The Future of Wireless System Design

Gerd Ascheid

RWTH Aachen University
Institute for Communication Technologies and Embedded Systems

Outline

Performance vs.
Processing Power

Wireless Communication
and Moore's Law

Conclusions for Future
Wireless System Design

Performance vs. Processing Power

- On-chip power consumption in mobile devices is critical
  - Battery operating time
  - Heat dissipation without forced cooling
- Wireless communication performance comes with a processing power cost
- Tradeoff example: iterative receivers

Outline

Performance vs.
Processing Power
- Iterative MIMO detection
- Implementation and Analysis

Wireless Communication
and Moore's Law

Conclusions for Future
Wireless System Design

OFDM MIMO Link Model

- For subcarrier ν the model for a multi-antenna system using OFDM is*

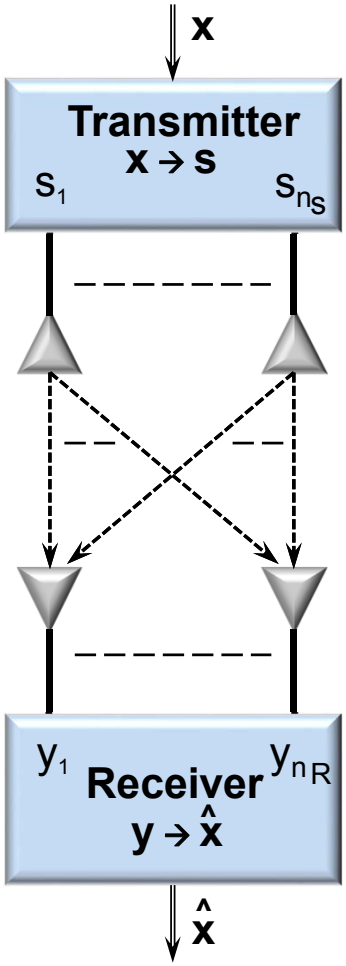
$$\mathbf{Y}(\mu,\nu) = \mathbf{H}(\mu,\nu)\,\mathbf{x}(\mu,\nu) + \mathbf{W}(\mu,\nu)$$

$$[n_R \times 1] = [n_R \times n_S][n_S \times 1] + [n_R \times 1]$$

with vectors and channel matrix

$$\mathbf{Y}(\mu,\nu) = \begin{pmatrix} Y_1(\mu,\nu) \\ \vdots \\ Y_{n_R}(\mu,\nu) \end{pmatrix} \quad ; \quad \mathbf{x}(\mu,\nu) = \begin{pmatrix} x_1(\mu,\nu) \\ \vdots \\ x_{n_S}(\mu,\nu) \end{pmatrix}$$

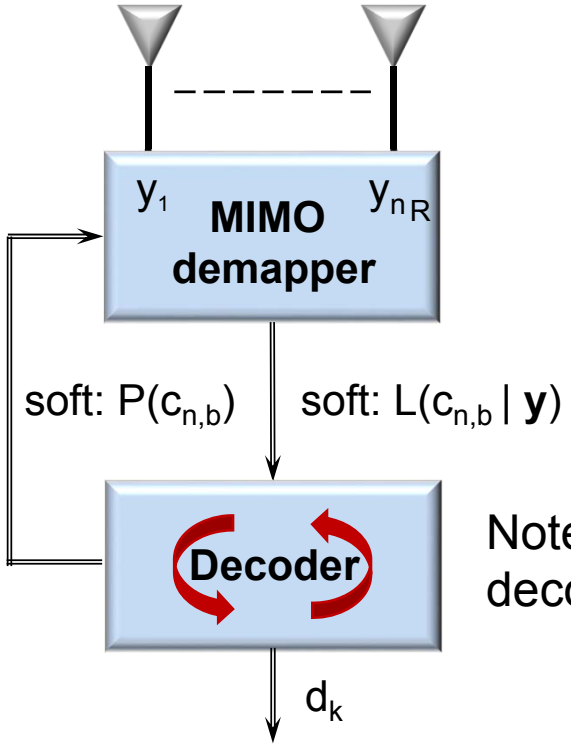
$$\mathbf{W}(\mu,\nu) = \begin{pmatrix} W_1(\mu,\nu) \\ \vdots \\ W_{n_R}(\mu,\nu) \end{pmatrix} \quad \mathbf{H}(\mu,\nu) = \begin{pmatrix} h_{1,1}(\mu,\nu) & \cdots & \cdots & h_{1,n_S}(\mu,\nu) \\ \vdots & & & \vdots \\ h_{n_R,1}(\mu,\nu) & \cdots & \cdots & h_{n_R,n_S}(\mu,\nu) \end{pmatrix}$$



* using μ for μT_{sym} and ν for $\nu \Delta f$

Iterative MIMO OFDM Receivers

- **Minimal BER/FER** is achieved with iterative MIMO Systems using soft demapping and soft feedback



Note that for LDPC / Turbo Codes decoder itself is also iterative

$c_{n,b}$ = *b*-th bit of the symbol transmitted by antenna *n* ; (B bit per Symbol)

L = Loglikelihood Ratio (LLR)

MIMO Demapping: Maximum Likelihood Detector

- The optimum detection scheme with the minimum error probability for equally likely symbols is Maximum Likelihood (ML)

$$\hat{\mathbf{x}} = \underset{\text{all } \mathbf{x}}{\arg\min}\left\{\lambda(\mathbf{x})\right\} = \underset{\text{all } \mathbf{x}}{\arg\min}\left\{(\mathbf{y}-\mathbf{H}\mathbf{x})^{H}(\mathbf{y}-\mathbf{H}\mathbf{x})\right\} = \underset{\text{all } \mathbf{x}}{\arg\min}\left\{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^{2}\right\}$$

- Efficient implementation as “Hard Sphere Decoder” (→ tree search) (based on the QR-decomposition: H = QR):
- Required metrik for soft output is the Log Likelihood Ratio (LLR)

$$L\left(c_{n,b}\middle|\mathbf{y}\right) = \ln\frac{P\left(c_{n,b}=0\middle|\mathbf{y},\mathbf{H}\right)}{P\left(c_{n,b}=1\middle|\mathbf{y},\mathbf{H}\right)} = \ln\frac{\sum_{\mathbf{x}\in S_{n,b}^{0}} e^{-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^{2}}{N_0}}}{\sum_{\mathbf{x}\in S_{n,b}^{1}} e^{-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^{2}}{N_0}}}$$

Soft Output

- The above LLR computation is very complex and has to be performed for each bit. Therefore, commonly the max-log approximation is used

$$L(c_{n,b}|\mathbf{y}) \approx \frac{1}{N_0}\left(\min_{\mathbf{x}\in S_{n,b}^{1}}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 - \min_{\mathbf{x}\in S_{n,b}^{0}}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right)$$

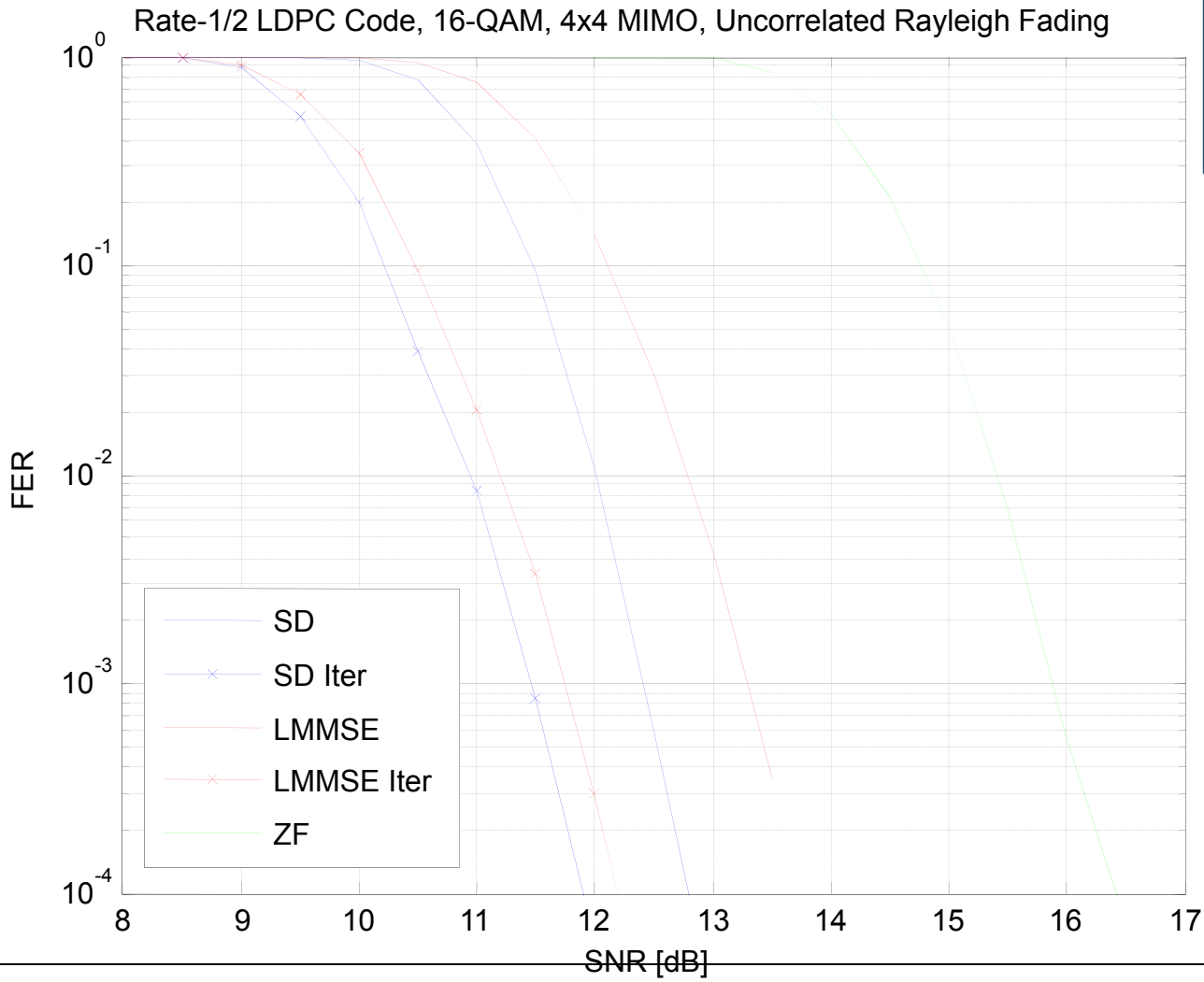
(→ requires two tree searches, can be combined into single tree search)

