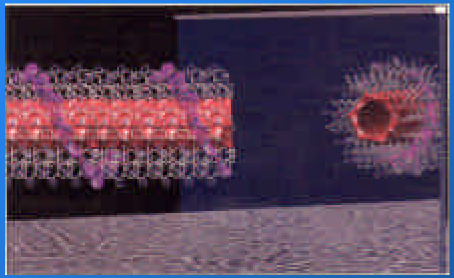
Non-digital content
Hetero Integration
Lots of processes

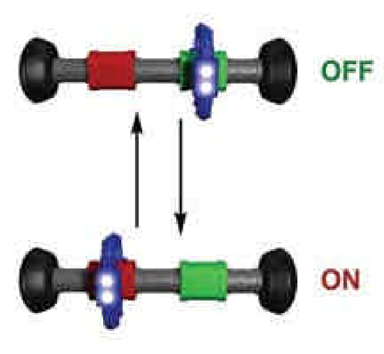
SoC can be a component of SiP

An array of new technologies

- Silicon nanowires
- Carbon nanotubes
- Single-electron transistor devices
- Molecular switches
- Quantum dots
- DNA computing
- …







Will a new nanoelectronic technology prevail?

- The skeptical view:
  - Investments in CMOS silicon are huge
  - We will not need localized computing power beyond what is achievable with a 1 cm^2 die in 25nm silicon CMOS
  - Wiring is the bottleneck: making transistor smaller does not help
- The optimistic view:
  - We will always need increasing computing power and storage capacity
  - We need to curb the increasing costs of manufacturing
  - We will invent new computing architectures, storage media and communication means

How is the transition path?

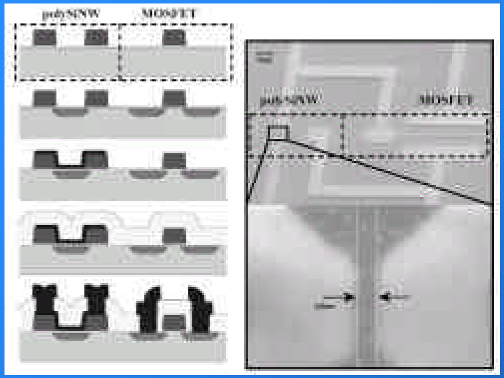
- When will current semiconductor technologies run out of steam?
- What factor will provide a radical change in technology?
  - Performance, power density, cost?
- Will the transition eliminate previous CMOS technologies?
  - Are the new nanoelectronic technologies compatible with standard silicon?
- How will we design nanoelectronic circuits:
  - What are the common characteristics, from a design technology standpoint?

Common characteristics of nano-devices

- Self-assembly can be used to create structures
  - Manufacturing paradigm is both top-down and bottom-up
  - Attempt to avoid lithography bottleneck
- Combined presence of micro and nano-structures
  - Interfacing and compatibility issues
- Significant presence of physical defects and higher failure rates
  - 10-15% defective devices according to recent estimates
  - Design must deal with nonworking and short-lived devices
- Competitive advantage stems from the high density of computing elements
  - Two orders up as compared to scaled CMOS

The nano comptatibility issue

- Combining nano with traditional technologies:
  - – Curse of dimensionality at interface
    - Geometry, driving strengths
- Examples
  - – Silicon nanowires within CMOS circuits
  - – Carbon nanotube interconnect on CMOS circuits
  - – Molecular electronic memories arrays with CMOS peripherals



Outline

- Nanotechnology and nanoelectronics
- Application domains and requirements
- Design technology challenges
- Architectures for nanocircuits
- Summary and conclusions

Application domains






Application domains

- Consumer electronics
  - Multimedia and game applications require increasingly higher performances for image rendering
- Wireless sensor networks
  - New array devices (e.g., image sensors) require high-throughput processing
- Medical electronics
  - Very large throughput requirement (e.g., image analysis) and very high resilience to transient errors (e.g., Xray)

System requirements

- Ultra low power operation
  - Because of untethered applications
  - But high-density computing arrays require significant power
- High reliability
  - Because embedded system applications may be life critical
  - But devices are more prone to fail
- High throughput
  - Because of massive data processing requirements
  - But high throughput must be compatible with ultra low power consumption
- *Can nanoelectronics provide us with a solution with these conflicting needs?*

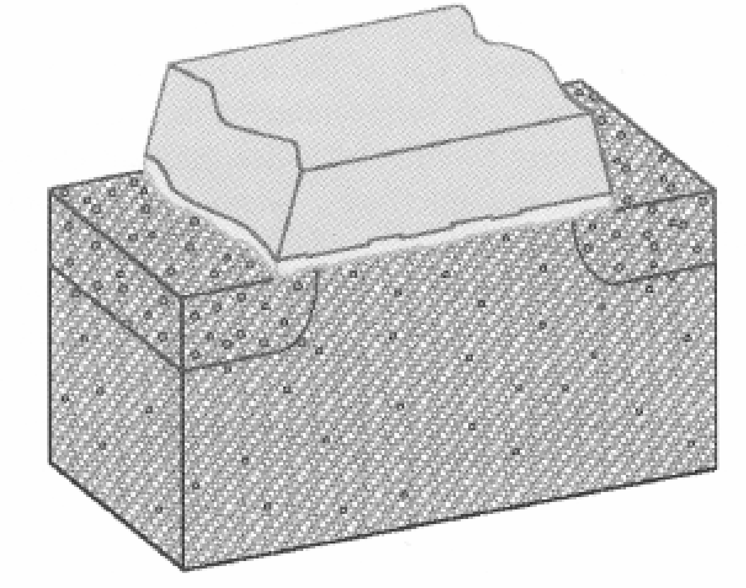
Outline

- Nanotechnology and nanoelectronics
- Application domains and requirements
- Design technology challenges
- Architectures for nanocircuits
- Summary and conclusions

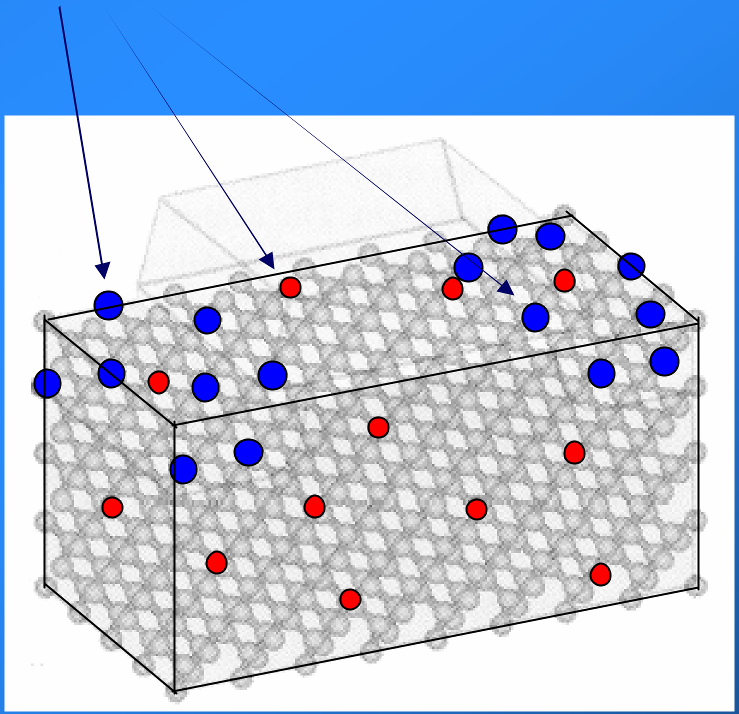
Design issues

- Variability
  - Physical parameter variation
  - Microscopic structural effects
- Reliability
  - Higher failure rate
  - Higher exposure to environmental effects
- Thermal management
  - Heat extraction and gradient avoidance
  - Temperature correlates with higher failure rates

