Communications-inspired Design for the Deep Nanoscale Era

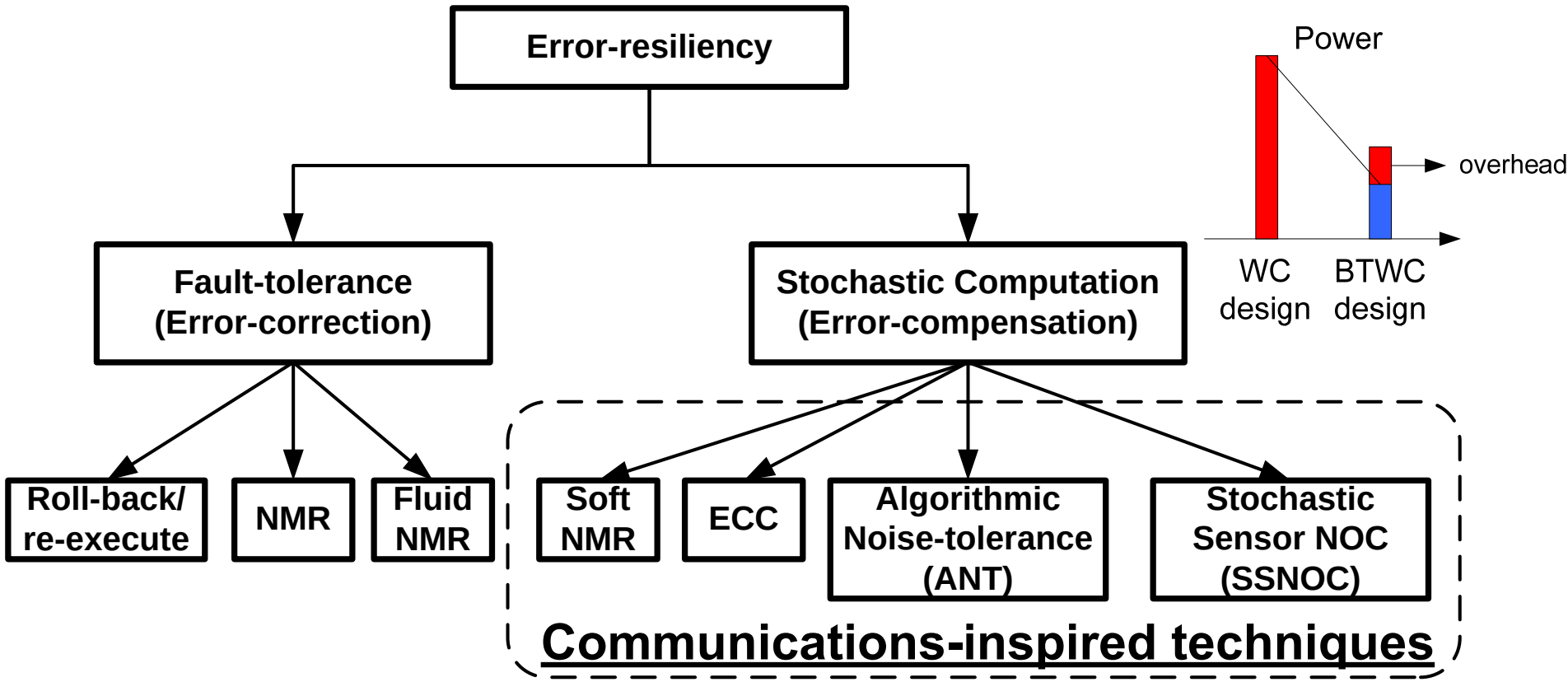
Naresh Shanbhag

Department of Electrical and Computer Engineering
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Assumptions

