Massive MIMO Processing at the Semiconductor Edge: Exploring the System and Circuit Margins for Power Saving

Yanxiang Huang [1,2], Claude Desset[1], André Bourdoux [1], Wim Dehaene [2,1], Liesbet Van der Perre [2]

1 imec, 2 KU Leuven, Leuven, Belgium

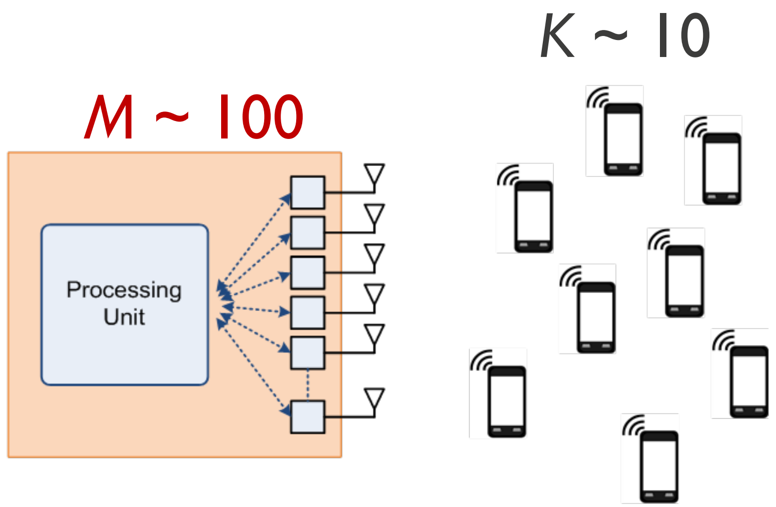
yanxiang.huang@imec.be

ICASSP 2017, New Orleans, 7th March.

Why Massive MIMO?

5G needs: more capacity/Watt/$

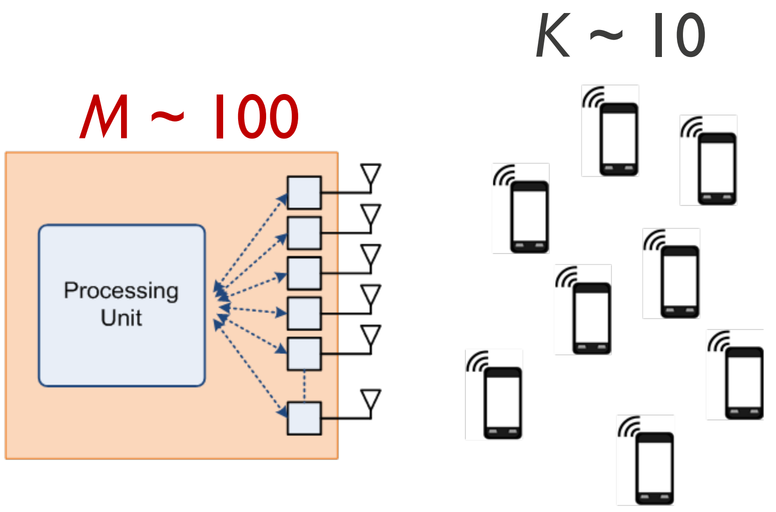
- Millimeter waves
- Heterogeneous networks
- Massive MIMO



Key features:

- M (base station antennas) >> K (users) >> 1
- MU-MIMO precoding (DL) or linear detection (UL)
- Uplink-based Channel-State Information
- Low-cost low-power low-accuracy hardware (per-antenna)

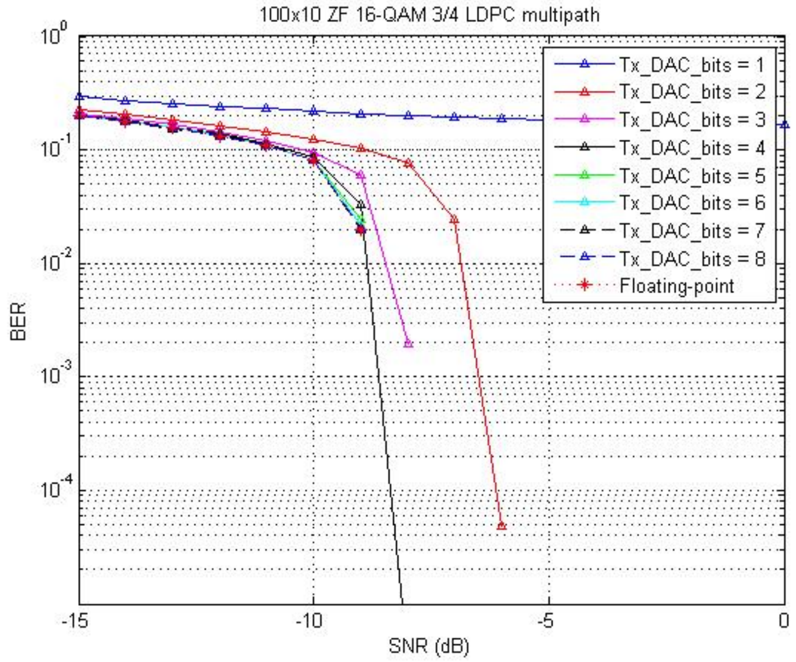
Is massive processing an issue?



	Subcomponent	DownLink (DL) [GOPS]	Uplink (UL) [GOPS]	Training [GOPS]
$K \sim 10$	Inner modern	175	520	290
	Outer Modern	7	40	0
$M \sim 100$	DFE incl. (I)FFT	920	920	920

BS (Base station) complexity of digital components for 100x10 Massive MIMO, with 20 MHz bandwidth, 3 bps/Hz (16-QAM, coding rate 3/4)

