Cross-layer studies: from technology to circuits and architectures

Ramon Canal

Universitat Politècnica de Catalunya

@RamonCanalUPC

Cross-layer studies: from technology to circuits and architectures

Shrikanth Ganapathy, Manish Rana, Zoran Jaksic, Josué Quiroga, Marc Riera, Martí Anglada, Marc Portavella, Antonio González
CLERECO partners: Polito, UoA, LIRMM, ABB, Yogitech, Thales

Ramon Canal
Universitat Politècnica de Catalunya
@RamonCanalUPC

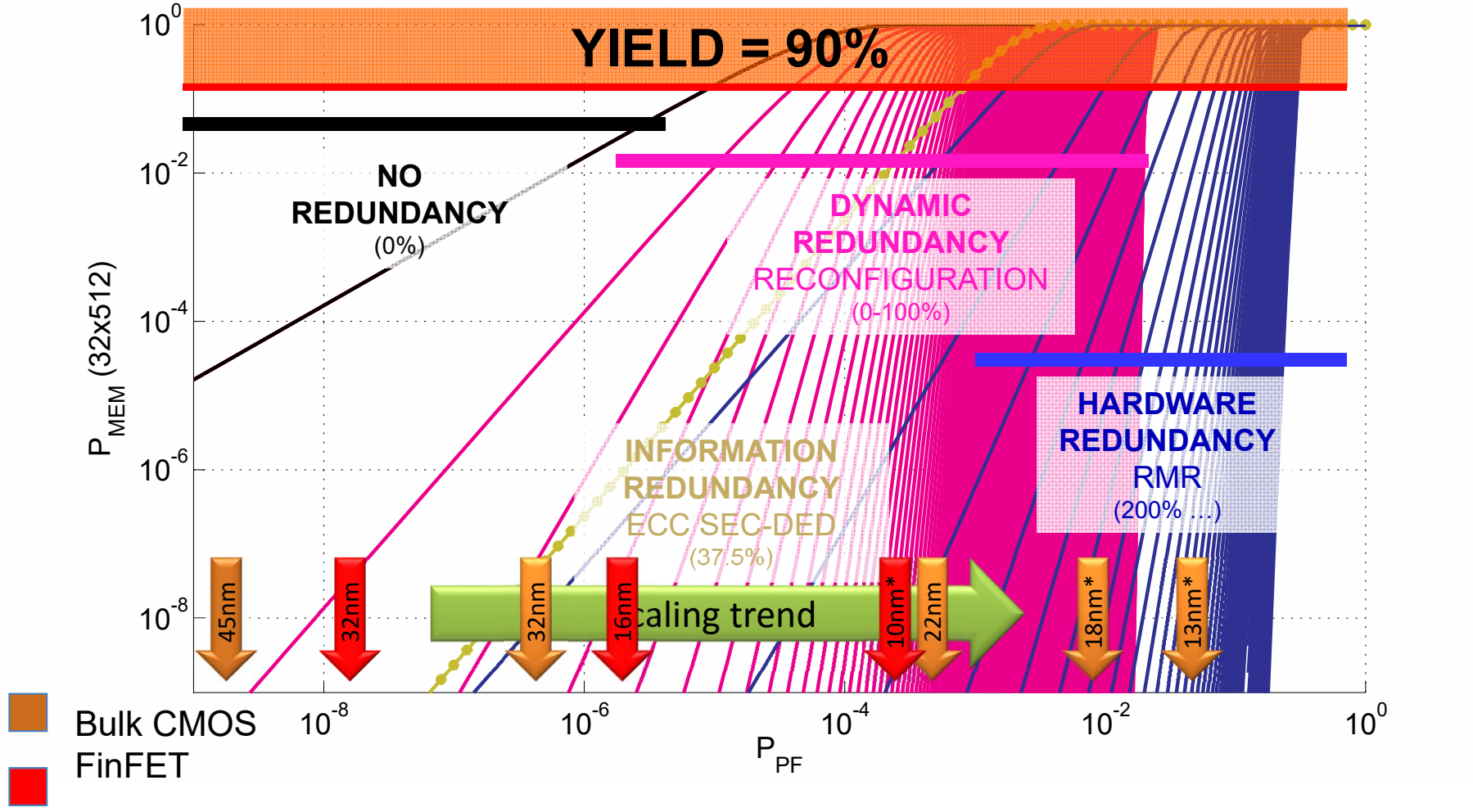
Organization

- **Part 1: Rapid change in silicon technology**
  - **Characterize the effects on memories and logic**
    - **Delay, Power and Robustness**
  - **Provide architecture good foundations on technology**
  - **Research on cross-layer approaches for better power/delay/robustness tradeoffs**

And.. the next time around..

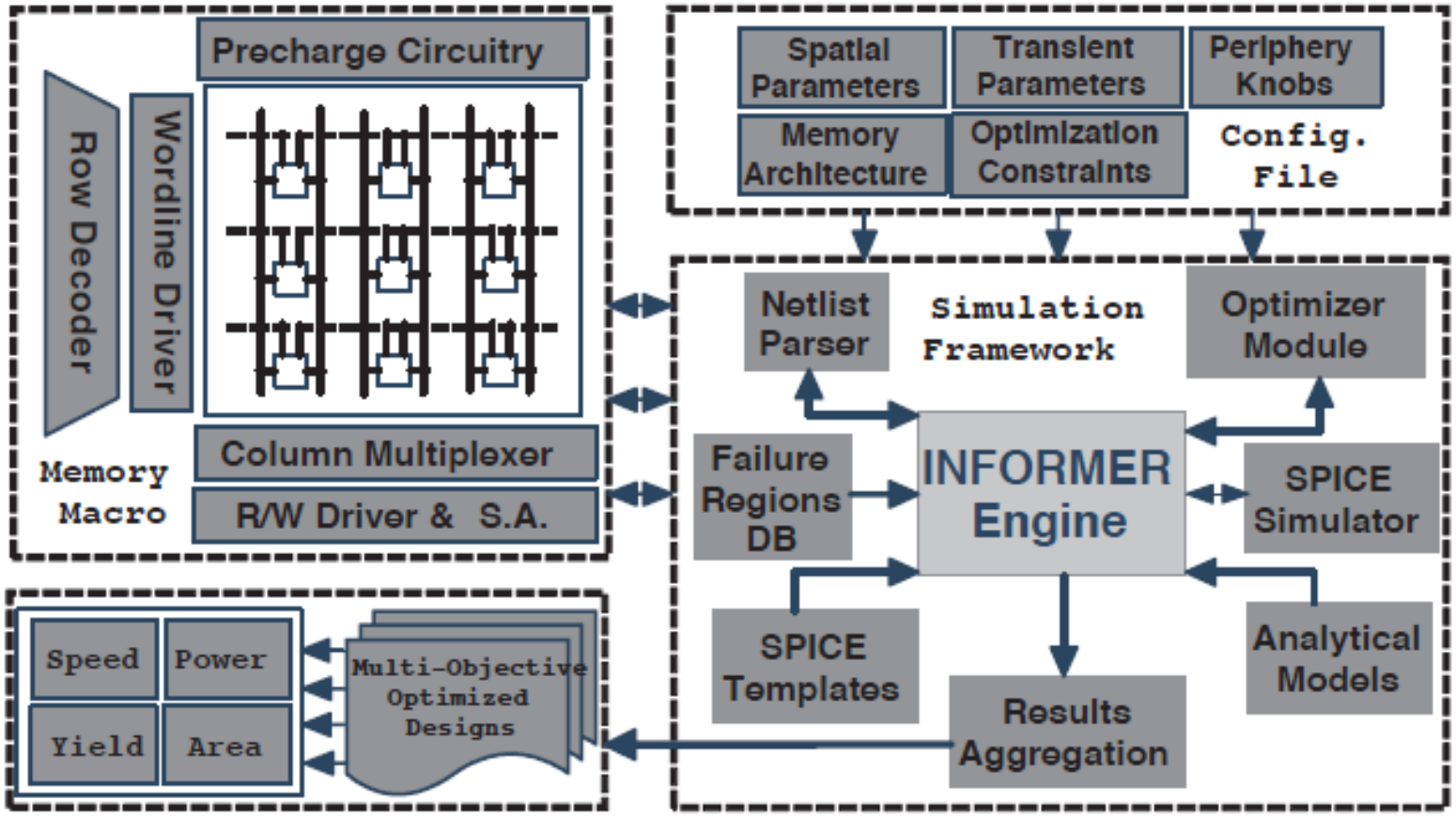
- **Part 2: Emerging new markets and applications**
  - **New memory organizations (SoC)**
    - **Heterogeneous Computing 2.0**
  - **Customizable accelerators for specific workloads**

