



Self-Healing ICs: Technologies from FCRP (C2S2 & GSRC) Research

Subhasish Mitra

Robust Systems Group

Departments of Electrical Eng. & Computer Sc.
Stanford University

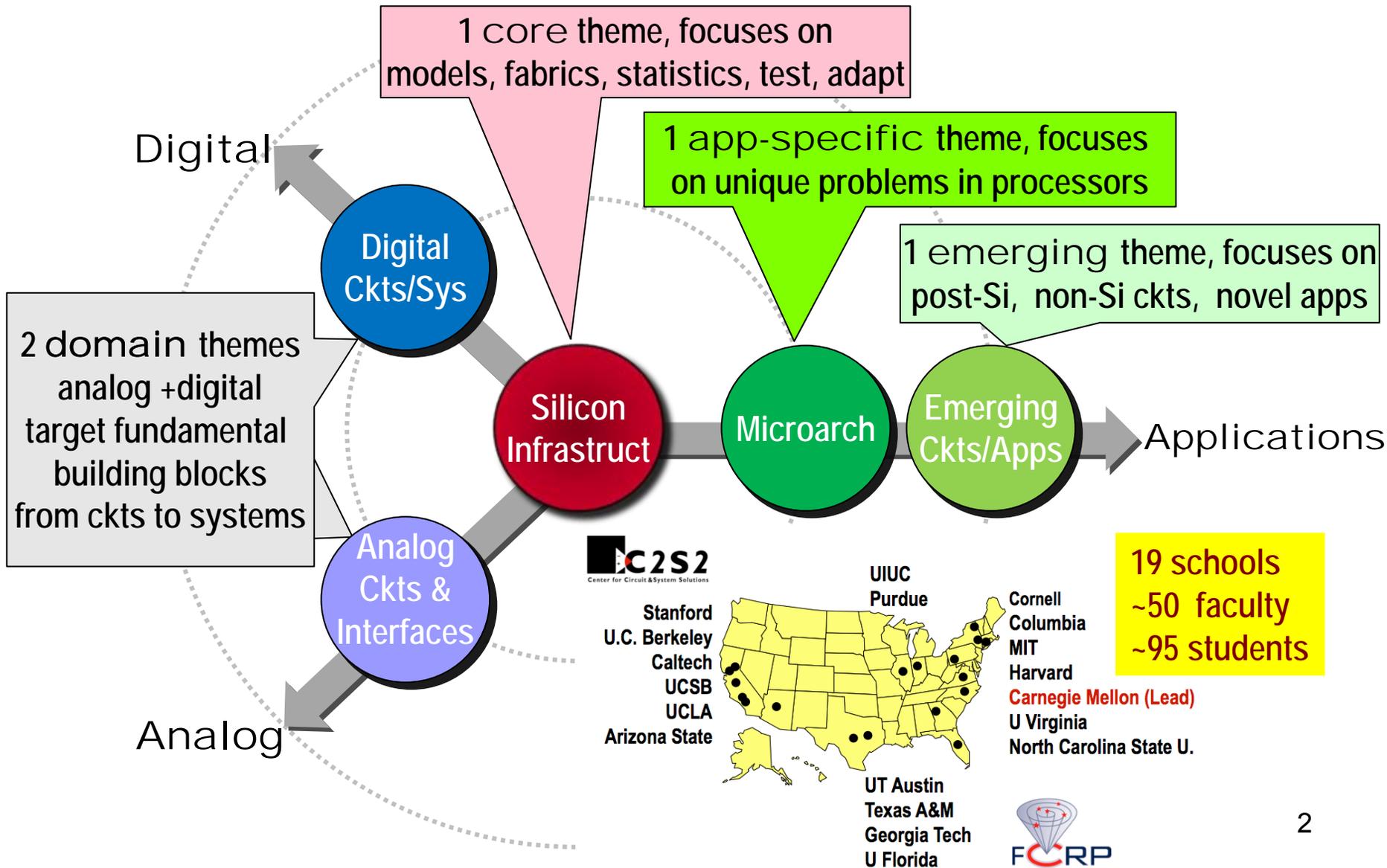
Email: subh@stanford.edu

www.stanford.edu/~subh

Acknowledgment: FCRP C2S2 & GSRC Researchers

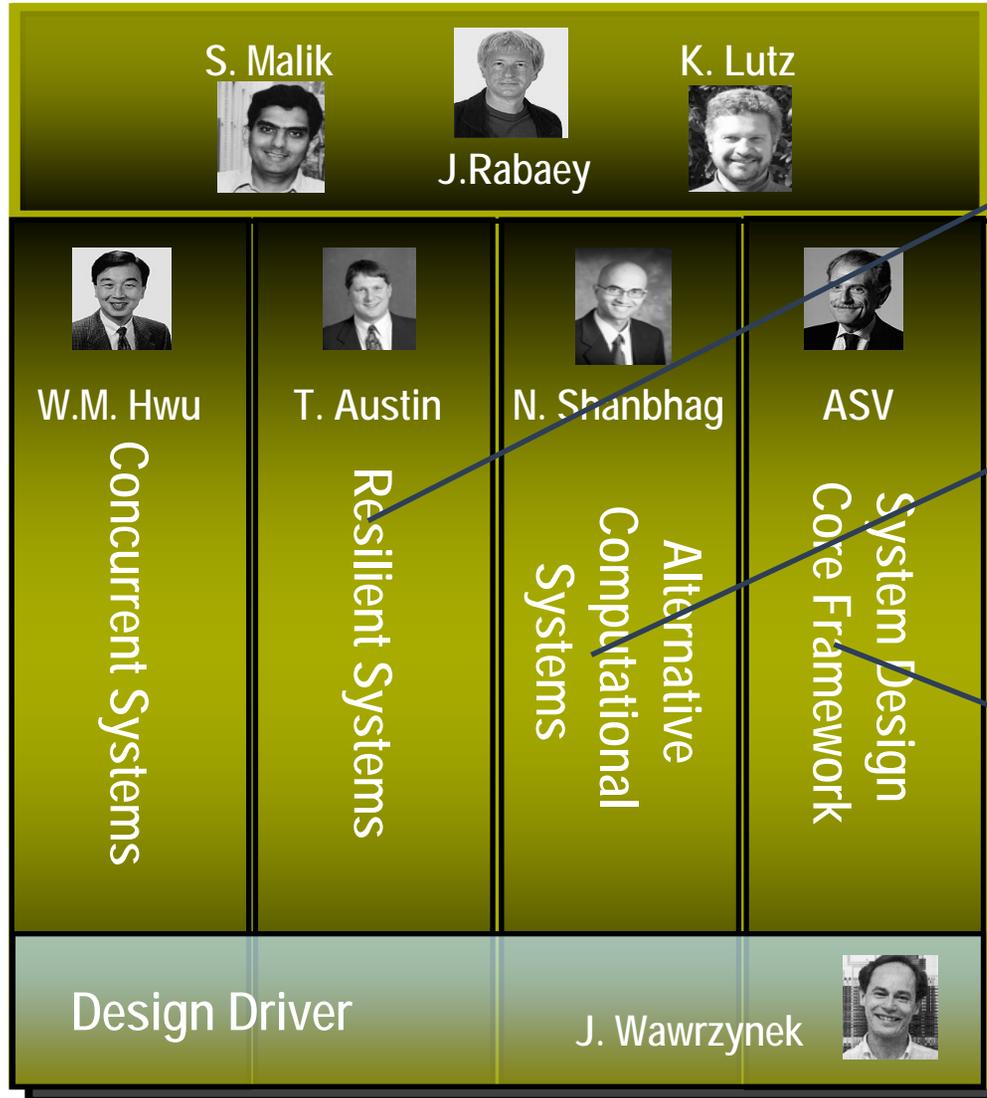
FCRP Center: C2S2

Center for Circuit and System solutions



FCRP Center: GSRC

Gigascale Systems Research Center



How to make “traditional architectures” reliable with minimal overhead

Depart from the traditional (and determinism)