Embracing low accuracy hardware (per-antenna) low-quantization

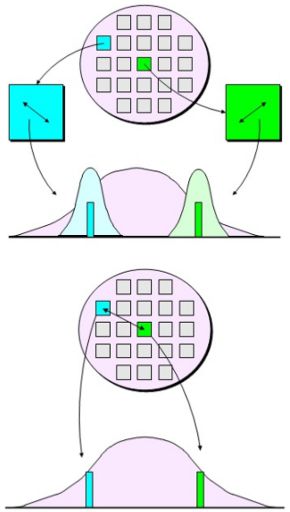


C. Desset, EuCNC, Validation of low-accuracy ...

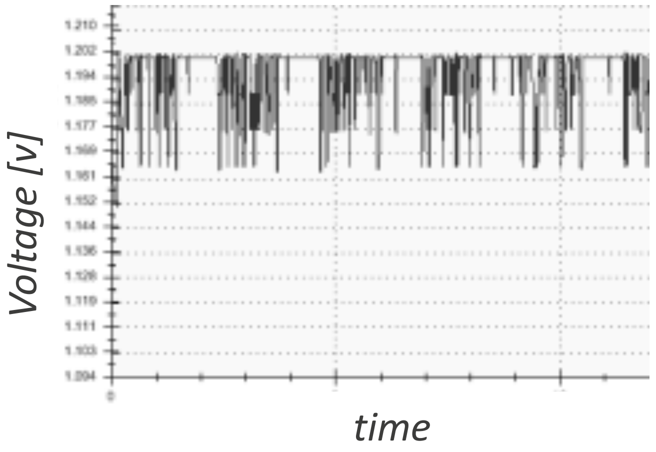
3-bit DAC for 16-QAM

It embraces noisy signal processing

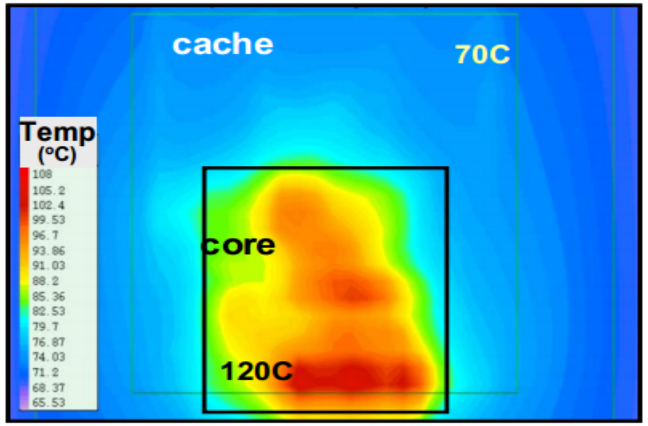
Semiconductors suffer from variations



Process Variations

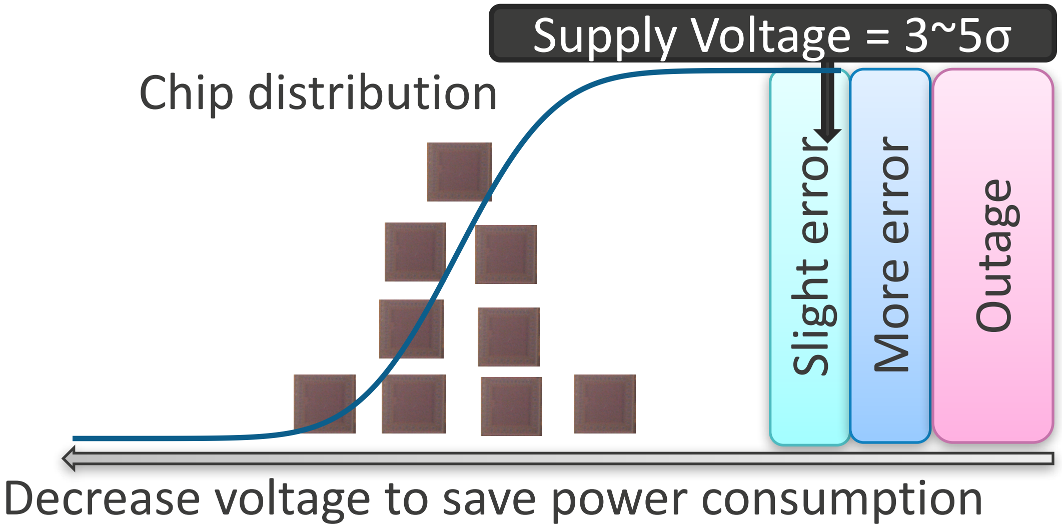


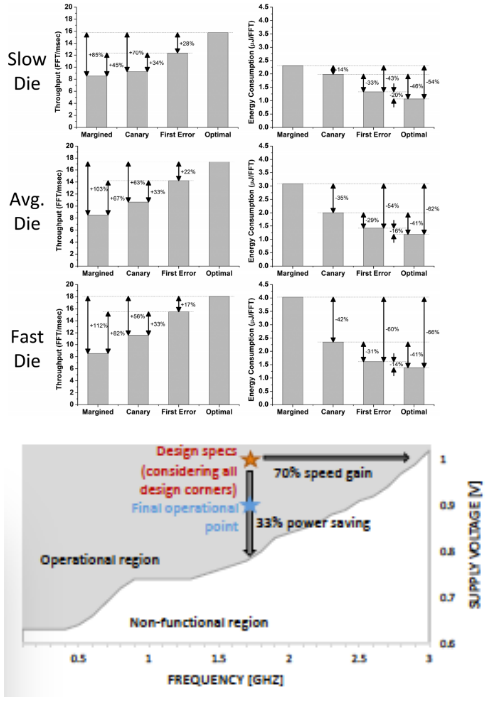
Voltage fluctuation



Temperature variation

Chips sacrifice power & cost to provide correctness



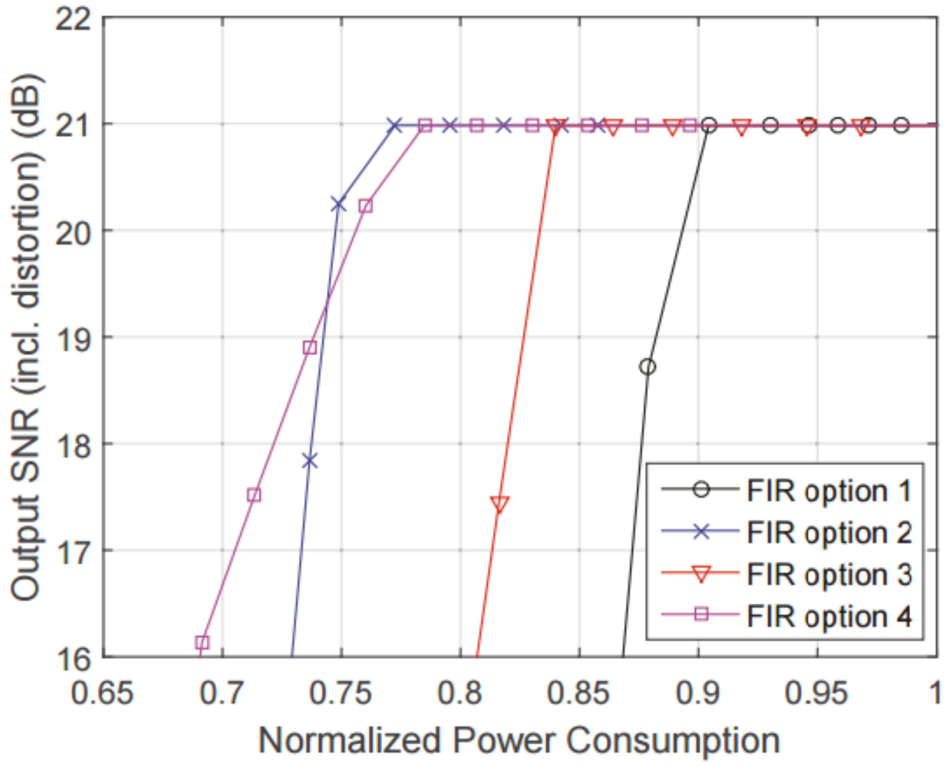


~40% power wasted to guarantee correctness

Refs

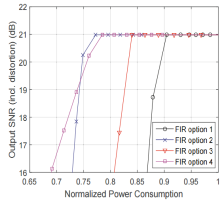