Part 1: Technology Analysis



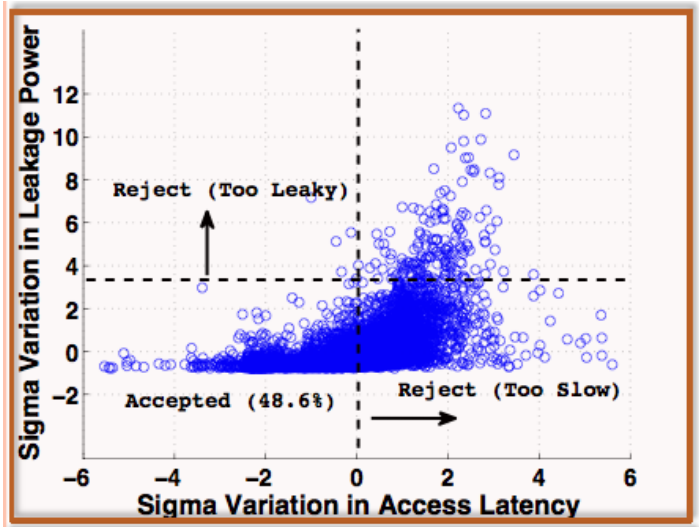
Past work

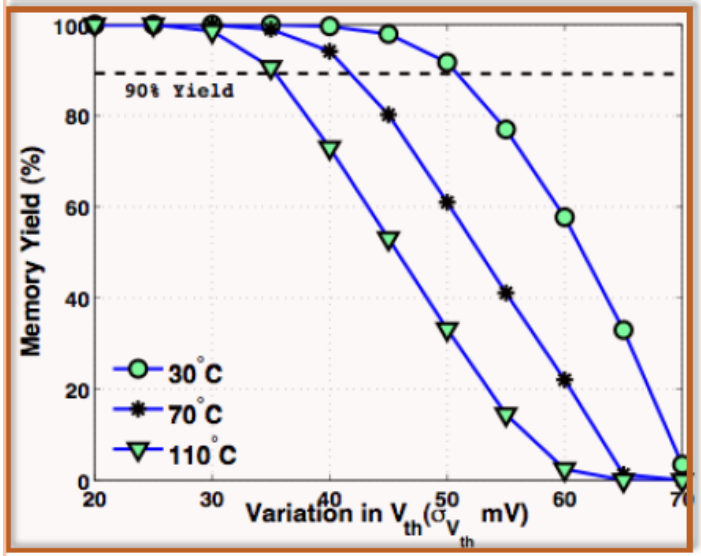
- **INFORMER: Integrated Framework (Date'14)**



Past work

- **INFORMER: Integrated Framework (Date'14)**

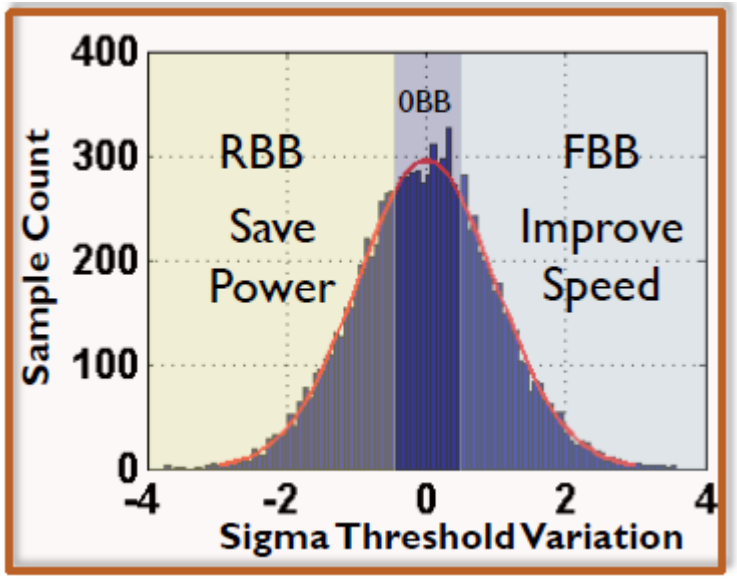


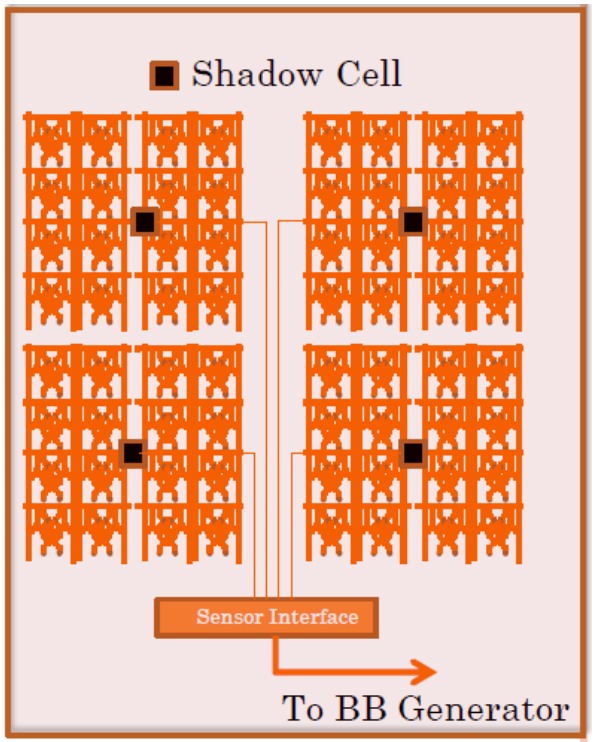


- Power/Performance binning of the memory
  - Provides deeper insight into yield distribution for post-silicon tuning
  - Better understanding of the economics governing chip yield
- Temporal yield of the memory
  - Evaluate indirectly the lifetime of the memory for a given threshold shift under temperature stress (BTI)
  - Can help design better proactive recovery techniques