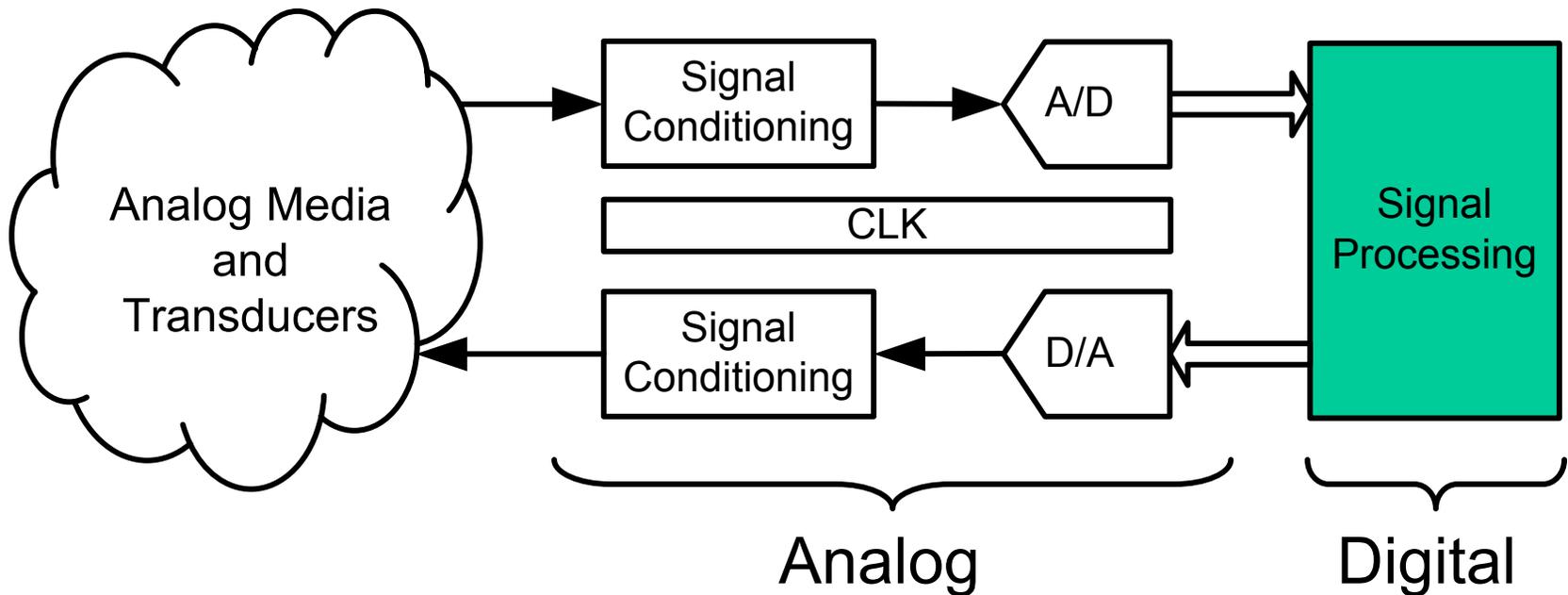
Computational models that deal with statistics

The (Bumpy) Road Ahead for Silicon



The lifetime of silicon will be determined by how cheaply and effectively we can make the substrate reliable.

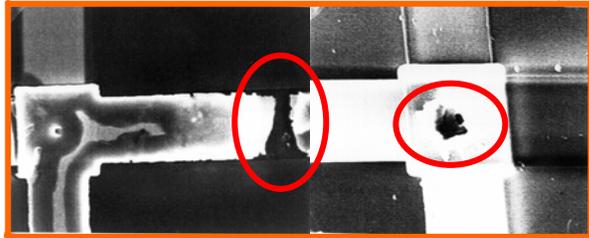
Typical Mixed-Signal System



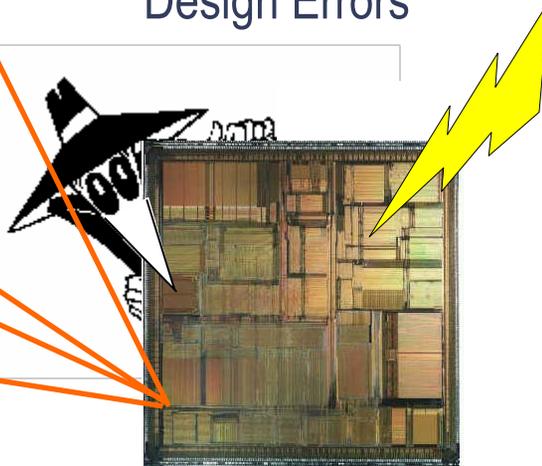
- Focus of this talk: Enabling technologies for
 - Self-healing sub-systems
 - Digital
 - Analog

Key Reliability Threats

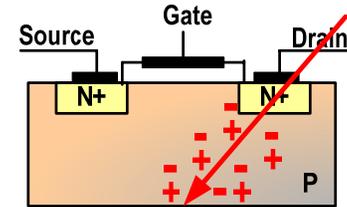
Silicon Defects
(Manufacturing defects and device aging)



H/W and S/W
Design Errors

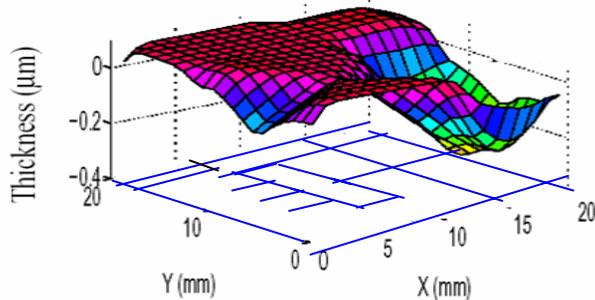


Transient errors
Cosmic Rays & Alpha Particles
(Exponential increase at chip-level)

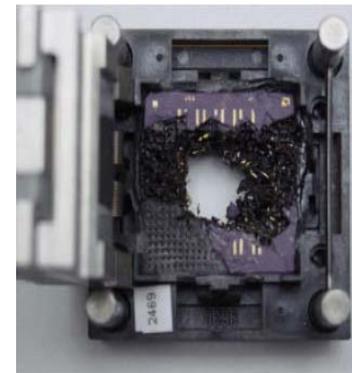


Parametric Variability
(Uncertainty in device and environment)

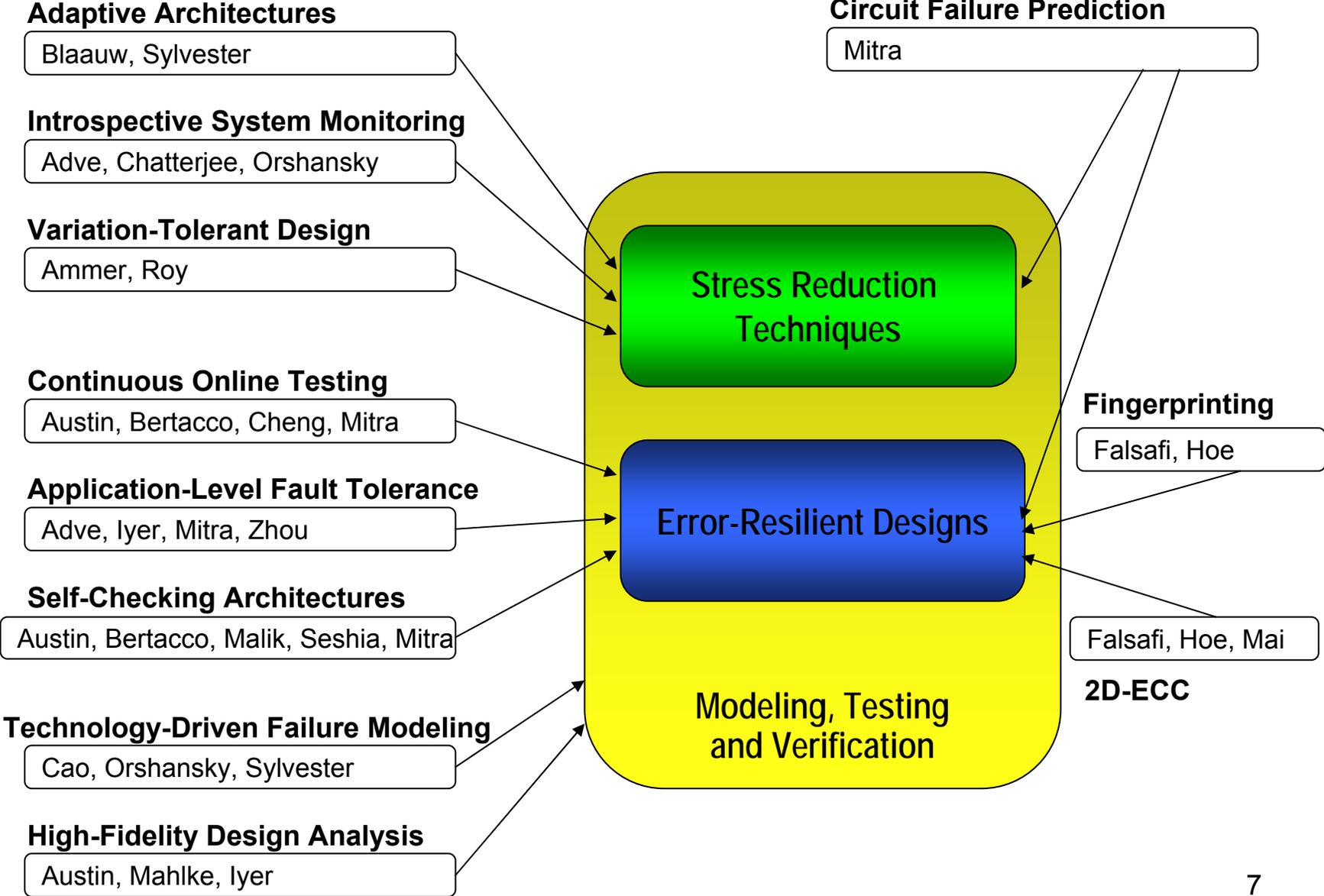
Intra-die variations in ILD thickness



Manufacturing Defects
(Inefficient Burn-in Testing)