Ref: 54% from 45nm M. Fojtik, JSSC13, Bubble Razor ...: Fig.12 Energy Slow die and avg. Die, first error power saving
33% from 28nm Y. Huang, ASSCC16, a 28nm ... : Fig. 7(b)

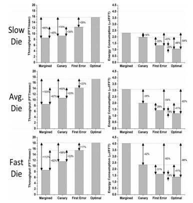
What if processing at the CMOS edge?
'Only' quality degradation

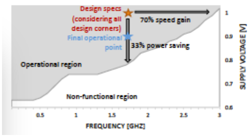


65nm, Y. Liu, TVLSI10, Computation error ...

Processing Massive MIMO at CMOS edge



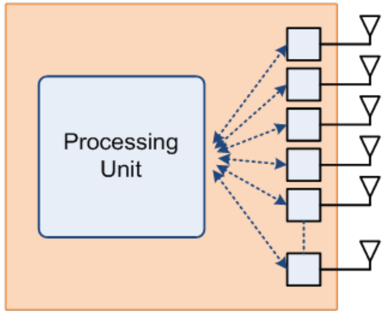




$$SDDR = 10 log \frac{\sigma_s^2}{\sigma_d^2}$$

Signal-to-Digital-Distortion

$M \sim 100$

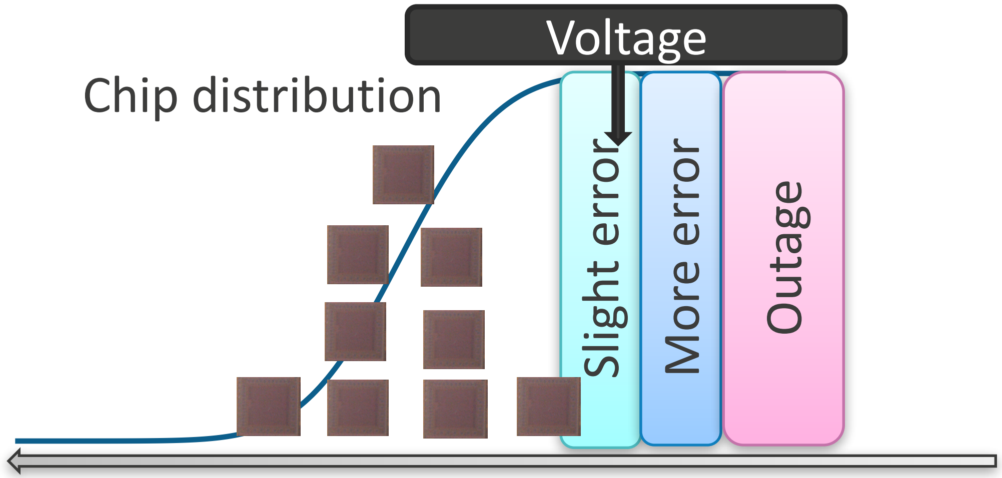


$K \sim 10$

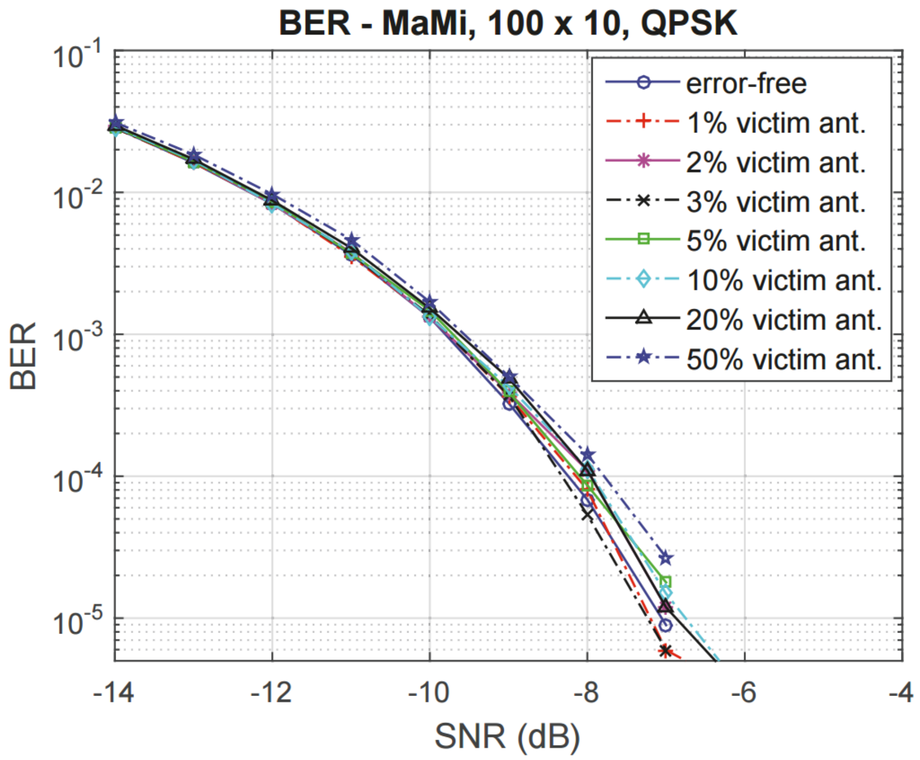
Massive MIMO, 100x10, DownLink, TDD, LTE-like