- Soft input extension (max-log approximation)

$$L(c_{n,b}|\mathbf{y}) \approx \min_{\mathbf{x}\in S_{n,b}^{1}}\left\{\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{N_0} - \sum_{\nu=1}^{n_T}\sum_{\beta=1}^{B}\ln P(c_{\nu,\beta})\right\} - \min_{\mathbf{x}\in S_{n,b}^{0}}\left\{\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{N_0} - \sum_{\nu=1}^{n_T}\sum_{\beta=1}^{B}\ln P(c_{\nu,\beta})\right\}$$

$S_{n,b}^{0/1}$ =set of symbols for which $c_{n,b}$ = 0/1

Performance Comparison



It is worth the effort: More complex algorithms yield lower FER !

Outline

Performance vs. Processing Power

- Iterative MIMO detection
- Implementation and Analysis*

Wireless Communication and Moore's Law

Conclusions for Future Wireless System Design

* Source: Filippo Borlenghi, Silicon Implementation of Iterative Detection and Decoding for Multi-Antenna Receivers, PhD Thesis, RWTH Aachen, 2015

Building Blocks

MIMO Detector

- Soft-input soft-output depth-first sphere decoding
- Max-log optimal performance
- Variable runtime

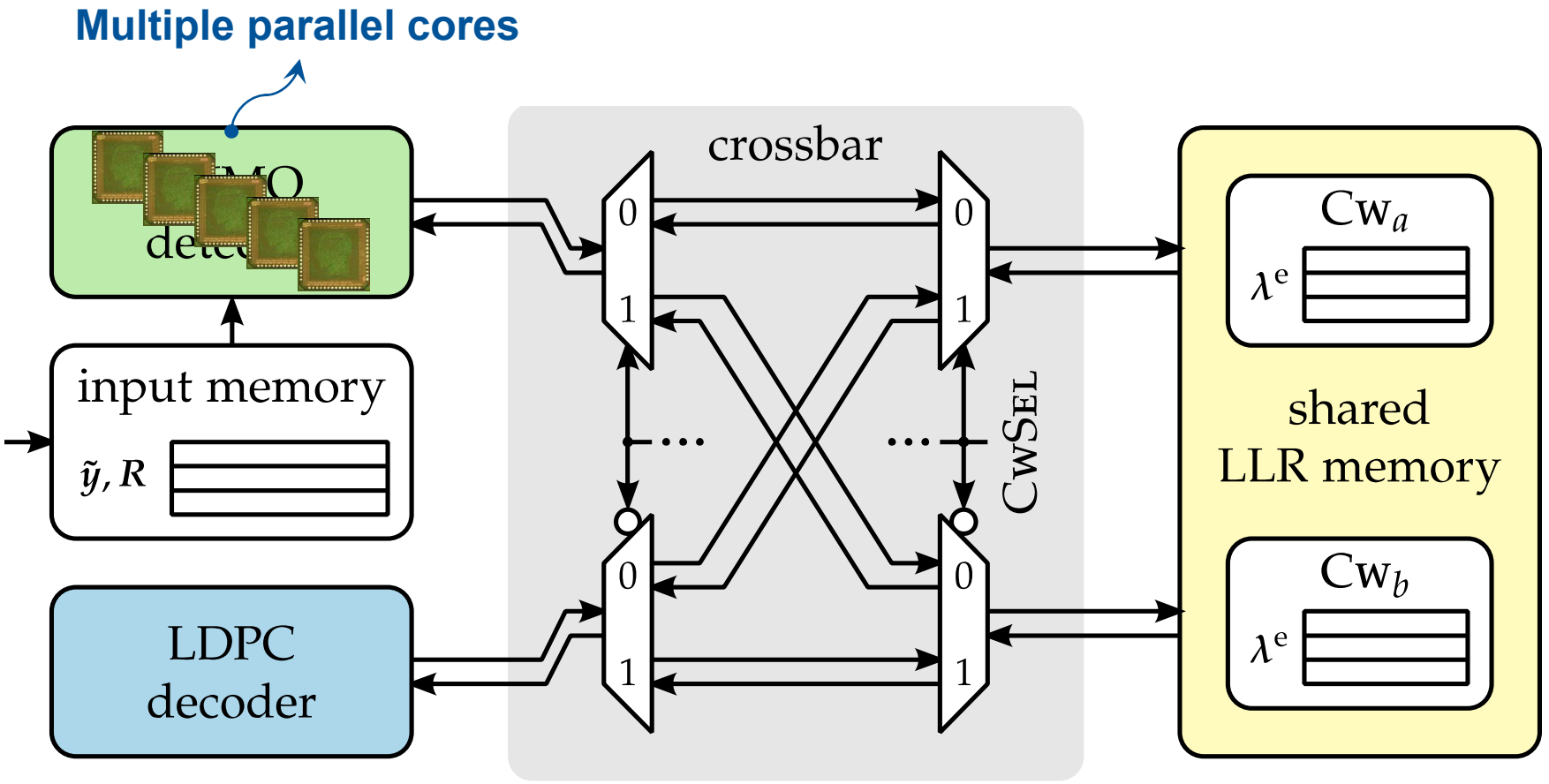
Witte et al., *A Scalable VLSI Architecture for Soft-Input Soft-Output Single Tree-Search Sphere Decoding*, TCAS-II, 2010
Borlenghi et al., *A 772 Mbit/s 8.81 bit/nJ 90 nm CMOS Soft-Input Soft-Output Sphere Decoder*, A-SSCC 2011

Channel Decoder

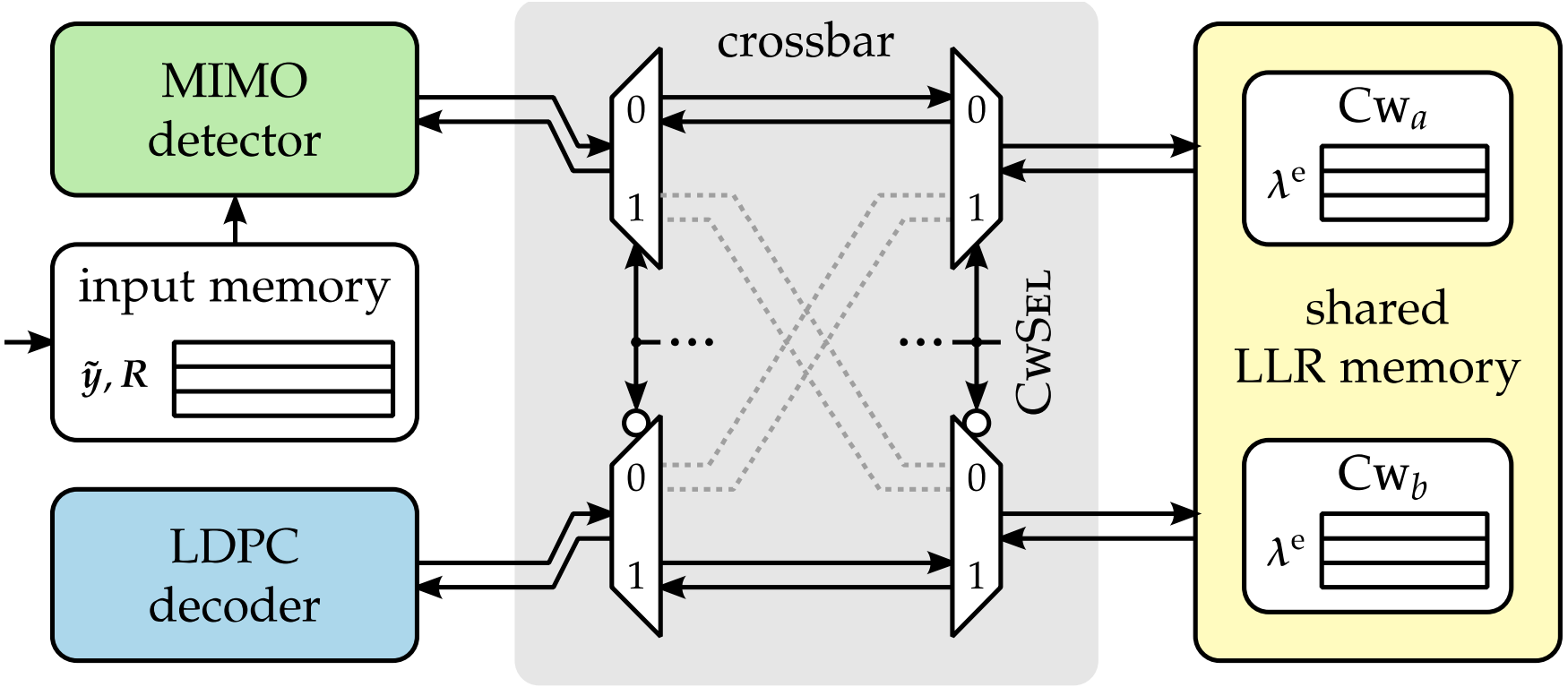
- IEEE 802.11n LDPC codes
- Layered offset min-sum (OMS) iterative decoding

