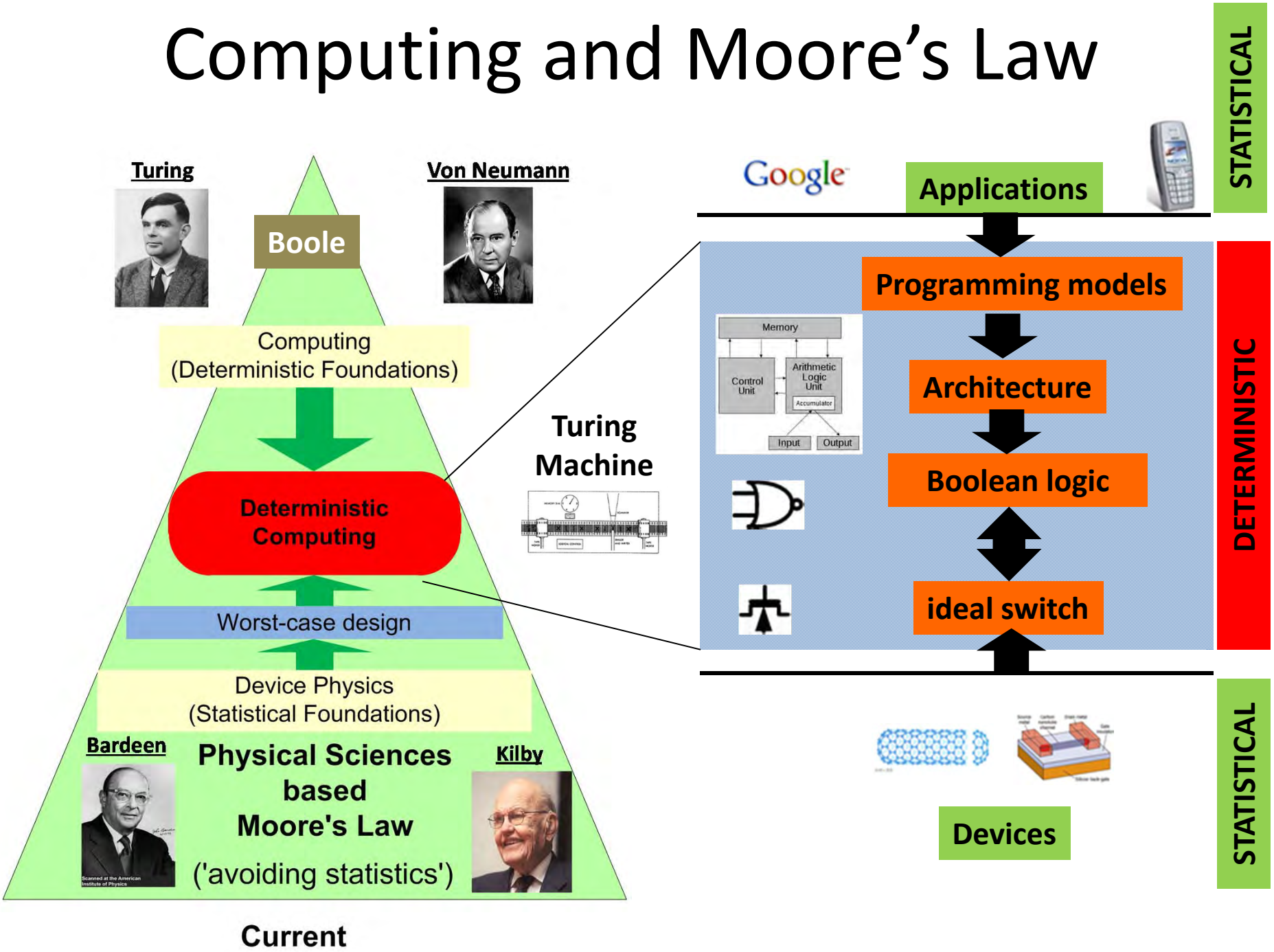
Stochastic Computing: A Design Sciences Approach to Moore's Law

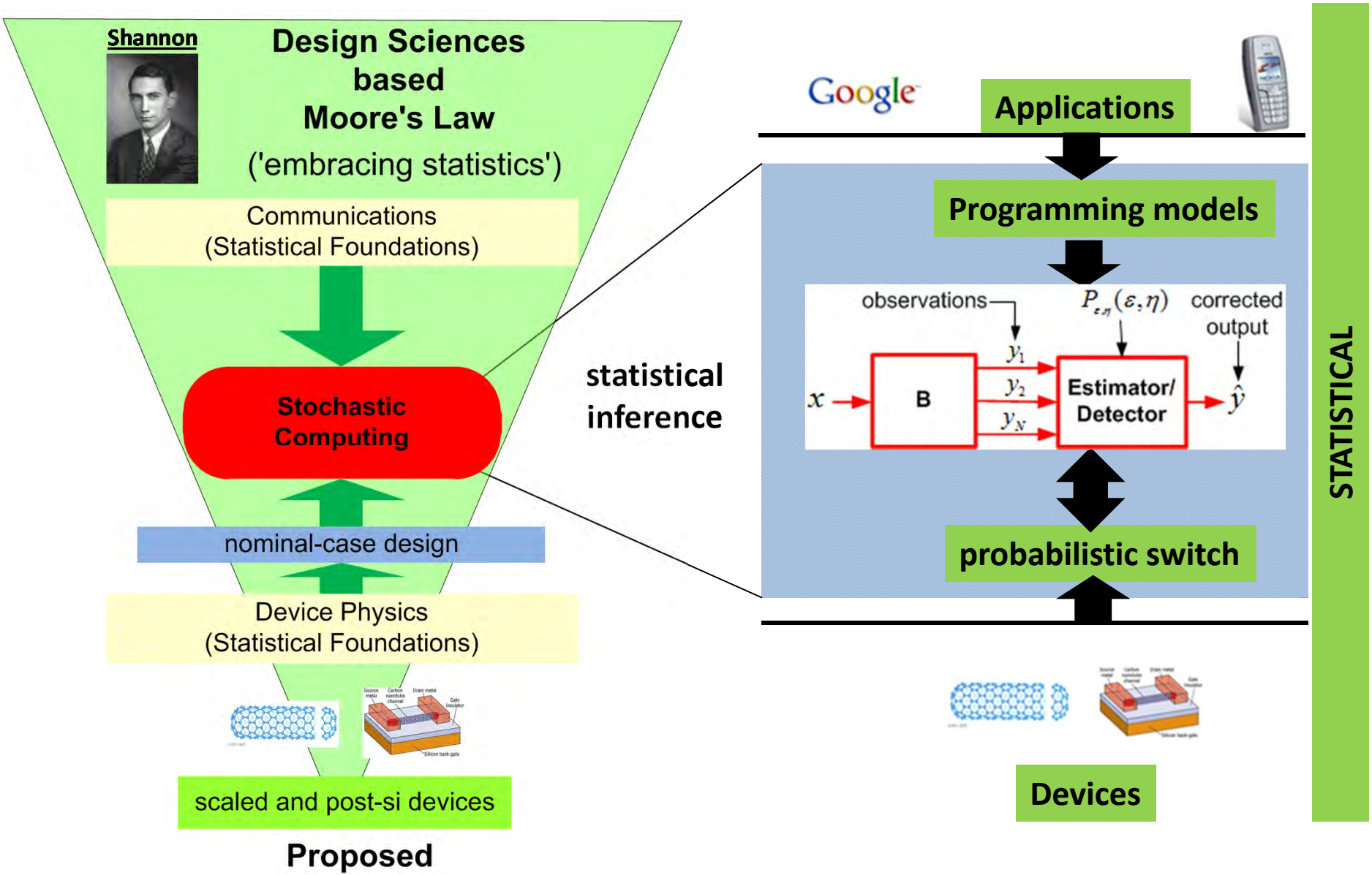
Naresh Shanbhag

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

Computing and Moore's Law



Stochastic Computing and Moore's Law



Von Neumann (1956)

J. Von Neumann, *Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components*, Princeton University Press (1956)

....treatment of error is *unsatisfactory and ad hoc*
*....error should be treated .....*as *information has been*, by the work of L. Szilard and C. E. Shannon.