Past work

- **Dynamic Fine-Grain Body Biasing (ICCD'11)**

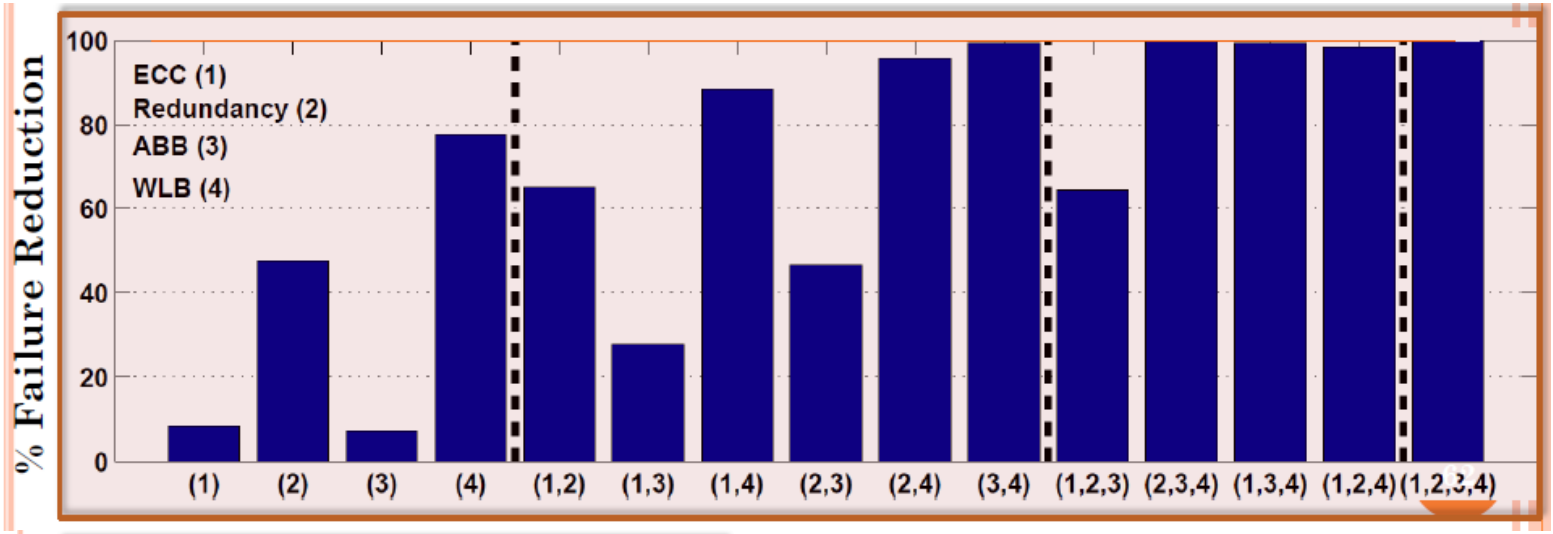


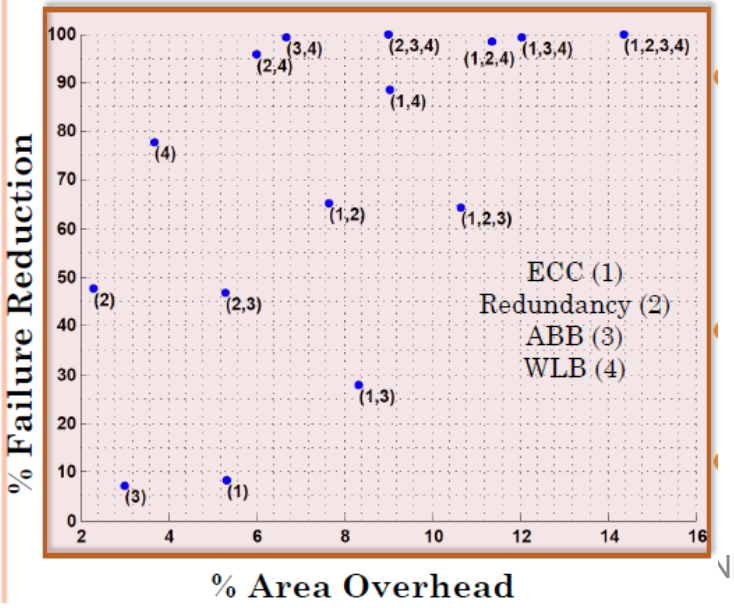


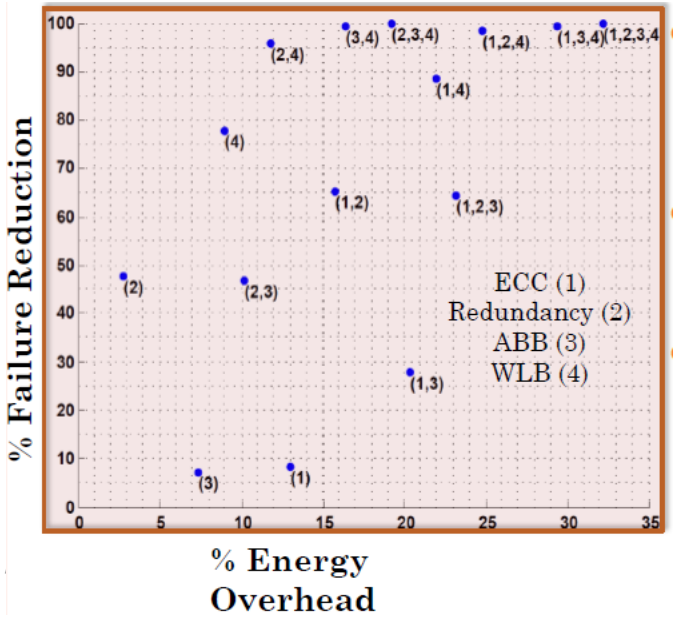
- **Use shadow cells (3T1Ds) as sensors**