Roth et al., *A 15.8 pJ/bit/iter Quasi-Cyclic LDPC Decoder for IEEE 802.11n in 90 nm CMOS*, A-SSCC 2010

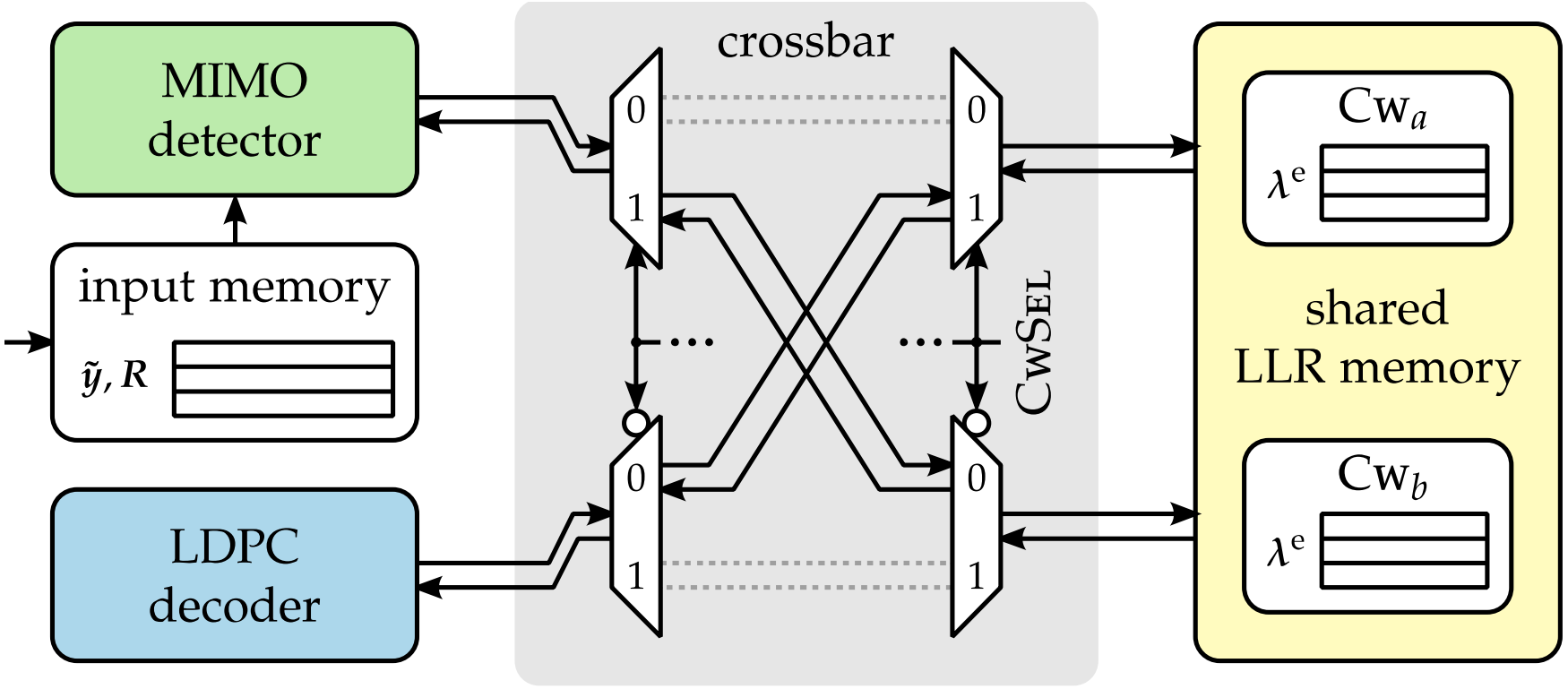
System Architecture



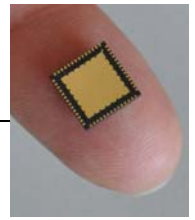
System Architecture



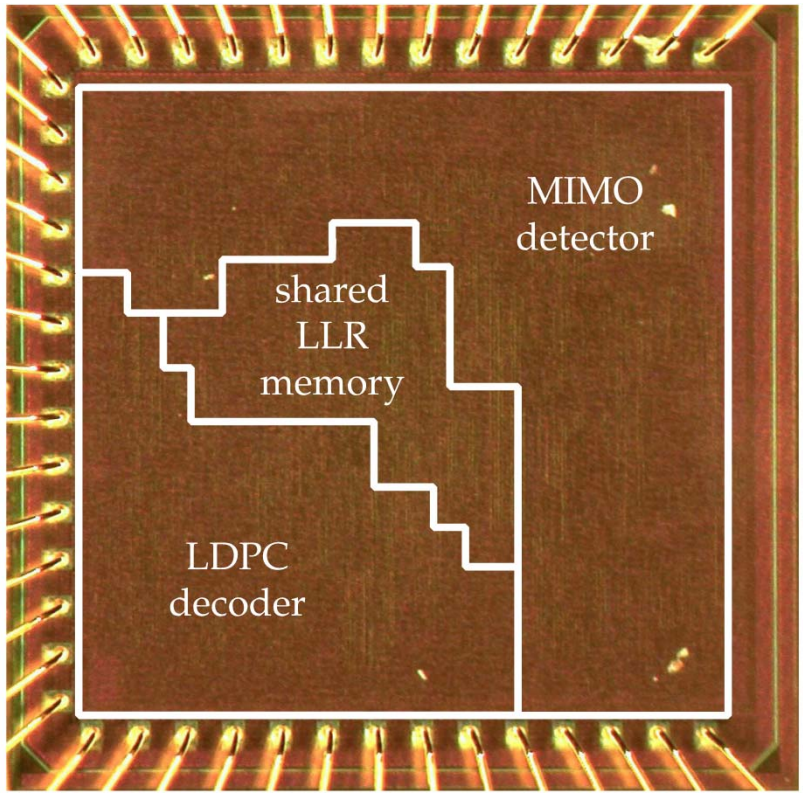
System Architecture



Implementation Results



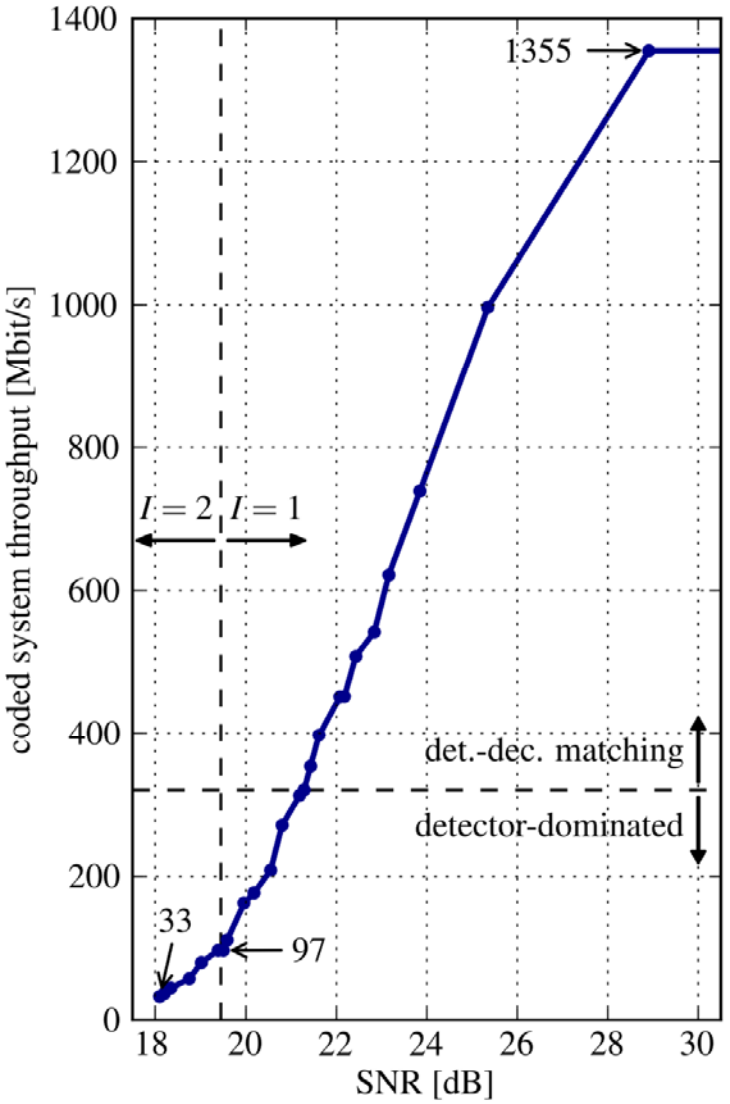
- Supports 2x2 / 3x3 / 4x4 MIMO with 4 / 16 / 64-QAM and all IEEE 802.11n LDPC codes
- 65 nm LL tech. @ 1.2 V
- Area: 2.78 mm^2 / 1.58 MGE
  - Detector (5): 872 kGE
  - Decoder: 447 kGE
  - Shared mem.: 210 kGE
- Max. frequencies:
  - Detector: 135 MHz
  - Decoder: 299 MHz
- Max. information throughput: > 1 Gbps