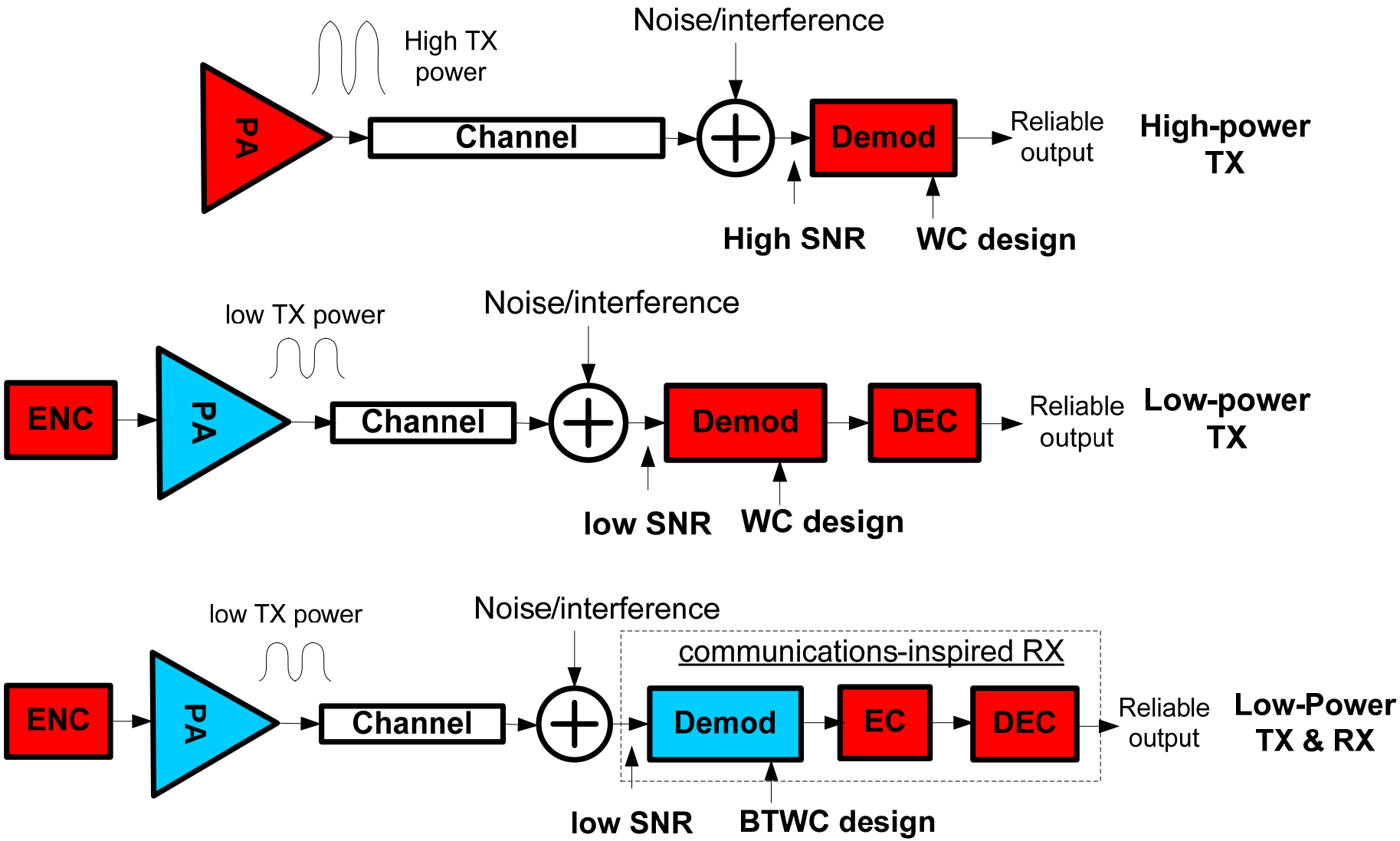
- There is a **Reliability-Power** Problem
  - Reliability → variations, noise, soft-errors, leakage
  - Power → complex apps, V_{dd} scaling floor
- An elegant solution to the Reliability-Power problem → ***error-resiliency*** (rather than error-avoidance)
  - Error-avoidance = worst-case design = expensive design
  - Error-resiliency = better than worst-case (**BTWC**) design
- Applications of interest tolerate **non-zero error-rate** (arbitrarily small) ***/finite* SNR** (arbitrarily large)
  - e.g. VIA, DSP/communications