Resilient Digital System Design (GSRC & C2S2)



Stress Reduction Techniques

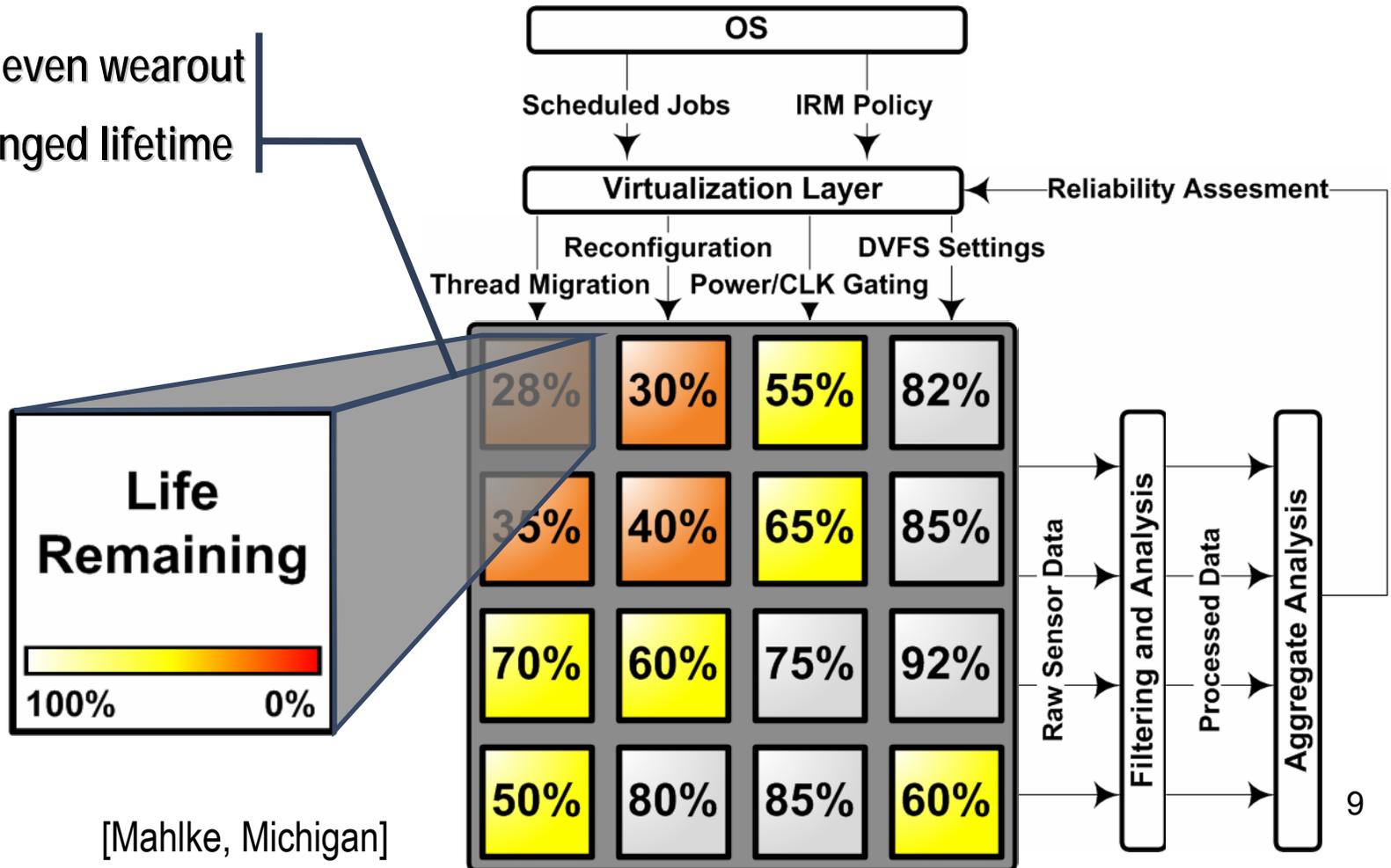
- Sensors monitor key aspects of system
 - Properties that accelerate failure mechanisms
 - Temperature, voltage, humidity, etc...
- System-level intelligence to reduce stress
 - Temperature control, resource assignment
 - Actively manage reliability trade-offs
- Tuning/repair to alleviate reliability threats
 - NBTI, in-field clock tuning



Proactive Systems Can Extend Lifetimes

- CMP workload variations → non-uniform aging
- Intelligent proactive reliability management
 - Guided by low-level sensors

- More even wearout
- Prolonged lifetime



[Mahlke, Michigan]

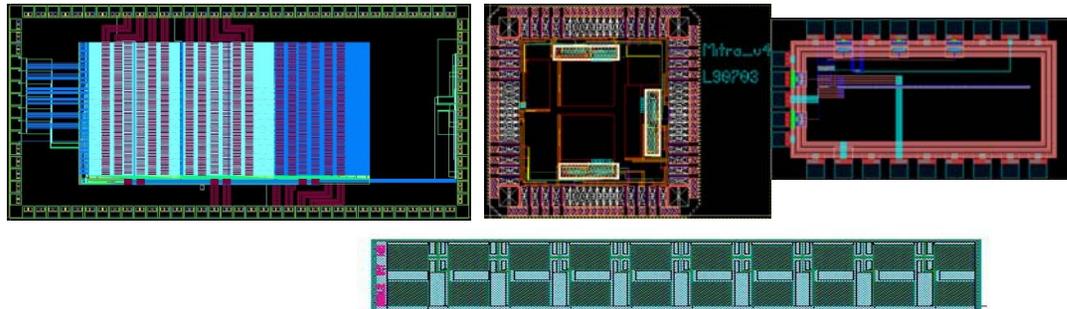
Failure Prediction for Self-Healing Systems

- Predict failures BEFORE errors appear

😊 Applicability: Transistor aging, early-life failures, variations

Failure Prediction	Error Detection
Before errors appear	After errors appear
+ No corrupt data & states	– Corrupt data & states
+ No errors	– High error rates difficult
+ Self-diagnosis	– Limited diagnosis

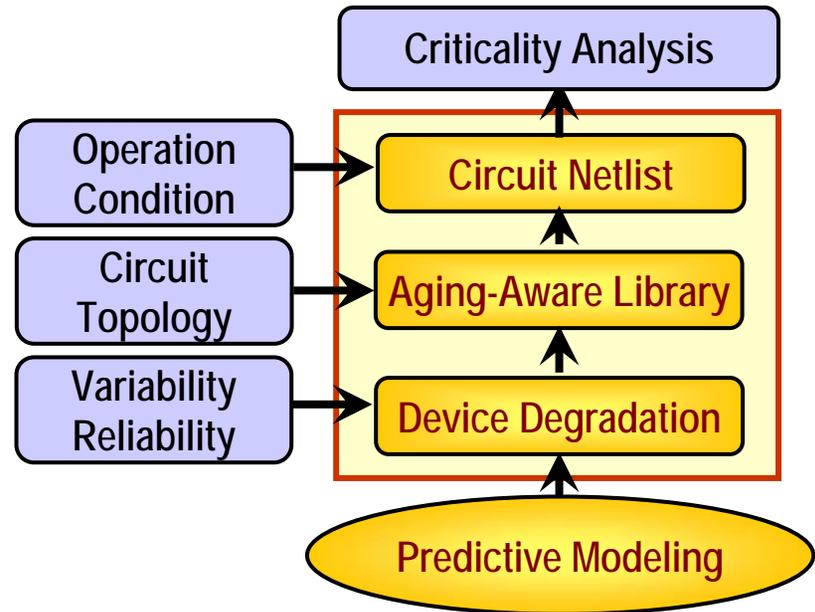
Both can be efficiently combined



- Effectiveness demonstrated
 - 90nm test chips
 - 45nm experiments on-going
- [Mitra, Stanford]

Identify Vulnerable Units to Reduce Costs

- Small portion (<10%) vulnerable
 - Identify critical units
 - Maximize system reliability
 - Minimize cost
- Predictive failure modeling
 - Hierarchical circuit analysis
 - Efficient hotspot identification



Demonstration (with S. Mitra, Stanford)

- 90nm Ethernet controller

Area overhead of failure prediction circuits	
w/o criticality analysis	10 %
w/ criticality analysis	2 %