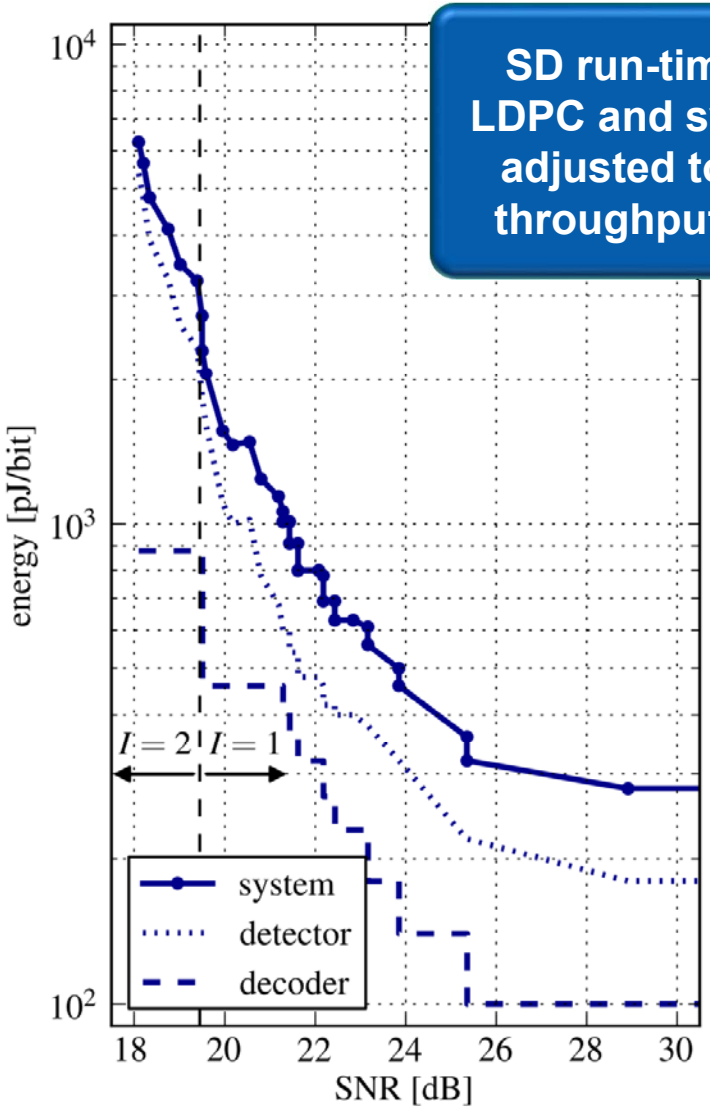


Is it **feasible**? **YES** ➔ Now, does it **make sense**?

Borlenghi et al., *A 2.78 mm^2 65 nm CMOS Gigabit MIMO Iterative Detection and Decoding Receiver*, ESSCIRC 2012

Throughput and Energy Efficiency

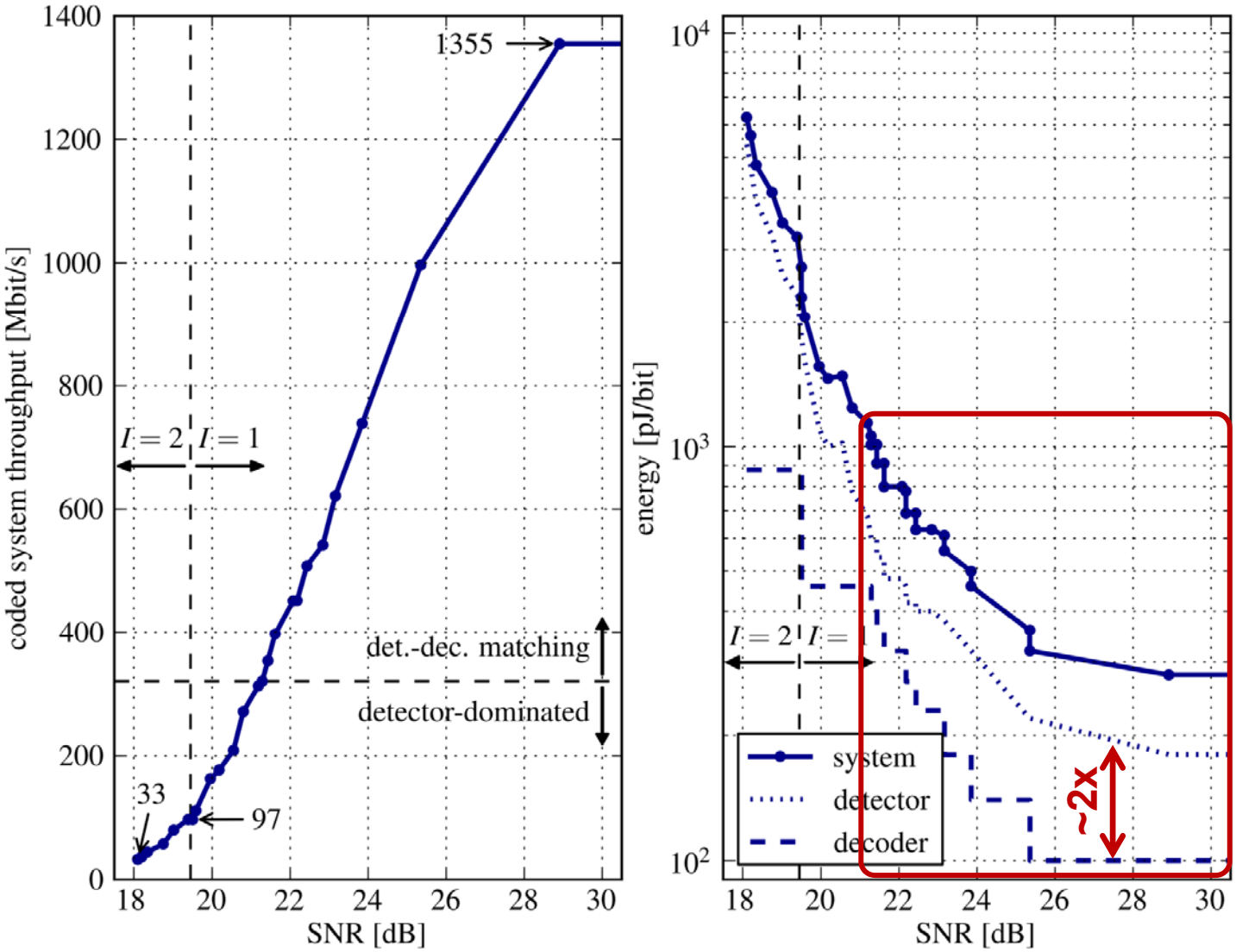




SD run-time constraints, LDPC and system iterations adjusted to achieve max. throughput @ BLER ≤ 1%

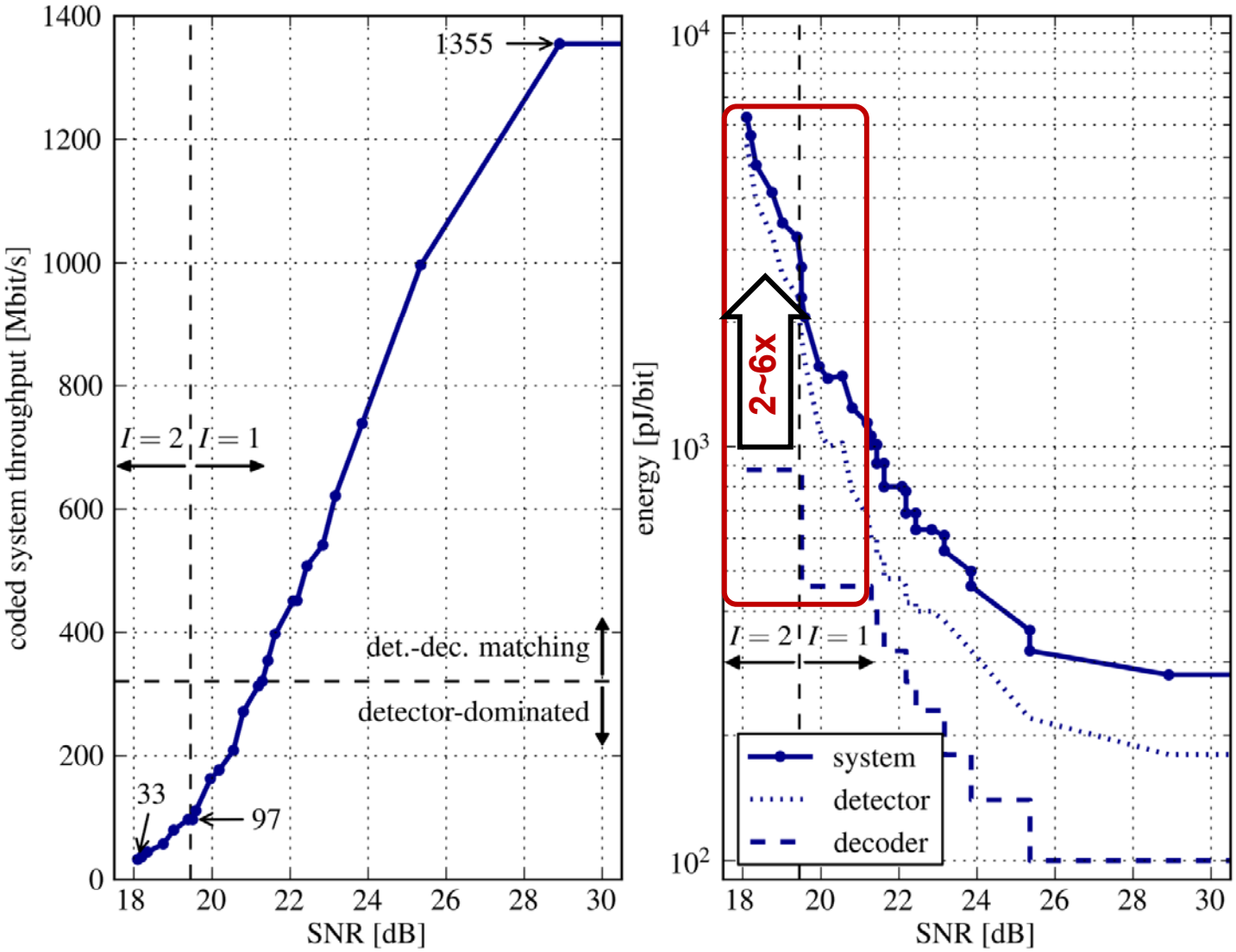
4x4 64-QAM, block length 1944, code rate ½