Past work

- **Hybrid-Recovery Techniques (ISQED'13)**





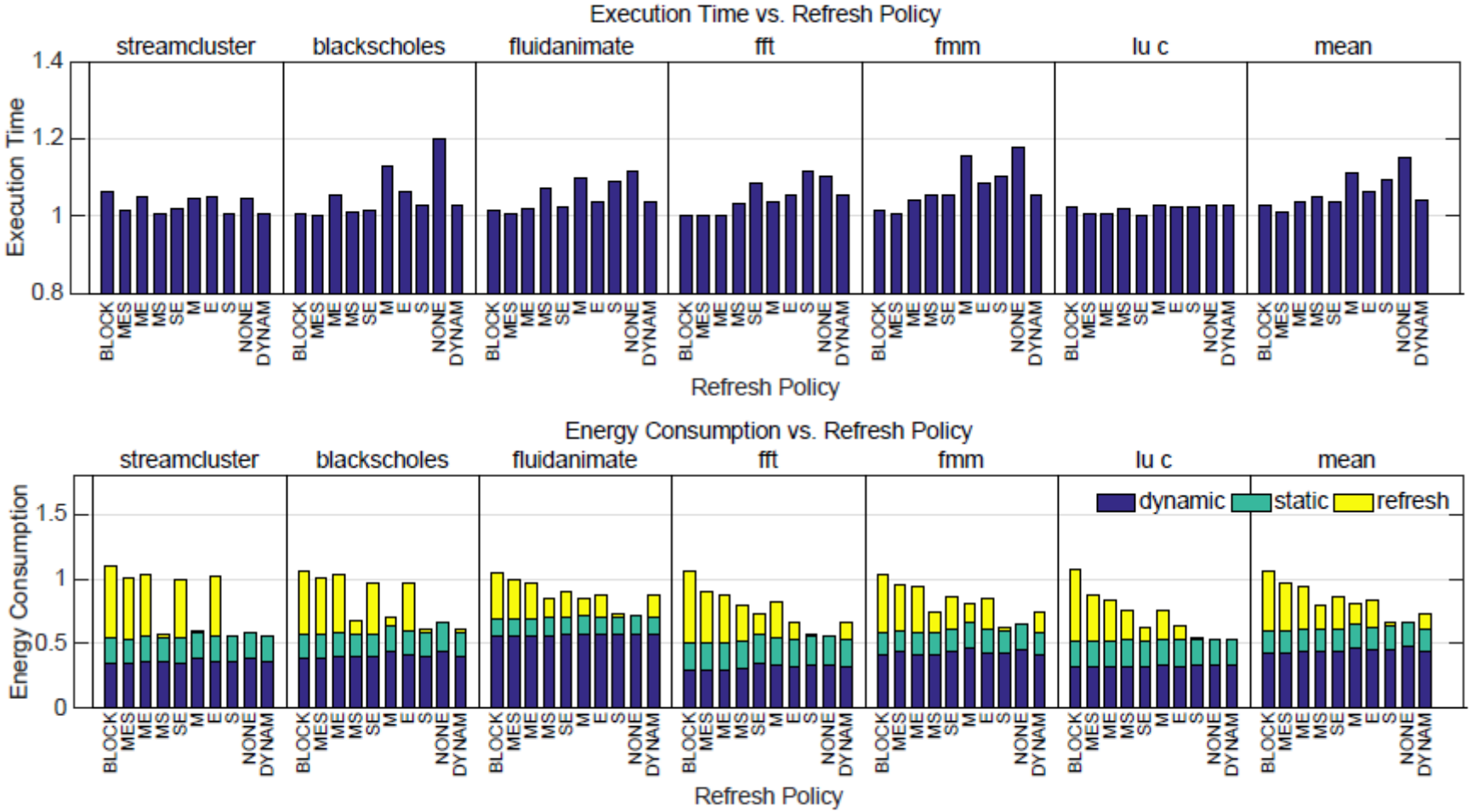


Recent work

- **BTI reduction**
  - **iRMW (invert, read-modify-write) for SRAM cells**
- **DRAM-Memories**
  - **4T DRAM robust cell**
  - **PMOS, NMOS only/mixed cell analysis**
  - **DRAM-based L2 coherent caches**

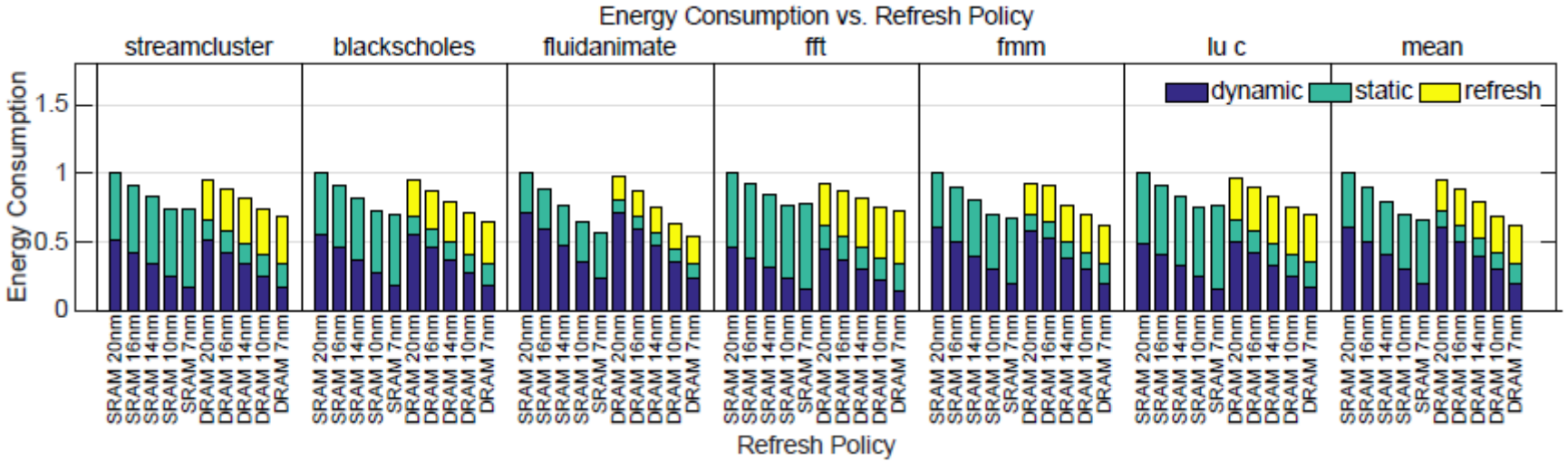
Recent work

- **DRAM-based Coherent Caches (Date'14)**
  - **Refresh only in certain coherence states**