Error-Resiliency

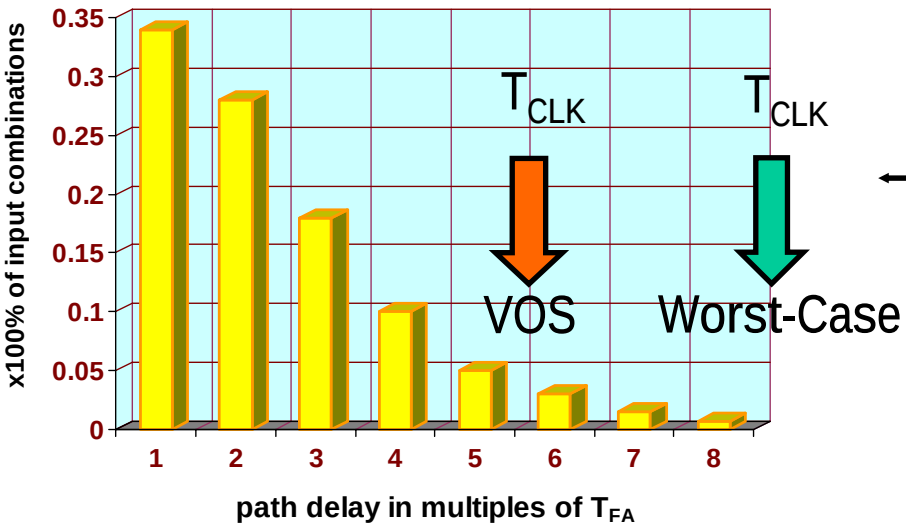


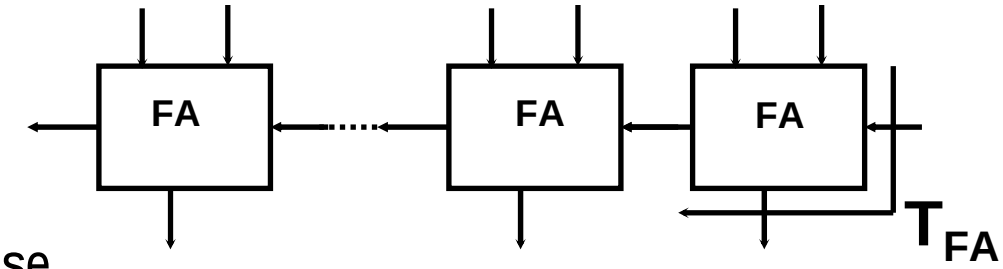
- ER → Power, latency and area overhead
- Communications-inspired techniques → view nanoscale SOCs as communication networks