Throughput and Energy Efficiency



4x4 64-QAM, block length 1944, code rate ½

Throughput and Energy Efficiency



4x4 64-QAM, block length 1944, code rate ½

Different Optimization Goals

Wireless communication performance is most expensive (in terms of processing power) for low SNR

- Optimize communication performance

or

- Optimize processing energy efficiency

Target Metrics

Only correctly received information matters

- Goodput : $G = B_S Q M_T R \, (1 - BLER)$ [Mbit/s]
- Spectral efficiency : $\eta_S = Q M_T R \, (1 - BLER) = G/B_S$ [bit/s/Hz]
- Energy efficiency: $\eta_{e,idd} = G/P_{idd}$ [bit/nJ]

with: B_S : symbol rate

Q : bits per QAM symbol

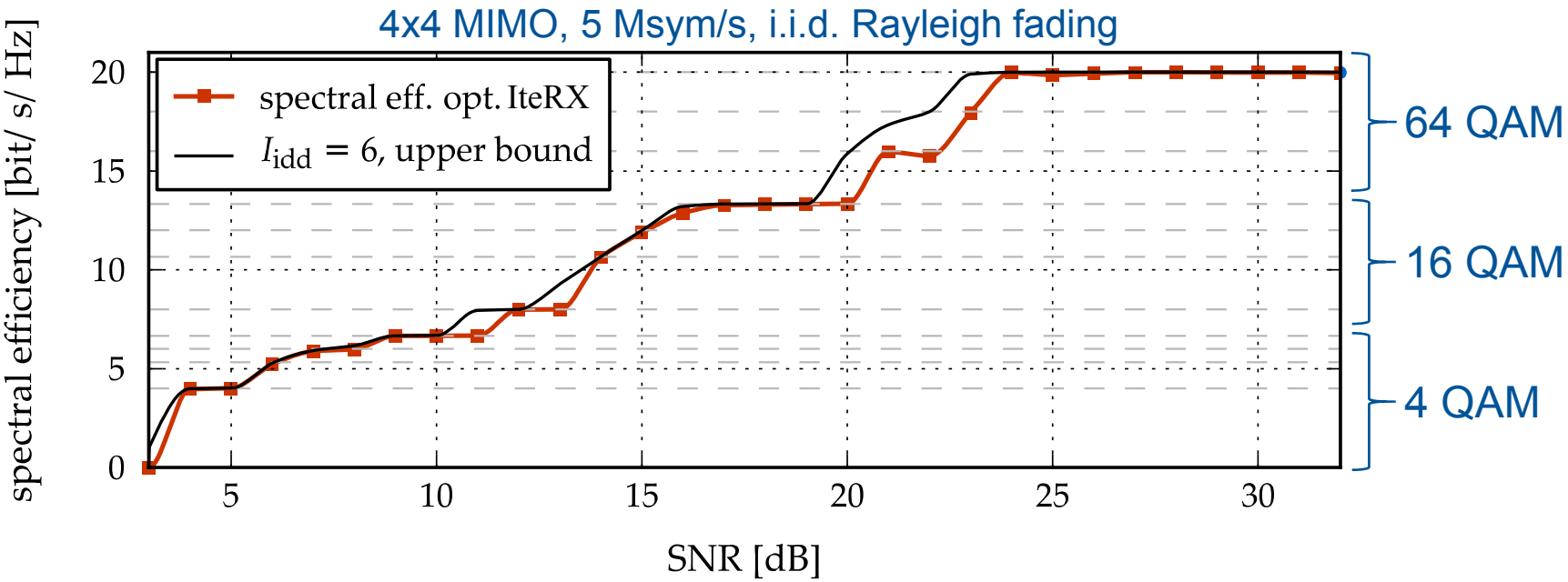
M_T : number of antennae

R : code rate

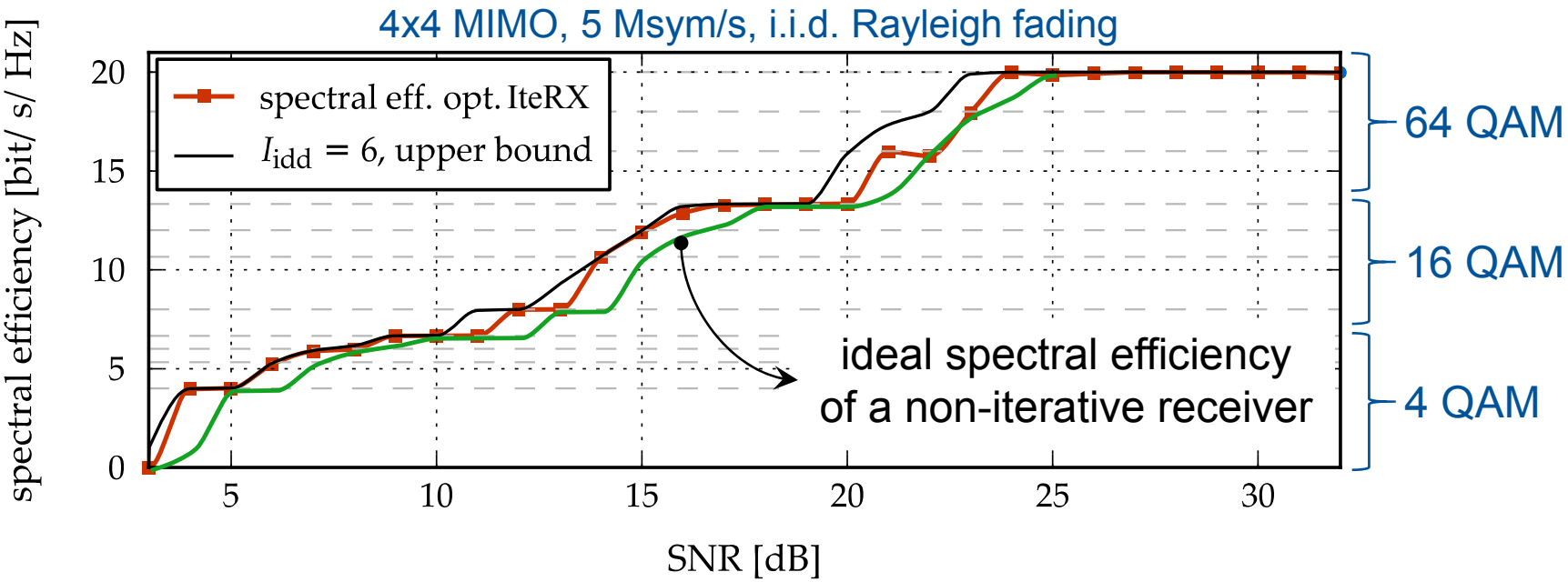
BLER : block error rate (block = code word)