Variability: physical motivation



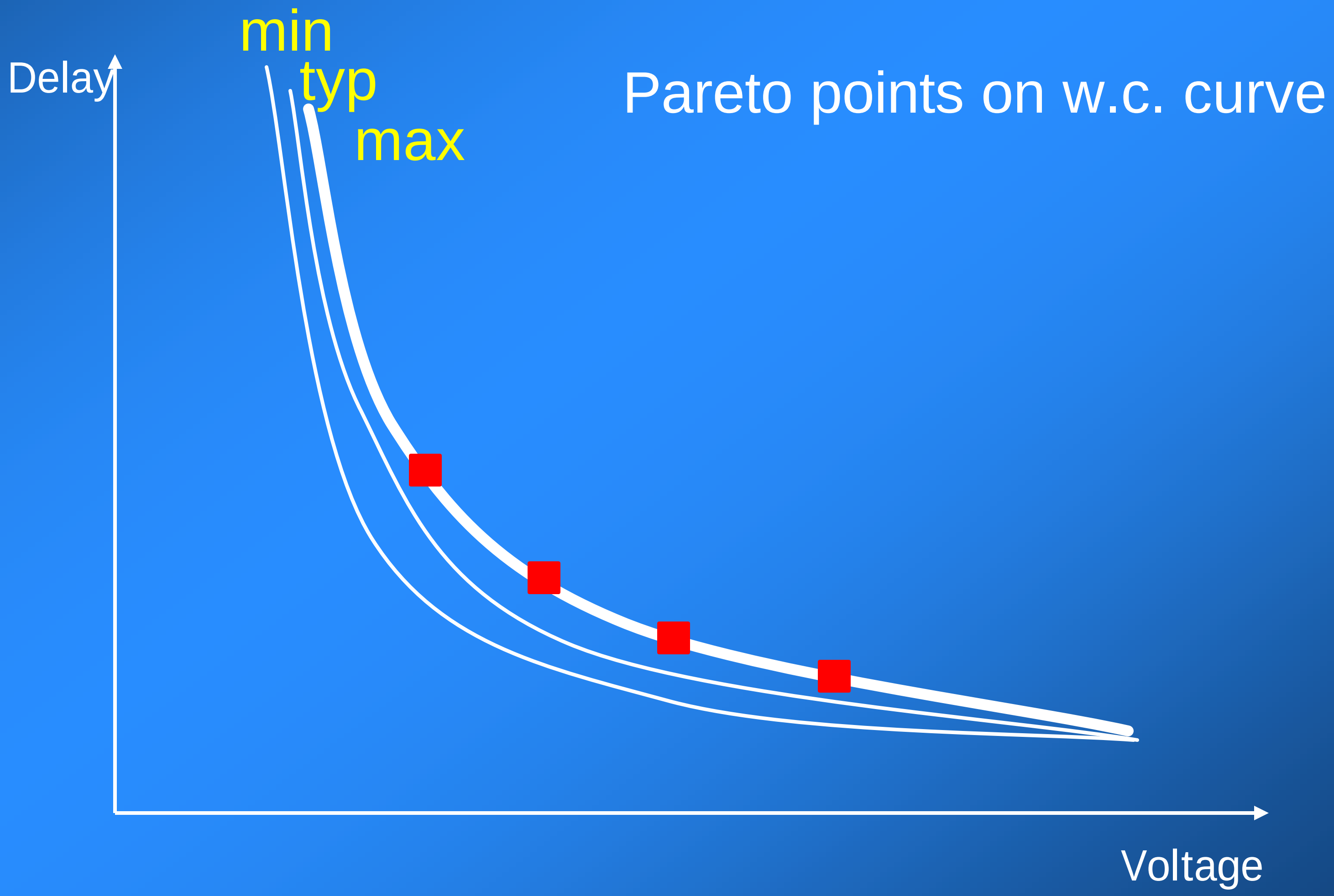
20 nm MOSFET (2010 ?)
50 Si atoms along the channel



4 nm MOSFET (2020 ?)
10 Si atoms along the channel

Design space exploration

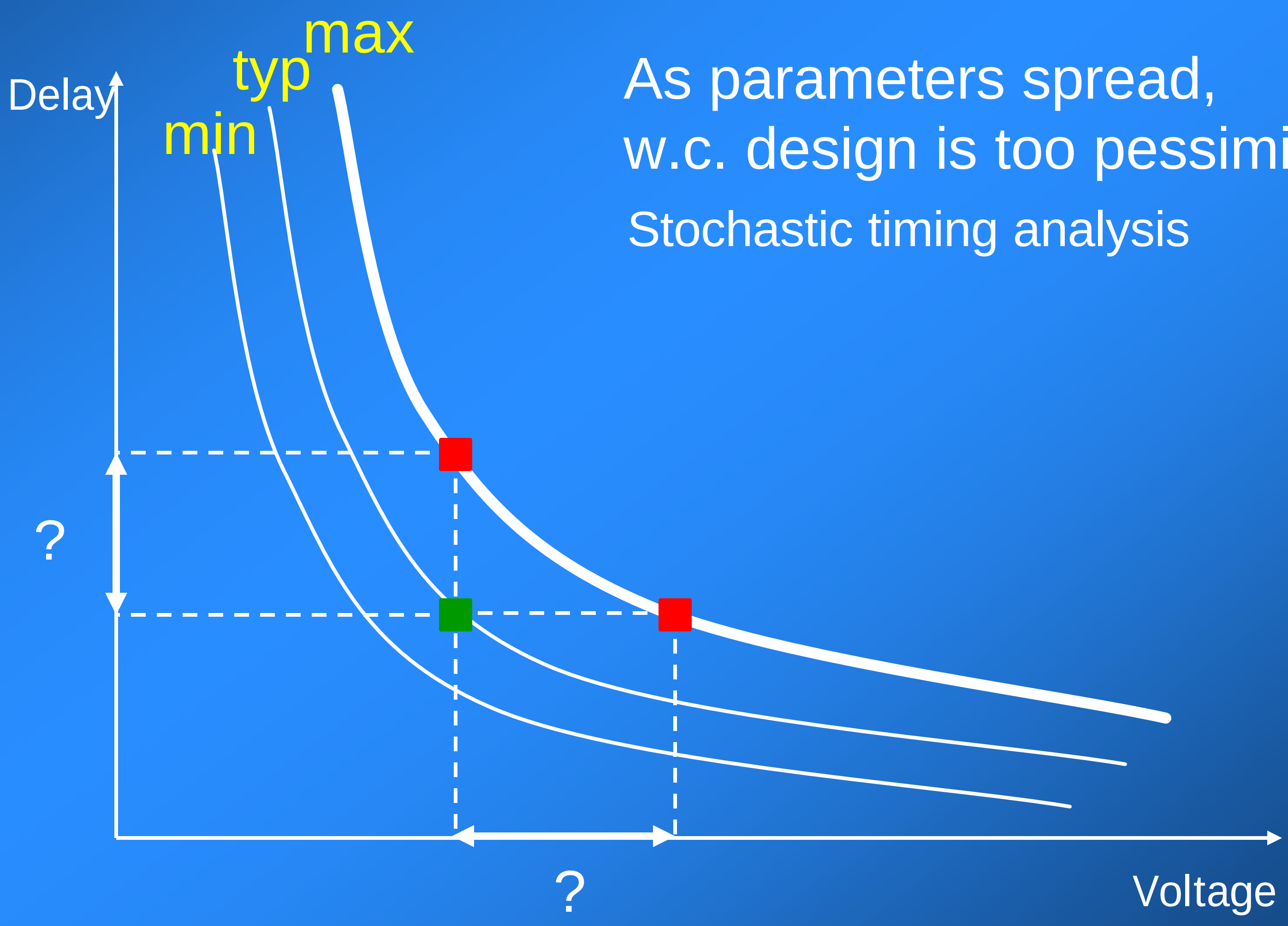
worst case analysis



Pareto points on w.c. curve

Adaptive design space

worst case analysis



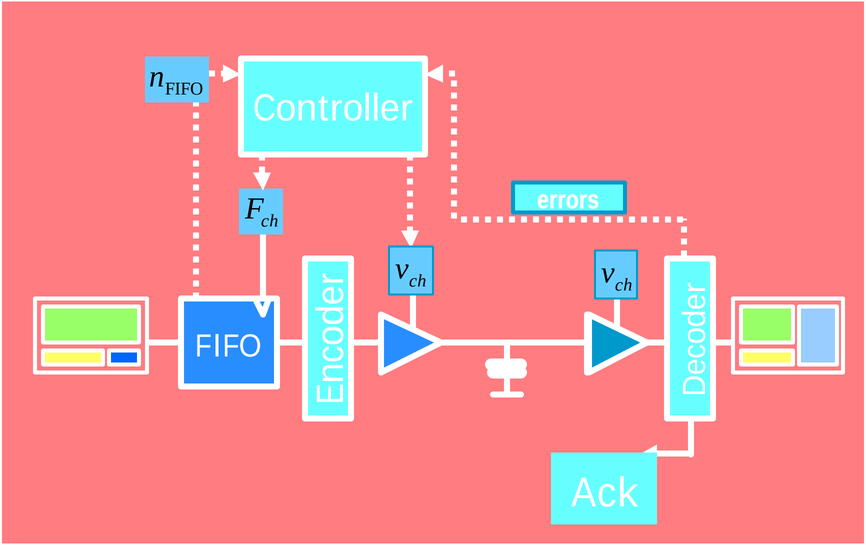
As parameters spread,
w.c. design is too pessimistic