Communication (-inspired) Design



VOS: Emulating BTWC Design



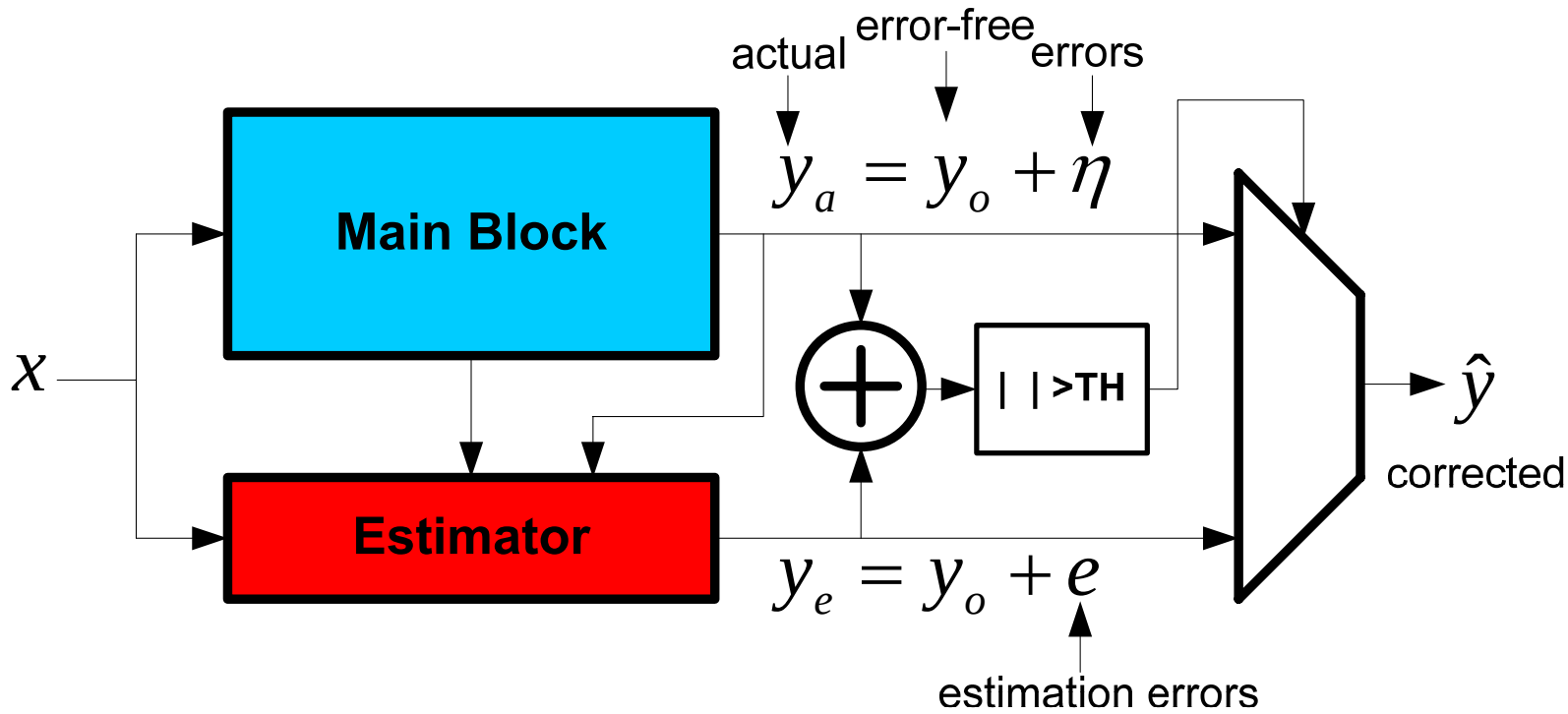


Ripple-carry Adder

Path Delay Distribution (PDD)

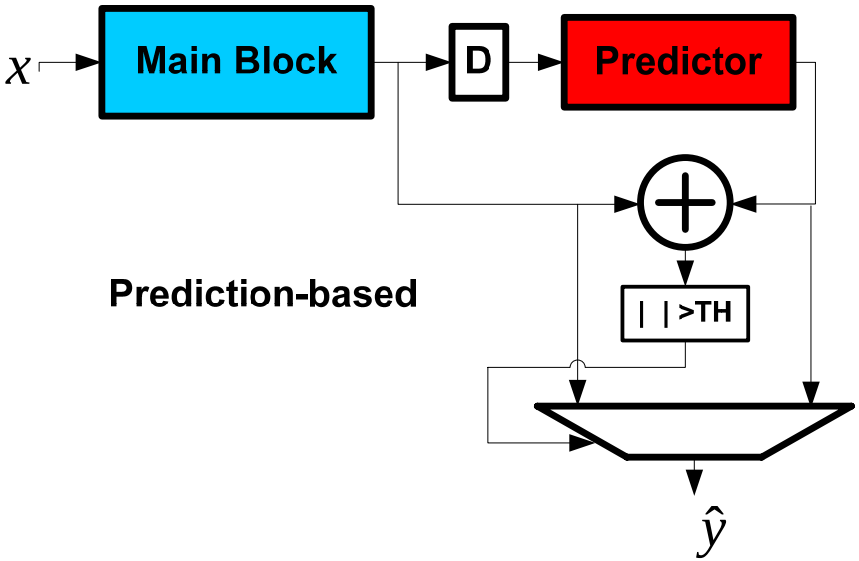
- Worst-case design is expensive → power-hungry
  - Supply voltage is limited by critical path delay
- Voltage overscaling (VOS): permit a few paths to fail
- Low-power but erroneous outputs
- Long-tailed PDD is desirable → low error rate
- Error-compensation friendly architectures

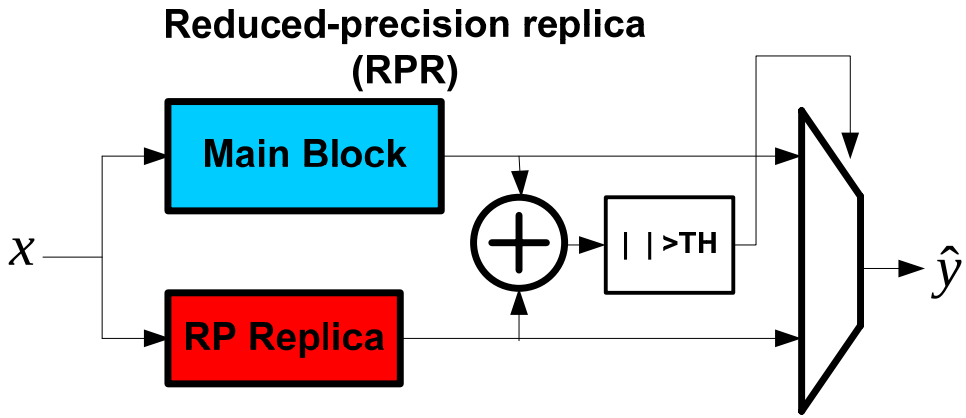
Algorithmic Noise-Tolerance (ANT)