Only slight degradation

when embracing slight digital distortion error

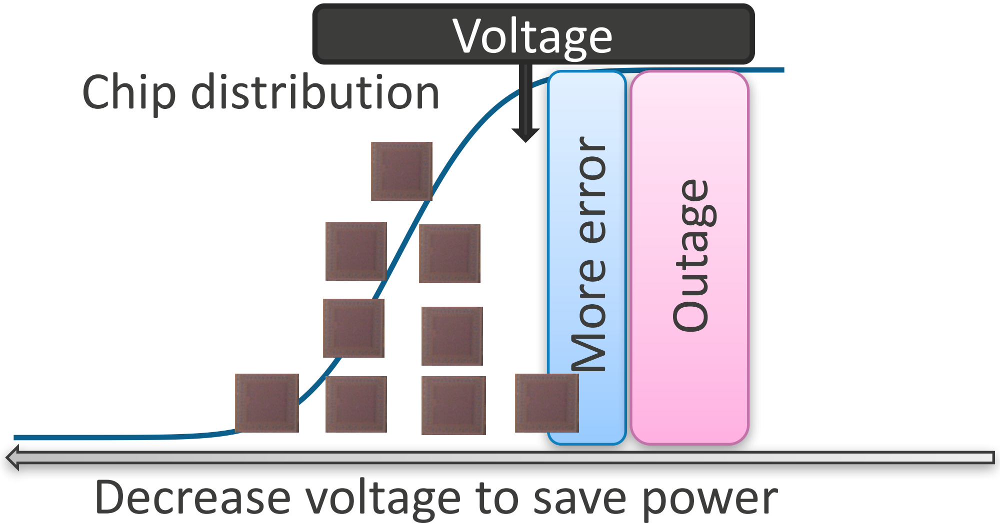


When Voltage decreases, more antennas facing slight error, or even outage

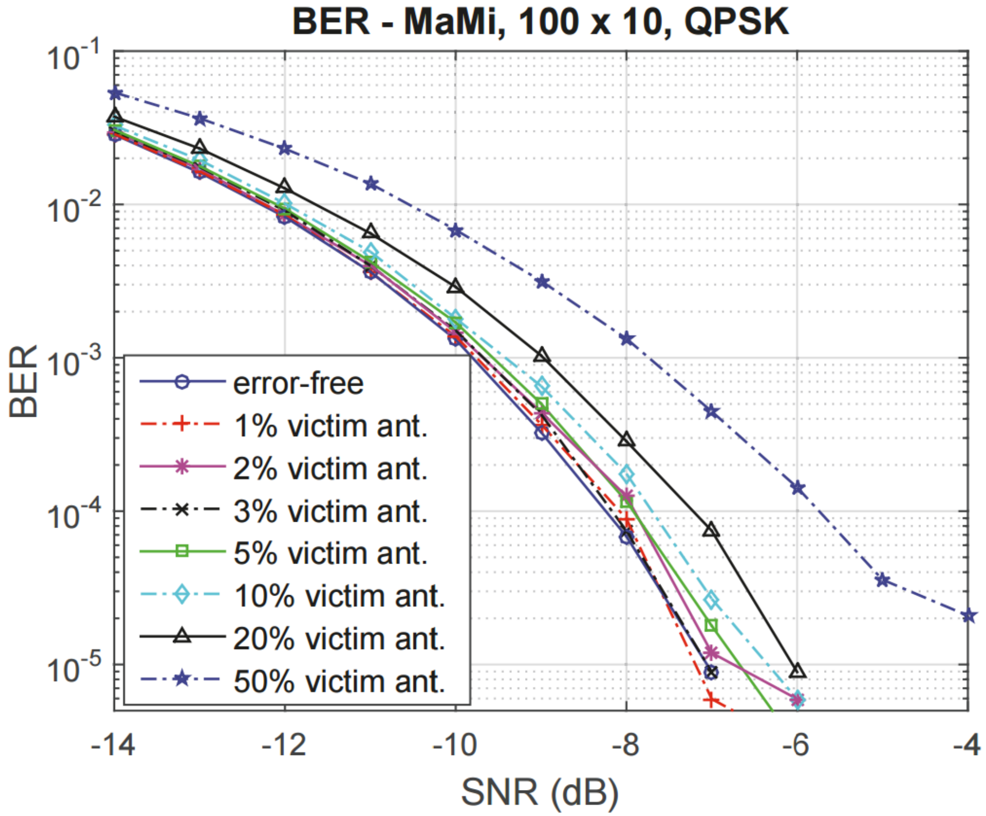


Slight error,
Signal-to-Digital Distortion Ratio = 10dB

Acceptable degradation even with heavy distortion



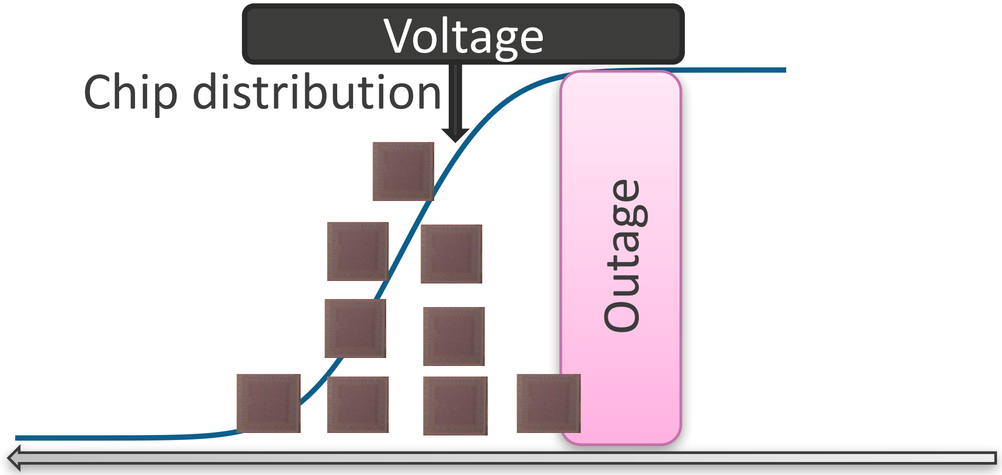
When Voltage decreases, more antennas facing more errors