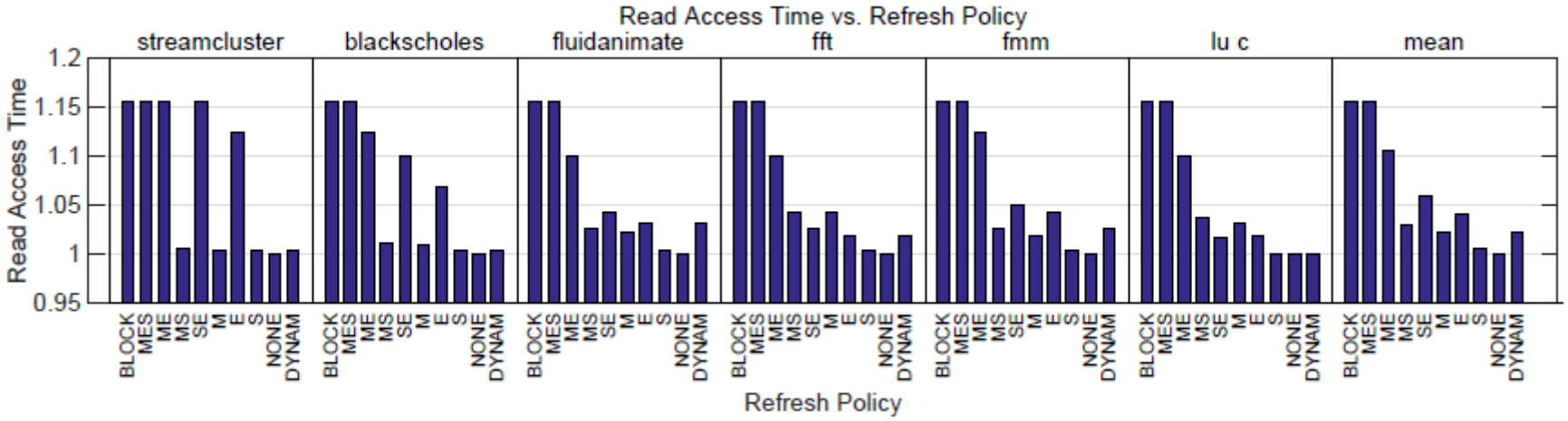


Recent work

- **DRAM-based Coherent Caches vs. technology**

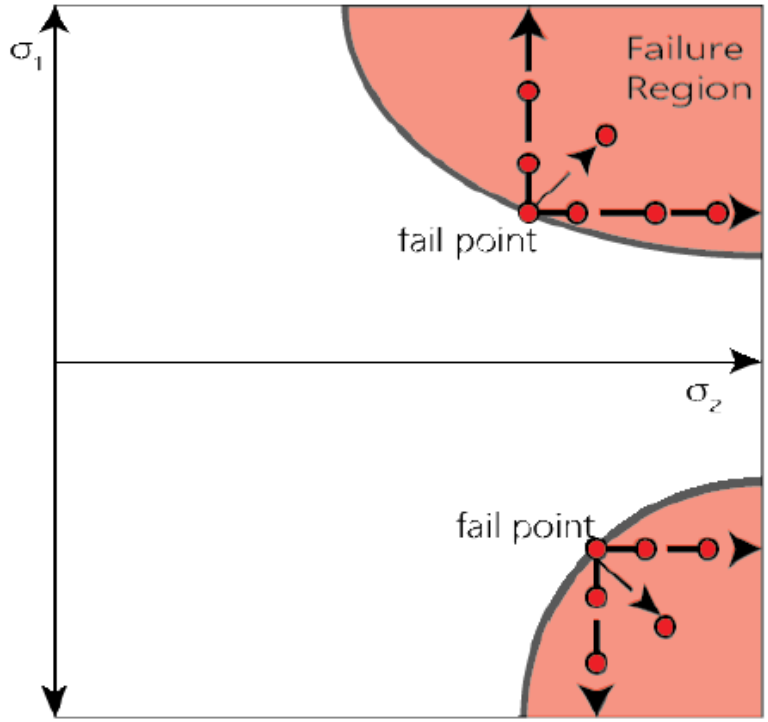


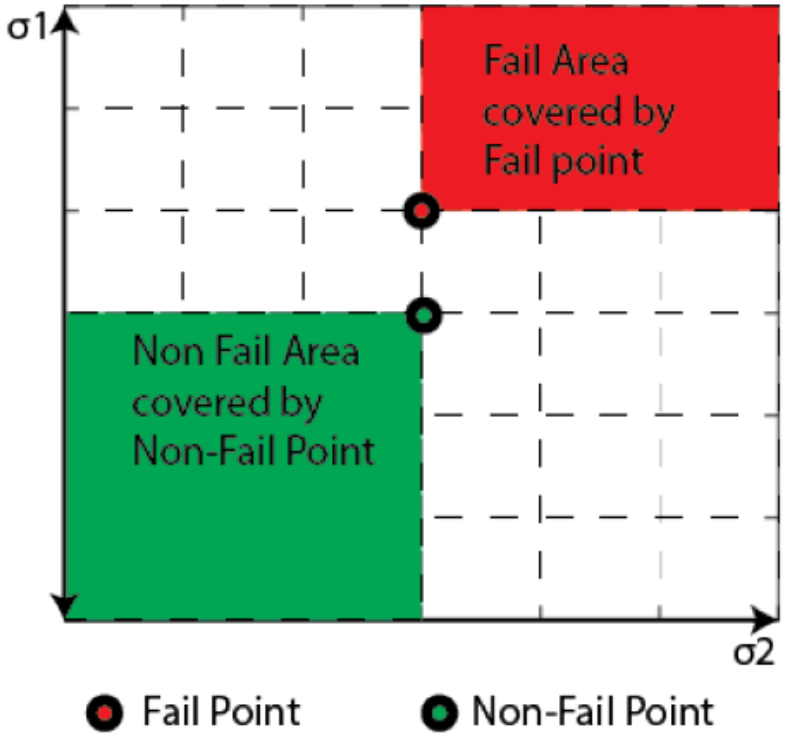
- **DRAM-based Coherent Caches + BTI** **(3 years, RD model, S=50%)**



Current work

- **SRAM Yield Analysis (DATE'14, ICCD'14)**

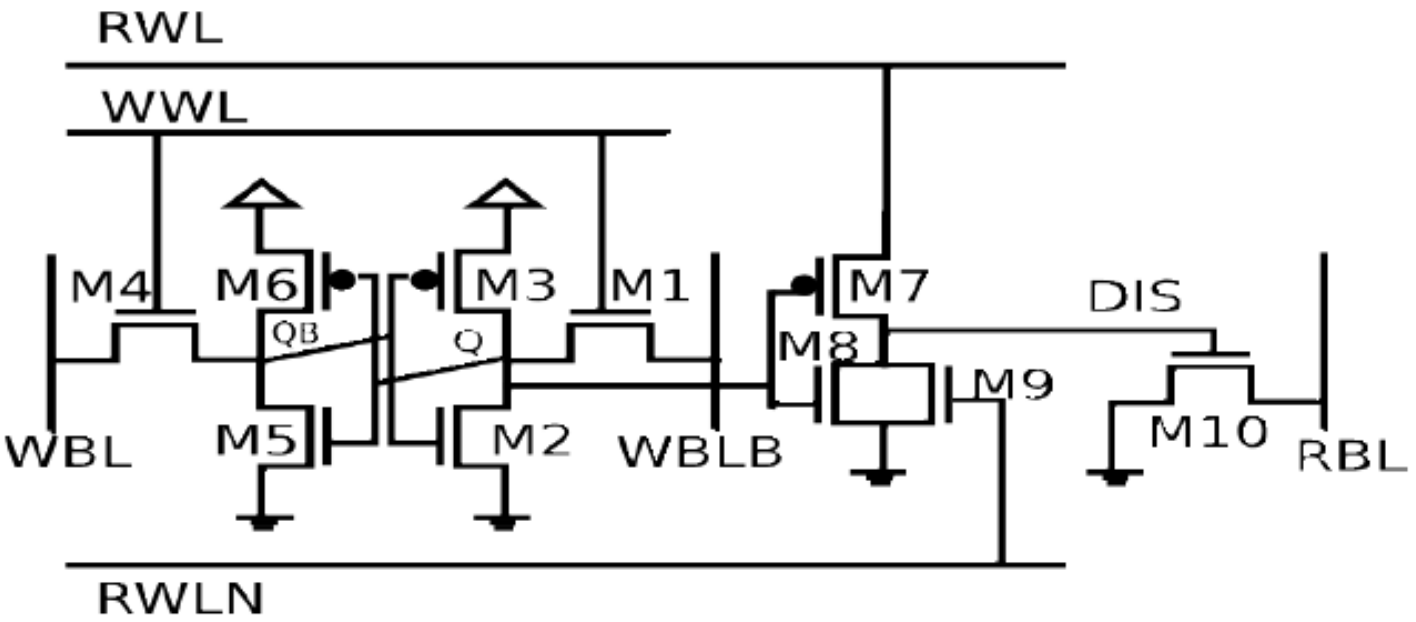




	Simulations		
	MPFP	IS (4e+4)	Total
REEM	624	9118	9742
Mean-Shift IS	6e+4	4e+4	1e+5