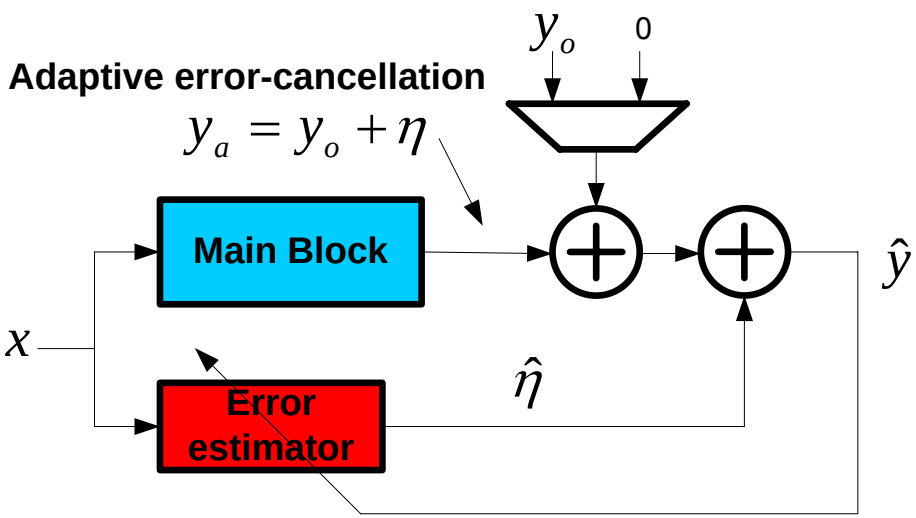


- Employs statistical signal processing techniques
- Main Block is designed for average case
  - Makes intermittent errors
- Estimator approximates Main Block output
- Error-correction: Compare and replace