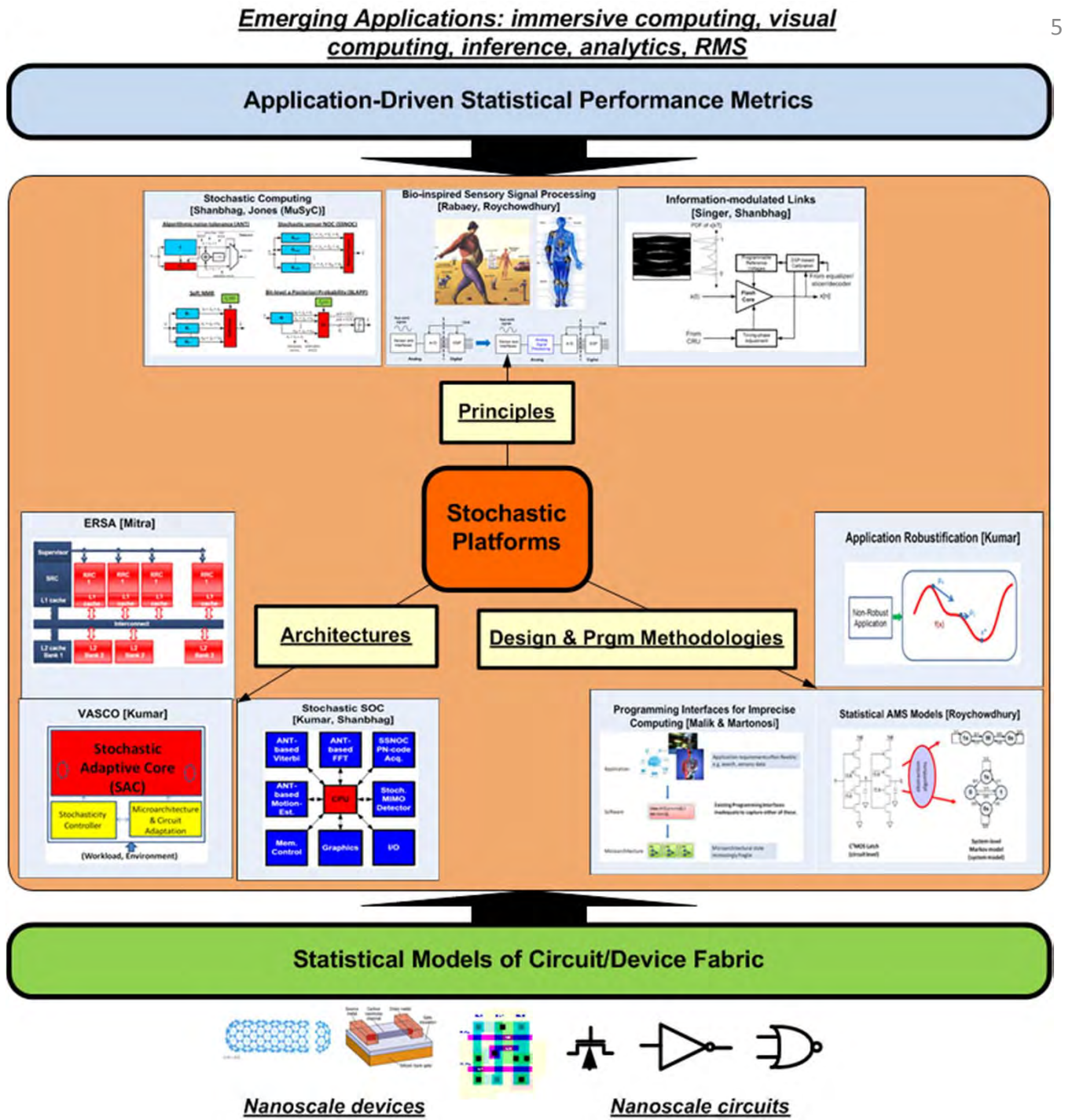
The present treatment *falls short of achieving this*.....

communications-inspired (stochastic) computation

Alternative Computational Models

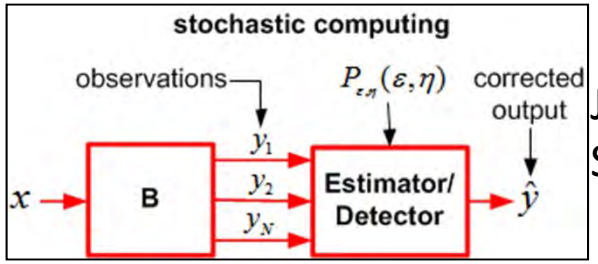
Research theme in the Gigascale Systems Research Center (GSRC)

Sponsors: DOD & SRC



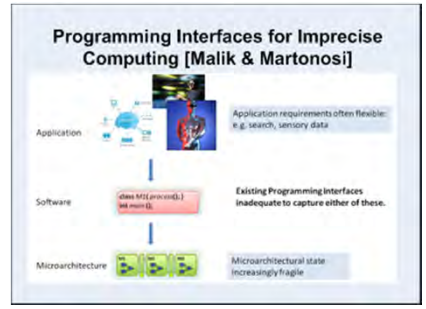
Stochastic Computing Panorama

Theoretical Foundations



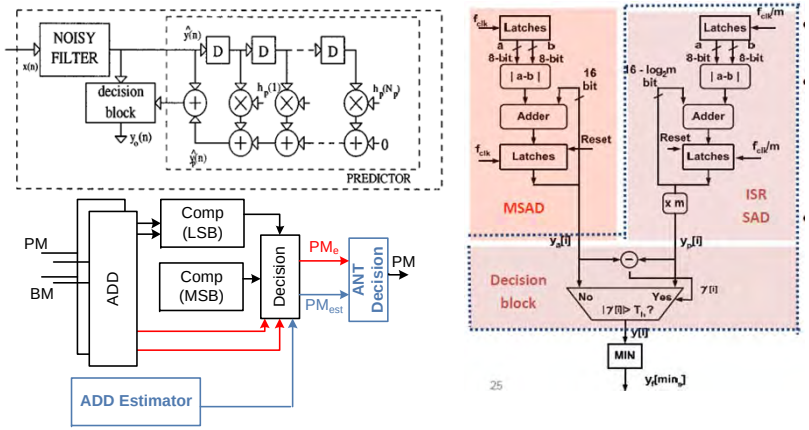
Jones (UIUC)
Singer (UIUC)

Design Methodologies



R. Kumar (UIUC),
Roychowdhury (UCB)
Malik (Princeton)

Application-specific Kernels

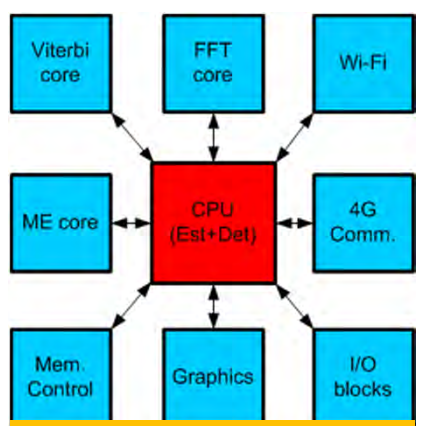


Kernels:FIR, FFT, Viterbi, video, comm., ML

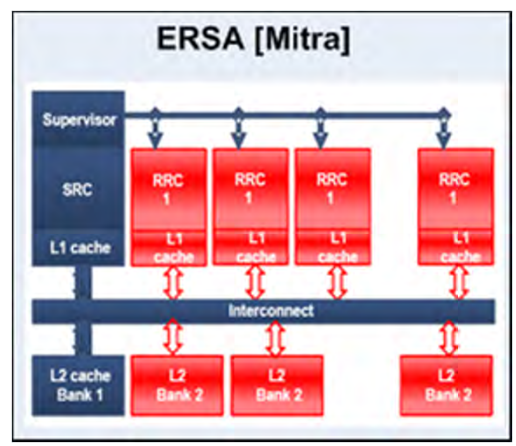
FPGA Prototypes

BEE2- ERSA

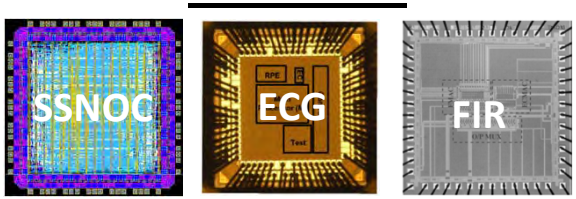
Stochastic Processors

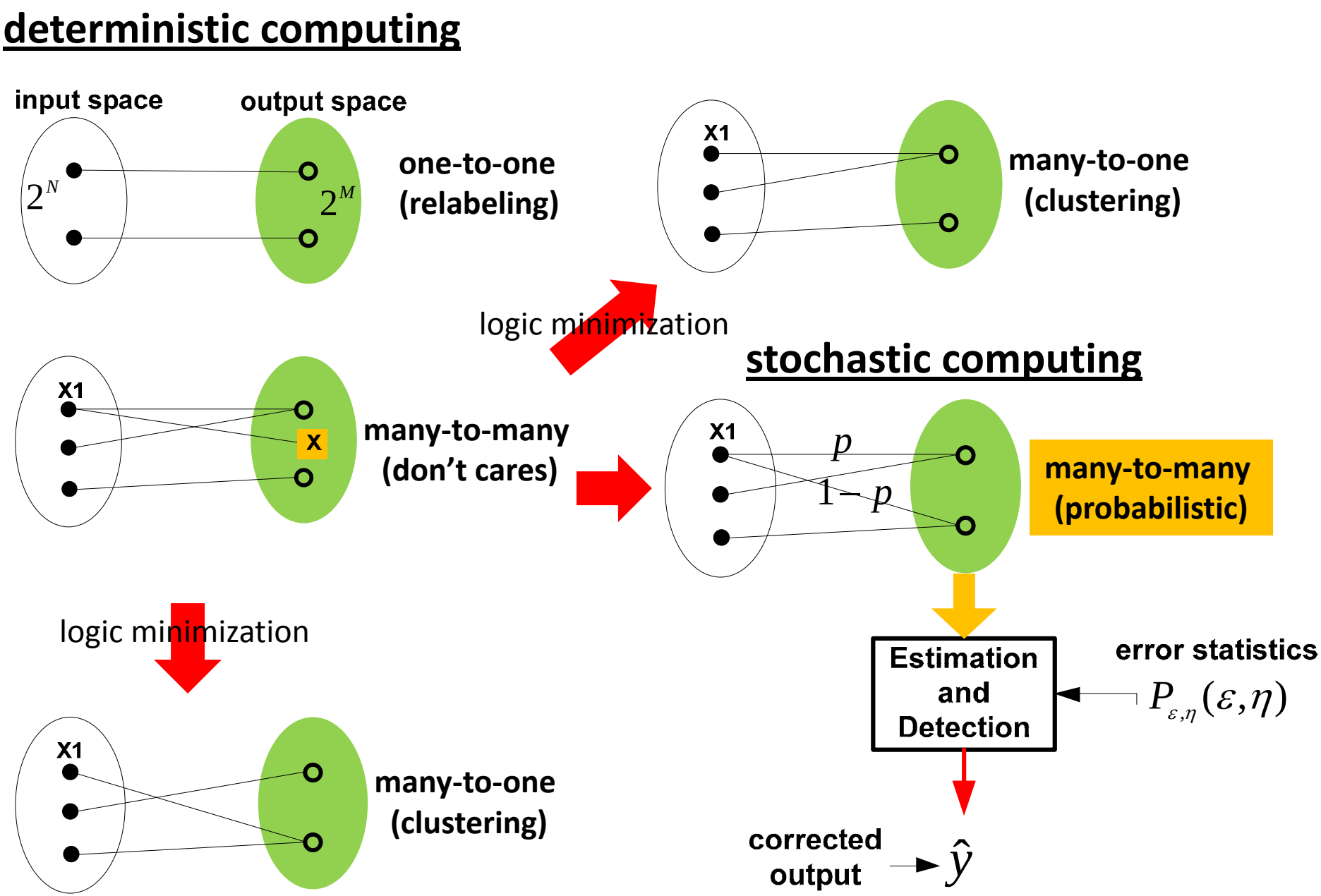


Stochastic SOC

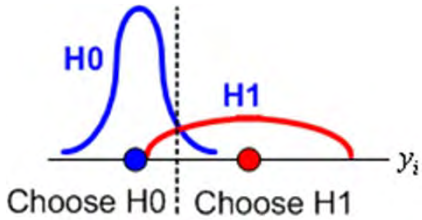


Custom ICs




deterministic computing
input space
output space
2^N
2^M
one-to-one
(relabeling)
X1
many-to-one
(clustering)
logic minimization
stochastic computing
X1
X
many-to-many
(don't cares)
X1
p
$1-p$
many-to-many
(probabilistic)
logic minimization
Estimation
and
Detection
error statistics
$P_{\varepsilon,\eta}(\varepsilon,\eta)$
X1
many-to-one
(clustering)
corrected
output
$\hat{y}$