Error-Resilient Systems

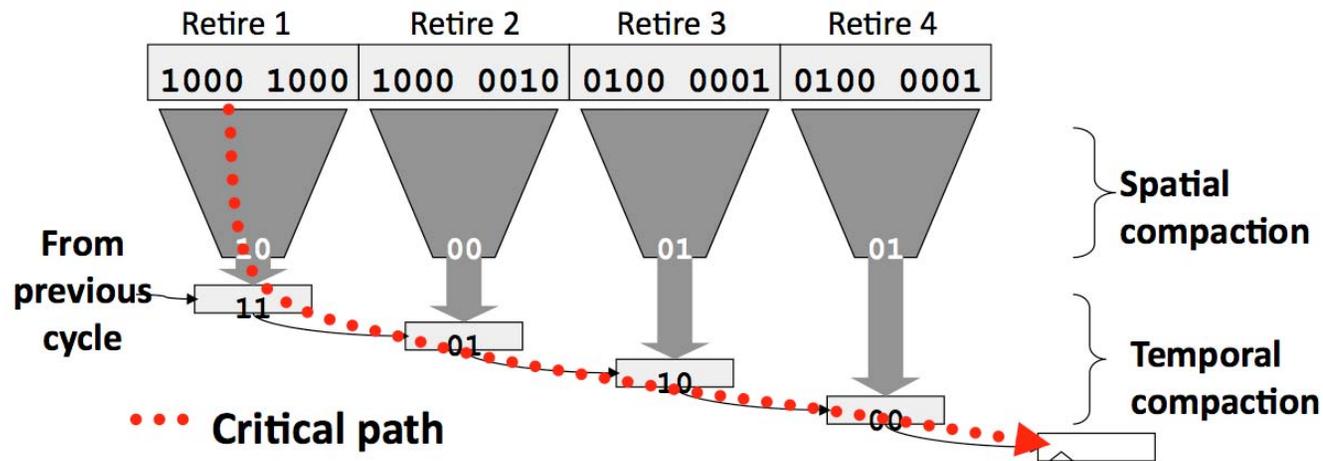
- Error detection technologies
 - Fingerprinting, online testing, continuous verification
- Fault diagnosis
 - Fine-grained testing, on-line testing

- System
– Mi
tol
- Phys
- **Resil**



Processor Error Detection with Fingerprinting

- Problem: Efficiently deduce “something is *wrong*”
- Challenge: Detect quickly
- Approach: *Architectural fingerprints* (Efficient “hash”)
 - Processor execution tracked
 - Compared among cores or w.r.t. golden reference



Compaction scheme: “X-compact + comb. MISR

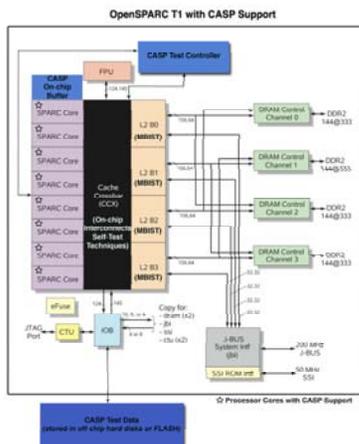
Result: Aliasing \approx *ideal CRC* (almost no false matches)

Cost: Cost/latency of 64-bit CLA adder \rightarrow $<4\%$ HW for OpenSparcT1

Online Self-Test Eliminates H/W Replication

- Traditional Logic BIST – not scalable
- CASP: **C**oncurrent, **A**utonomous, **S**tored **P**atterns
 - Extremely thorough tests
 - No system-level downtime
 - Virtualization-assisted CASP
- **Negligible system performance impact**

OpenSPARC T1 with CASP (8 cores)



Extremely thorough: 99.5% stuck, 96% Transition, 94% True Time
Inexpensive Flash: 60MB
Small area impact: 0.01%

ARM Multi-core System Results
Virtualization Assisted CASP

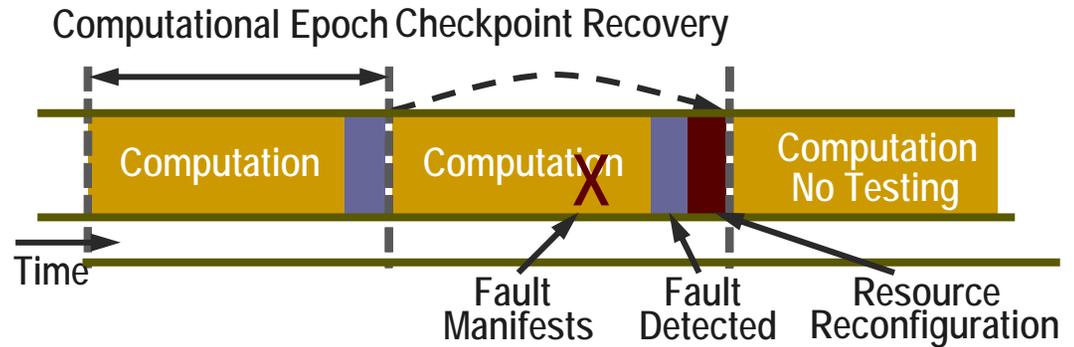


Negligible system performance impact

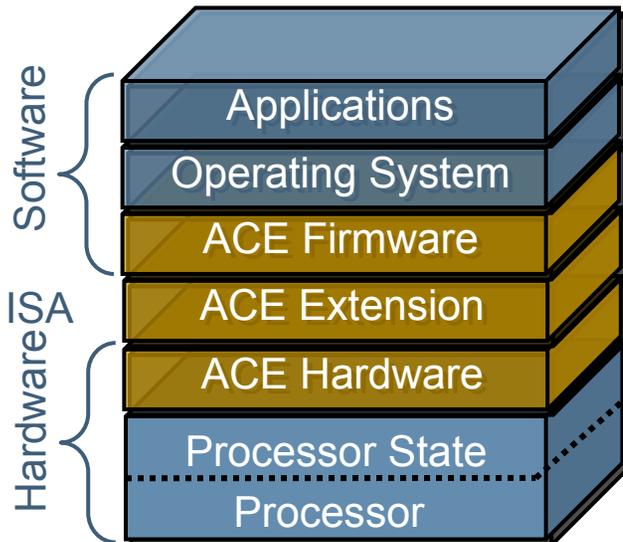
Migrate Testing Into S/W for Ultra-Low Cost

BulletProof

1. Execute and protect state
2. Test s/w periodically checks for underlying faults
3. **Test fails** → roll back, disable component, restart

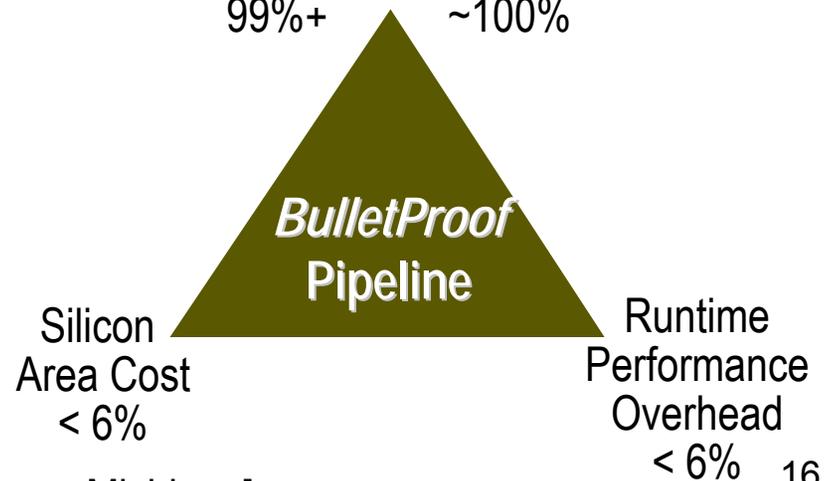


BulletProof Architecture



Measured Results

Provided Reliability:
Silicon defects + SER tolerance
99%+ ~100%



[Austin & Bertacco, Michigan]