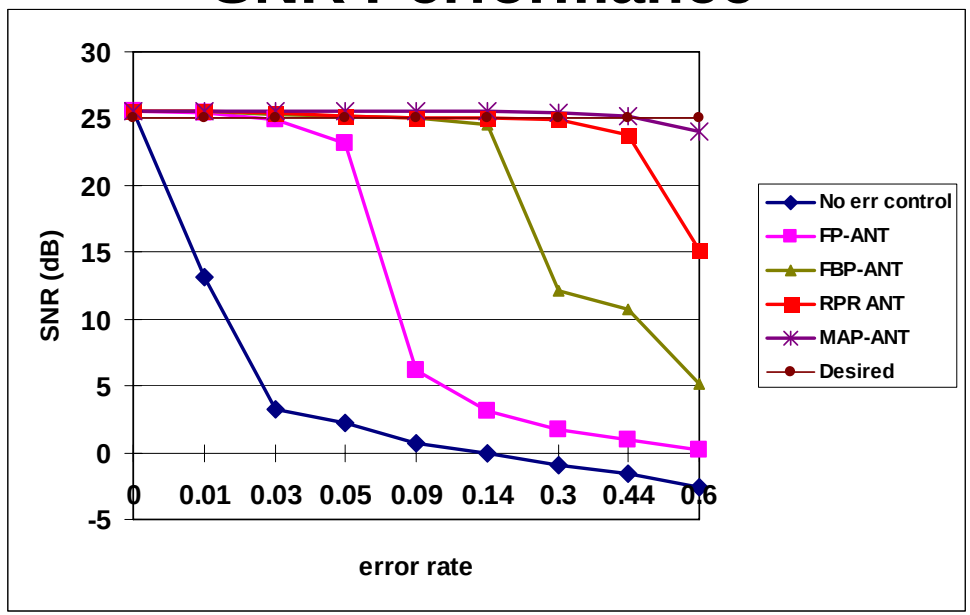
ANT Techniques







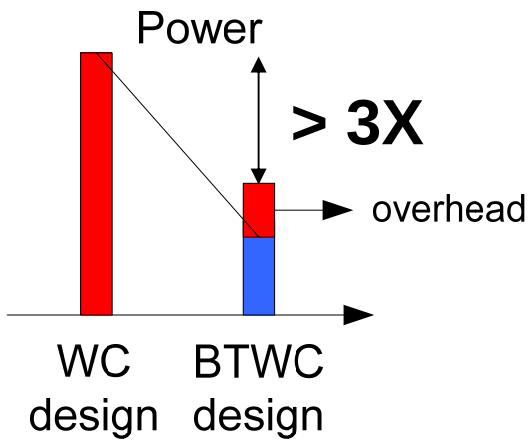
SNR Performance



Power-Latency Overheads

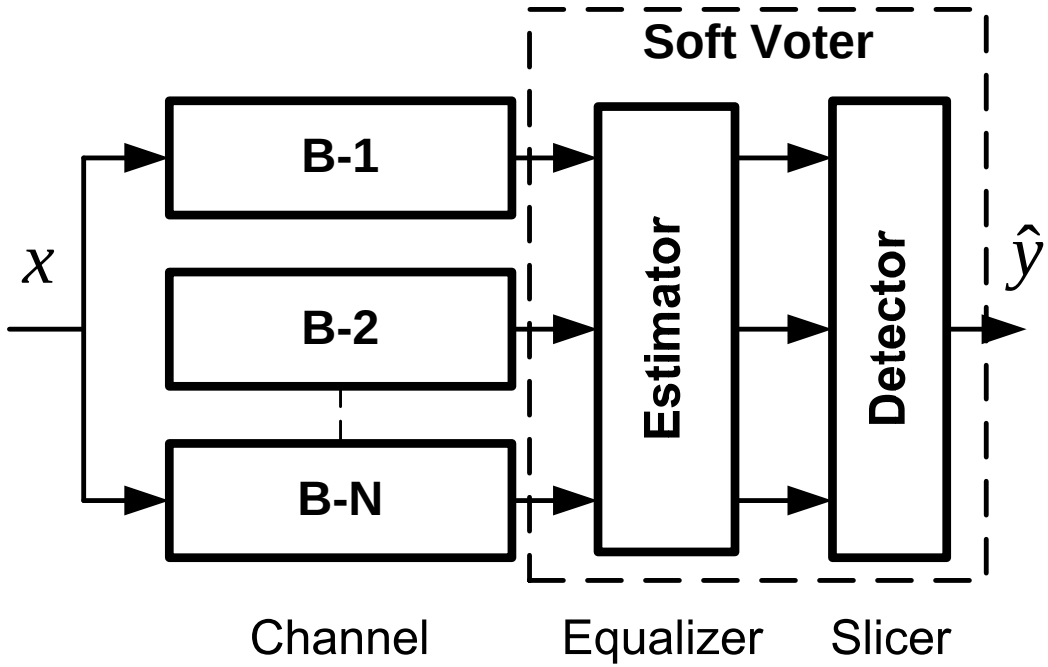
Block	Taps	Mult	Add	Gate – Count	Complexity Overhead%
Main Block	32	16x16	33	34944	0
FP ANT	4	8X8	17	1184	3.4
FPB ANT	5	8X8	17	1480	4.2
RPR ANT	28	8X8	17	7696	22
MAP ANT	28	8X8	17	7800	22.3

FIR filtering



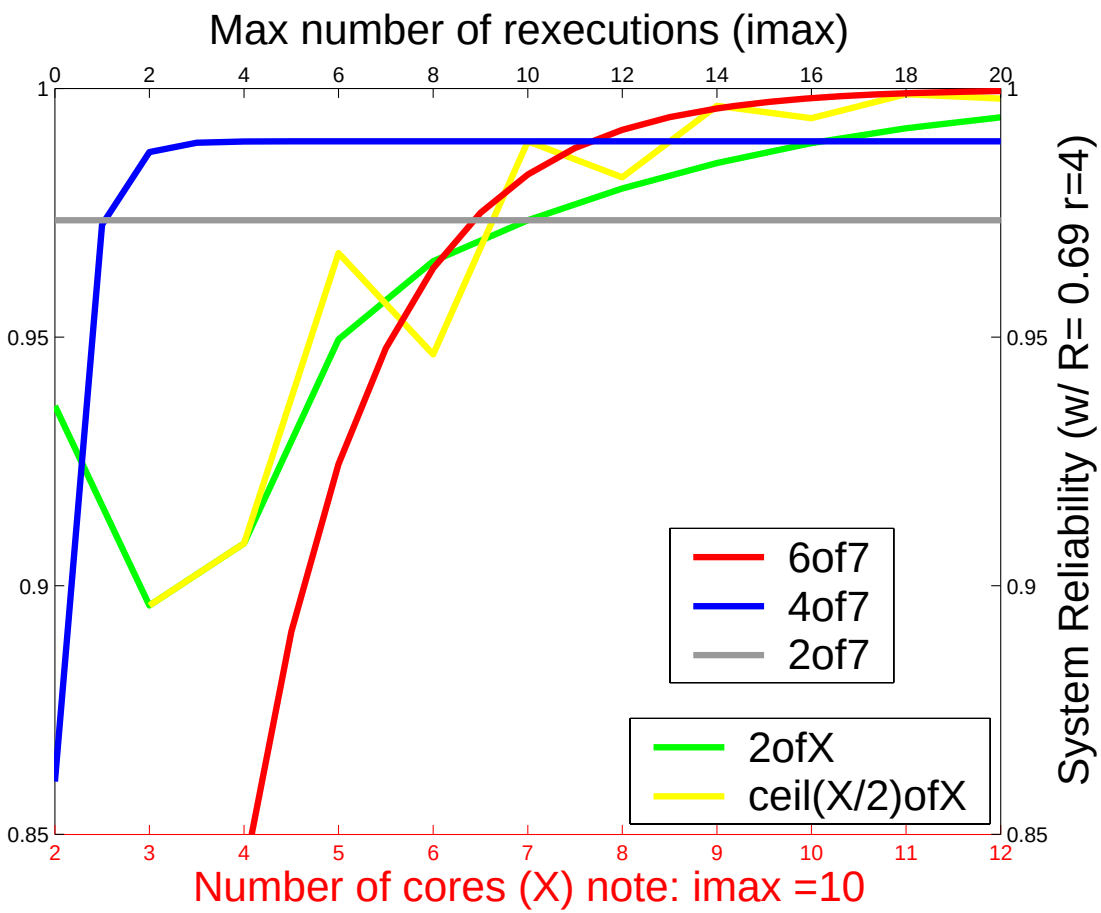
- Complexity overhead 5%-22%
- Power savings: 3X or more
- Latency overhead: $t_{adder}+t_{comp}+t_{mux}$