Statistical Estimation & Detection



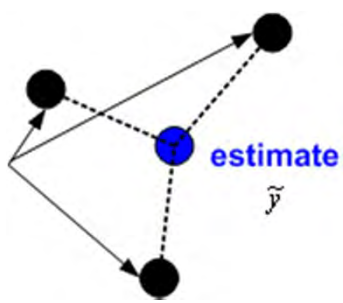
$$\hat{y} = \arg\max_{H_i} P(y_1, y_2, \ldots, y_N \mid H_i)$$

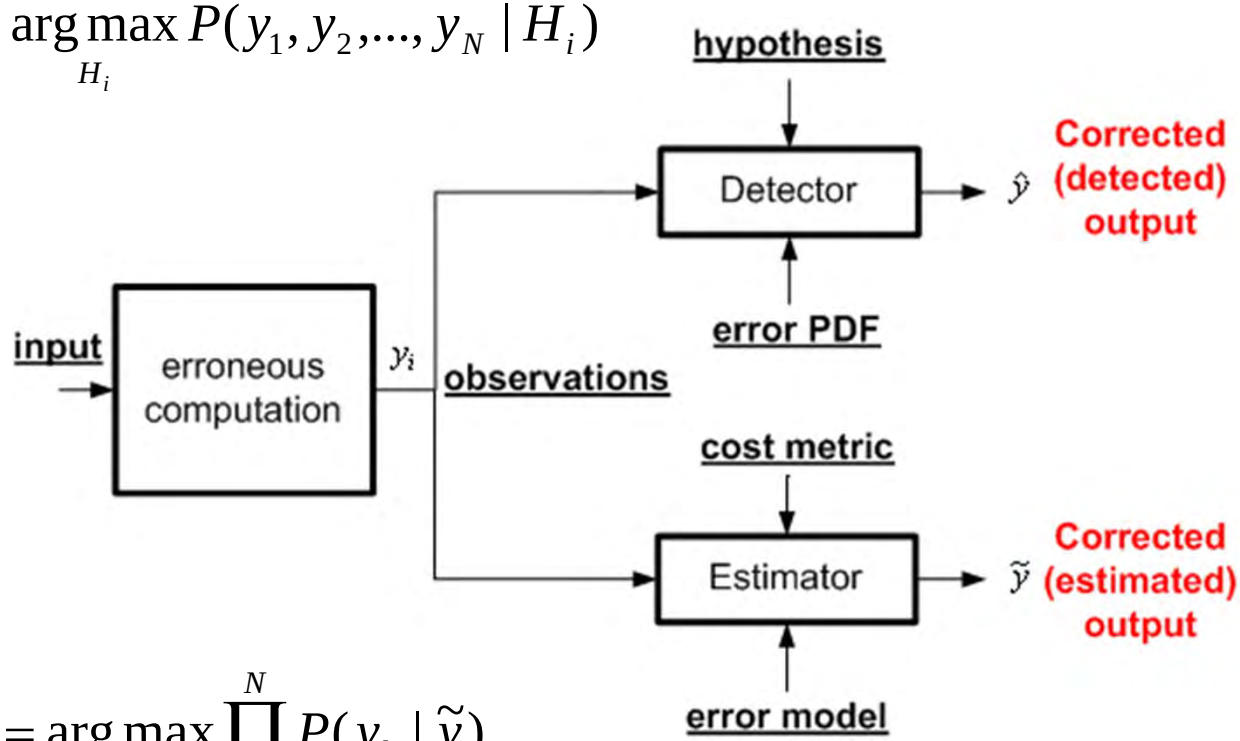
correct output

error

$$y_i = y_0 + e_i(x, y_0, V_{dd}, f)$$

$$P(e_i) : \text{PDF of } e_i(x, y_0, V_{dd}, f)$$



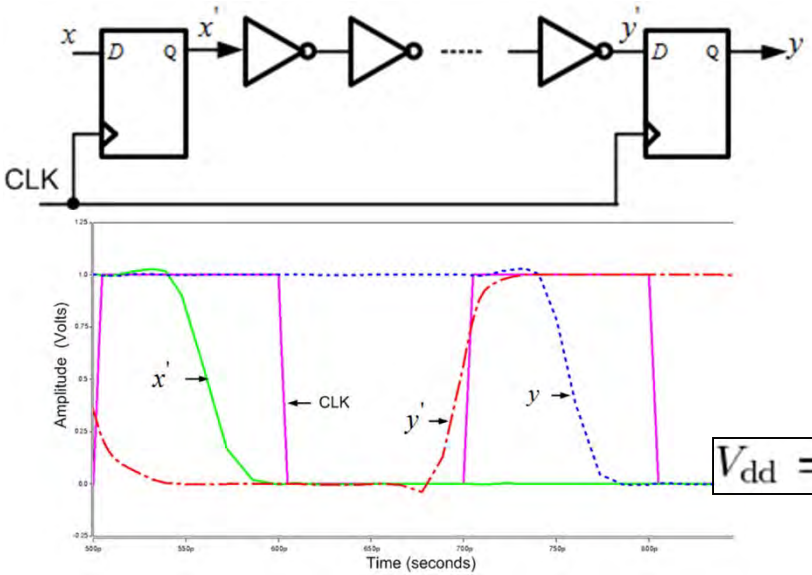


$$\tilde{y} = \arg\max_{\tilde{y}} \prod_{k=1}^{N} P(y_k \mid \tilde{y})$$

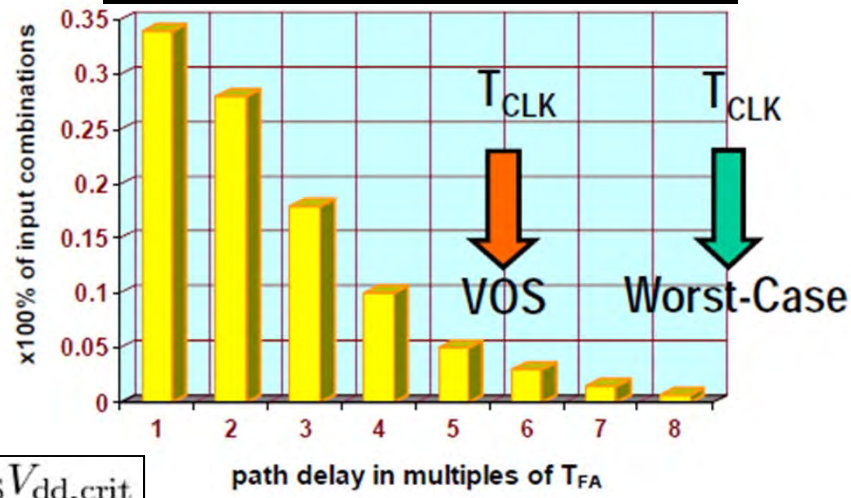
Metrics: maximum a posteriori probability (MAP), maximum likelihood (ML), minimum mean-squared error (MMSE), minmax, minimum absolute error

Error Statistics

VOS induced timing violation



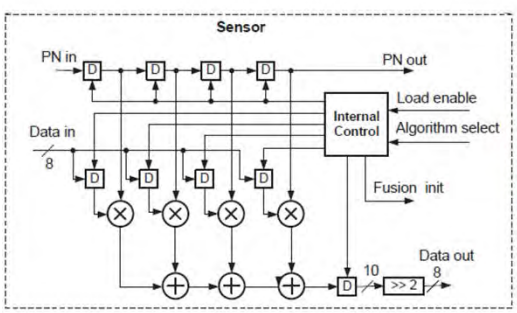
Path delay distribution (8b RCA)



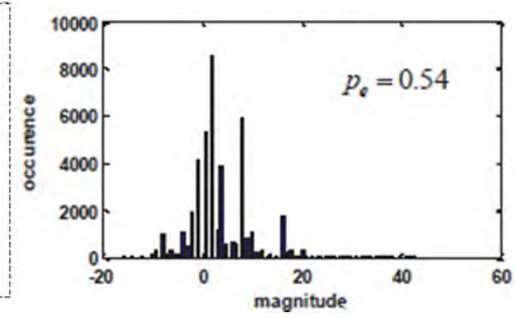
$$V_{dd} = K_{VOS} V_{dd,crit}$$

effective error rate vs. energy trade-off
engineer circuit error statistics
prefer long-tailed PDDs