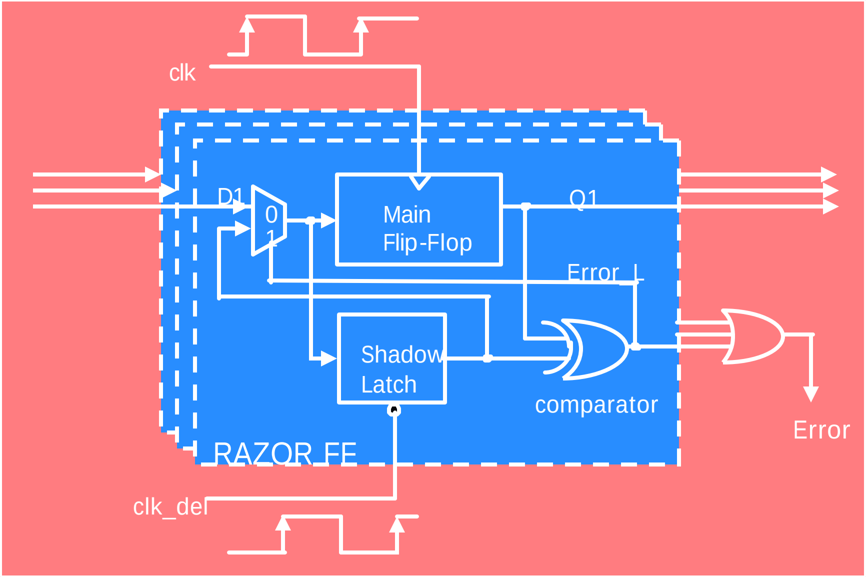
Stochastic timing analysis

De Micheli

Self-calibrating circuits

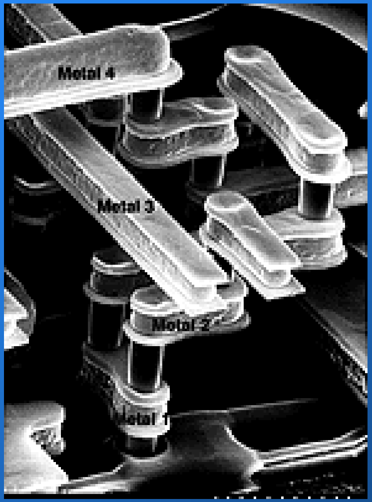
- Design self-calibrating circuits operating at the edge of failure
- Examples:
  - Dynamic voltage scaling of bus swings [Worm,Ienne –EPFL]
  - Dynamic voltage scaling in processors
    - Razor [Austin – U Michigan]
  - Dynamic latency adjustment for NoCs
    - Terror [Tamhankar -Stanford]
- Autonomic computing
  - Systems that understand and react to environment [IBM]



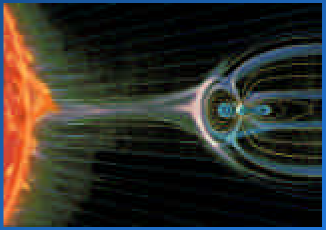


Dealing with transient malfunctions

- Soft errors
  - Data corruption due external radiation exposure
- Crosstalk
  - Data corruption due to internal field exposure
- Both malfunctions manifest themselves as timing errors
  - Error containment

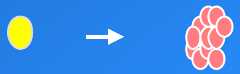


Sources of soft errors



Solar wind
(source)

Neutrons
@ sea level



Silicon

Ions

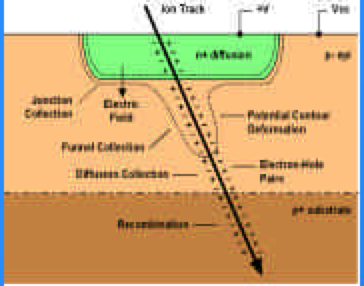
25Mg+a

28Al+p

24Mg+n+a

Si

Transistor



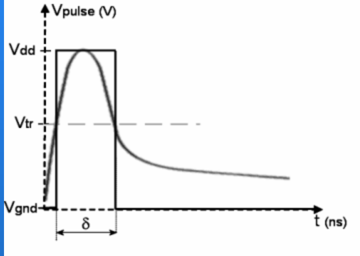
Electrons

e^-

e^-

e^-

Transient pulse

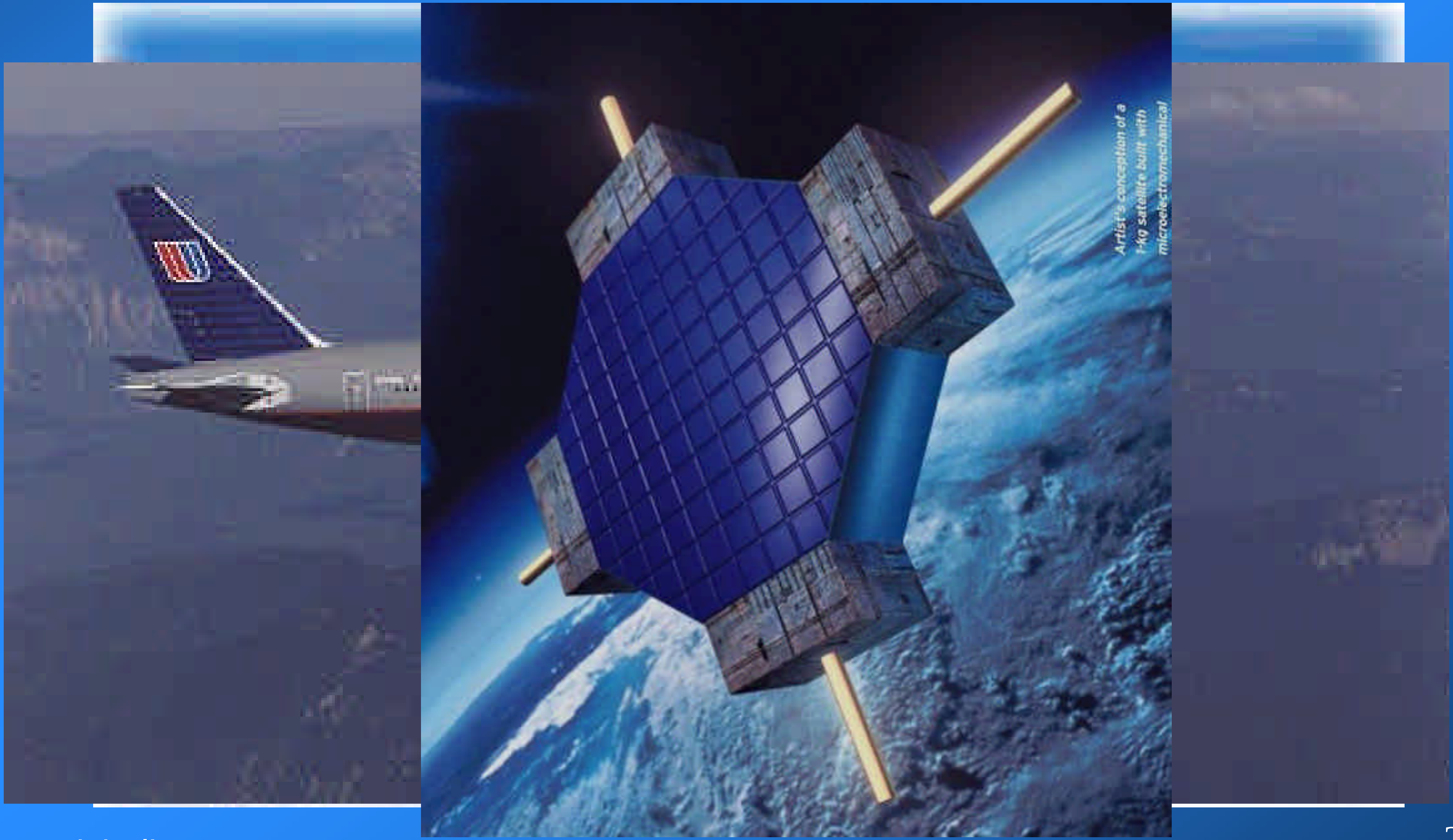


Strong (nuclear)
interaction

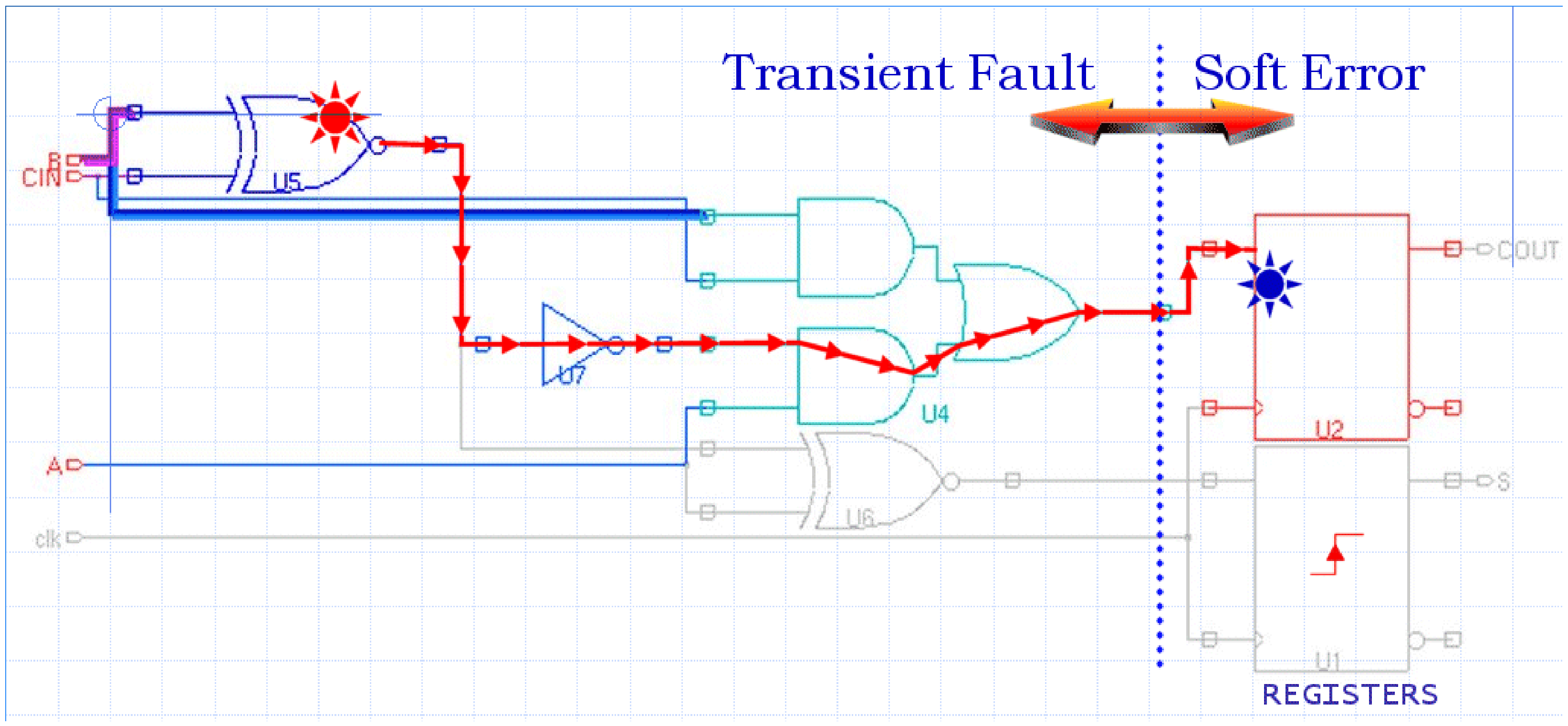
Electromagnetic interaction
(Silicon reaction)