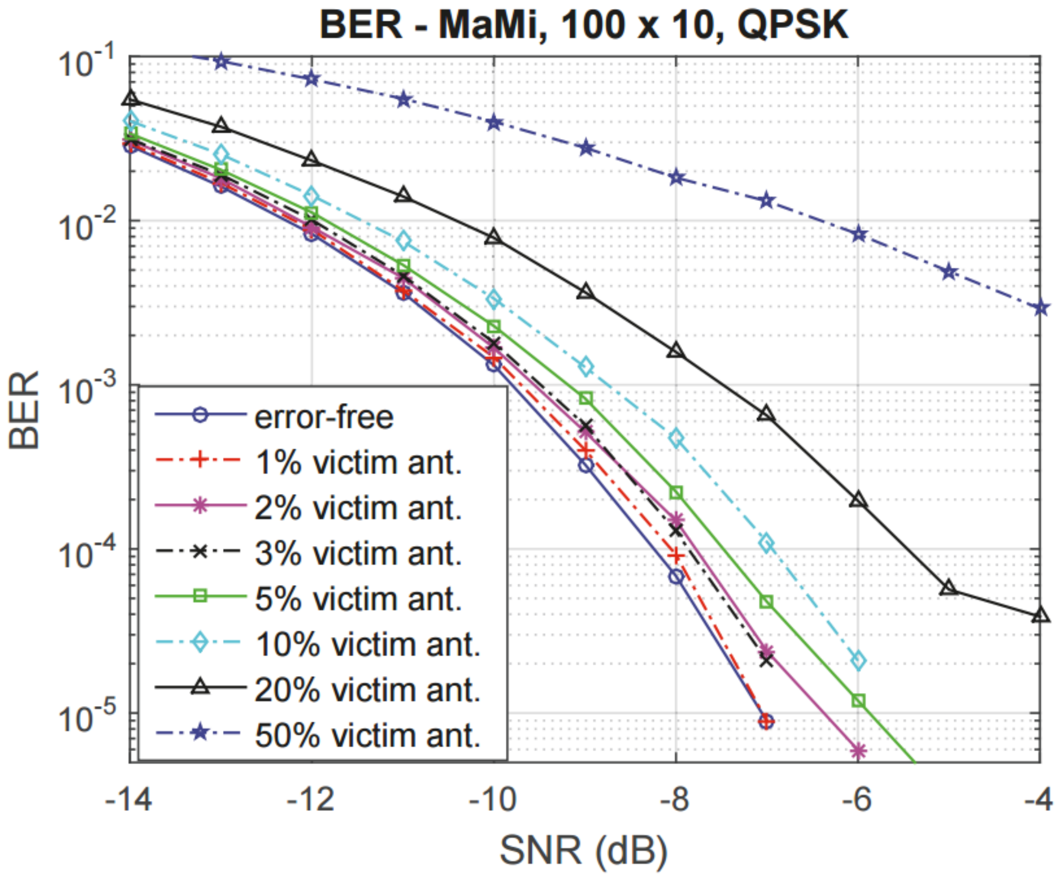


Extensive error,

Signal-to-Digital Distortion Ratio = 0 dB

Manageable even with completely outage





When Voltage decreases, antenna outage occurs

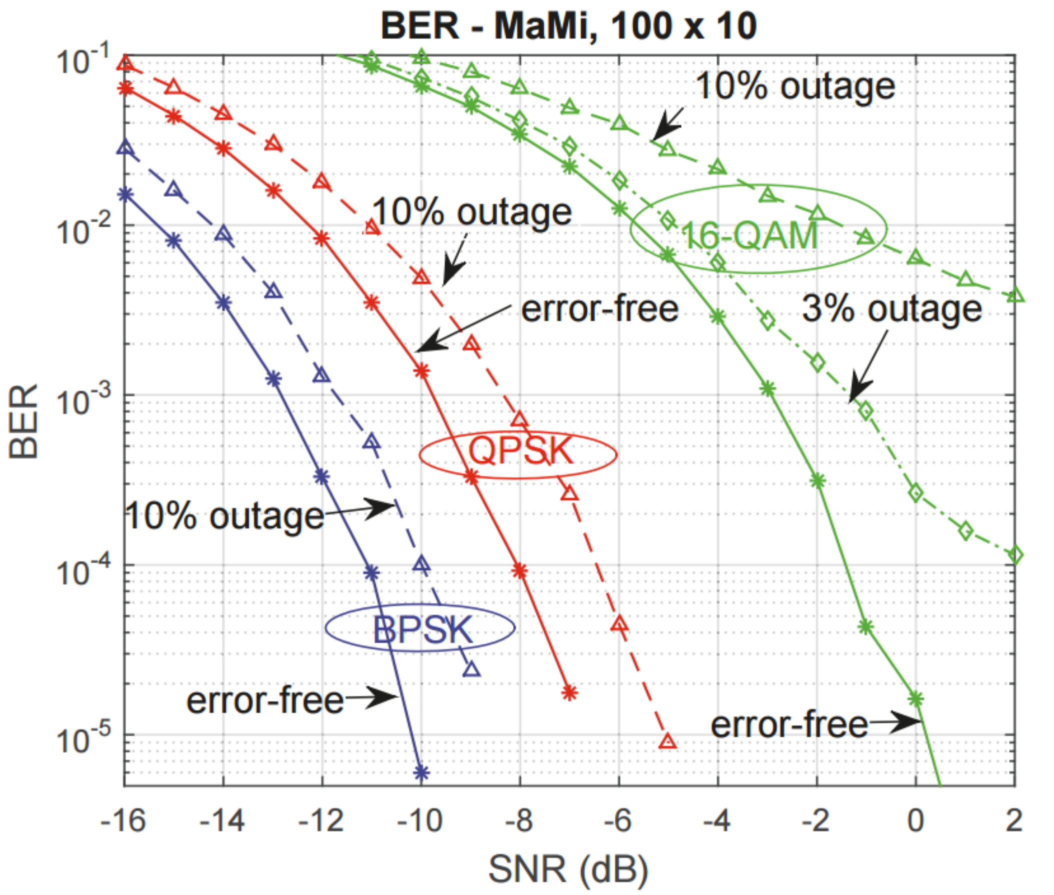
It tolerate 10% outage, (1.3σ variability margin), 43%~54% power saving [ref]

Antenna completely outage, output stuck at a fixed value

[Ref] M. Fojtik, JSSC13, Bubble Razor ...: Fig.12 Energy Slow die and avg. Die, first error power saving