FIR filter in 180nm

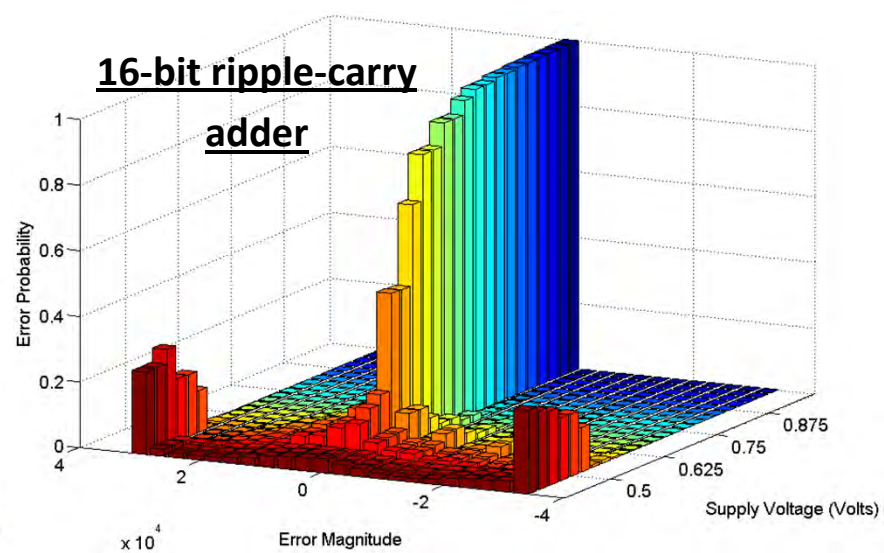


Measured error PMF at Vdd = 0.76V



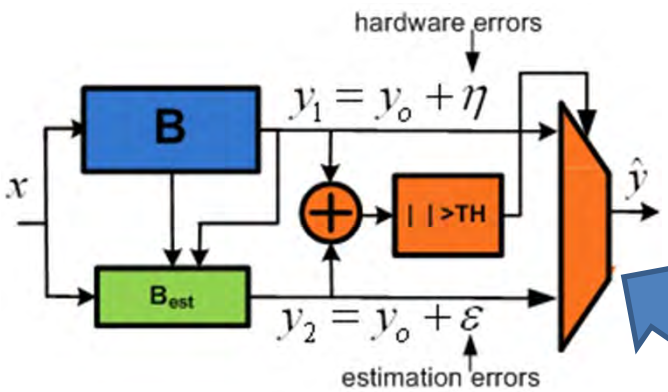
Error Probability Mass Function (PMF)

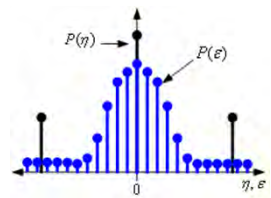
16-bit ripple-carry adder



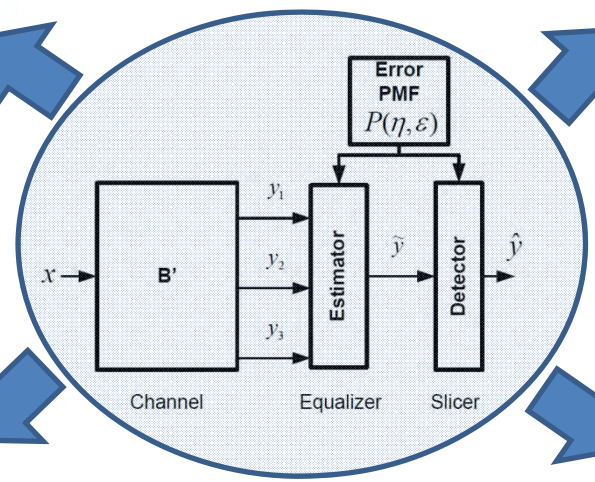
Stochastic Computing Techniques

Algorithmic noise-tolerance (ANT)

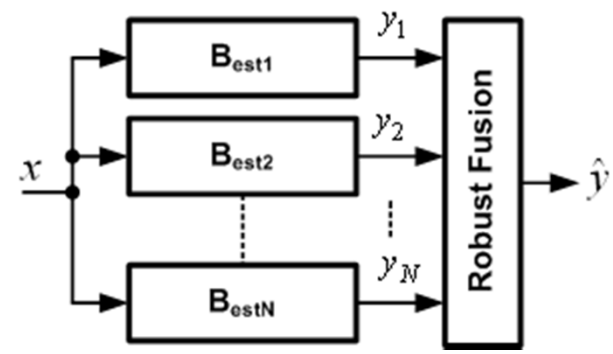




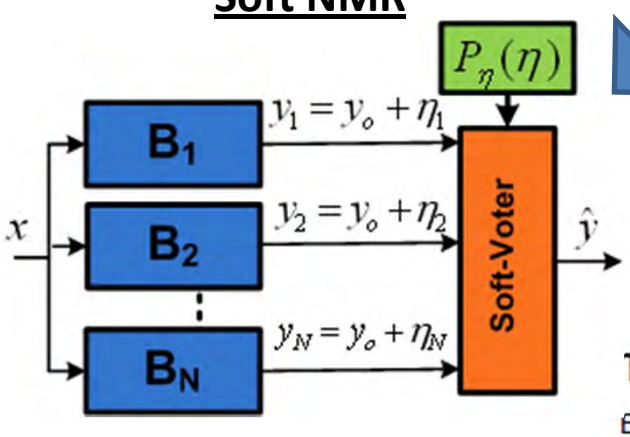
Stochastic Computing Framework



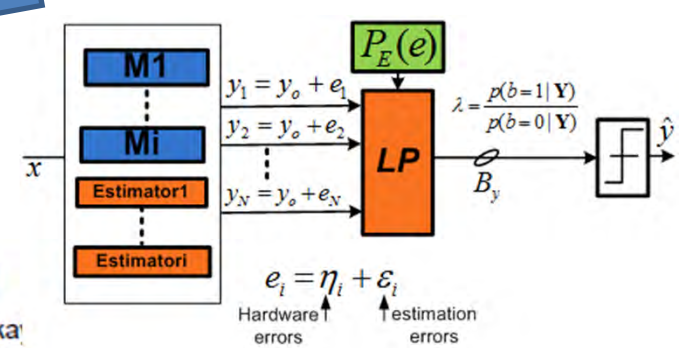
Stochastic sensor NOC (SSNOC)



Soft NMR



Likelihood Processing



IEEE Spectrum, Nov. 2010

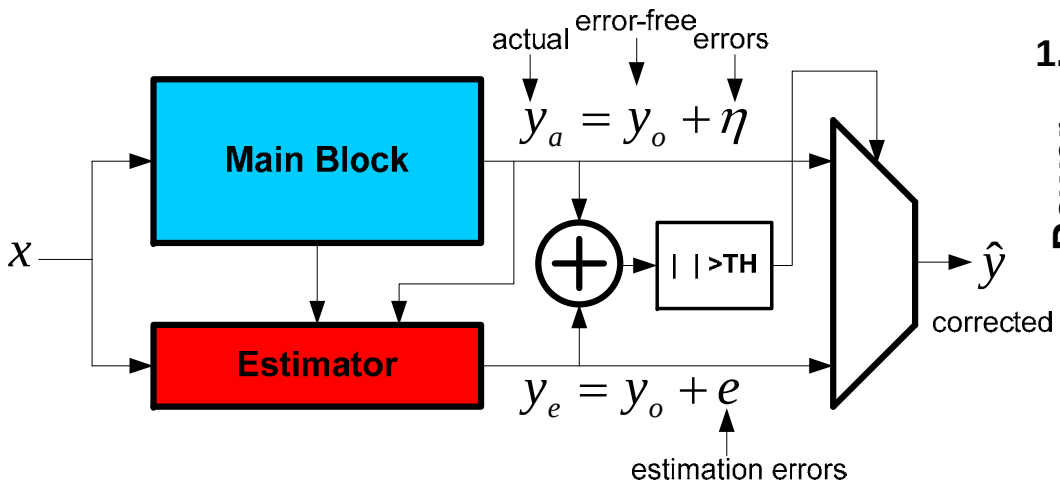
The Era of Error-Tolerant Computing

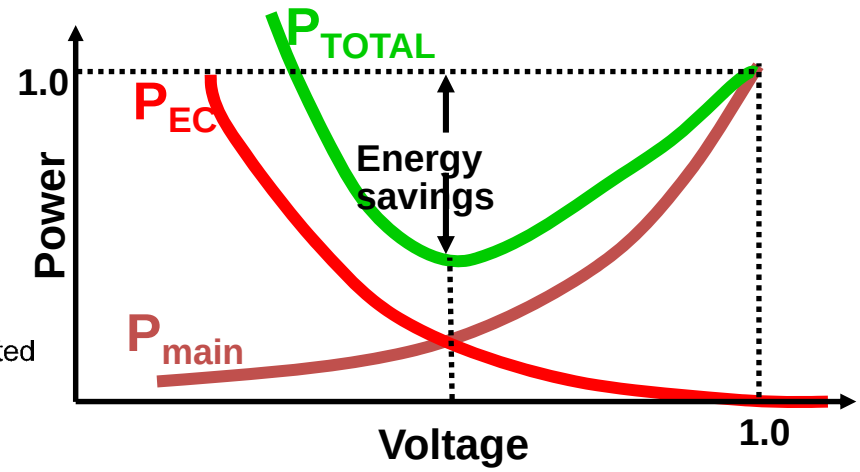
Errors will abound in future processors...and that's okay

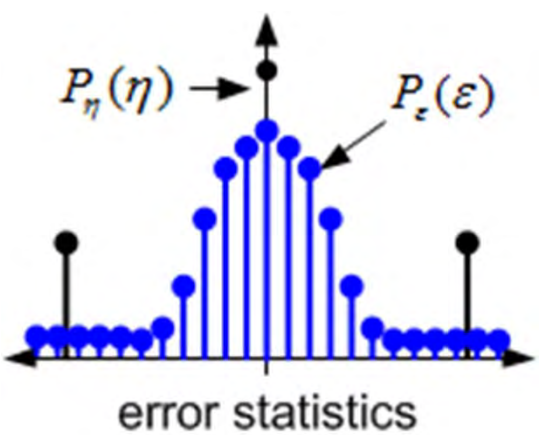
By DAVID LAMMERS / NOVEMBER 2010

Algorithmic Noise-Tolerance (ANT)