Soft NMR



- Soft NMR: System reliability 100X better than NMR
  - Makes soft decisions (estimation) then hard (detection)
  - Exploits additional information (e.g., error PDFs)
- Power overhead: (N-1)X+voter
- Latency ovh: t_{voter} (NMR), $t_{estimator}$+$t_{detector}$ (soft NMR)

Fluid NMR

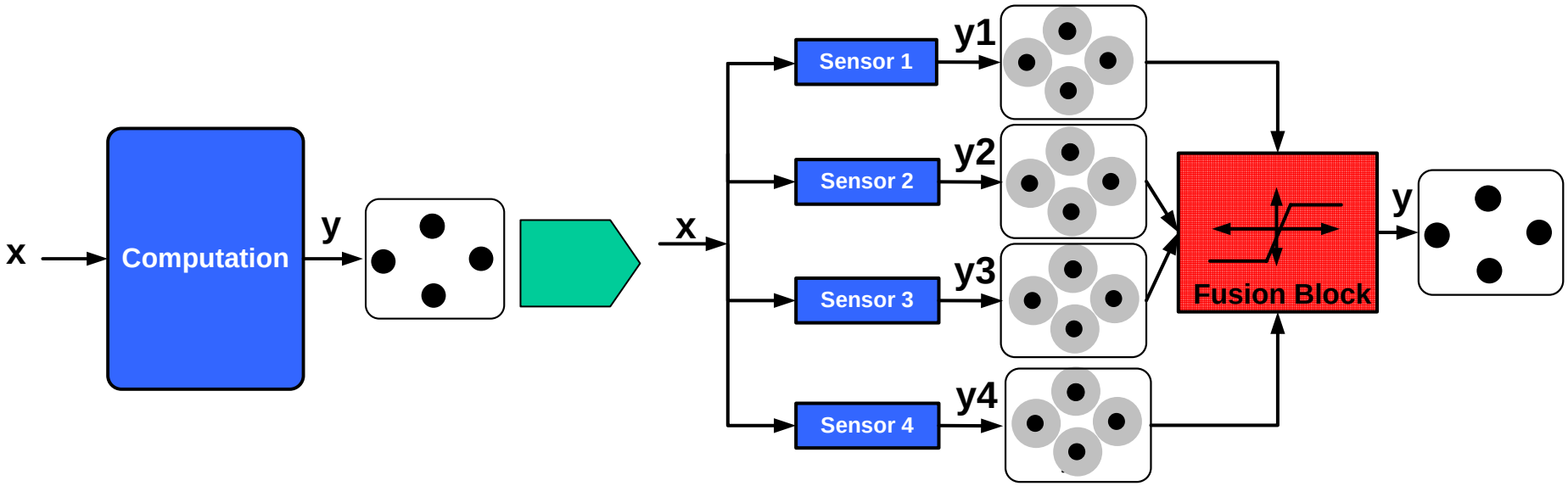


Rakesh Kumar (UIUC)

- Dynamic voting strategy: F(component reliability, latency, size output space)