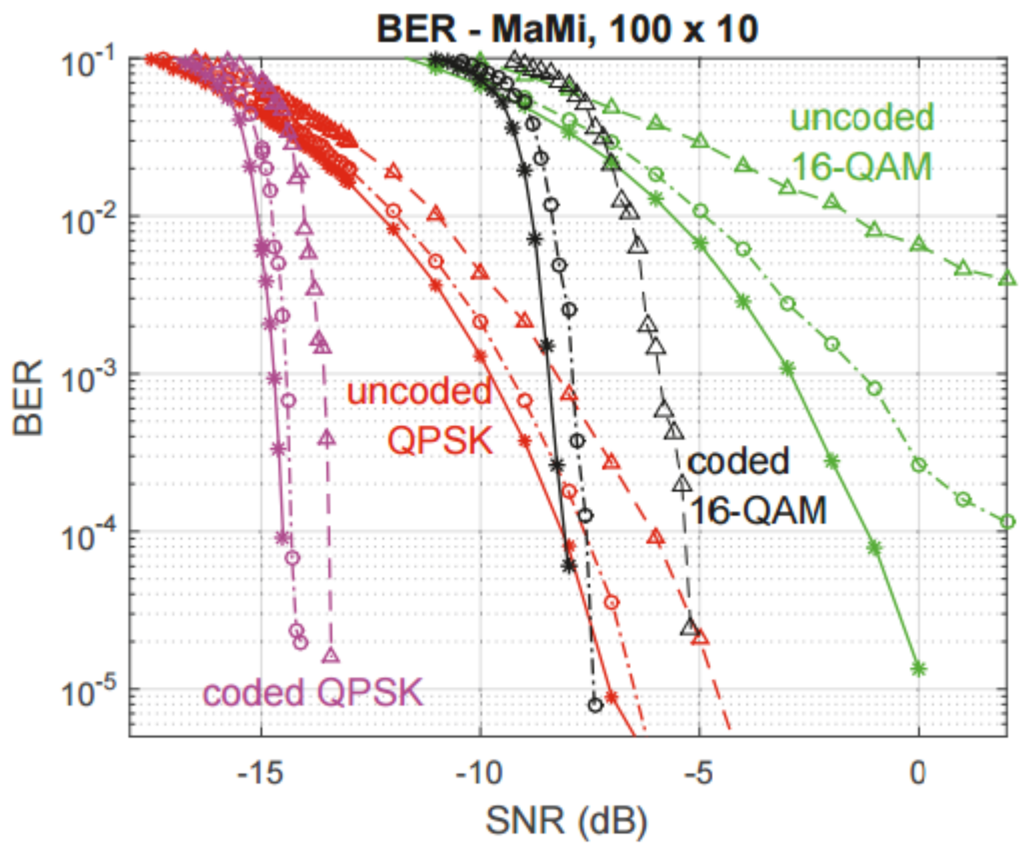
Simpler modulations are more robust

to embrace distortions



Antenna outage distortion

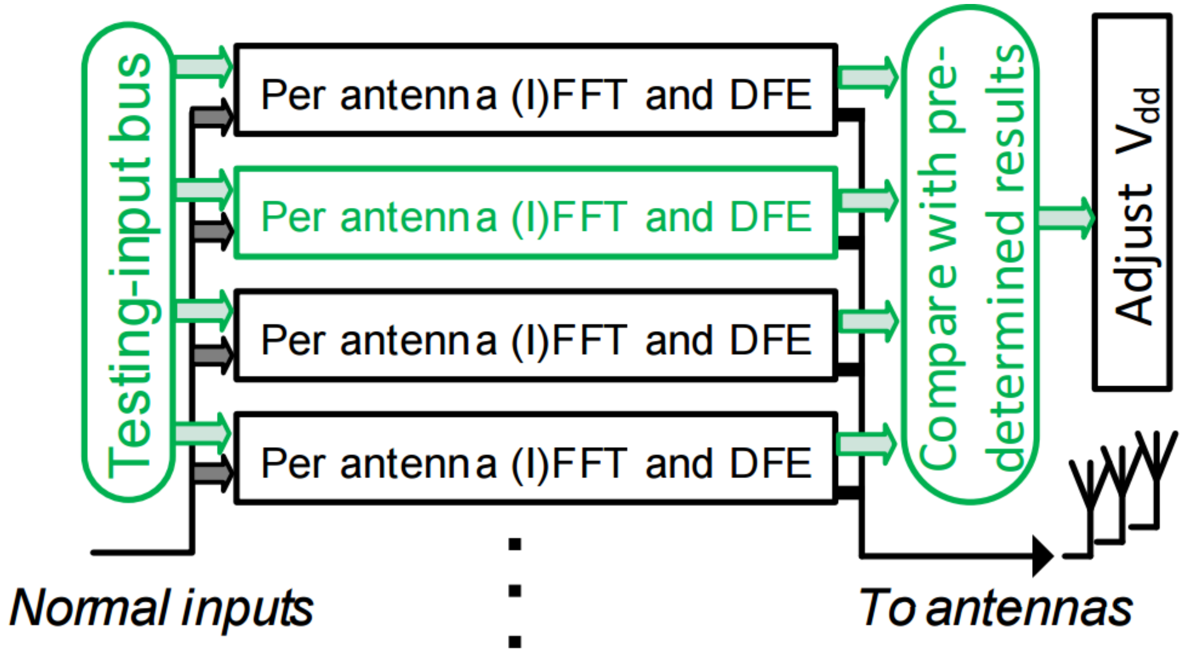
On error-coded performance
SNR degradation is even smaller



Antenna outage distortion (¾ LDPC coding)