[Hegde, Wang, Shim, Varatkar, Abdallah]







$$SNR_{uc} \ll SNR_e \ll SNR_{ANT} \cong SNR_o$$

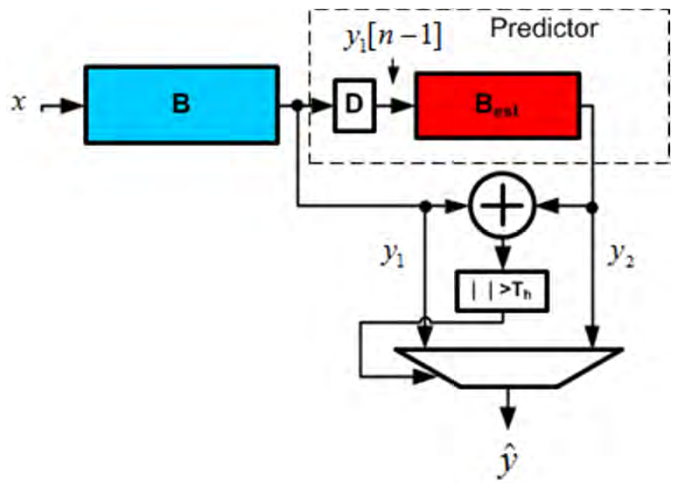
high error-rates (up to 60%)

overhead (gate-count): 5%-22%

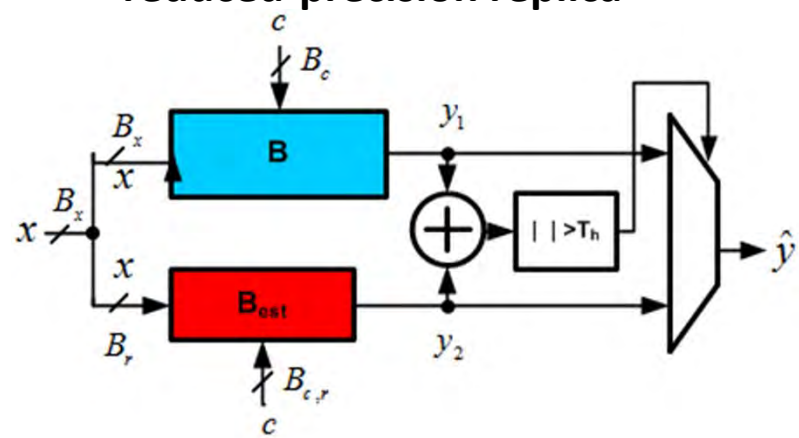
energy savings: 40%-70% (<1dB SNR loss)

ANT Techniques

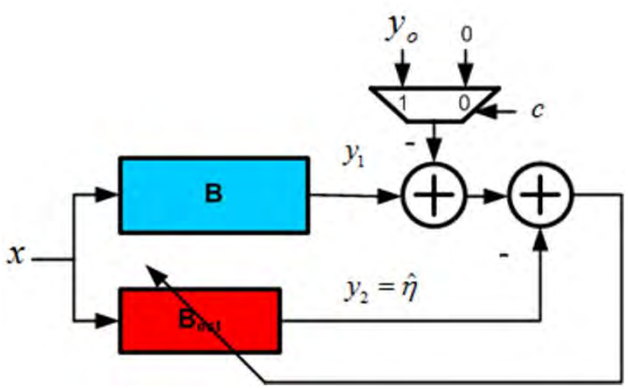
prediction-based



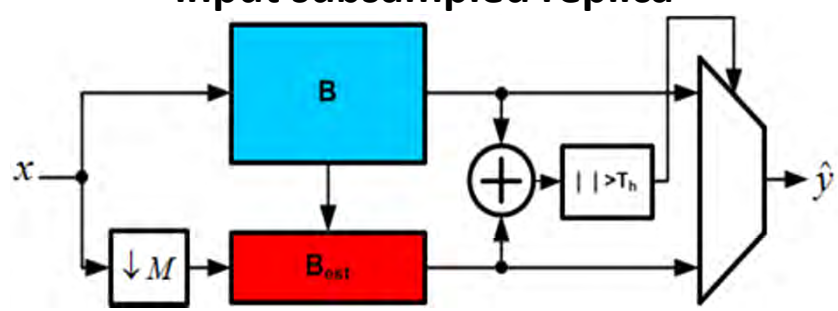
reduced-precision replica



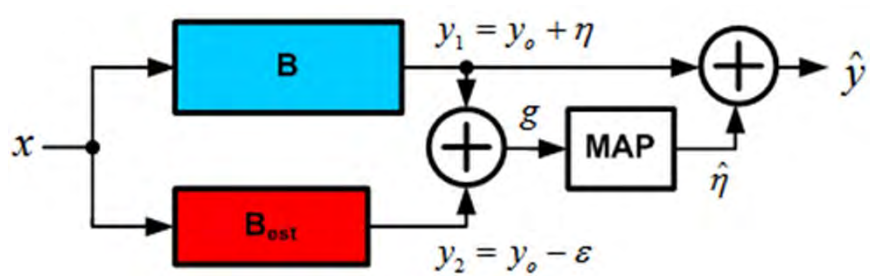
adaptive error-cancellation



input subsampled replica

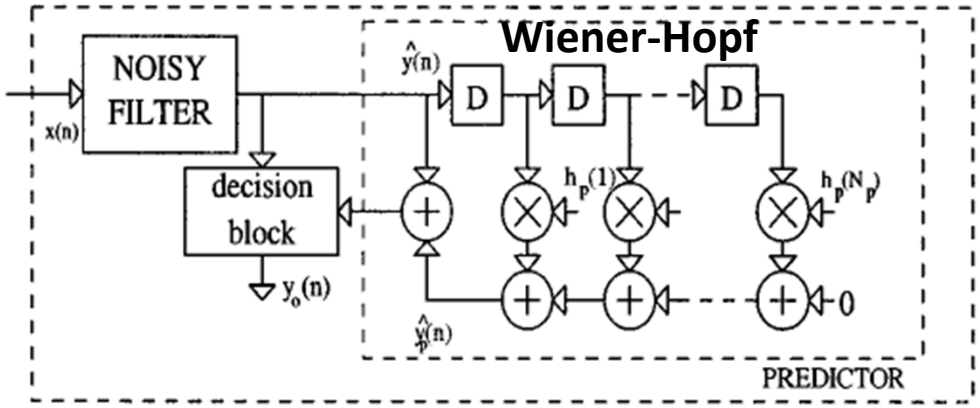


maximum a posteriori (MAP)



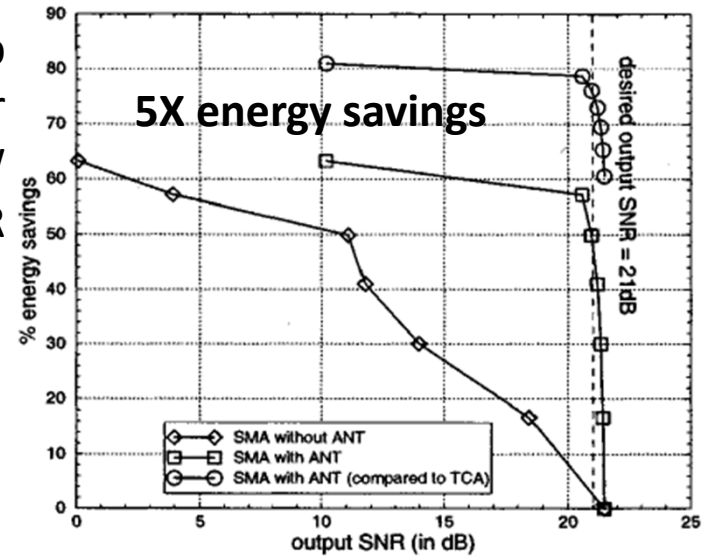
ANT-based Error-resilient FIR Filter

Prediction-based ANT (Hegde)