Better-than-worst-case: “Aggressive” Deployment

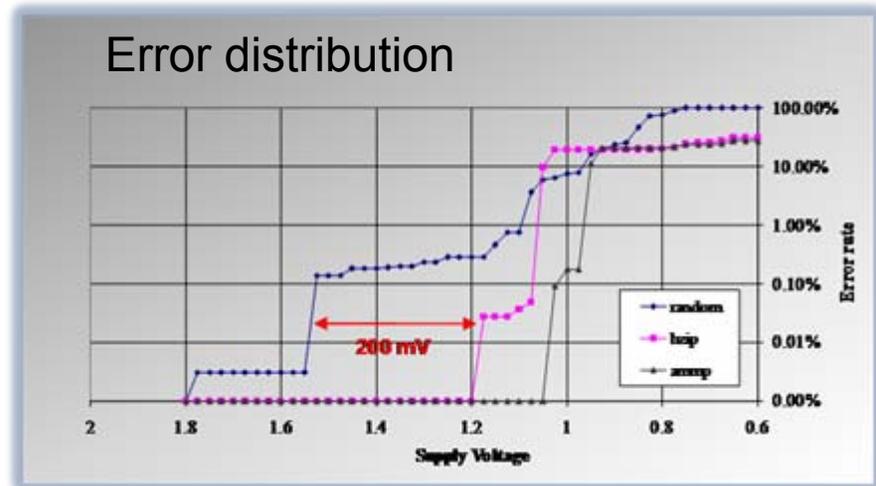
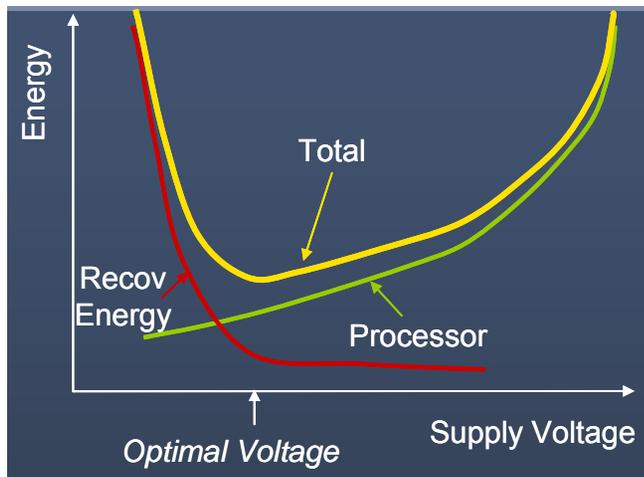
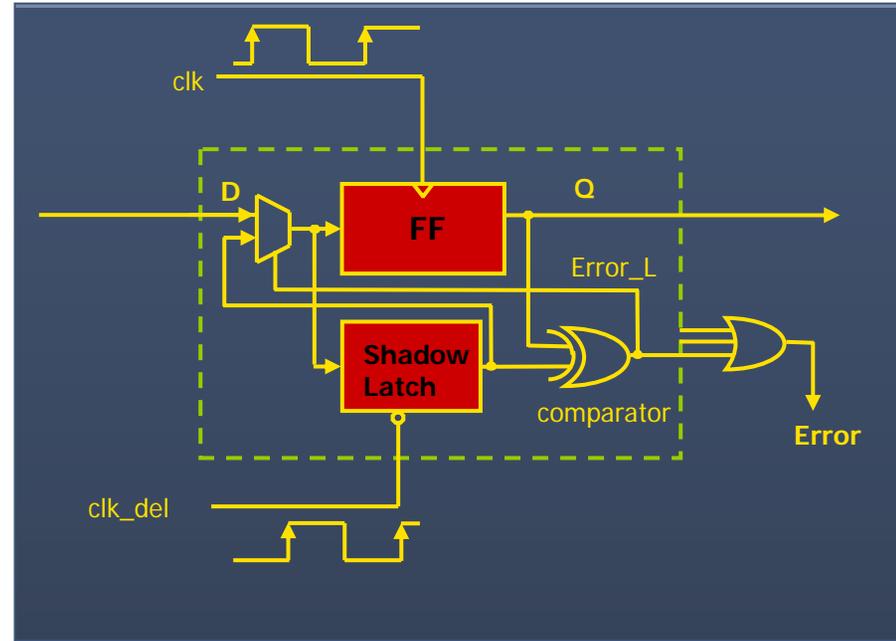
Scale voltage more than is allowable, deal with consequences

Example: “Razor”

“Pseudo-synchronous”

Process variations addressed

Circuit / architecture combined



“Aggressive” Deployment at Architecture-Level

Globally Optimized Resilient Systems



Recognition, Mining, Synthesis (RMS)

- Google MapReduce, Bayesian nets, etc.

☹️ RMS on unreliable h/w → DOESN'T WORK

Key observations:

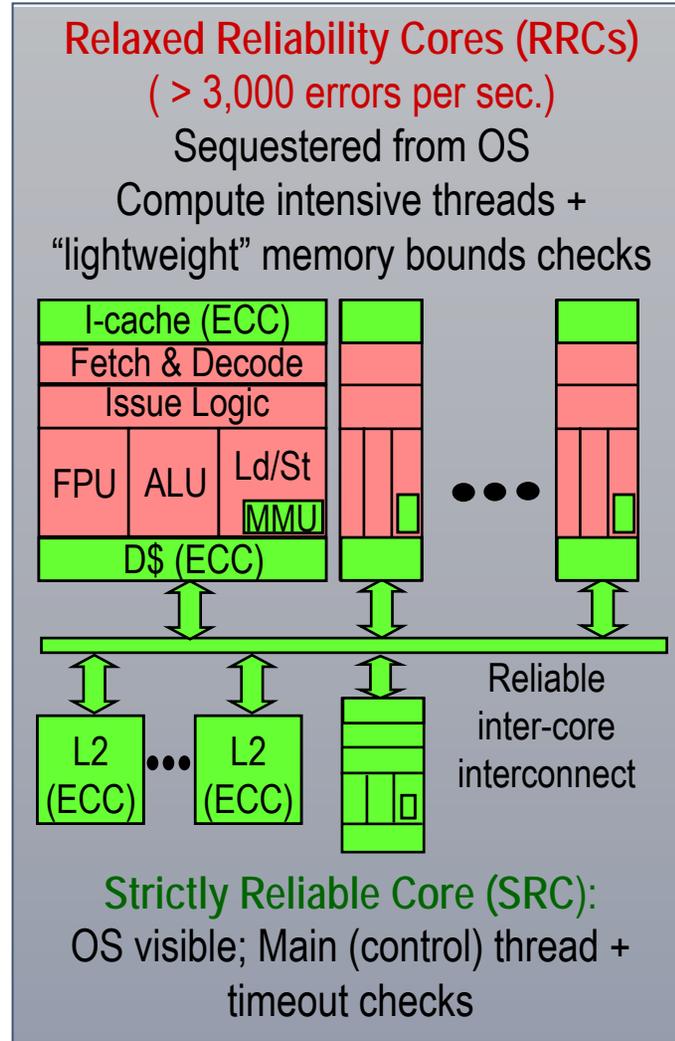
- Compute intensive threads → Errors OK
- Main (control) thread → No errors allowed
- **Error Resilient System Architecture (ERSA)**

ERSA Hardware prototype



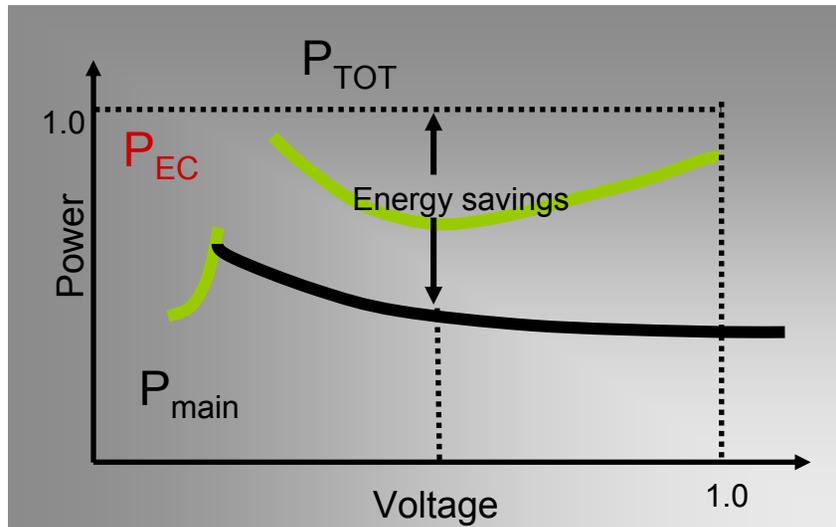
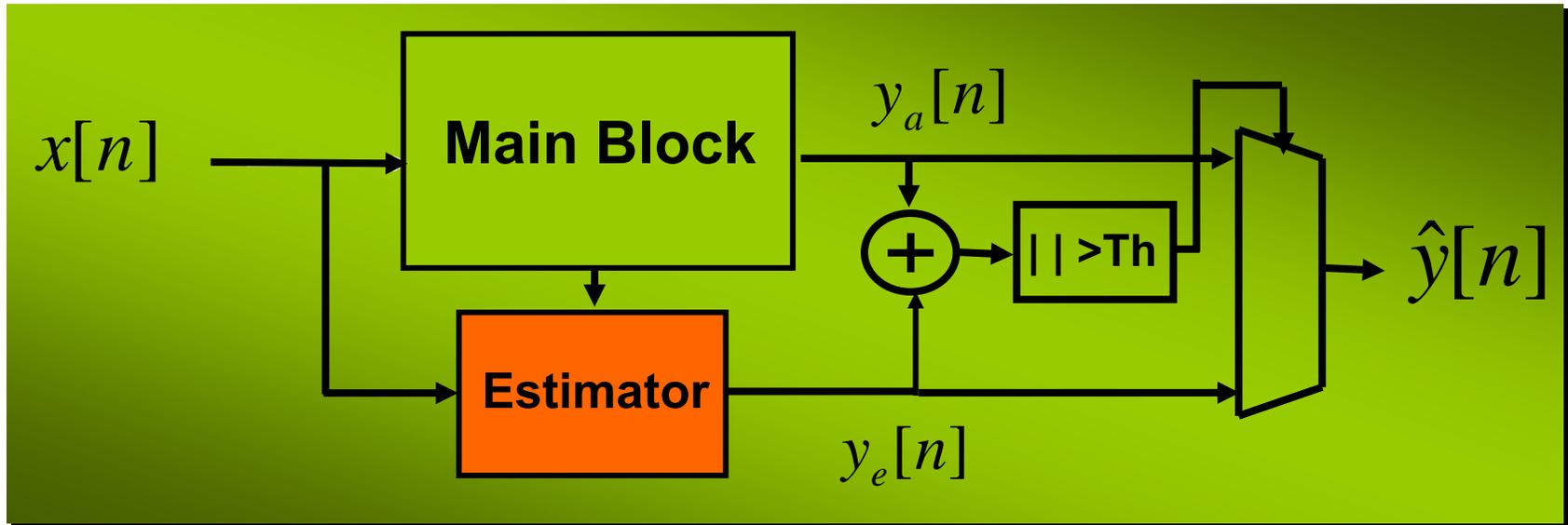
- **Resilient to $> 10^{16}$ FITs**
- Accurate results, useful throughput maintained