Leftmost error-free, middle 3% victim, right 10% victim

Solutions to harness variation (error)
Monitor variation on-the-fly

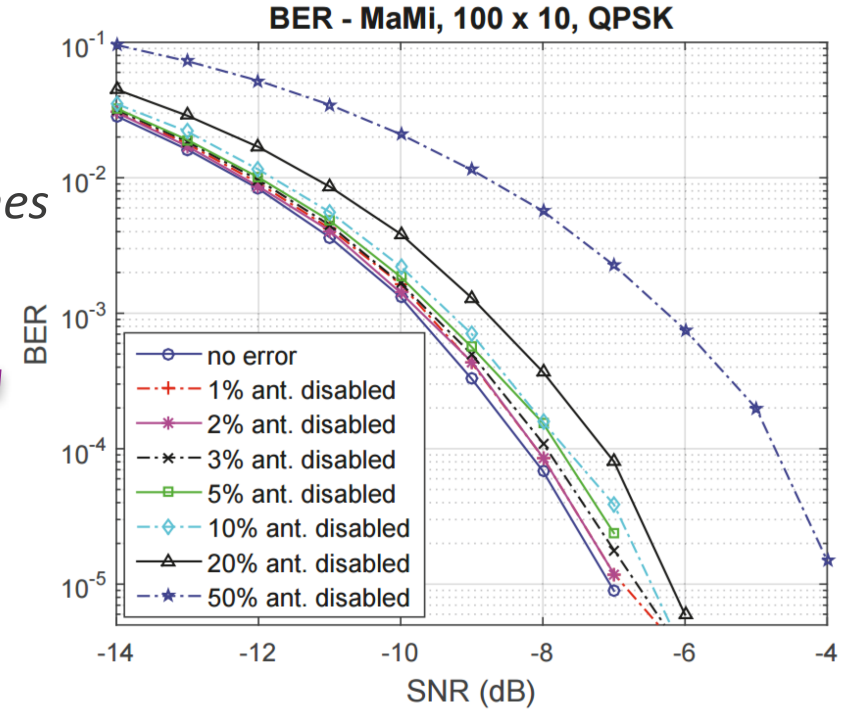


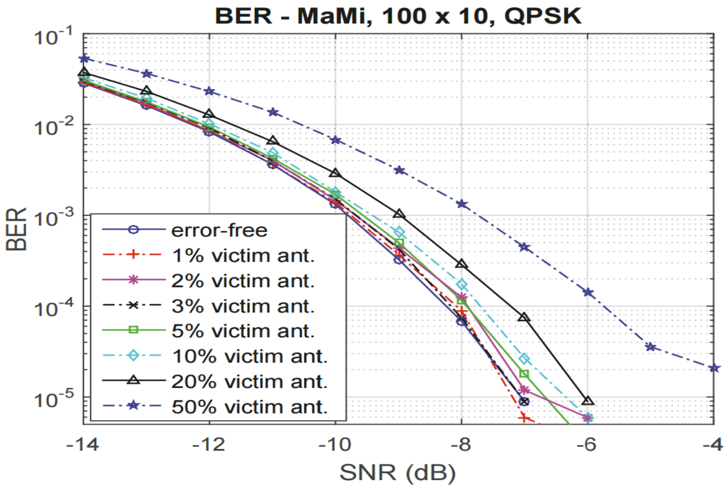
Monitor per-antenna functionality, to decide whether to adjust environment (V_{dd}), continue, or discard antenna ...

Case study:

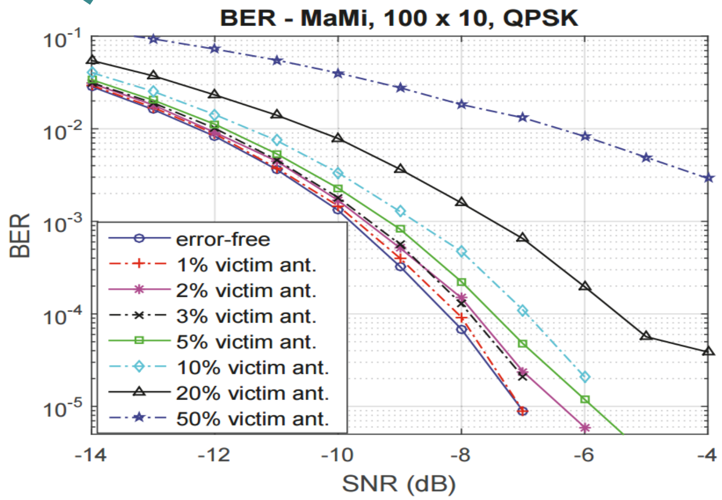
When facing error, continue using erroneous antenna?

A possible solution:
Discard erroneous antenna, and only use remaining error-free ones





Extensive error, SDDR = 0 dB



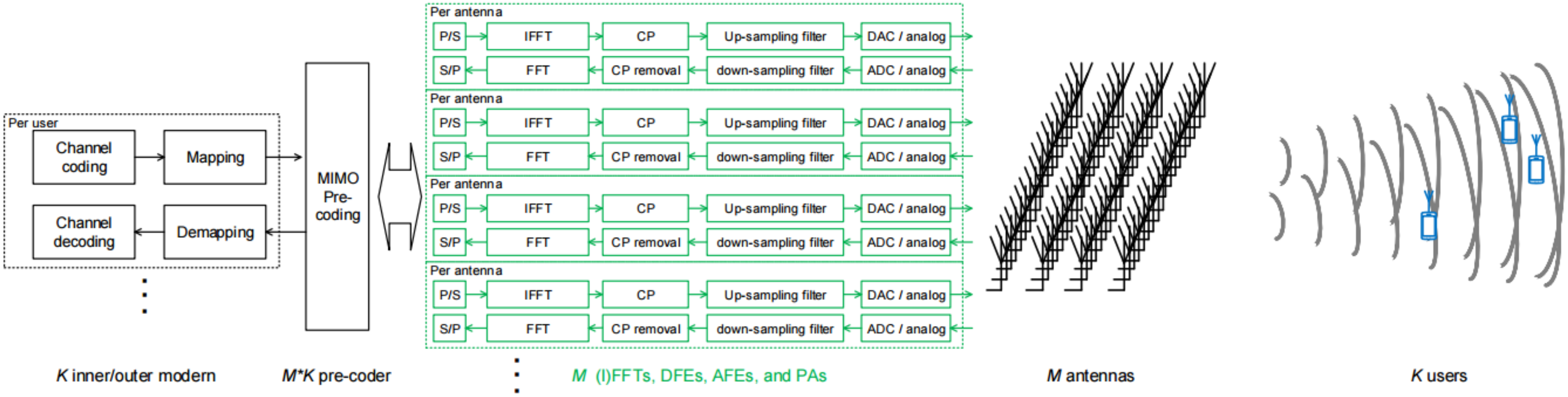
Antenna outage

Conclusions
Massive MIMO absorbs distortions from digital error

- 40% power consumption saving possible with 1dB degradation.
- Use on-chip monitors to detect & adjust hardware failures on-the-fly.
- When hardware error distortion is
  - −low, e.g. 0 dB, should continue using the erroneous antenna;
  - −high, adjust the environment, or discard the erroneous antenna.