GSRC Project: Stochastic Sensor Network-on-a-Chip

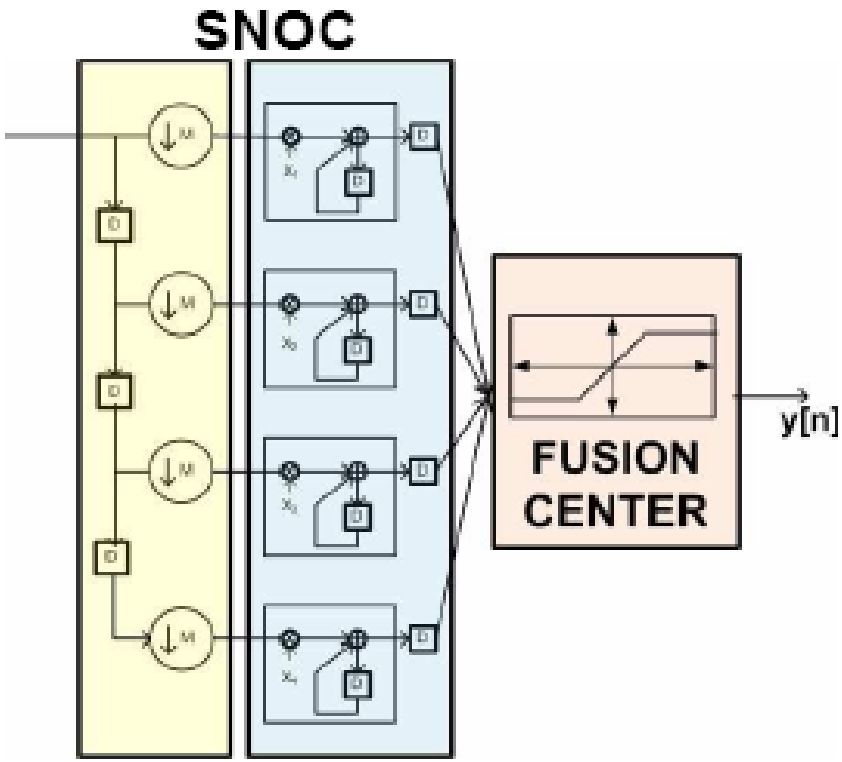
With Doug Jones (UIUC)

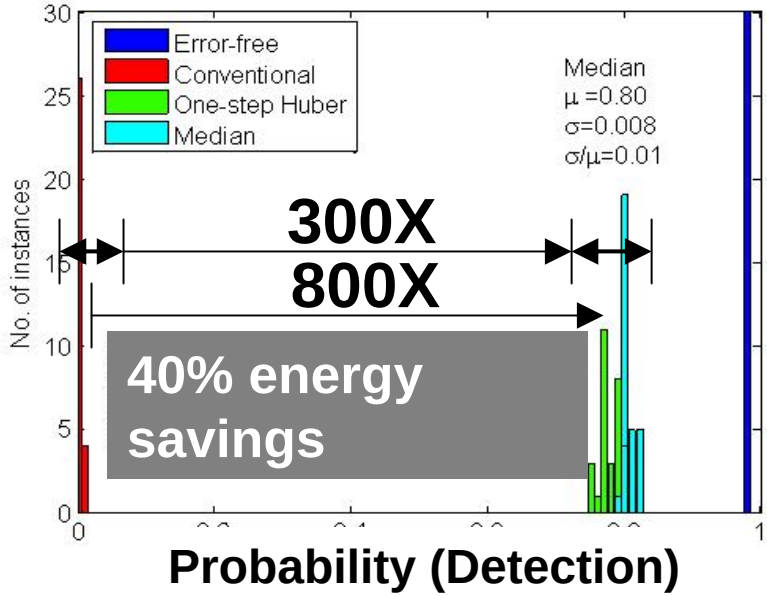


Centralized **Networked**

- Treat computational cores as sensors-on-a-chip
- **Statistically similar decomposition (SSD)**
- Robust fusion
- Overhead = fusion block complexity

SNOC-based PN-Code Acquisition





- Commonly wireless CDMA receiver kernel
- Polyphase decomposition
- 800X (better performance), 300X (reduced performance variation), 40% (energy savings)

Summary

- Communications-inspired design paradigm
  - Treat SOCs as miniature communication networks
  - Energy-reliability-latency trade-offs via error-compensation
  - Applications to on-chip computation, communication, storage
- Algorithmic noise-tolerance (ANT)
  - Statistical signal processing for error-compensation
- Stochastic Sensor Network-on-a-Chip (SSNOC)
  - GSRC research → distributed computation
- Soft NMR, Fluid NMR → stochastic processors for VIA