ERSA Architecture



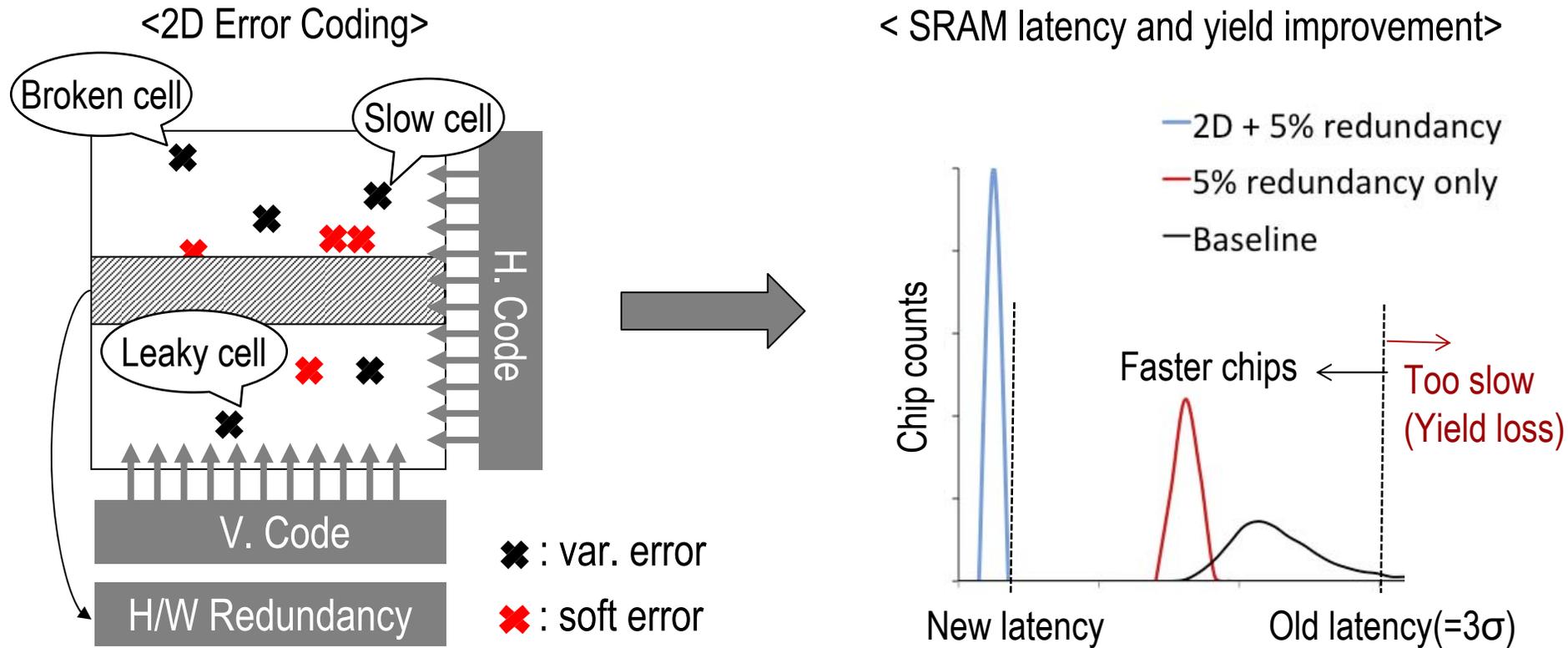
[Mitra, Stanford]

“Aggressive” Deployment at Algorithm-Level



Voltage **overscale** in main block
Correct errors using Estimator.
Purposeful degradation of SNR
(better performance vs. power savings)
Algorithmic Noise tolerance

SRAM Error Recovery: 2D Error Coding



2D error coding: Mask out bad cells, correct soft errors

→ High variability tolerance + high reliability + low cost

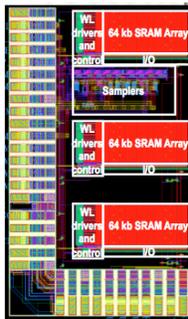
New Memories in C2S2

SRAM in C2S2

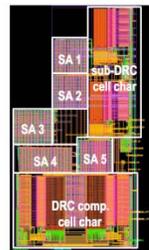
90nm
Canary



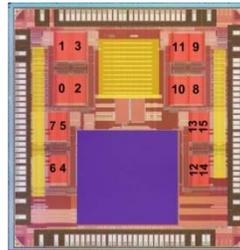
90nm
cRBL



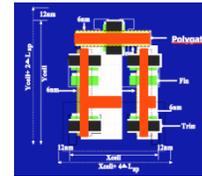
45nm
Canary/Chr



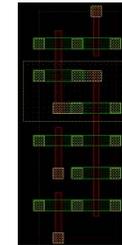
45nm
Var Study



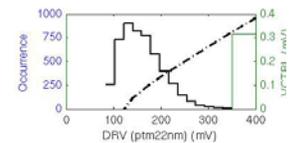
45nm
FinFET



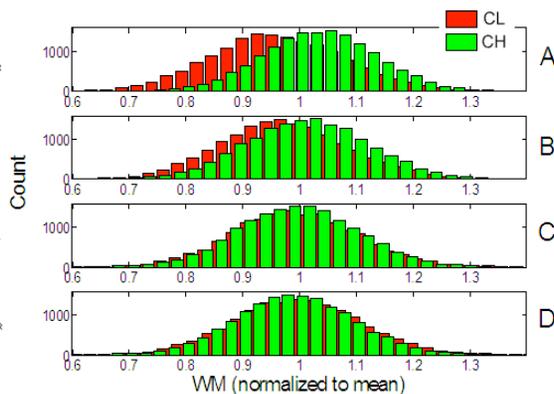
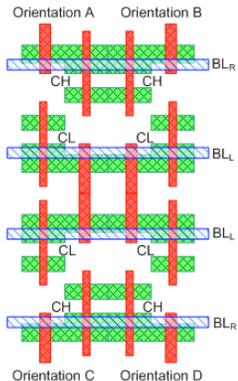
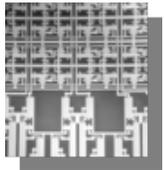
32nm
Regular



22nm
DRV study



Organic
Prototype

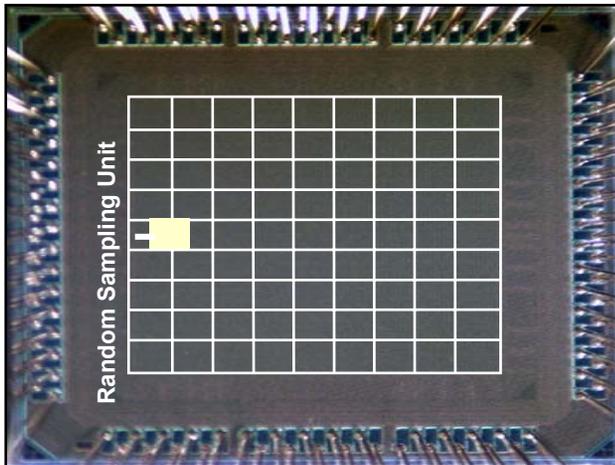


- Correlate simulation vs. silicon for SRAMs
- Understand *limitations of stochastic SRAM bitcell design*
- Analyze *impact of regularity* and neighboring patterns on transistor variability

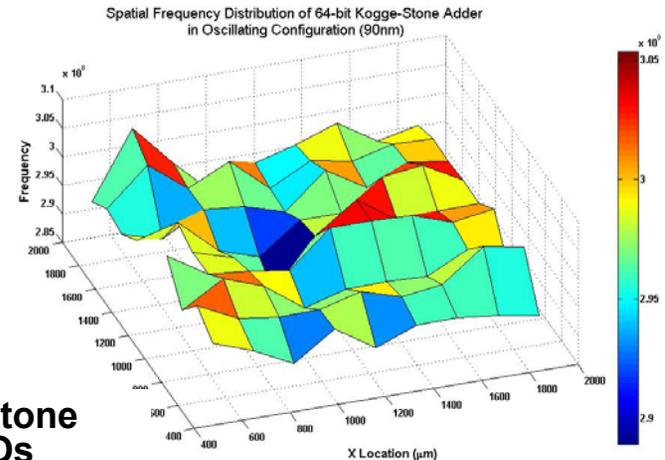
[Nikolić, Berkeley]