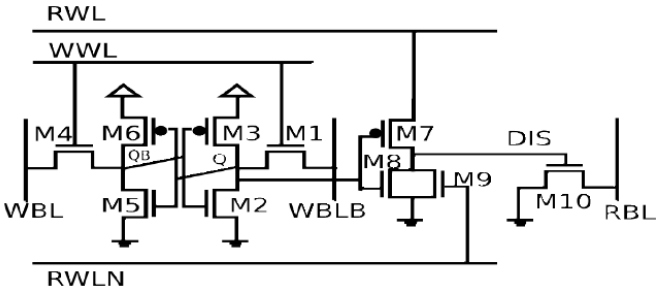
Current work

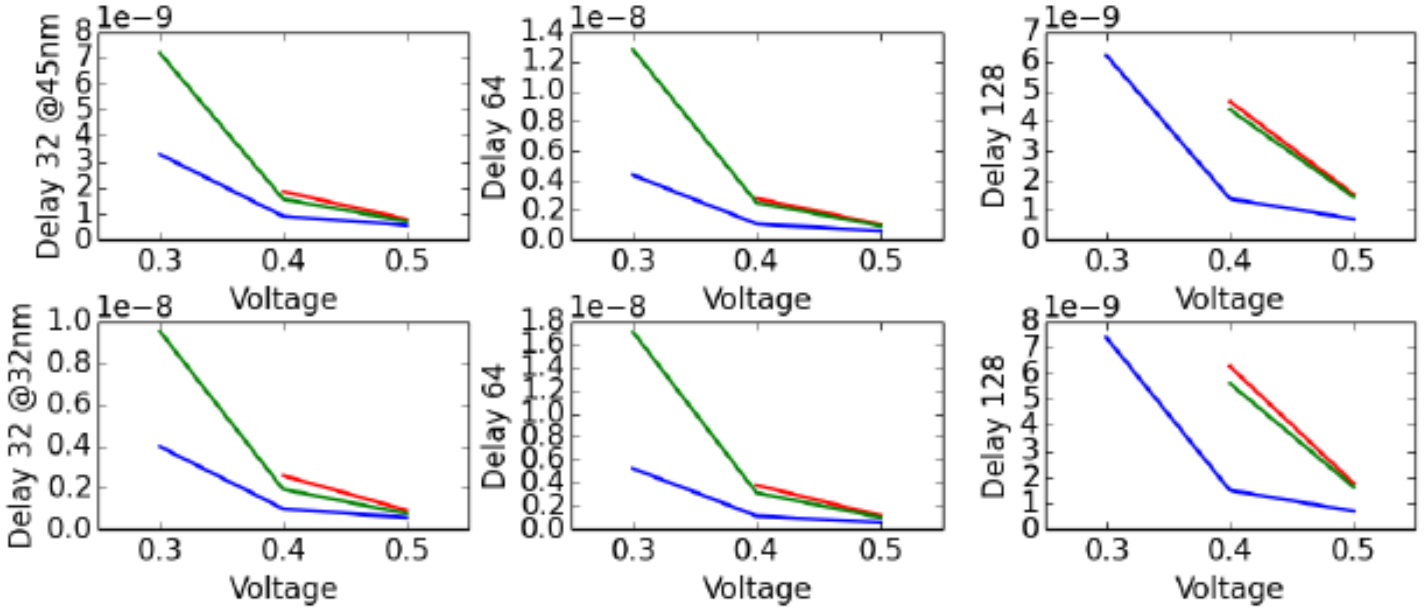
- **Novel 10T Subthreshold SRAM cell**



Current work

- **Novel 10T Subthreshold SRAM (**

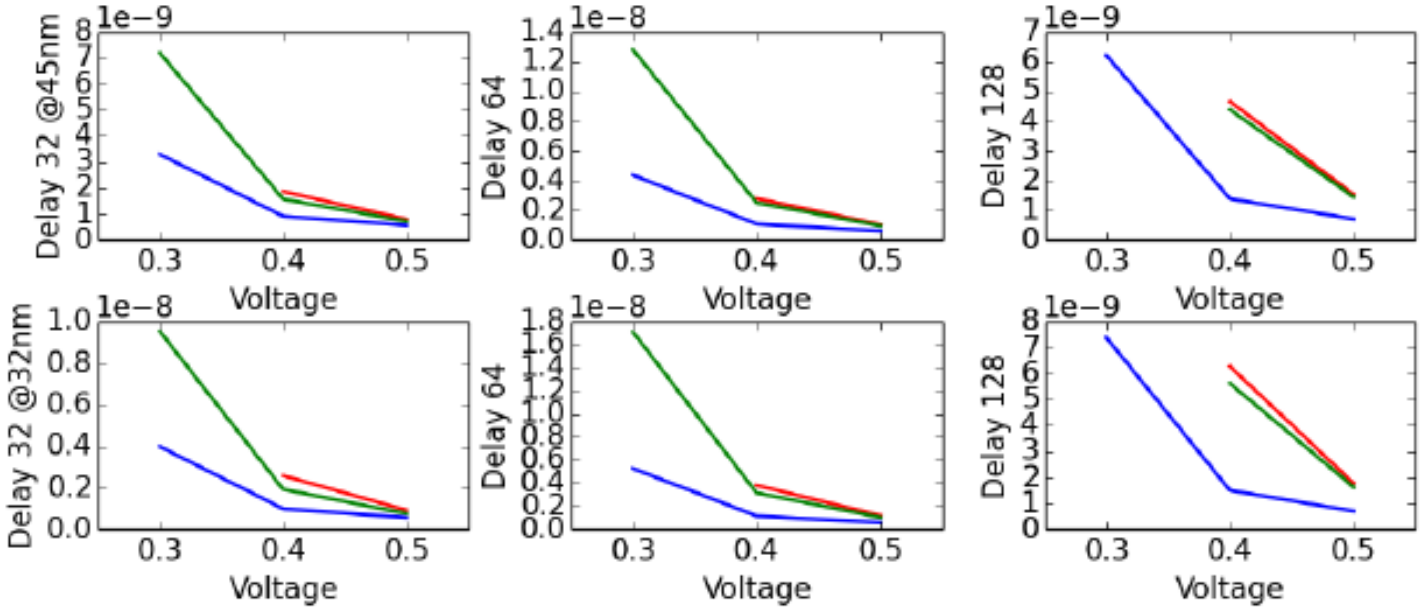




Read-0 delay comparison for a) 45nm b) 32nm technology

Current work

- **Novel 10T Subthreshold SRAM cell**



Read-0 delay comparison for a) 45nm b) 32nm technology

Work completed

- **Energy/Delay modelling**
- **Variability modelling**
- **Integrated Framework (INFORMER)**

- **Speed-up failure probability estimation through statistical models**

- **Novel memory cells**

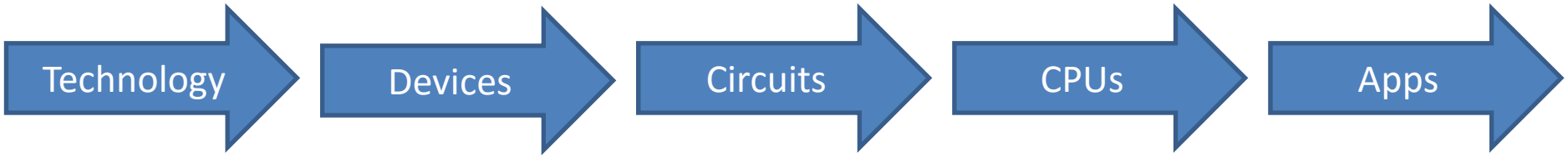
- **Next step, integration in a ~commercial tool**

CLERECO FP7 Project

- **Avoid worst-case design.**
  - **Take advantage of masking effects.**
- **Reduce the device to system characterization**
  - **Define cross-layer interfaces**
    - **vulnerability factors, probability of failure, …**
  - **Speed-up the computation of metrics**
    - **Statistical models, IA techniques**
- **Ultimate goal**
  - **Design space exploration of future systems, bearing in mind the underlying technology.**