Soft error rates

- Vary with altitude and latitude

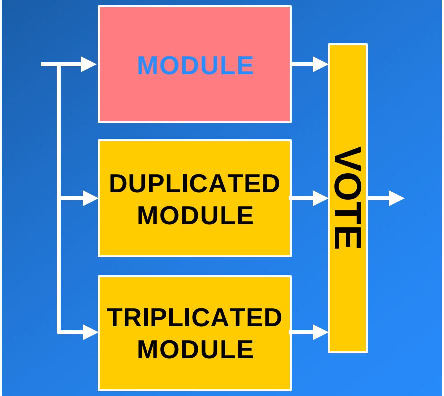


Propagation of soft errors



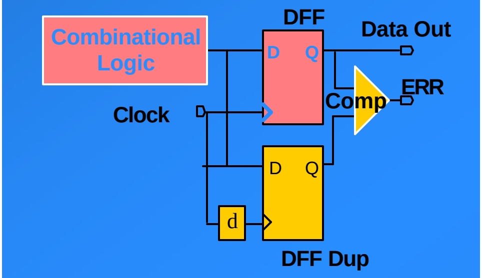
Logic protection techniques

Redundancy (TMR)



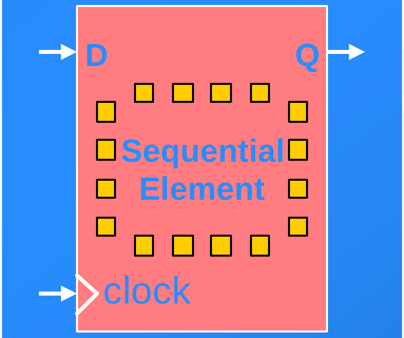
100% to 200% overhead

Detection + System Correction



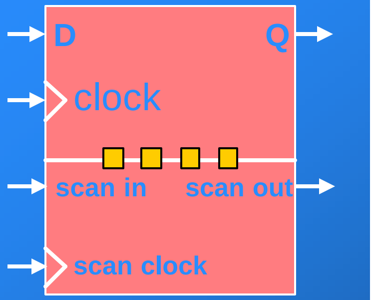
ERR signal used by system (Hardware or Software) to correct the error

Hardened Libraries



🟨 Protection Transistor embedded in the cell

Other Techniques



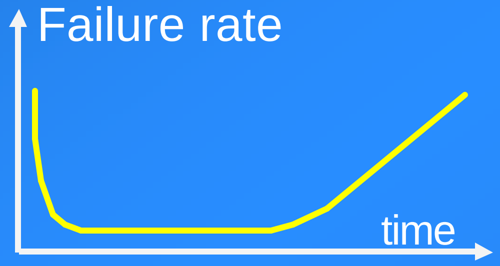
Scan Sequential Element

SCAN hardware used as DETECTION hardware in functional mode

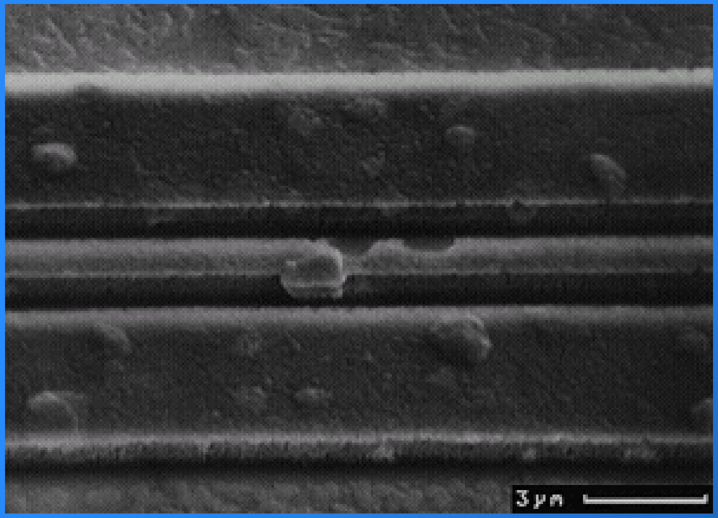
Redundancy ***Shielding*** ***Others***

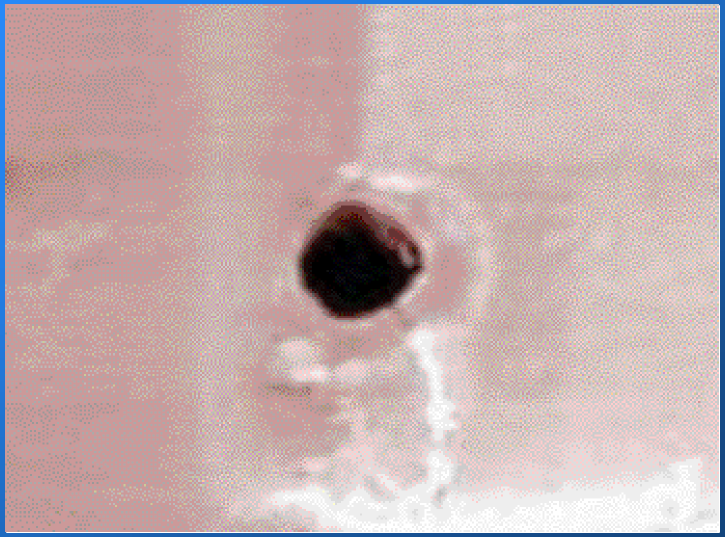
[Source IROC]

Aging of materials



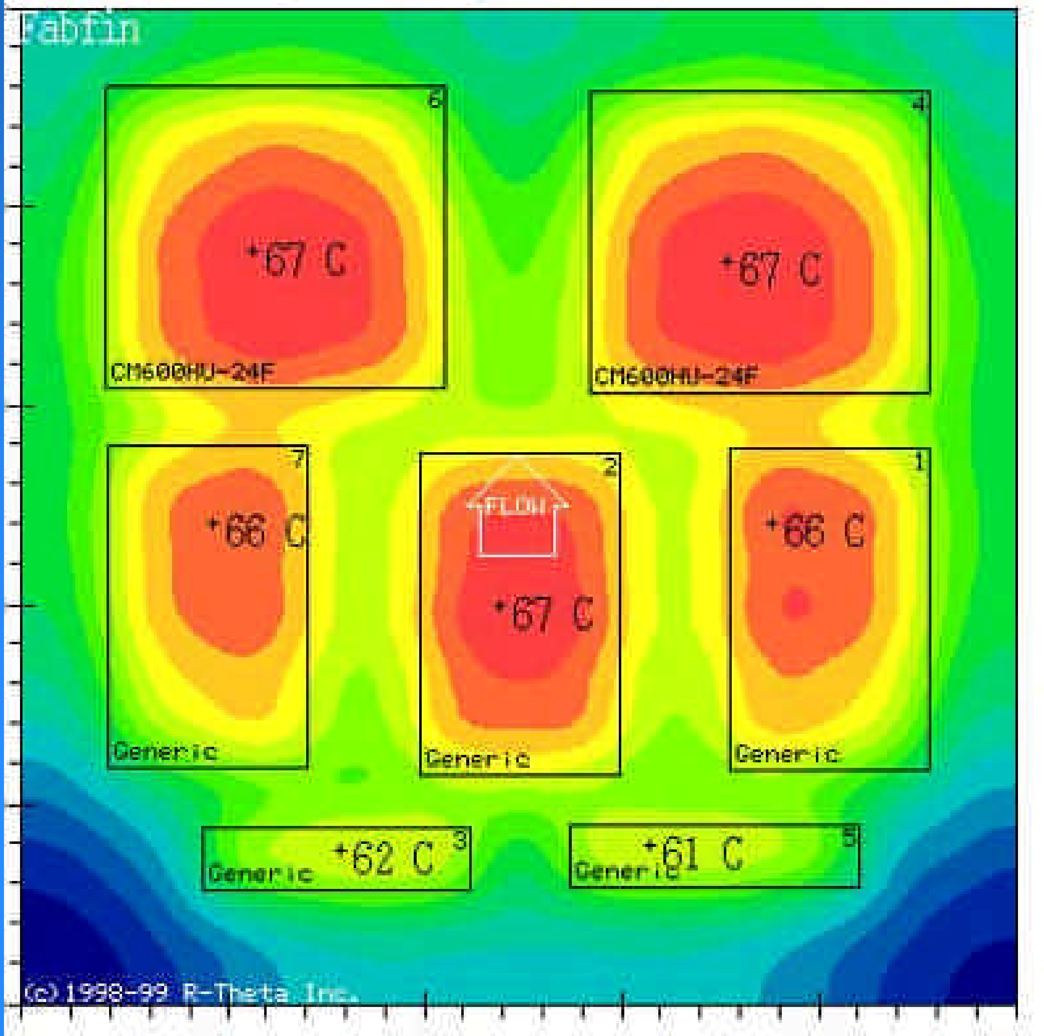
- Failure mechanisms
  - Electromigration
  - Oxide Breakdown
  - Thermo-mechanical stress
- Temperature dependence
  - Arrhenius law