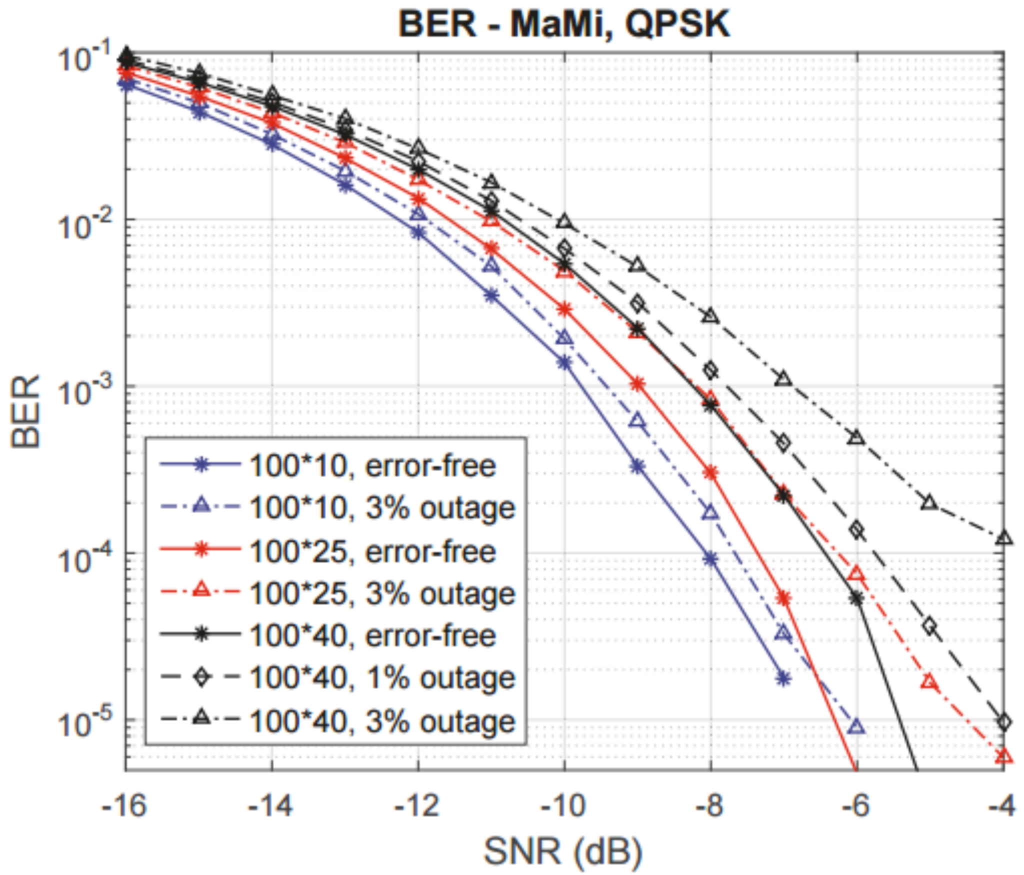
Q & A

Backup system



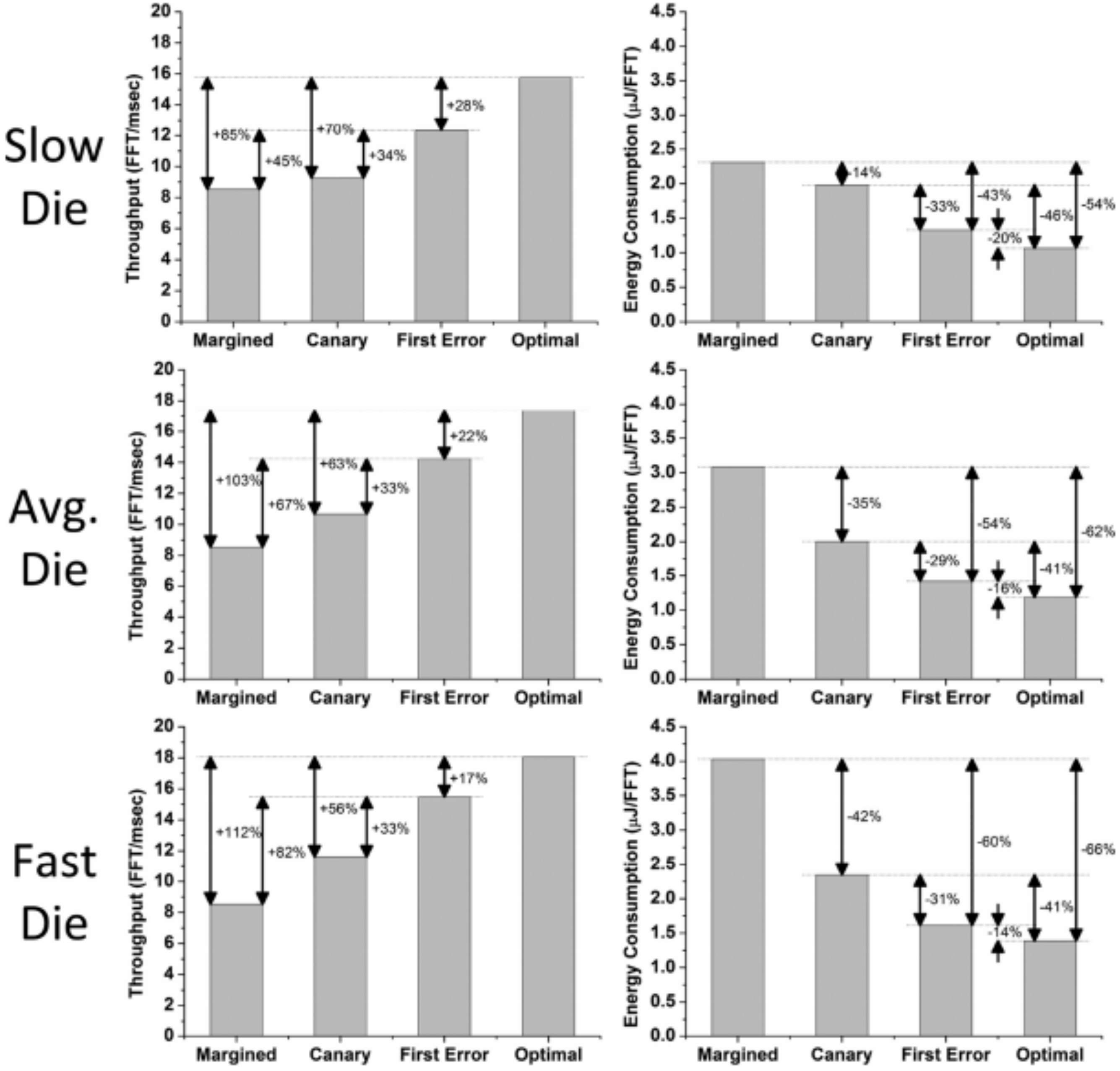
System model of the MaMi system. Each BS is equipped with M antennas and serves K users. Typically, $M = 100$ and $K = 10$.

Backup load



BER - MaMi, QPSK
BER
SNR (dB)
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
-16
-14
-12
-10
-8
-6
-4
100*10, error-free
100*10, 3% outage
100*25, error-free
100*25, 3% outage
100*40, error-free
100*40, 1% outage
100*40, 3% outage

Backup load



45nm, M. Fojtik, JSSC13, Bubble Razor ...