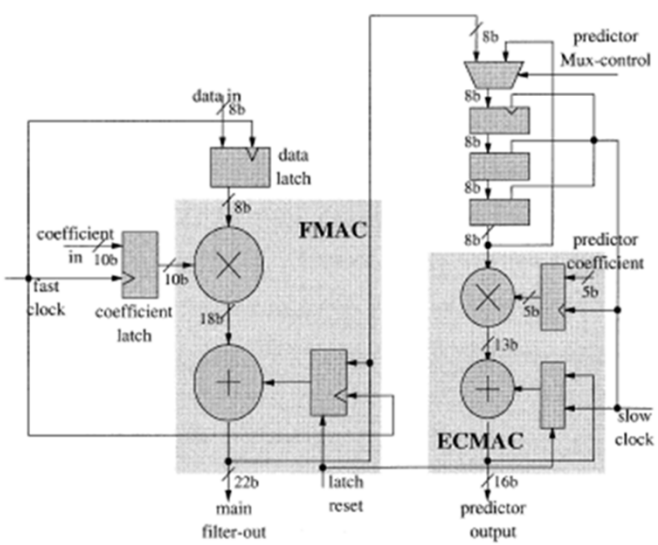


3-tap predictor
0.5π BW
29-tap FIR

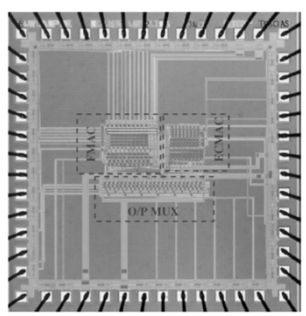
Simulation results



Chip architecture



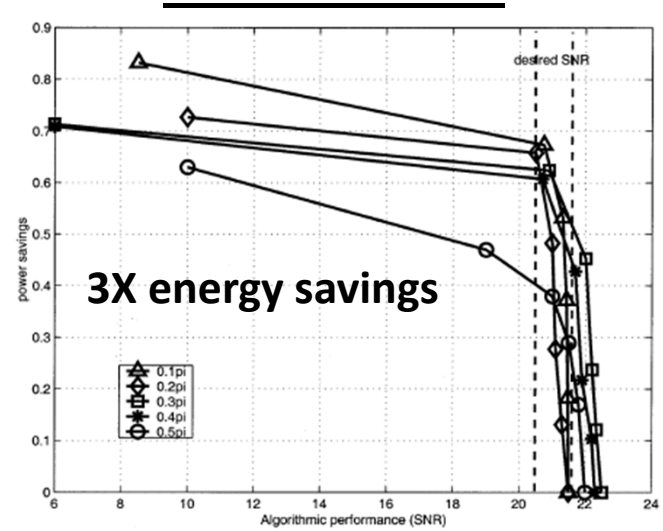
microphotograph



0.35μm
3.3V CMOS
32-tap FIR

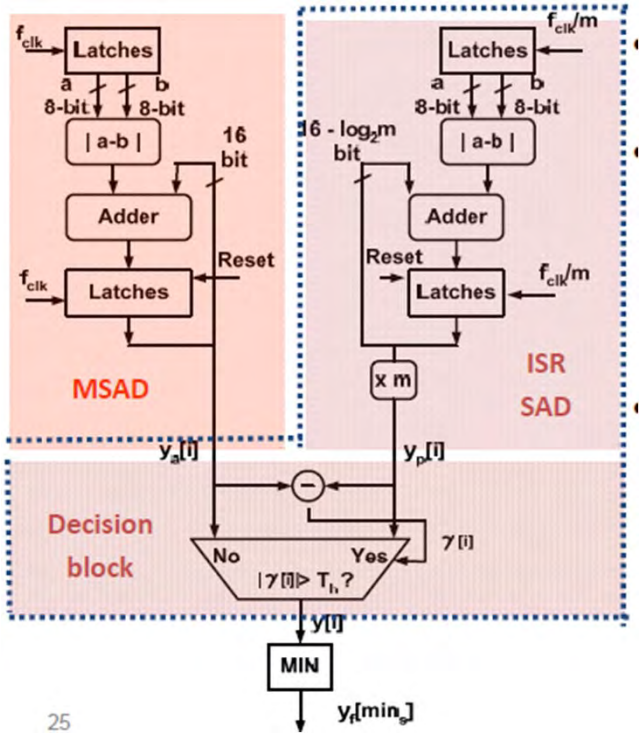
FMAC: 88MHz;
ECMAC: 11MHz;
$V_{dd\text{-}crit}$ =3.55V;2.25V
$V_{dd\text{-}min}$ =2.32V

measured results

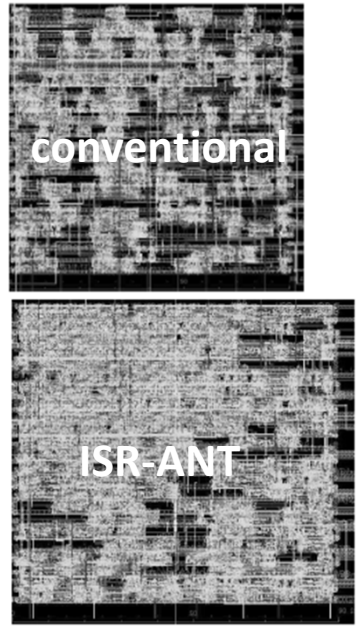


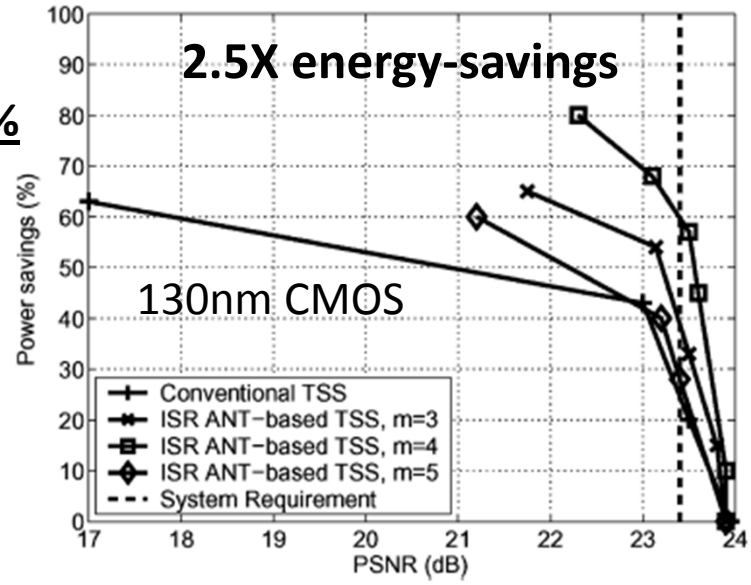
Error-resilient Motion Estimation

<u>input sub-sampled replica (ISR-ANT)</u> (Varatkar)

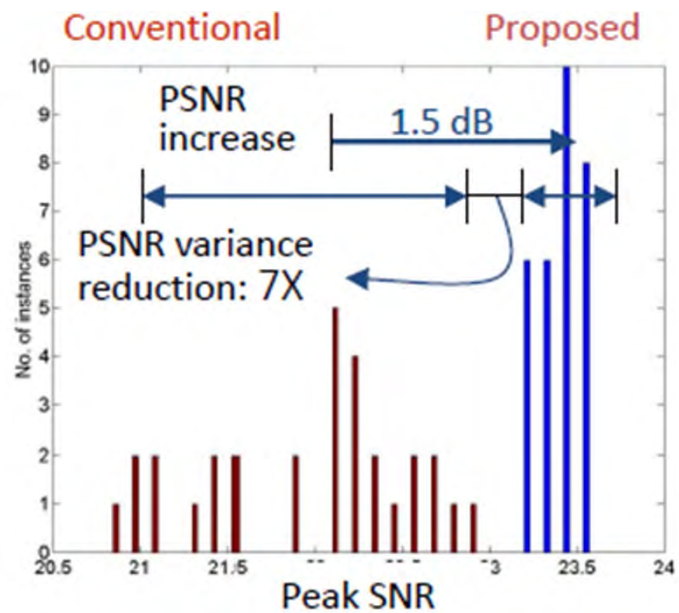


<u>area overhead = 26%</u>

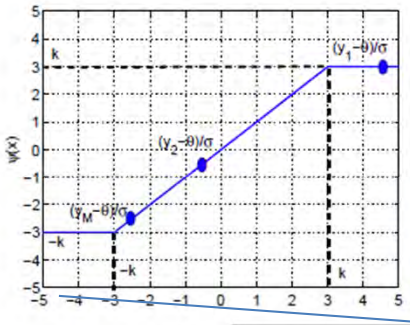




ideal conventional proposed



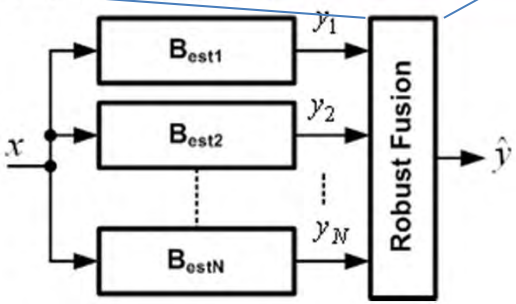
Stochastic Sensor Network-on-a-Chip



robust estimation

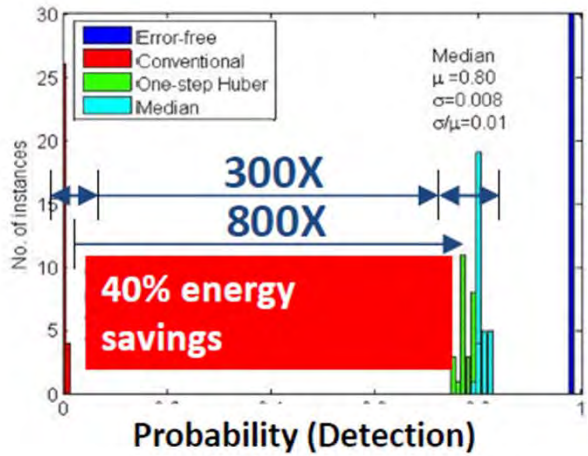
observations → $Y_i = \theta + \eta_i$ ← noise