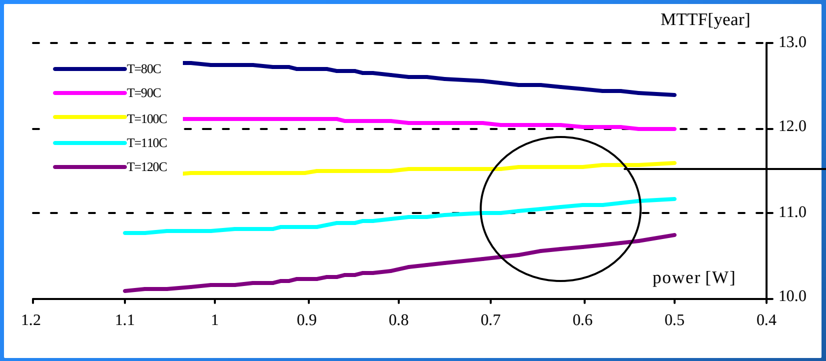
De Micheli

Thermal map: multiprocessosr



Thermal effects

- Keep chip as cool as possible
  - Reduce failure rates and power consumption
- In multi processor (core) system, power management shuts down idle cores
  - The temperature distribution will change in time
  - Thermal stress may increase
- Balance temperature reduction and thermal stress



Positive effect of DPM
EM and TTB dominate
High T – small gap

[Simuninic, UCSD]

Outline

- Nanotechnology and nanoelectronics
- Application domains and requirements
- Design technology challenges
- Architectures for nanocircuits
- Summary and conclusions

Architectural solutions

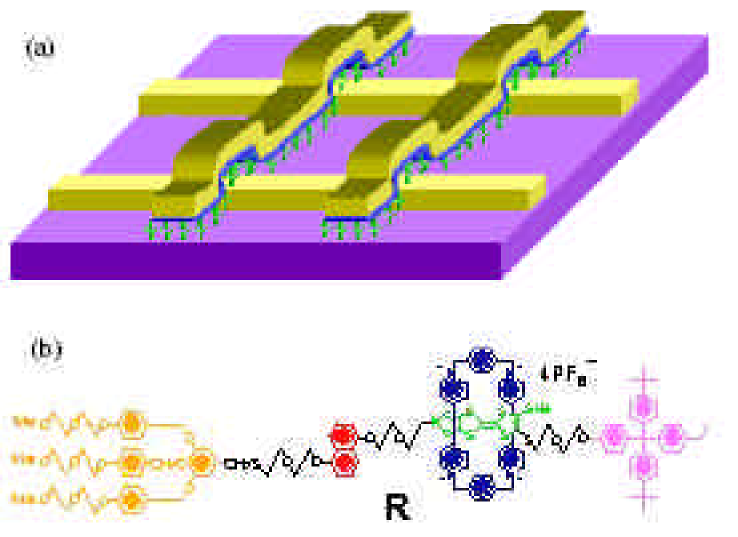
- Computation
  - Array logic has predictable wiring delays
- Storage
  - Array organization
  - Co-planar or 3D integration to cope with technology compatibility
- Communication
  - Structured and scalable communication means
  - Networks on Chips

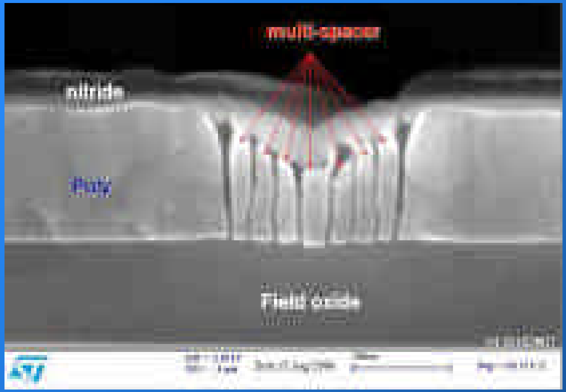
Array logic

- ROMs and LUTs
  - Realize combinational functions by using the input as an address.
  - For n input variables, table requires 2^n rows
  - For m outputs, table requires m columns
- PLAs
  - Realize combinational logic functions as *sum of products*
  - In CMOS, *SoPs* are realized as *NORs of NORs*
  - Two planes (input and output) with n+m columns
  - More compact than ROM, as *SoPs* are more compact than *sums of minterms*
- *PLAs differ from ROMs because they are designed to implement specific functions*

Nano storage arrays

- Nanowire Xbar functionalized with molecular switches



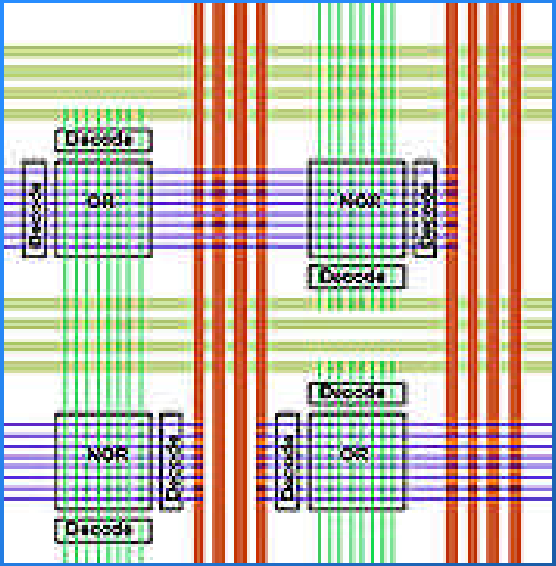


[Source: Williams, HP]

[Source: Cerofolini, ST]

Nano PLAs

- Nano planes connected to microwires
    - Registers/latches can also be embedded
- Regular layout
    - – Wiring delay are predictable
    - – Regular structure support redundant logic design
        - Additional rows/columns
- Planes can be sparse or made sparse
    - – Low active cross-point density



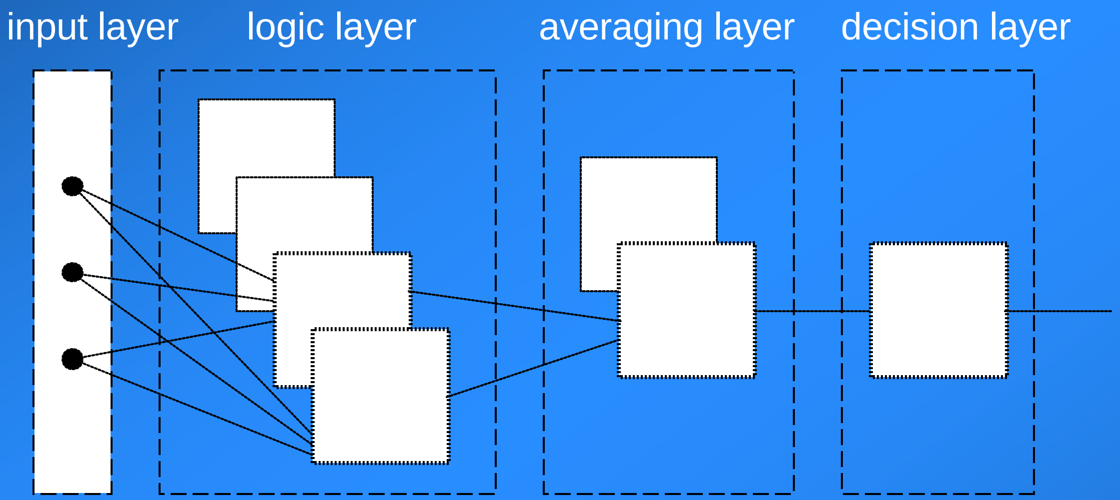
Reliable nano-design:

Logic synthesis

- Device-level redundancy
  - Duplicate transistors/switches to achieve broader coverage
  - Cover Boolean minterms more than once
- New paradigm for testing
  - Circuit with faulty devices may still work
  - Exploit, rather than remove, redundancy
- Objective is enhancing overall yield

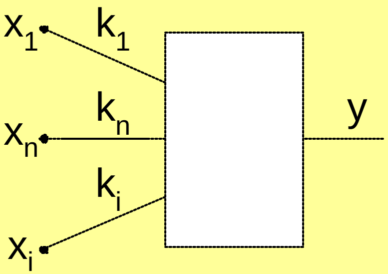
Reliable nano-design

Weighted averaging



input layer logic layer averaging layer decision layer

Fault tolerant architecture based on multiple layers

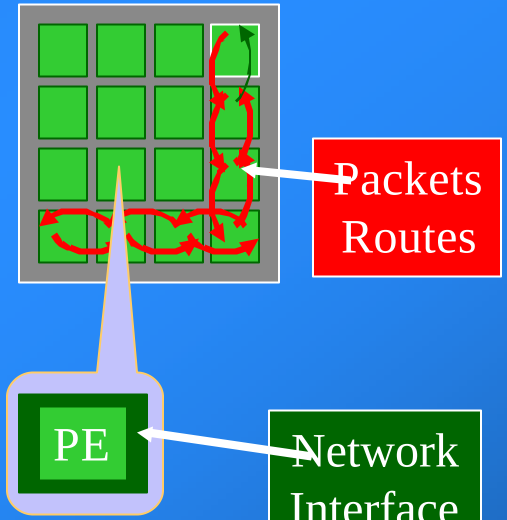


$$y = \frac{V_{fs}}{\sum_i k_i} \sum_i k_i \; x_i$$

General weighted averaging and re-scaling function used in the third layer

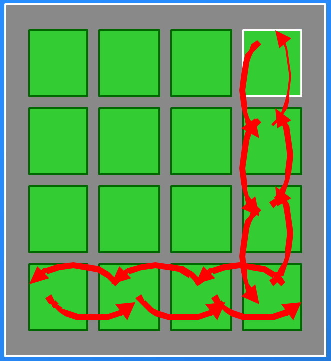
[Source: Leblebici and Schmid]

On-chip networks



- Provide a structured methodology for realizing on-chip communication
  - Modularity
  - Flexibility
- Cope with inherent limitations of busses
  - Performance and power of busses do not scale up
- Support reliable operation
  - Layered approach to error detection and correction

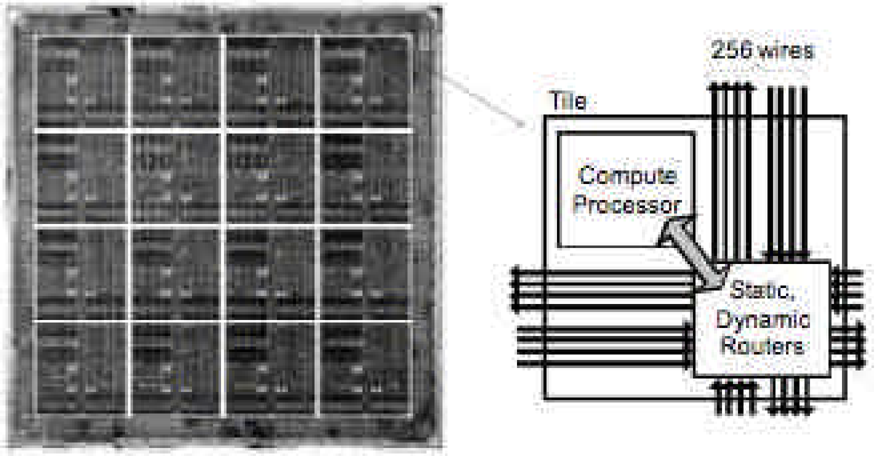
Hierarchical circuit view



- System:
  - Modules (PE) are processors

- Modules:
  - Submodules are nano-arrays
- The NoC provides the communication means
- What is the right functionality for a module in a nano-environment?
  - Look up table (FPGA)
  - Finite state machine (with stack)
  - Program state machine
  - Data-flow element (possibly syncronous)
  - Processor (ARM)

NoC multi-processors: the RAW architecture

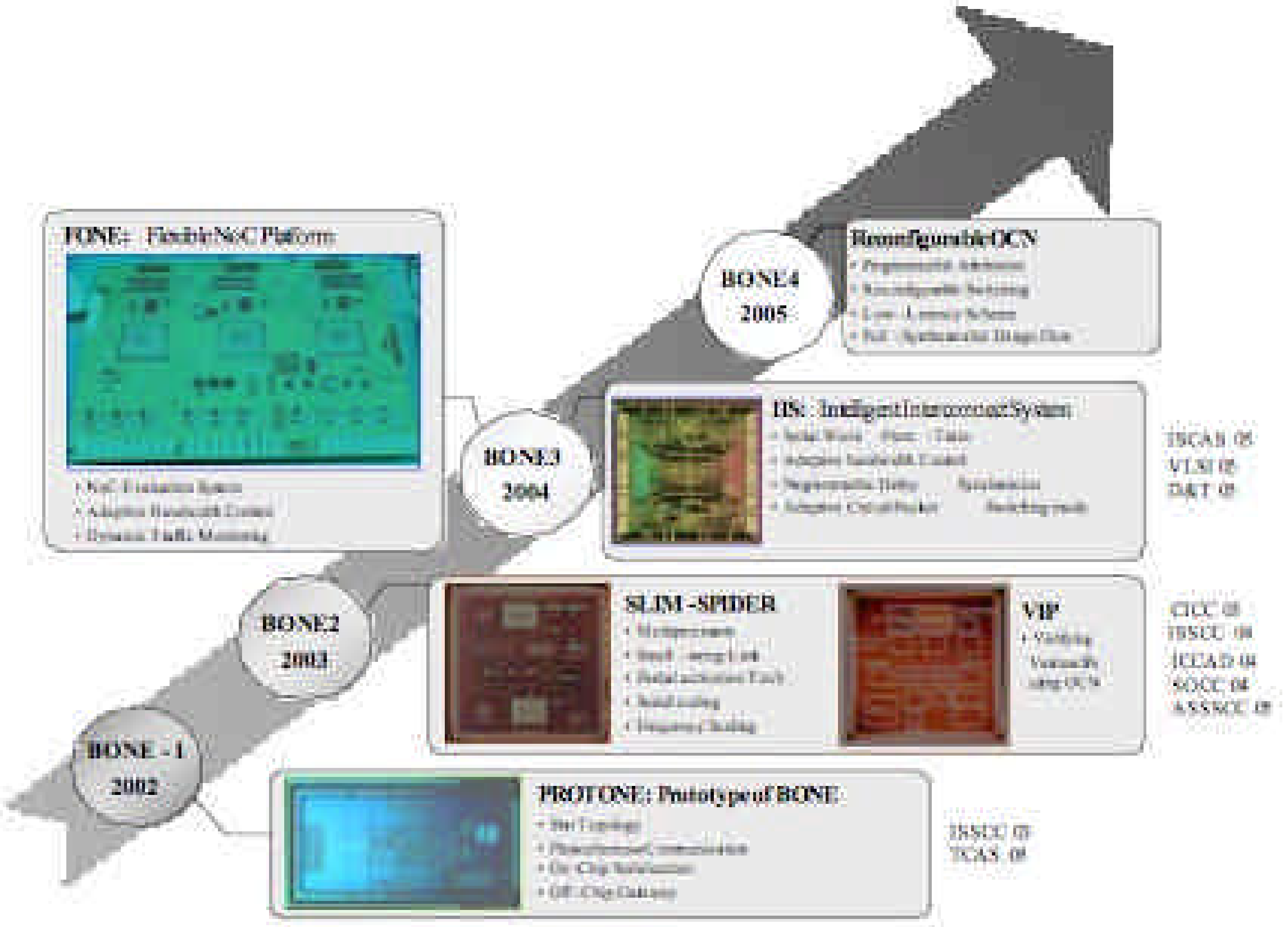


- Fully programmable SoC
  - Homogenous array of tiles:
    - Processor cores with local storage
    - Each tile has a router

[Agrawal MIT]

- The raw architecture is exposed to the compiler
  - Cores and routers are programmable
  - Compiler determines which wires are used at each cycle
  - Compiler pipelines long wires