Variation Measurement, Modeling & Design

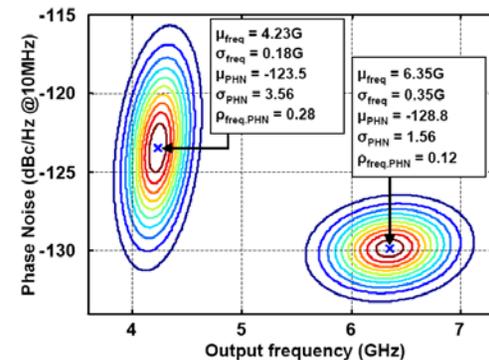
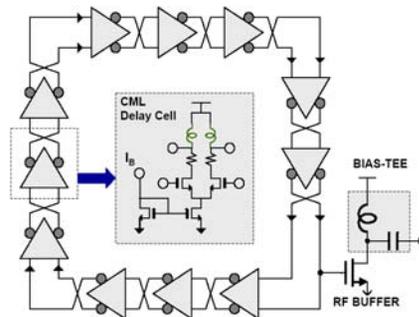
- Measurement indicating very *weak spatial correlation* between identical digital circuits and ring oscillator monitoring circuits



64-bit Kogge-Stone Adders & ROs



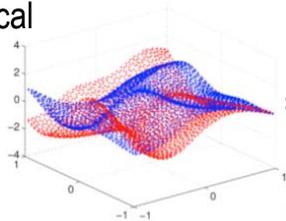
- *Ring VCOs* for improved variation estimation and control



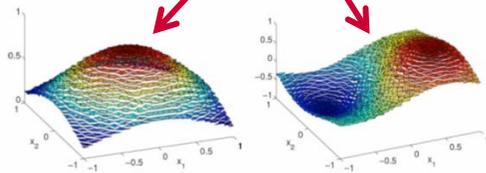
Fast Monte-Carlo for Variability Characterization

- Statistical *problem size increasing* with variables and spatial correlation
 - Reduce dimensionality
 - Fast MC sampling
 - A Karhunen-Loeve transform + Galerkin expansion/basis to get a small set of independent RVs

2d statistical variation (eg, L, Vth)

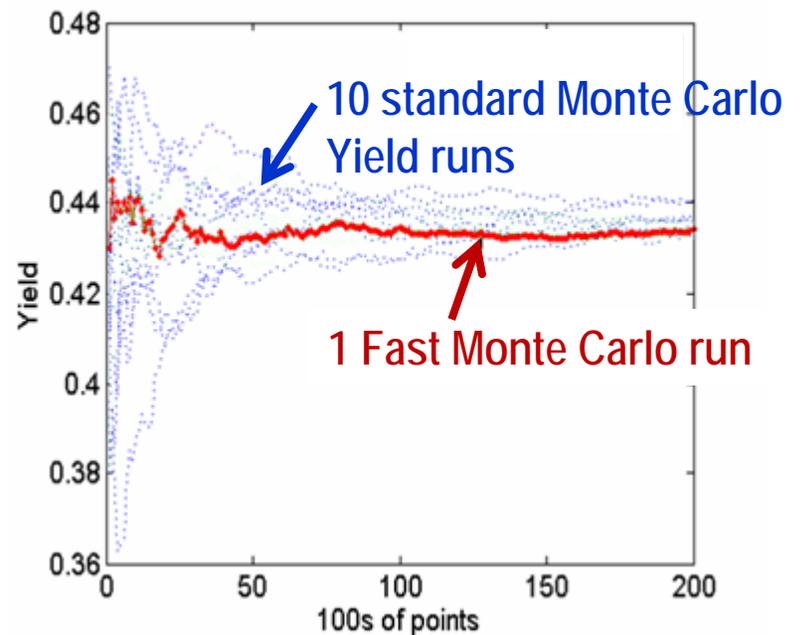


$$p(\mathbf{x}, \theta) = \sum_{j=0}^{\infty} \sqrt{\lambda_j} \xi_j(\theta) f_j(\mathbf{x})$$

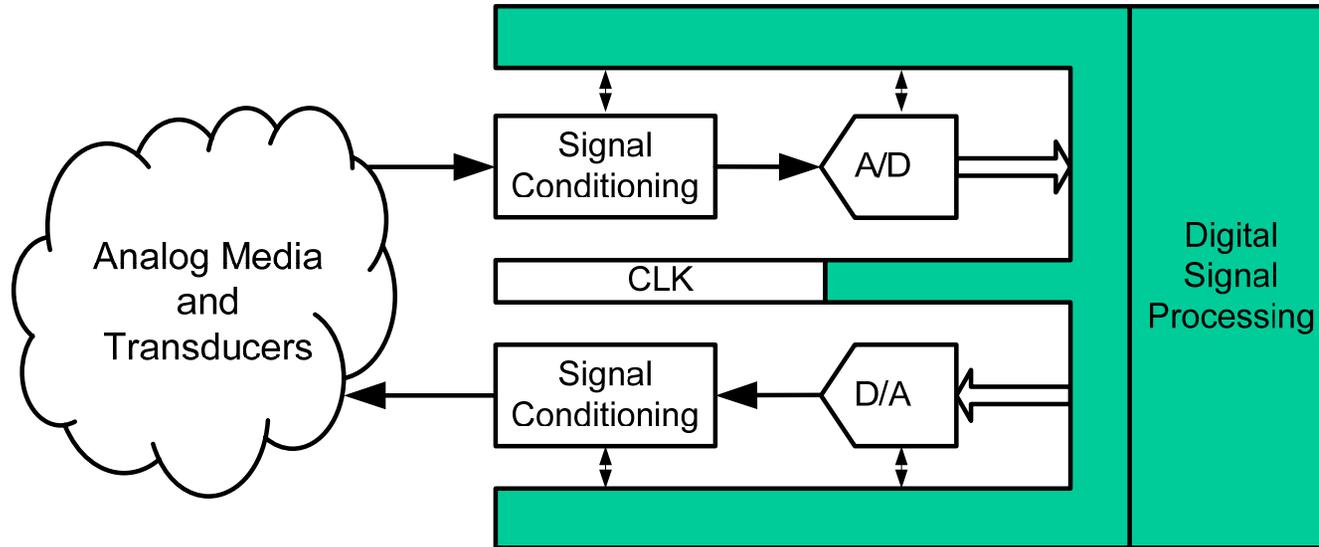


[Rutenbar, CMU]

- Redeploy *fast methods from comp finance* for efficient sampling
 - Quasi-Monte Carlo (QMC) deterministic sampling
 - Faster than MC, Latin Hyper Square, all common MC tricks



Digitally Enhanced Analog



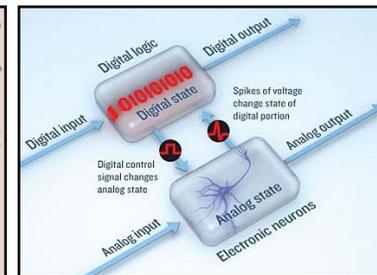
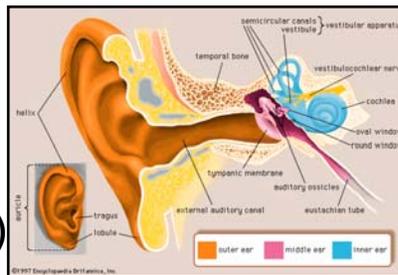
- Synergistic **glue logic**: additional digital processing
 - Enhanced analog performance
 - Speed, precision, power

Human ear example:

$P=14\mu\text{W}$,

DR=120dB,

1GFLOPS (spectral analysis)



[Mурmann, Stanford]

[Sarpeshkar]

Mixed-Signal System Design – New Paradigm

- Old thinking: Gates to improve isolated block specs
 - e.g., ADC with self-contained calibration logic
- **New paradigm:** Embrace system complexity
 - Across block boundaries, system attributes
 - Interplay between blocks
 - Understand system constraints
- Motivation: Digital vs. analog scaling trends
 - Energy per gate vs. per ADC
 - 300x vs. 35x reduction (over last 10 years)
 - Digital gate density follows Moore's law