P_{idd} : average power consumed by receiver

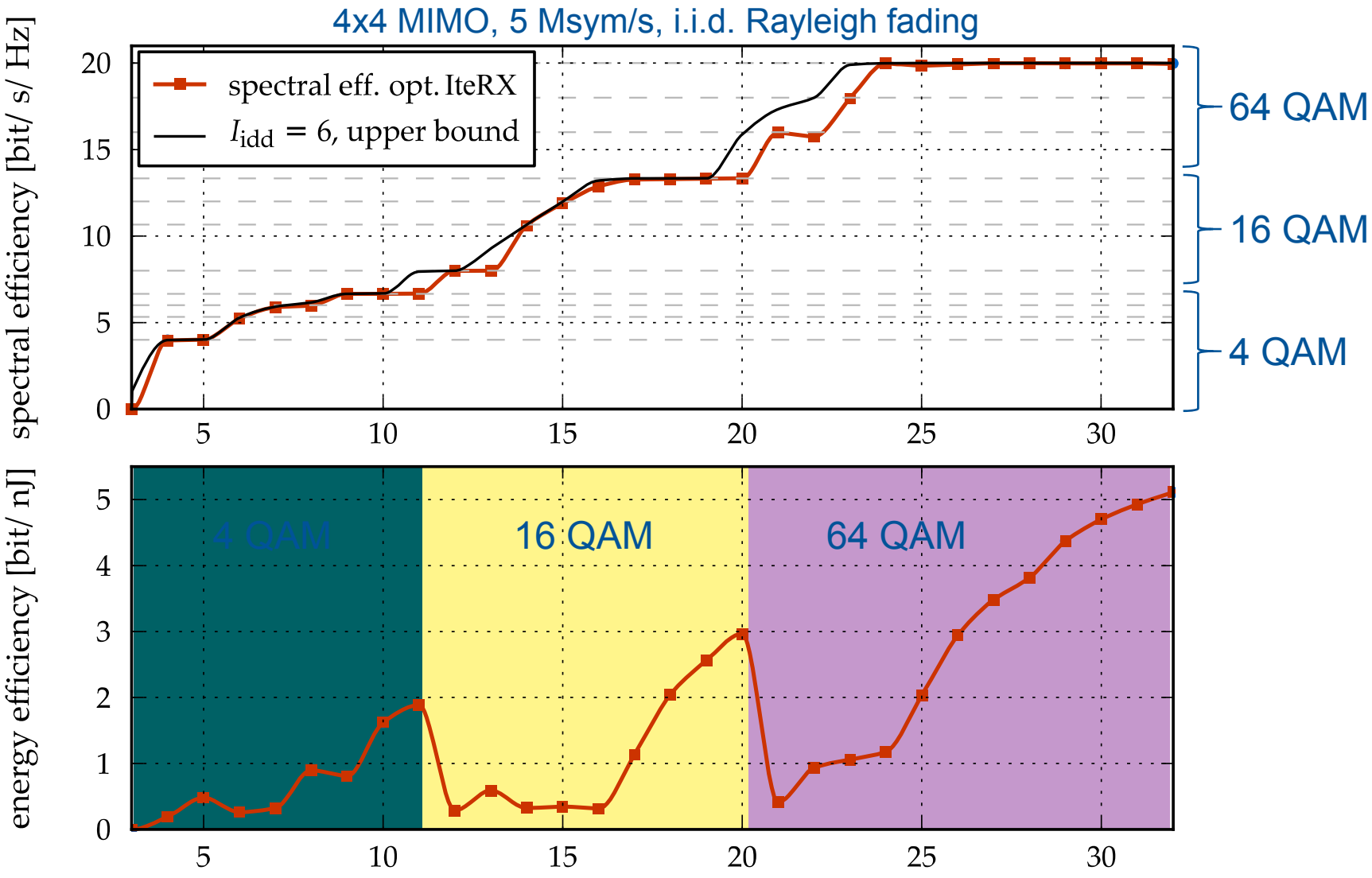
Optimizing for Spectral Efficiency



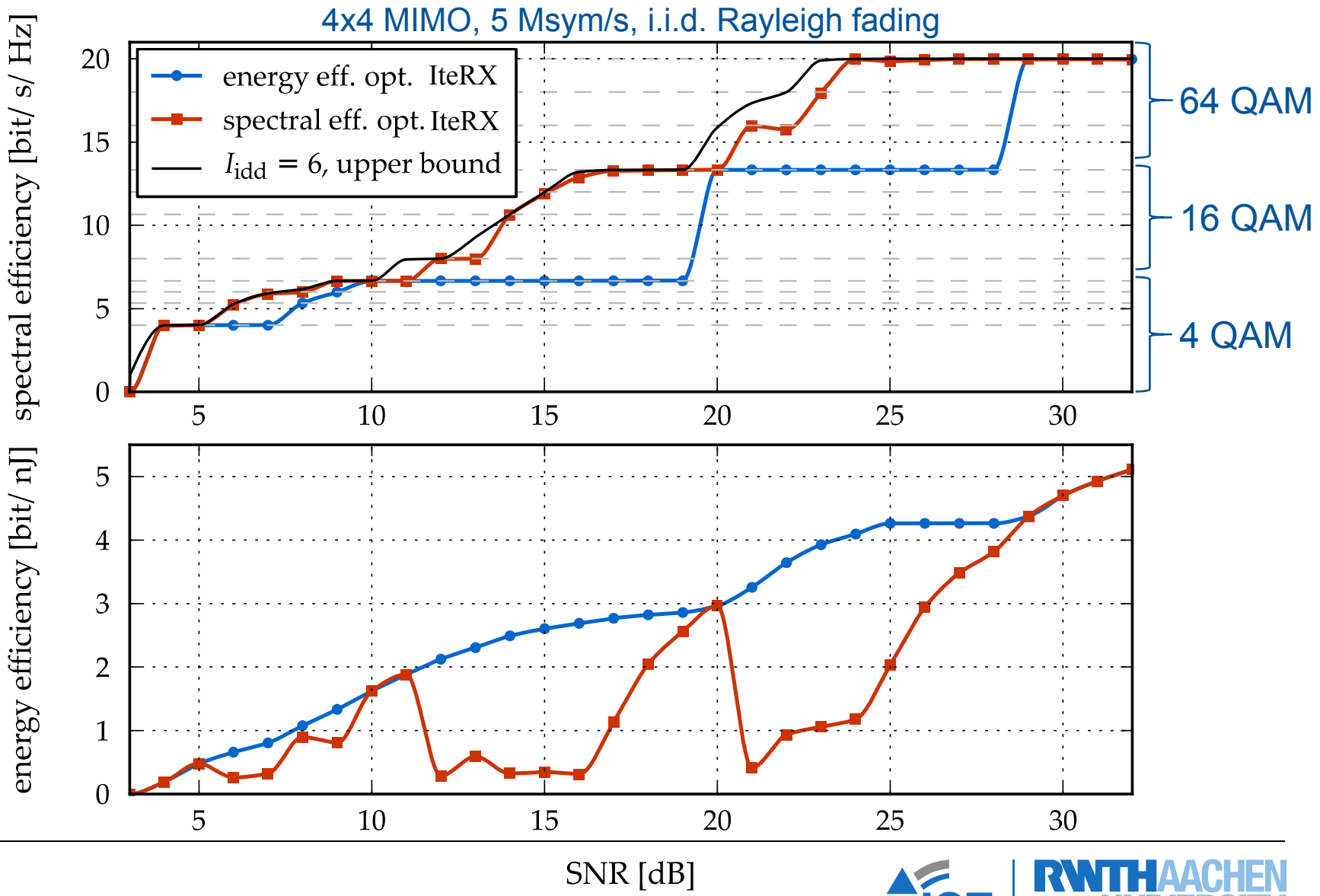
Optimizing for Spectral Efficiency



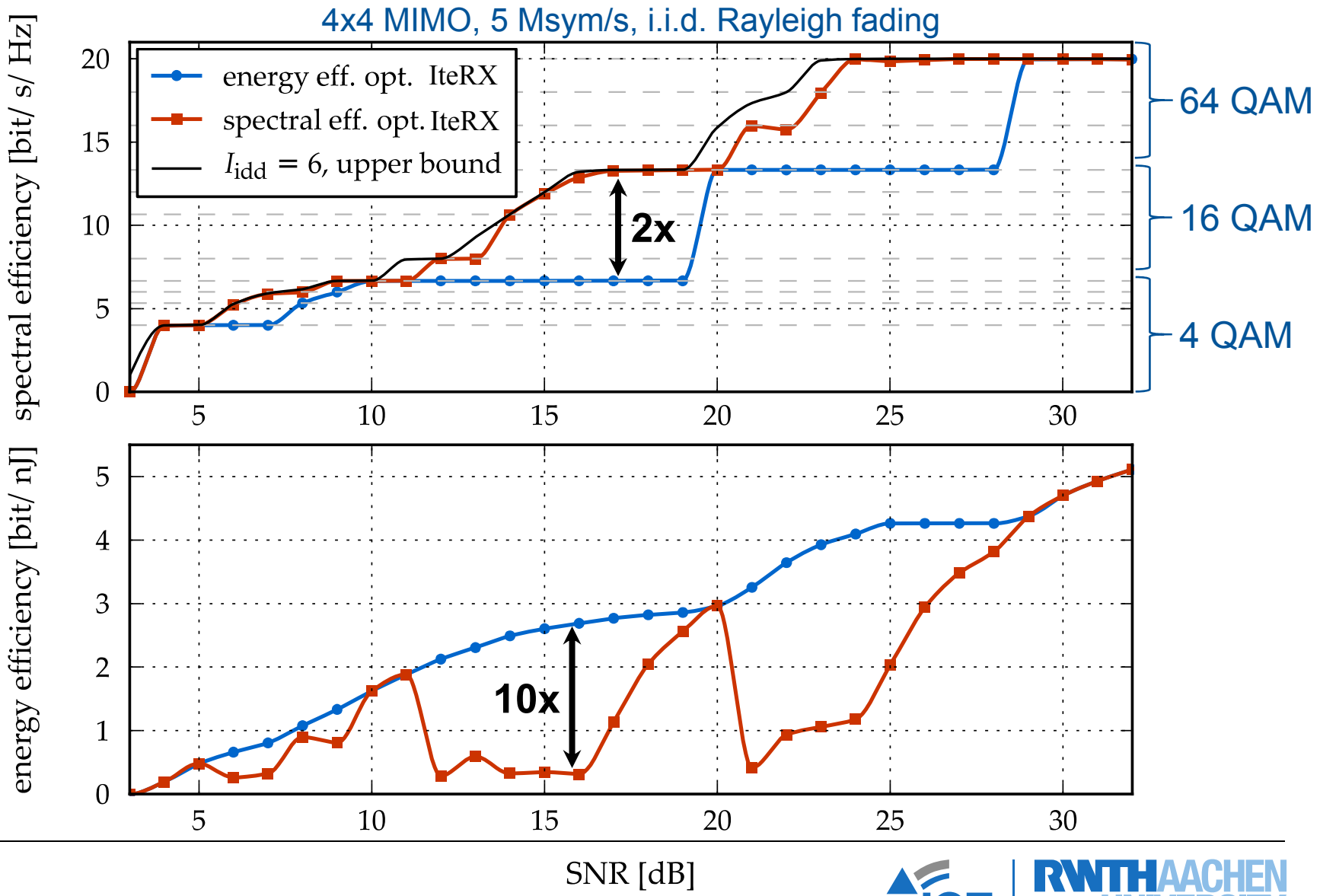
Optimizing for Spectral Efficiency



Targets: Spectral Efficiency vs. Energy Efficiency



Targets: Spectral Efficiency vs. Energy Efficiency



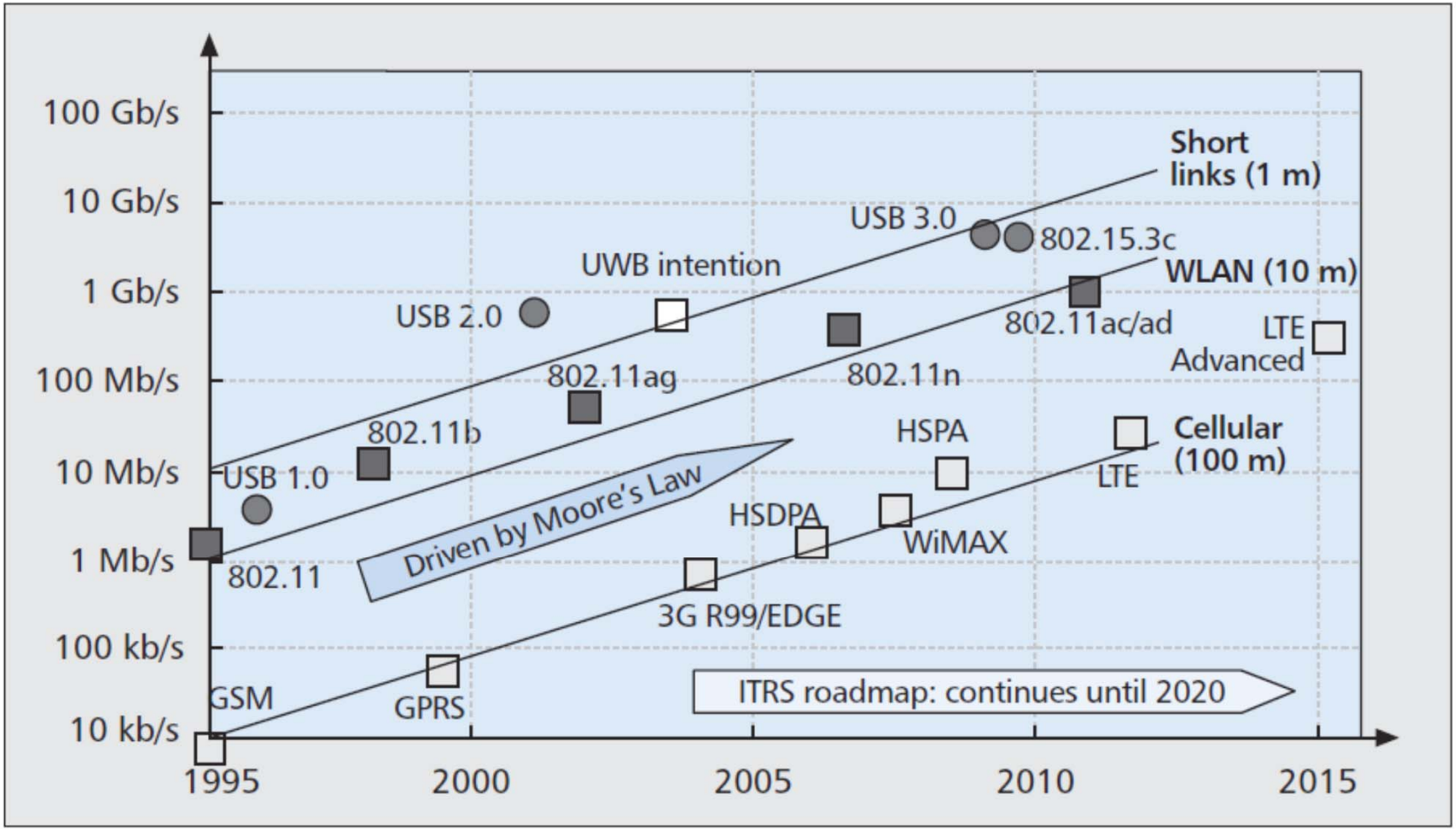
Outline

Performance vs.
Processing Power