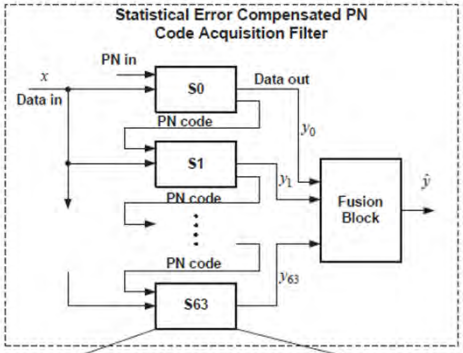
θ ← estimate

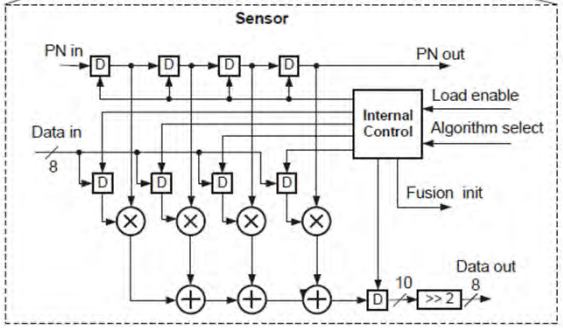


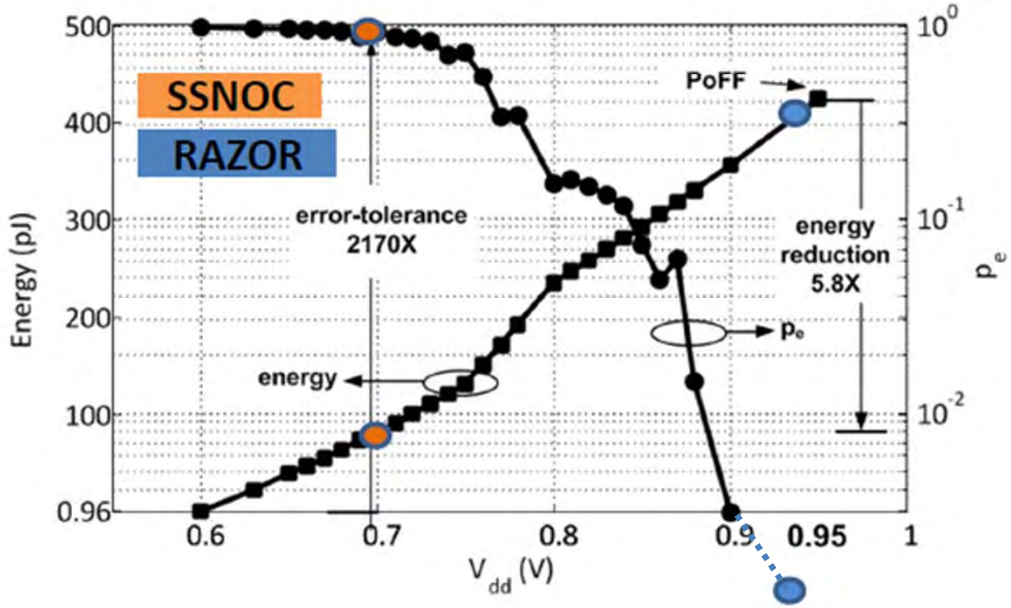
[Varatkar,Narayanan,Jones]

(sim) 130nm CMOS
WID process variations



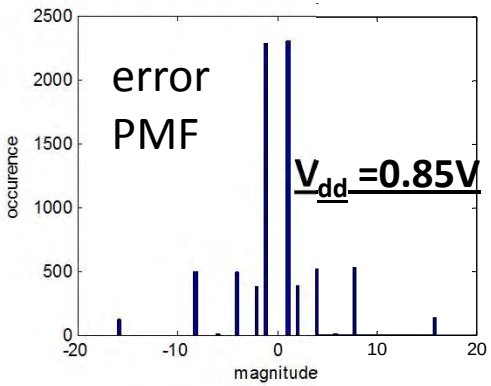


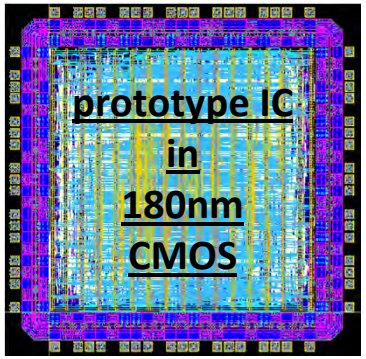




5.8X energy reduction

86% error-rate handling, P_{det} > 90%

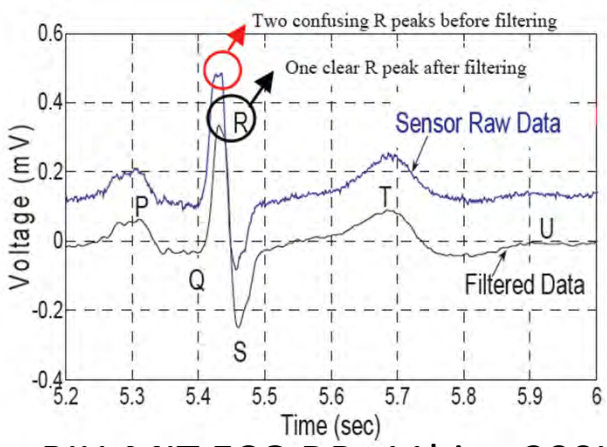




ECG Analysis IC @ MEOP

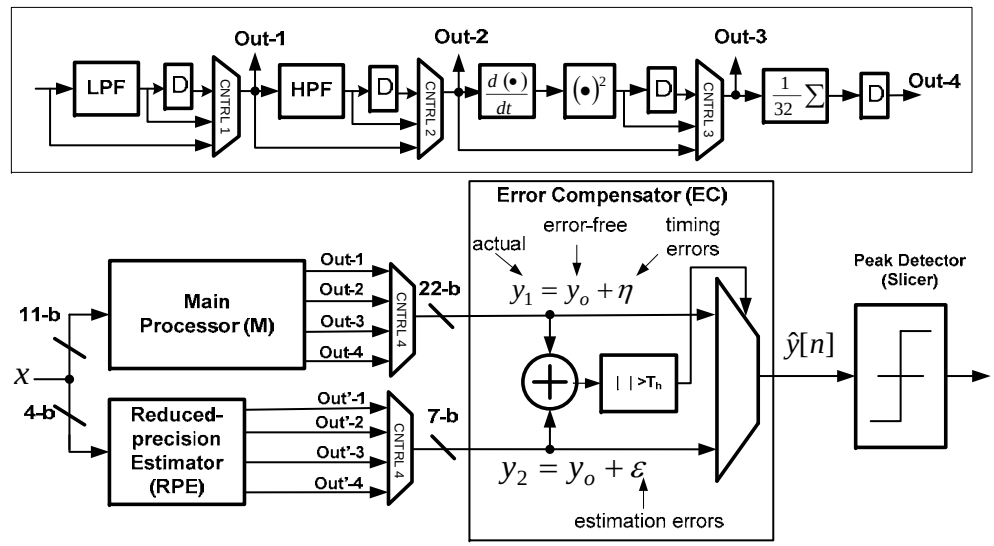
[Abdallah]

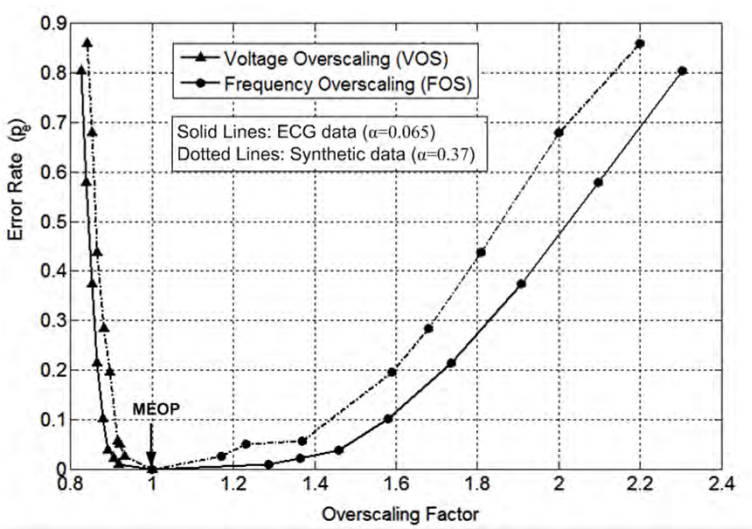
ECG waveform

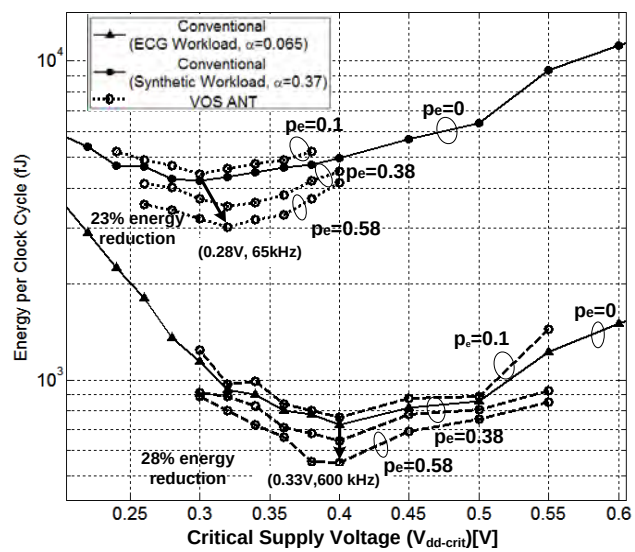


BIH-MIT ECG DB: 11bits, 200Hz

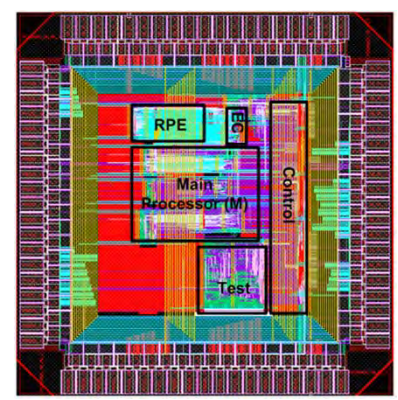
PAM-Tompkin Algorithm







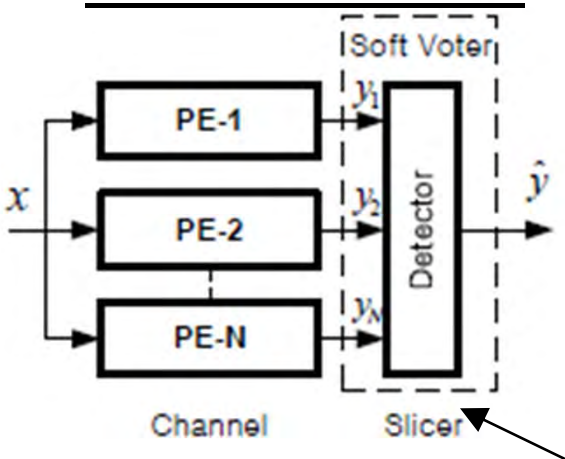
45nm, IBM process



Soft NMR

[Kim]