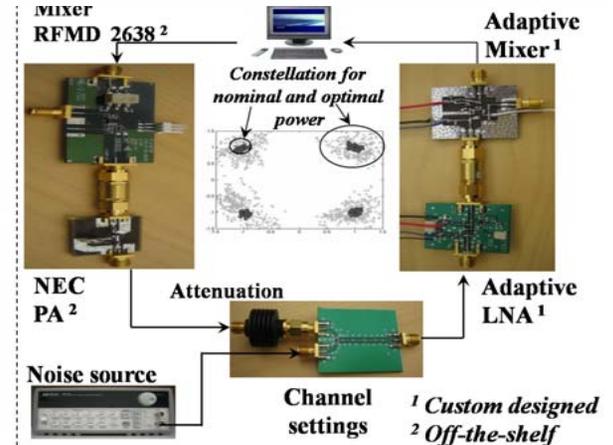
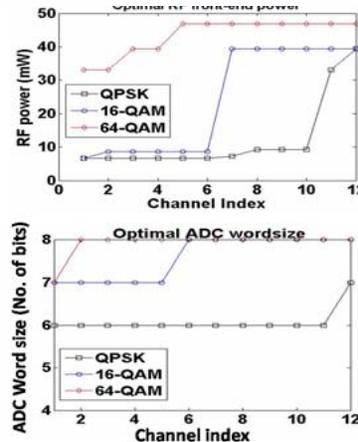
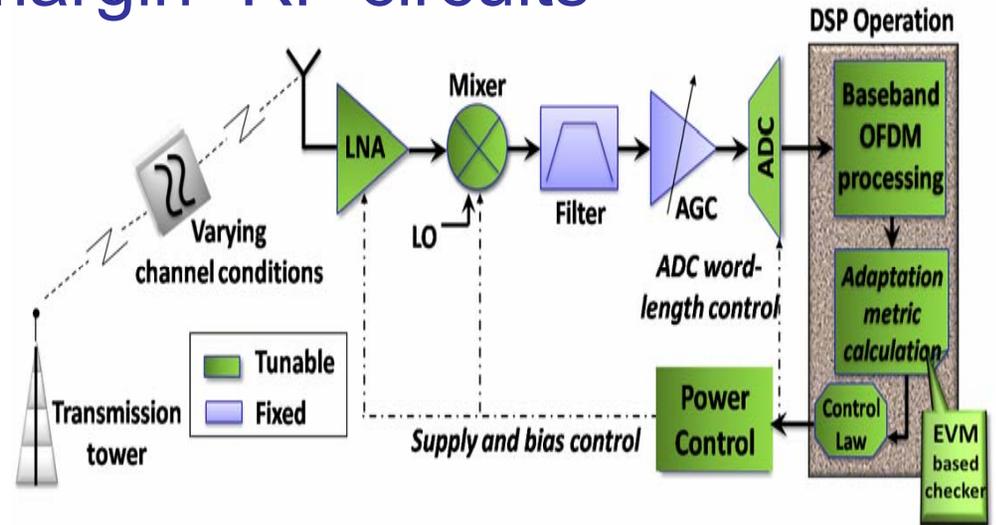
Error-Resiliency for Improved Radio Performance

VIZOR: Virtually “zero-margin” RF circuits

- Adapt to environment
- Adapt to variations
- Co-adapt across circuit-software boundaries

Approach

- ‘Hot-checking’ and error feedback
- Built-in test
- Tight performance feedback control



Power savings (for best channel and QPSK) > 4 X

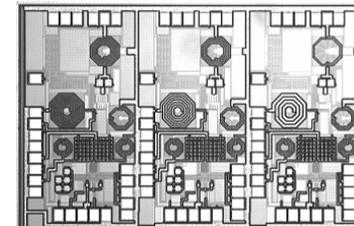
Si Infrastructure for Configurable Analog/RF

- Optimally specify physical patterns (fabric)
 - For lithography
 - For configurability / tunability

- Application
 - Post-manufacturing tuning
 - Process & environmental variations

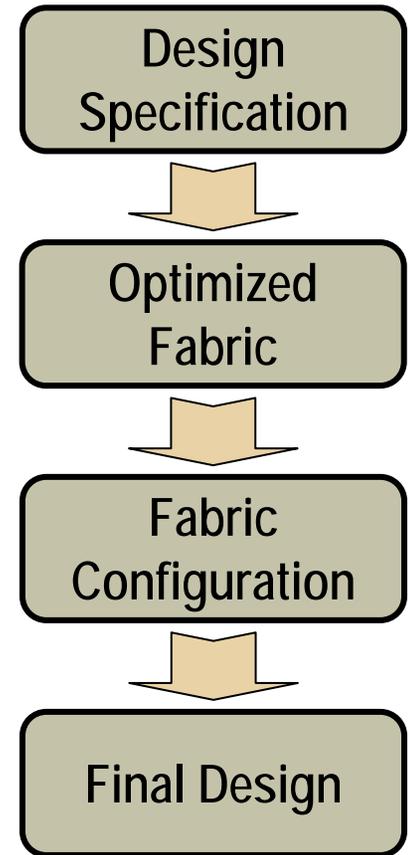
- Demonstration: Configurable RF fabric
 - IBM SiGe 0.25 μm BiCMOS
 - *Customized top 3 metal layers*

- 1.5GHz GPS, 2.1GHz WCDMA, 5GHz WLAN



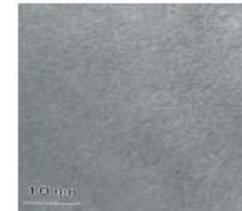
Multi-stage
Stochastic
Optimization

Post-Silicon
Measurement



Field Programmable Analog / RF

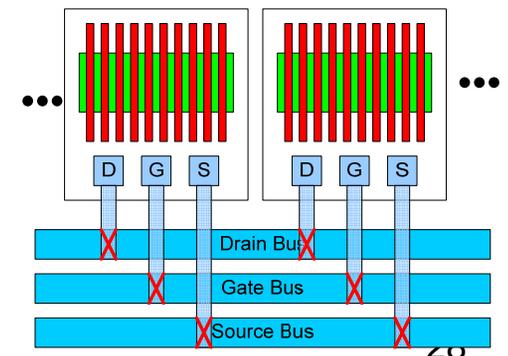
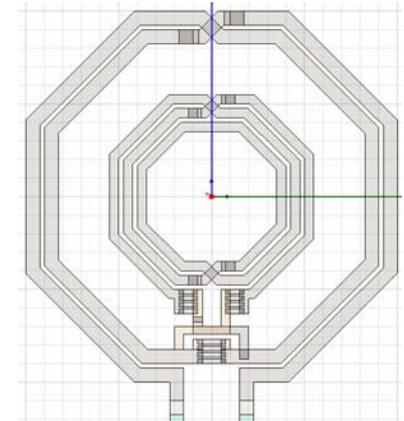
- Phase change (PC) materials
 - Configurable RF vias
 - Lower parasitics
 - Challenges
 - On-chip heaters
 - Power delivery to switch PC material
- Configurable inductors
- PC vias to “program” transistor sizing
 - PC-configurable RF circuit design
 - 90nm IBM CMOS
 - In progress



Amorphous PC



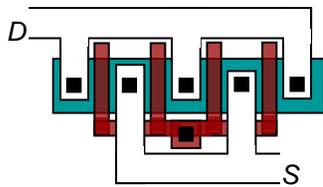
Crystalline PC



Exploiting Variability

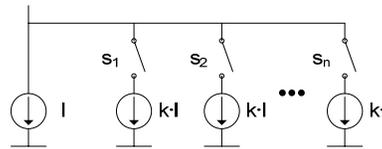
- Analog / RF designs: Live / die based on device matching
- Systematic variations: Regular, symmetric pattern
- Random variations – *Exploit* for post-Si tuning
 - Statistical Element Selection (SES)

- *Exponential mismatch reduction* (function of area)



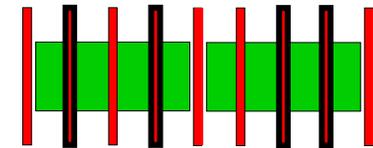
Pelgrom style sizing:

$$\sigma(V_{TH}) \sim \text{Area}^{-1/2}$$



Digital Trim Fix:
Switch in N sources

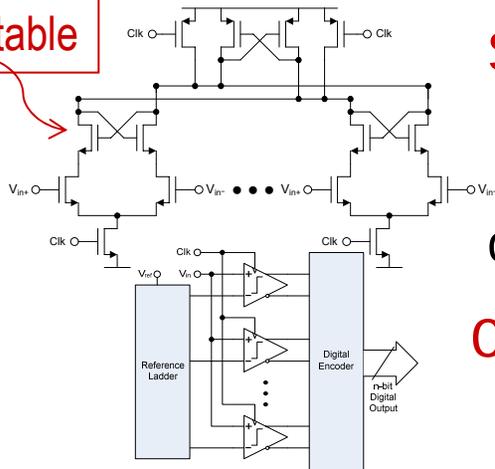
$$\sigma(V_{TH}) \sim \text{Area}^{-1}$$



SES: Regular fabric,
post-Si choose best subset

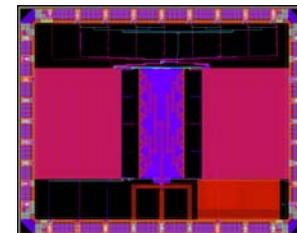
$$\sigma(V_{TH}) \sim e^{-k \cdot \text{Area}}$$

Selectable



Self-healing 8-bit Flash ADC

Over 60x improvement in bandwidth and power as compared to Pelgrom sizing



Other SES Examples: Self-healing SRAM read path

[Pileggi and Li, CMU]

Conclusion

- Self-healing systems required for advanced scaling
 - Optimize for
 - Power
 - Performance
 - Reliability
- FCRP research
 - Core technologies
- Complete self-healing mixed-signal systems
 - HEALICs program
 - Exciting research opportunity