The BONE roadmap



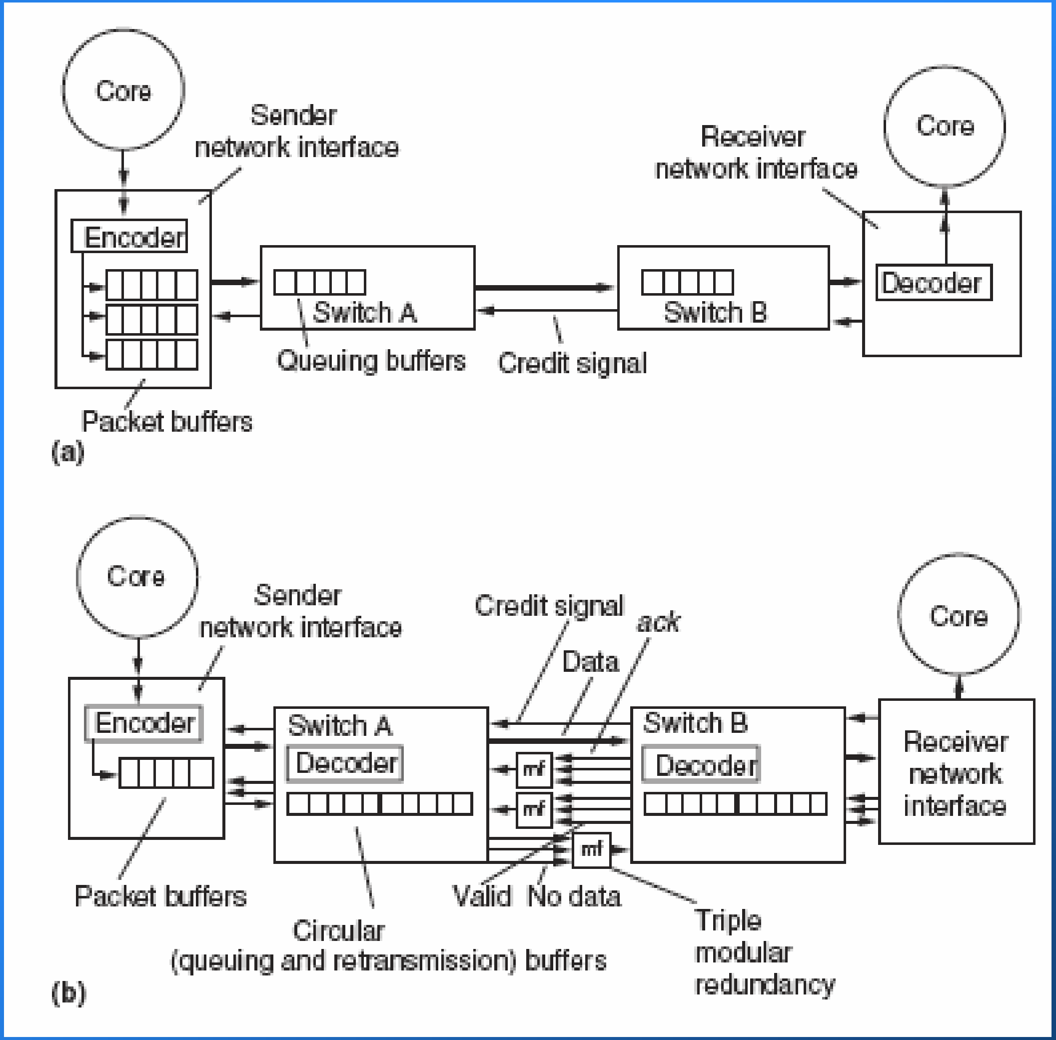
[Source: KAIST]

Metrics for NoC design

- Low communication latency
  - Streamlined control protocols
  - Data and control signals can be separate
- High communication bandwidth
  - To support demanding SW applications
- Low energy consumption
  - Wiring switched capacitance dominates
- Error resiliency
  - To compensate/correct electrical-level errors
- Flexibility and programmability

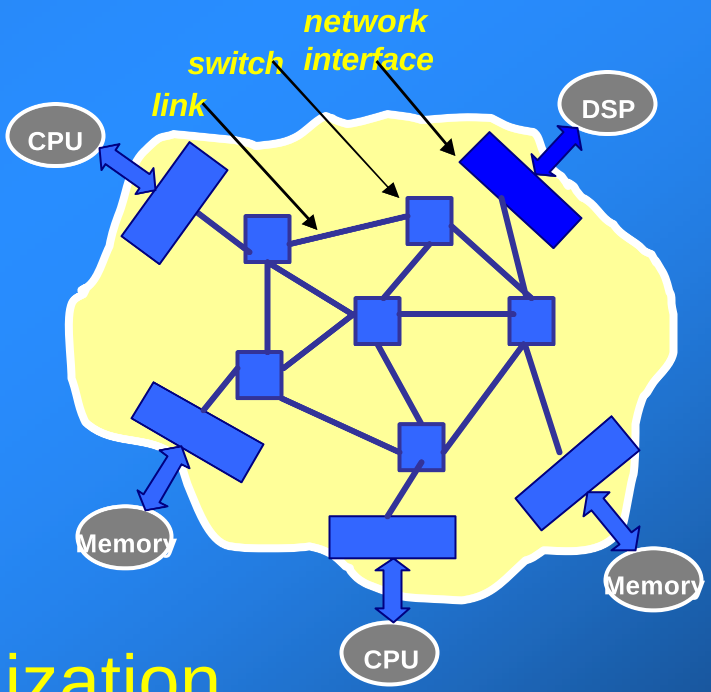
Error resiliency

- Several implementation styles:
  - – Local link-level
    - ECC in switches
  - – Global *end to end*
    - ECC at core interfaces
  - – Transaction level
    - Software approach



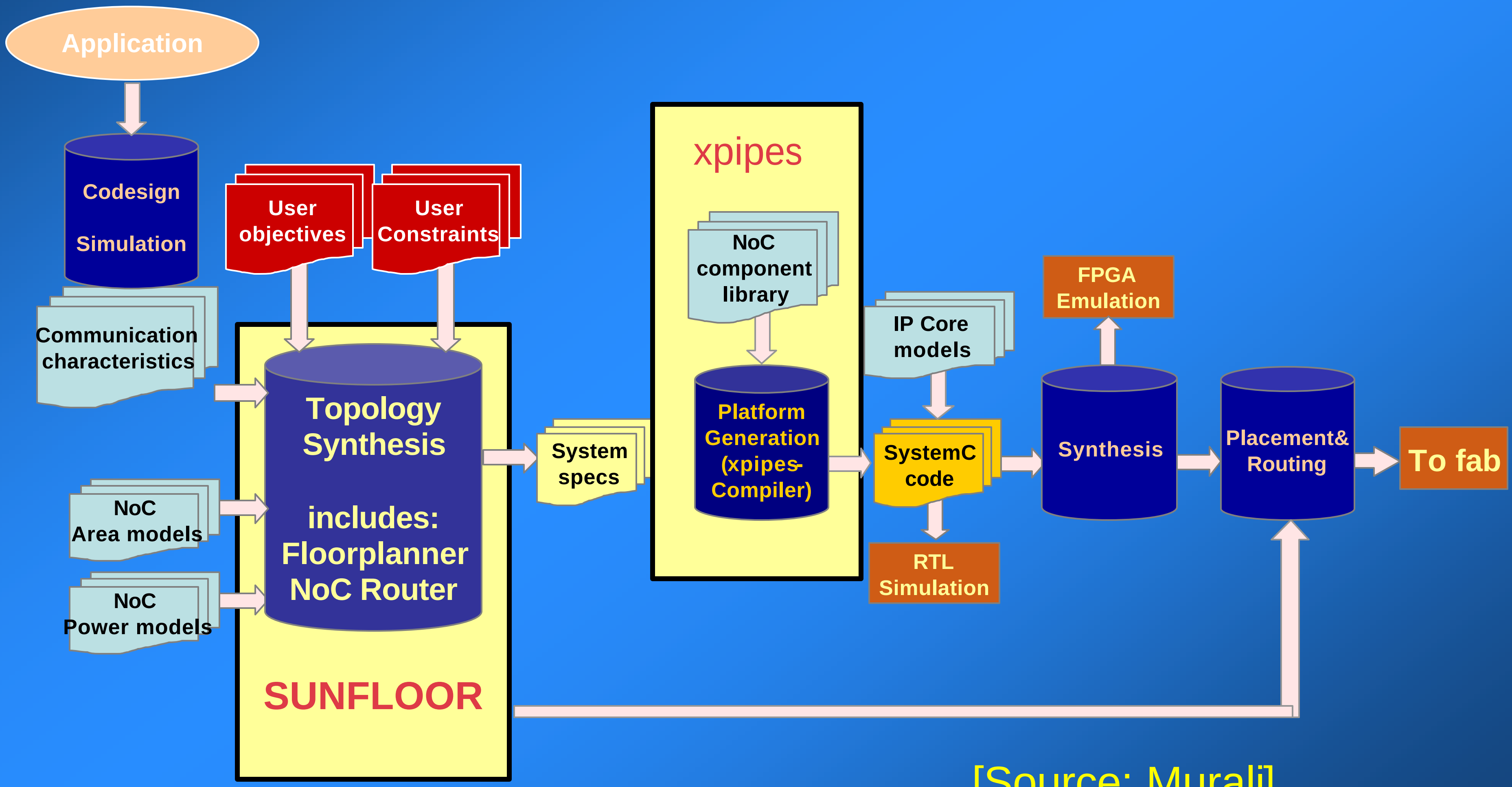
Flexibility in NoC design

- NoCs have modular structure
  - Core interfaces
  - Switches/routers
  - High-speed links
- NoCs can be tailored to applications
  - Topology selection
  - Switch/link sizing
  - Protocols
- Several parameters for optimization and a large design space
  - NoC synthesis and optimization