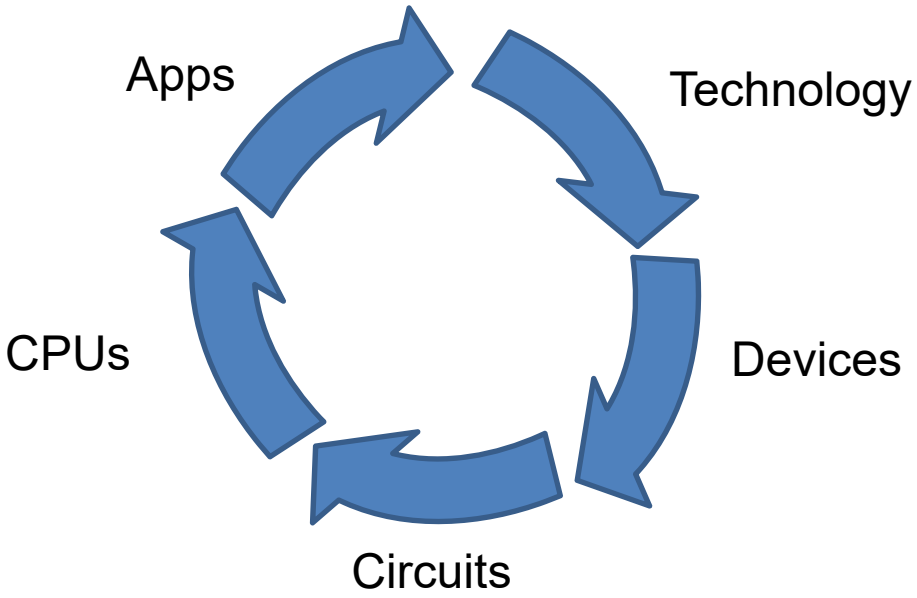
Positive Side-effects

- **Feedback loop made a reality**



Positive Side-effects

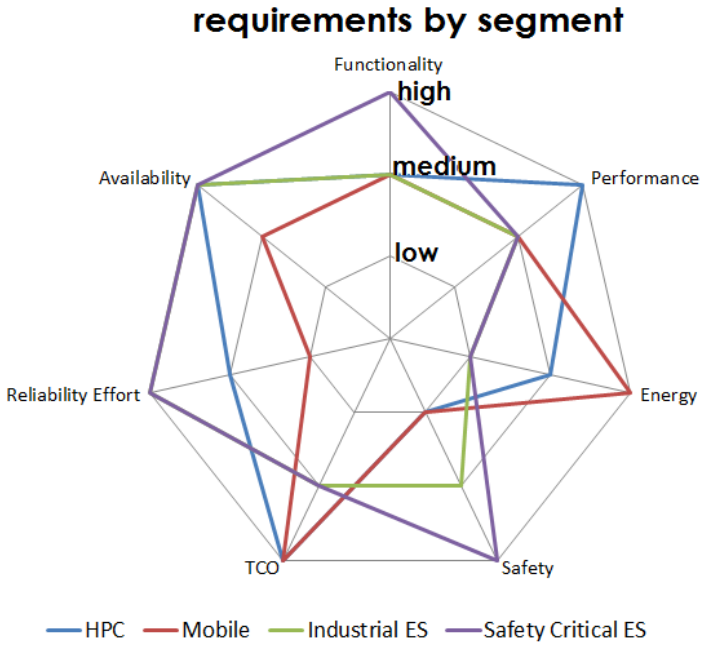
- **Feedback loop made a reality**



Answers for old questions

- **When running “Angry birds” will I see a performance difference with a:**
  - **10% higher fin height**
  - **2% thicker insulator**
  - **Fully-depleted SOI**
  - **My new memory cell**

Computer segment



- **Different Hardware Scenarios**
  - **Operating voltage**
  - **Operating conditions (Temperature, Radiation)**
  - **On top of traditional ones (performance, energy, power…)**

Next steps

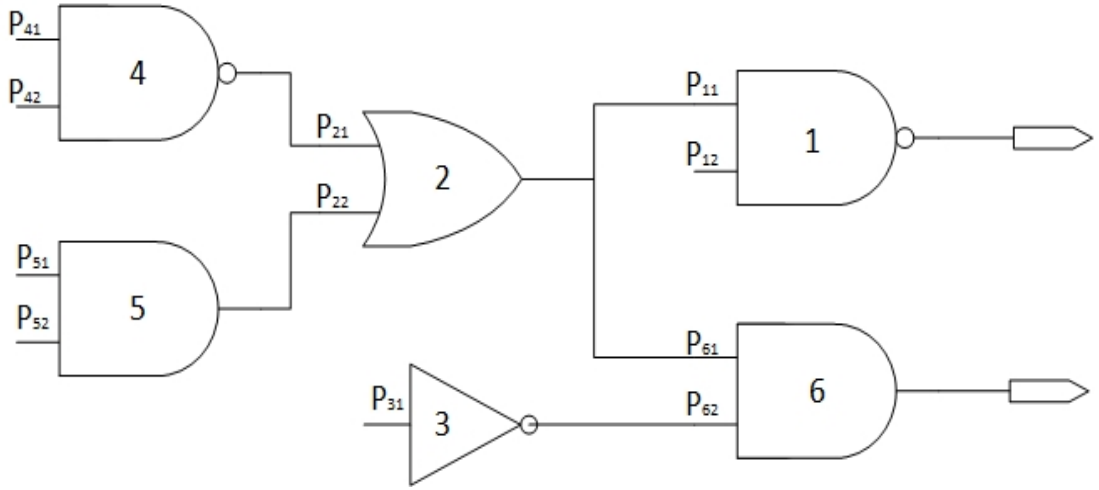
- **Analyze the behavior of DRAM memories in near-subthreshold scenarios (IoT, sensor networks)**

- **Particle strikes**
  - **Fast computation masking (derating) effects:**
    - **Logical (e.g. 0 AND x = 0)**
    - **Electrical (glitch is not latched)**
  - **Quantify the effect of system location (i.e. coordinates + altitude)**

Next steps – Logic Masking

- **Upsets may be masked by underlying logic gates**

Input 0	Input 1	NAND Output	Output if Input 1 changes
0	0	1	1
0	1	1	1
1	0	1	0
1	1	0	1



P(i)=Probability that value is a logic "1"

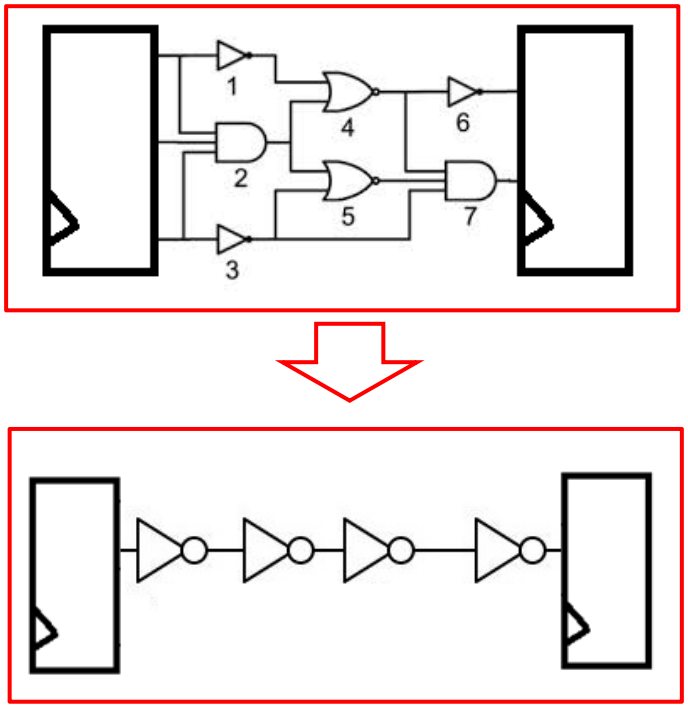
*P(NAND)=1-P(in1)*P(in2)*

P(NOR)=(1-P(in1))(1-P(in2))*

P(NOT)=1-P(in)