Wireless Communication
and Moore's Law

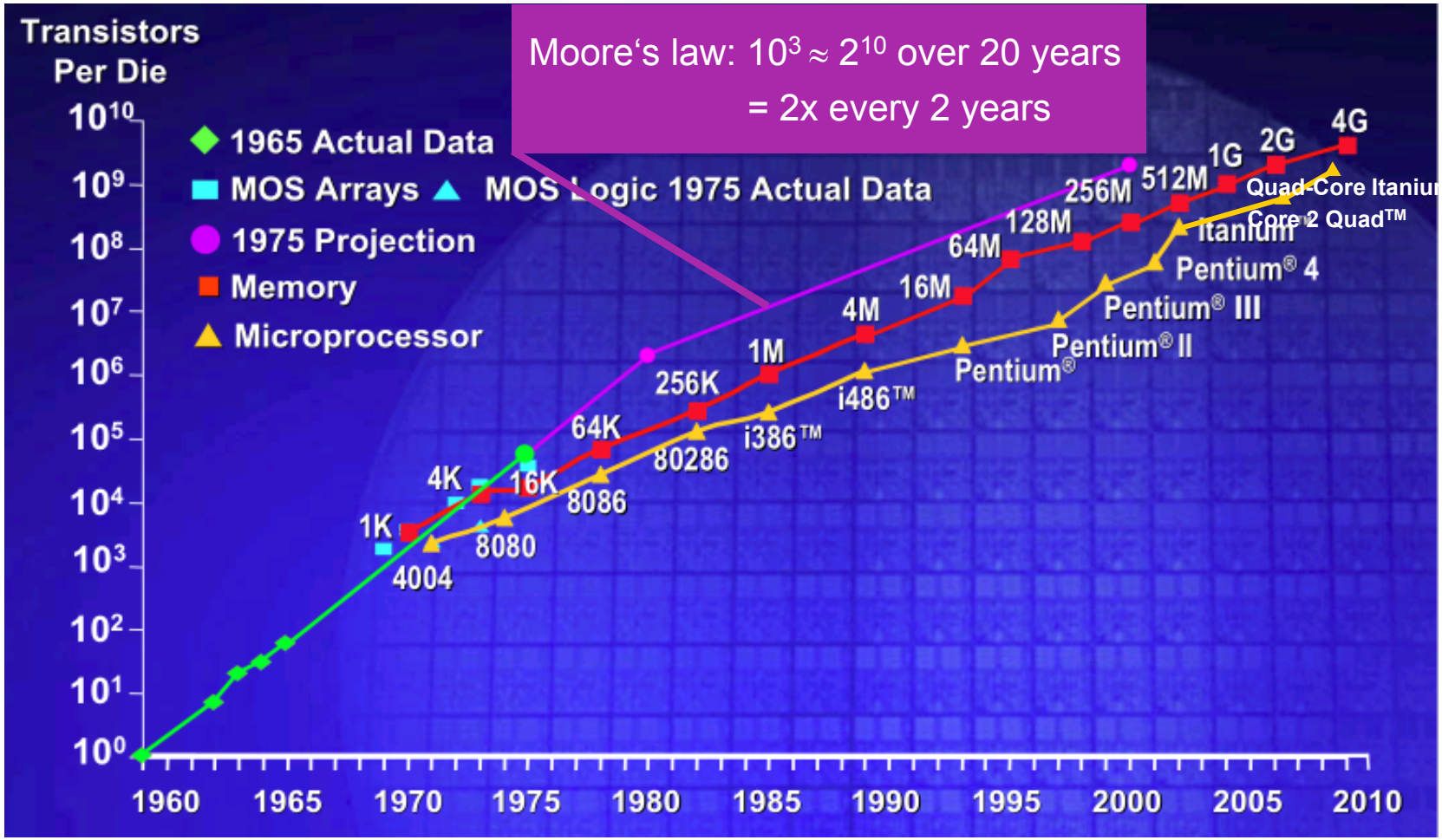
Conclusions for Future
Wireless System Design

Data Rates in Wireless Communication



Source: Fehske, A.; Fettweis, G.; Malmodin, J.; Biczók,G.: The Global Footprint of Mobile Communications: The Ecological and Economic Perspective, IEEE Communication Magazine, August 2011, pp. 55-62.

Silicon Technology Roadmap



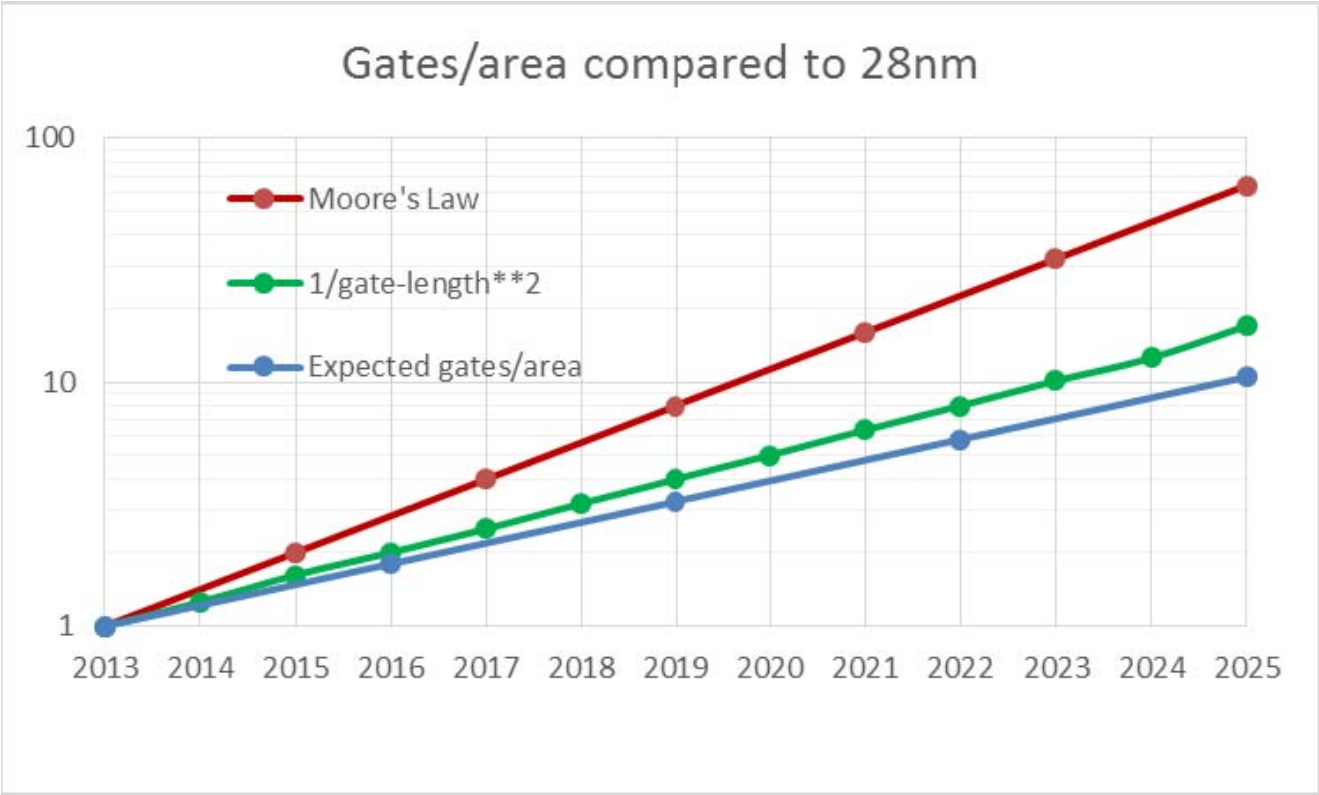
Source: thenextwavefutures.wordpress.com