De Micheli

Netchip tool flow

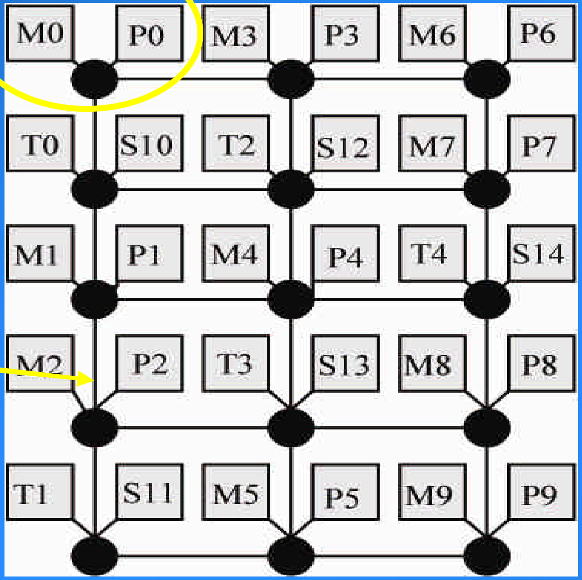


[Source: Murali]

SUNFLOOR vs. manual design

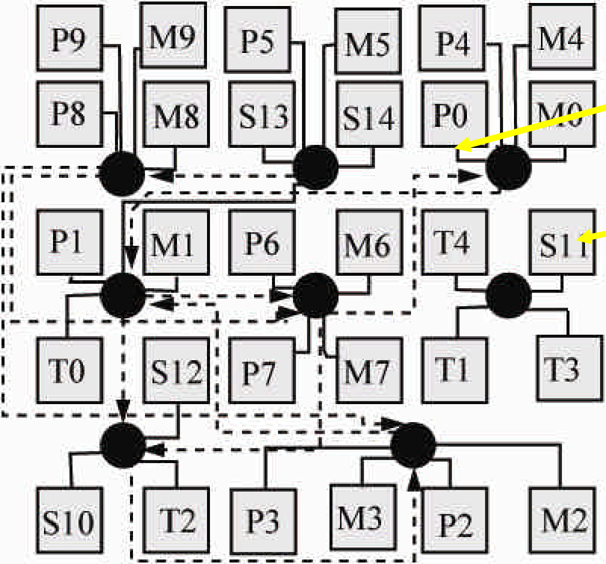
multimedia chip with 30 cores

Processor-memory cluster



Bi-directional links

Hand-mapped topology



Uni-directional links

Bi-directional links

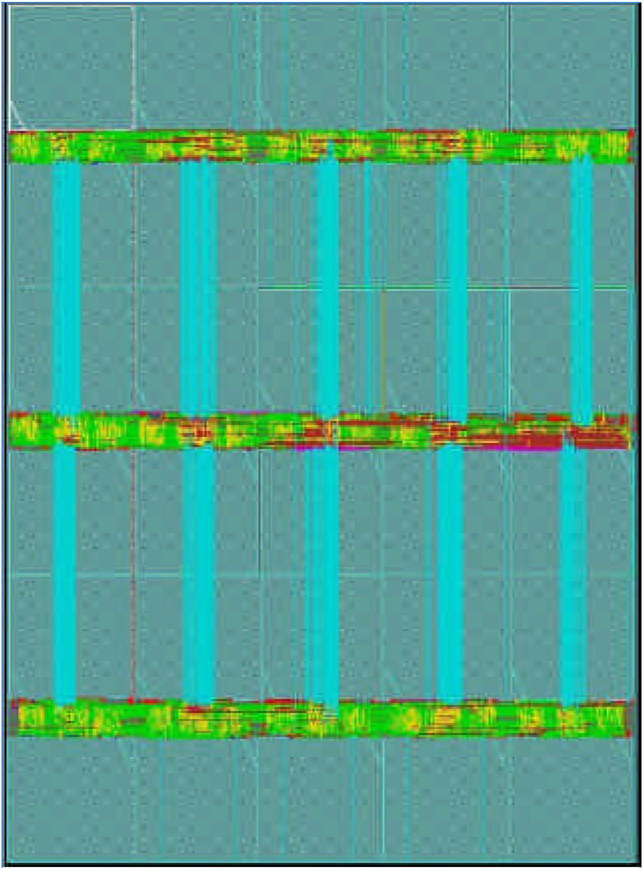
SUNFLOOR custom topology

P-processors, M-private memories,

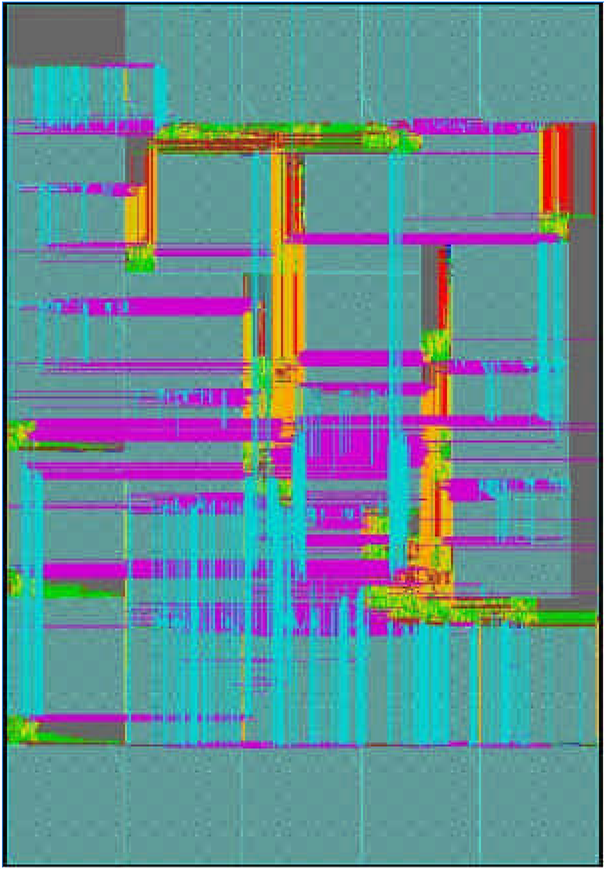
T-traffic generators, S-shared slaves

Design layouts

From Cadence
SoC Encounter



Hand-design (custom mesh)



SUNFLOOR Design

SUNFLOOR vs. manual design

Manual design:

- Topology: 5x3 mesh (15 switches)
- Operating frequency: 885 MHz (post-layout)
- Power consumption: 368 mW
- Floorplan area: 35.4 mm^2
- Design time: several weeks
- 0.13 µm technology

SUNFLOOR:

- Topology: custom (8 switches)
- Operating frequency: 885 MHz (post-layout)
- Power consumption: 277 mW **(-25%)**
- Floorplan area: 37 mm^2 **(+4%)**
- Design time: **4 hours design to layout**
- 0.13 µm technology

- Benchmark execution times comply with application requirements and, in fact, are even 10% better on the SUNFLOOR topology.

Putting it all together …

- ULP demands low-voltage operation
- High throughput requires parallel computation
- High reliability is achieved by redundancy
- New paradigm for computation:
  - Array-based computation (e.g., RAW)
  - Array-oriented communication (NoC)

System implications

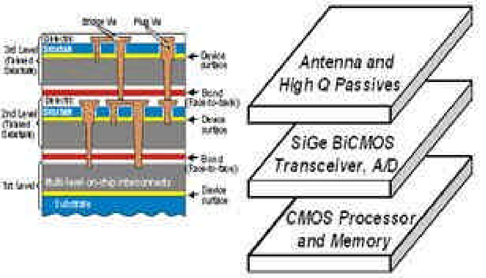
the sw side

- Modularity, redundancy, regularity
- Cellular approach to computation
  - Massive parallelism
  - Stream computing
- Programming paradigms:
  - Expose both computation and communication to Sw compiler
  - Designer need to think “parallel” to exploit these architectures at best

System implications

the package side

- Computing systems are limited by planar geometry:
  - Space usage and wiring
- Electrical constraints (e.g., voltage) and manufacturing constraints limit heterogeneous planar integration
- Die stacking allows designers to superimpose:
  - Computing arrays
  - Memory arrays
  - Analog and RF circuitry



- 3D on-chip networks provide effective and reconfigurable means of realizing communication

Outline

- Nanotechnology and nanoelectronics
- Application domains and requirements
- Design technology challenges
- Architectures for nanocircuits
- Summary and conclusions

Summary and conclusions

- Novel nanotechnologies will provide us with unprecedented levels of functional integration and performance
- High-performance, ultra low-power, reliable circuits will be required by distributed embedded systems
- Novel architecture will be needed to leverage the potentials of nanotechnology and satisfy system requirements
  - Parallel array-oriented computational logic blocks, with built-in fault tolerance and predictable delays
  - On chip networks to provide units with structured communication
  - 3-dimensional packages to support integration of different technologies
- Novel design tools and methodologies, to support array logic and NoC design and cope with variability, reliability and thermal issues