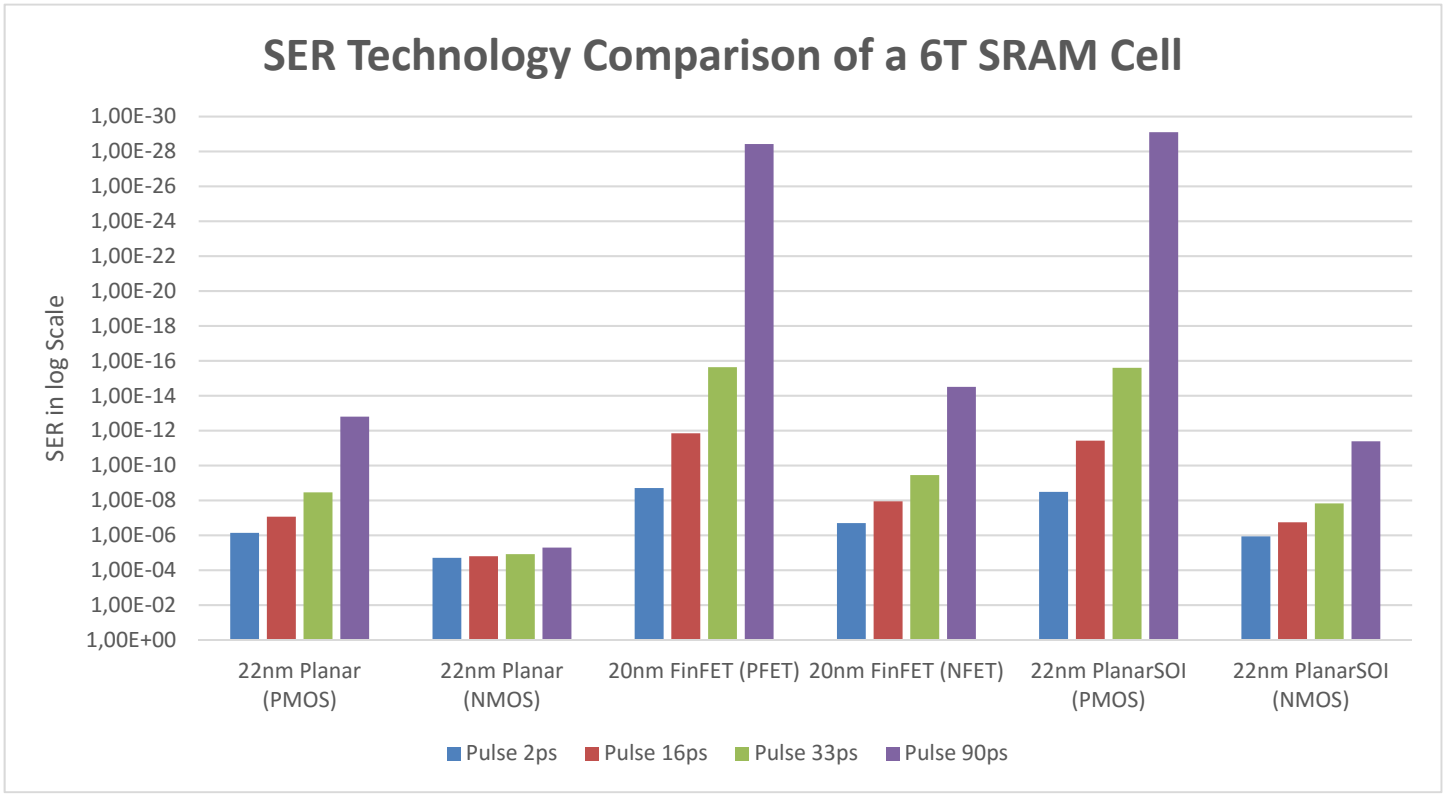
Next steps – Electrical Masking

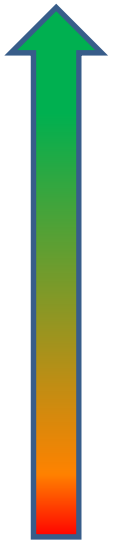
- **Define a vulnerability window for each gate**



Next steps – SER analysis

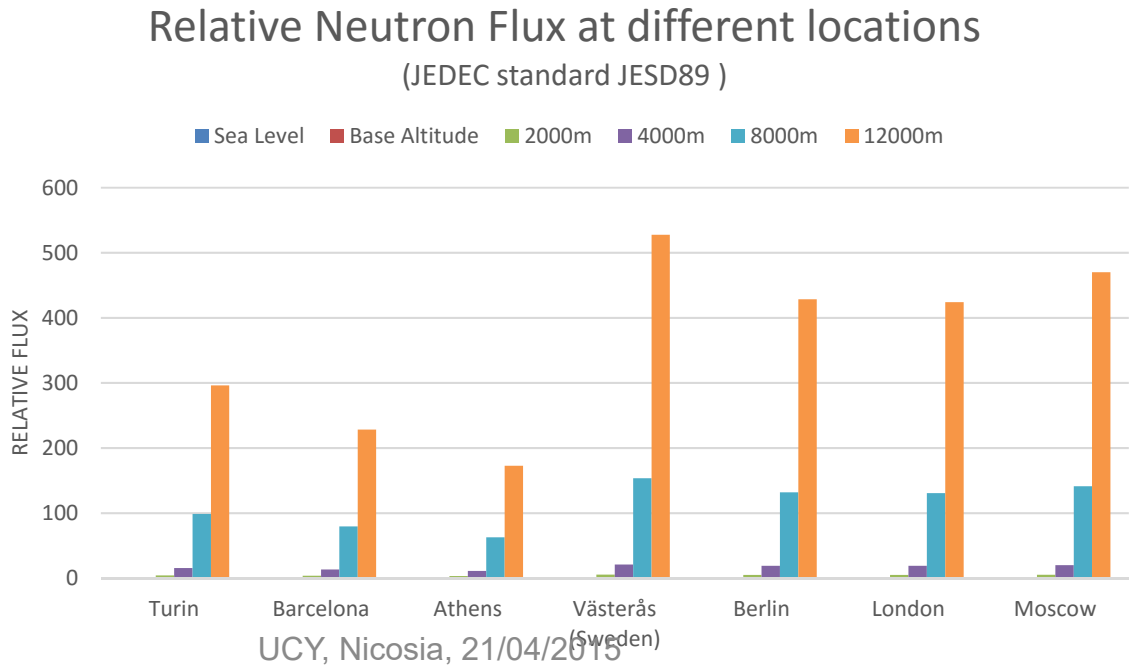
- **Analyze the vulnerability of memories and combinational circuits in front of SEUs**





Next steps – System location

- **Effect on soft-error rate is linearly dependent on neutron flux (and solar activity)**
  - **Small (~20% variation at sea level in Central-South Europe)**
  - **Larger at higher altitudes and closer to the poles**



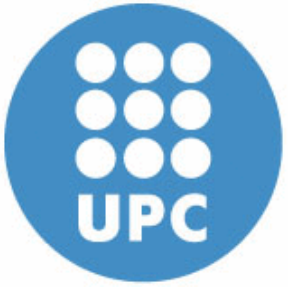
Conclusions

- **Technology-Aware Computer Architecture**
  - **Cross-area fertilization**
  - **Cristal ball to the future**

- **Need to define better "all-inclusive" models for architects**

- **Adapt to system specifications and requirements**
  - **Expand impact and applicability of our findings!**

Cross-layer studies: from technology to circuits and architectures



Ramon Canal

Universitat Politècnica de Catalunya

@RamonCanalUPC