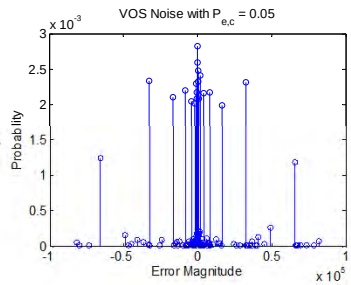
soft NMR architecture



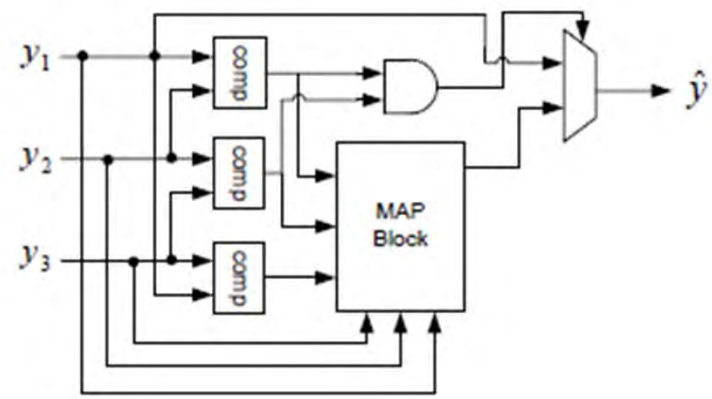
MAP rule

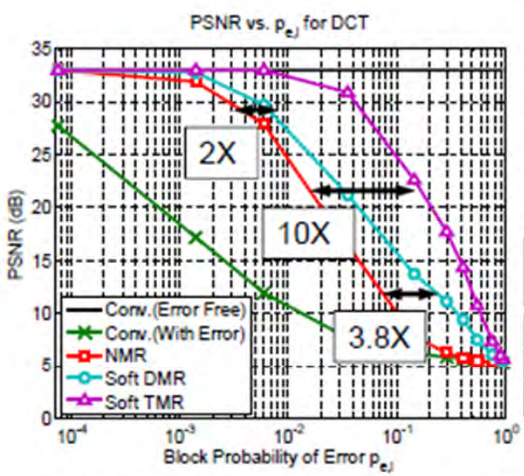
$$P(y_o \mid y_1, y_2, \ldots, y_N) = \frac{P(y_1, y_2, \ldots, y_N \mid y_0) \cdot P(y_0)}{P(y_1, y_2, \ldots, y_N)}$$

error statistics



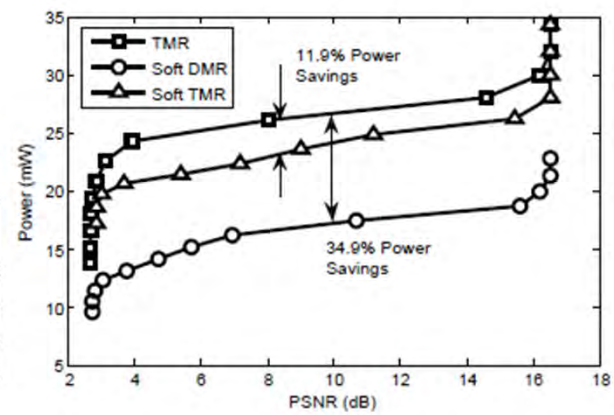
soft voter

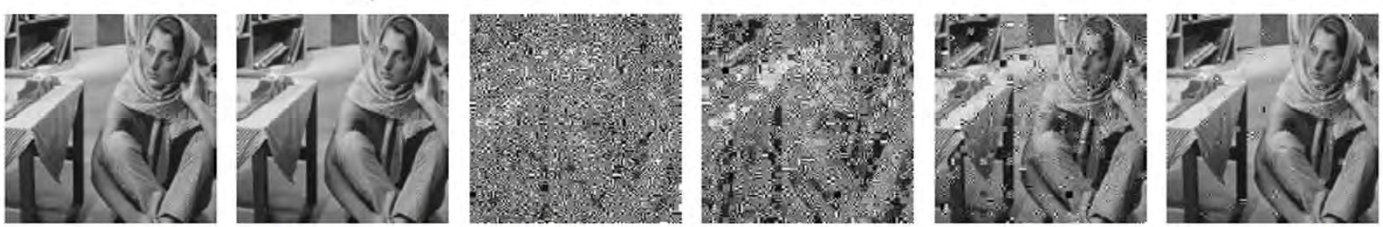




soft DCT

	Complexity (transistor count)	Complexity Overhead	Power Overhead
TMR	432	0.1%	0.07%
Soft DMR	35848	8.2%	0.25%
Soft TMR	64868	14.9%	1.49%

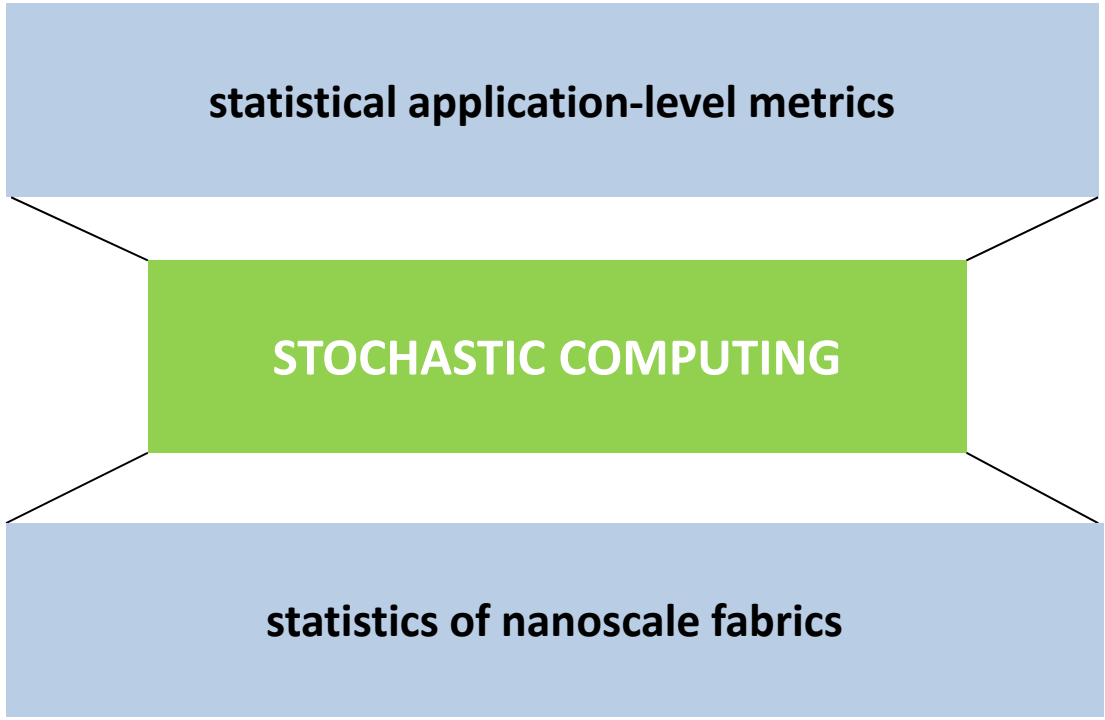




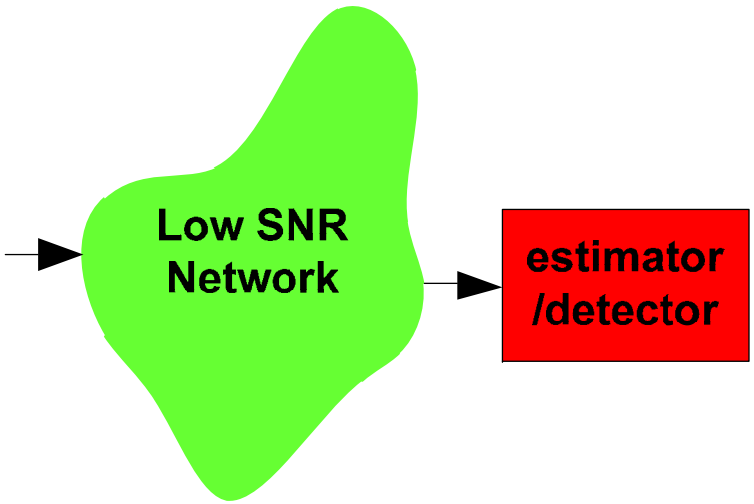
(a) Original Image (b) Error Free (c) With Errors (d) TMR (e) Soft DMR (f) Soft TMR

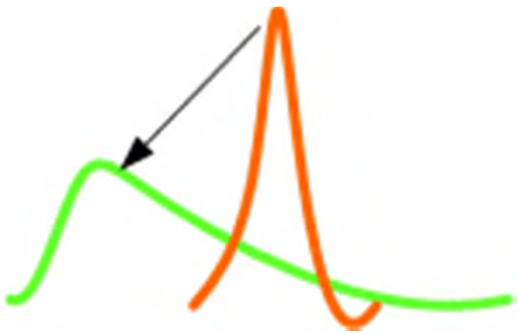
soft DMR vs. TMR
35% energy savings
2X more robust

Matching the statistical requirements of applications to the statistics of nanoscale fabrics



Implications on Device Design





- In return for energy-efficiency, we can handle
  - non-deterministic device behavior
  - improved average case behavior for worse corner-case → long-tailed distributions
  - 'low SNR switches' → smaller gap between 1-variable and a 0-variable
  - 'multi-state switches' (instead of two)

Summary

- Device design combined with stochastic computing can drive Moore's Law
- Device design for stochastic platforms
  - what are the device properties that result in favorable error statistics?
  - ideal switch model unnecessary
  - relaxed device specifications
- Design and programming methodologies
- New applications in biomedical, energy and security