Silicon technology progress is a key enabler of progress in wireless communication systems

Source: thenextwavefutures.wordpress.com

Moore‘s Law In The Next 10 Years

- Density growth has slowed down significantly (2x only every 3.6 years)



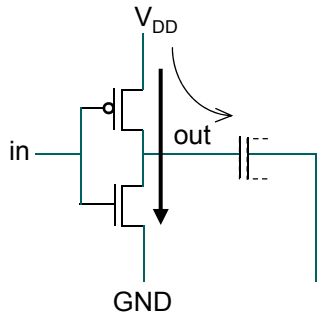
Sources:
1.) ITRS Report 2013
2.) Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Conclusion #1

Not more than 10x complexity
on same chip size by 2025

Energy and Power in CMOS

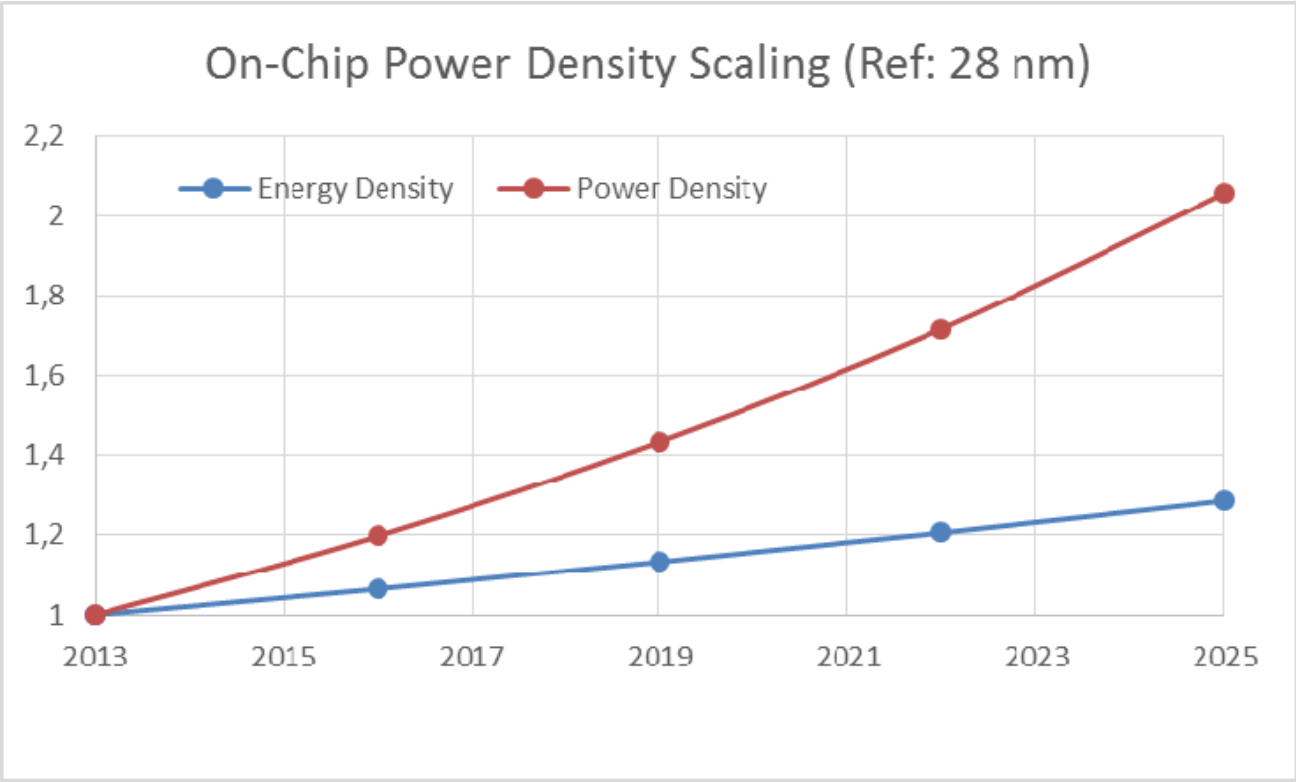
- Switching 0 → 1 charges next gate capacitance
  ⇨$E_d=V_{DD}Q$ [Ws=J]
- Frequency of switching yields power
  $P_d=E_d*f$ [W]



- A second power conversion occurs due to leakage
  $P_l=V_{DD}I_l$ (average power proportional to average on time)
  → will not be discussed here, addressed by execution scheduling

On-Chip Power Density

- Thermal energy and thermal power (Energy*Clock_Frequency)



Sources:
1.) ITRS Report 2013
2.) Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Conclusion #2

Energy efficient design required to avoid doubling of power density (→ heat radiation)

Outline

Performance vs.
Processing Power

Wireless Communication
and Moore's Law

Conclusions for Future
Wireless System Design

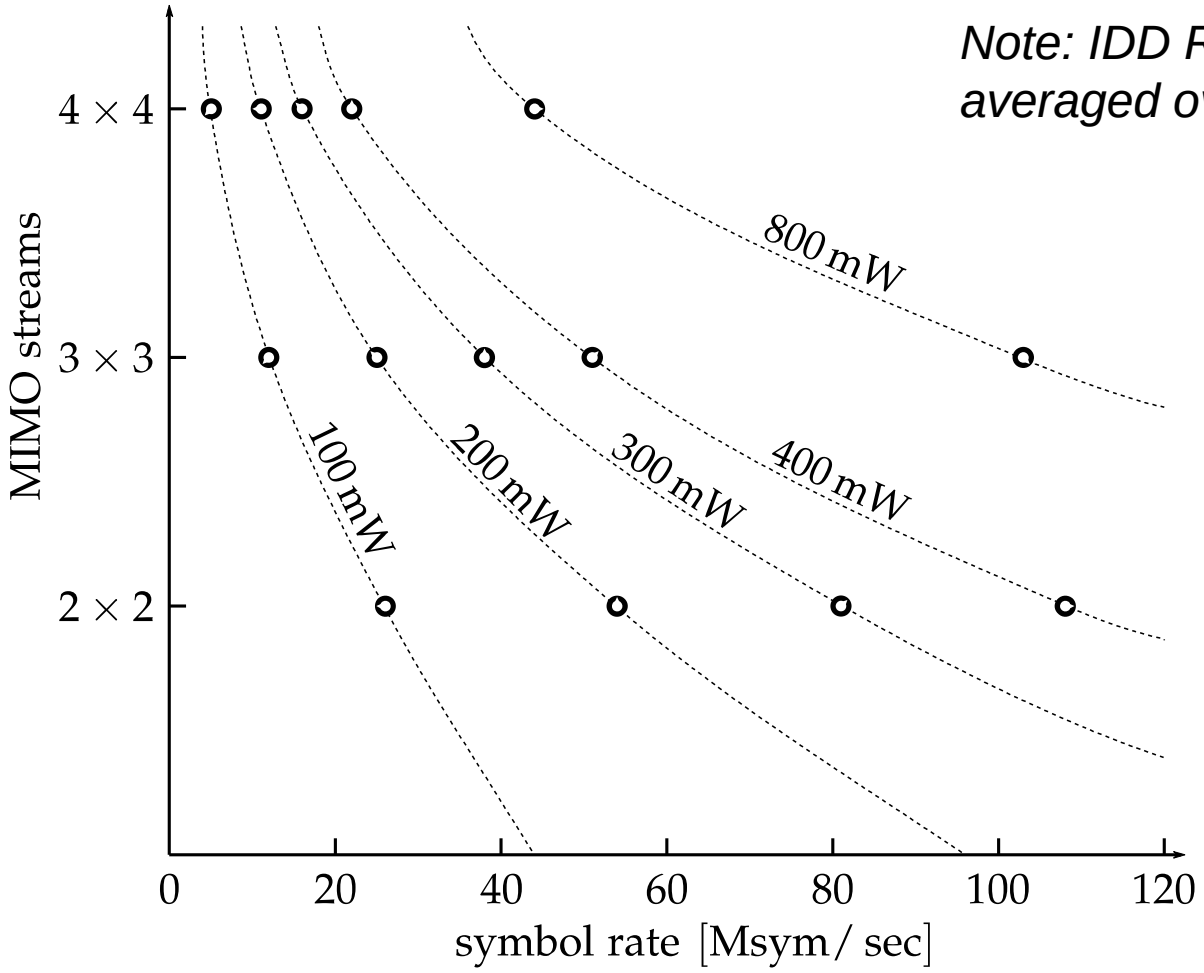
A New Design Paradigm

Wireless system architectures must be designed for processing energy efficient implementation

Processing Power vs. Symbol Rate

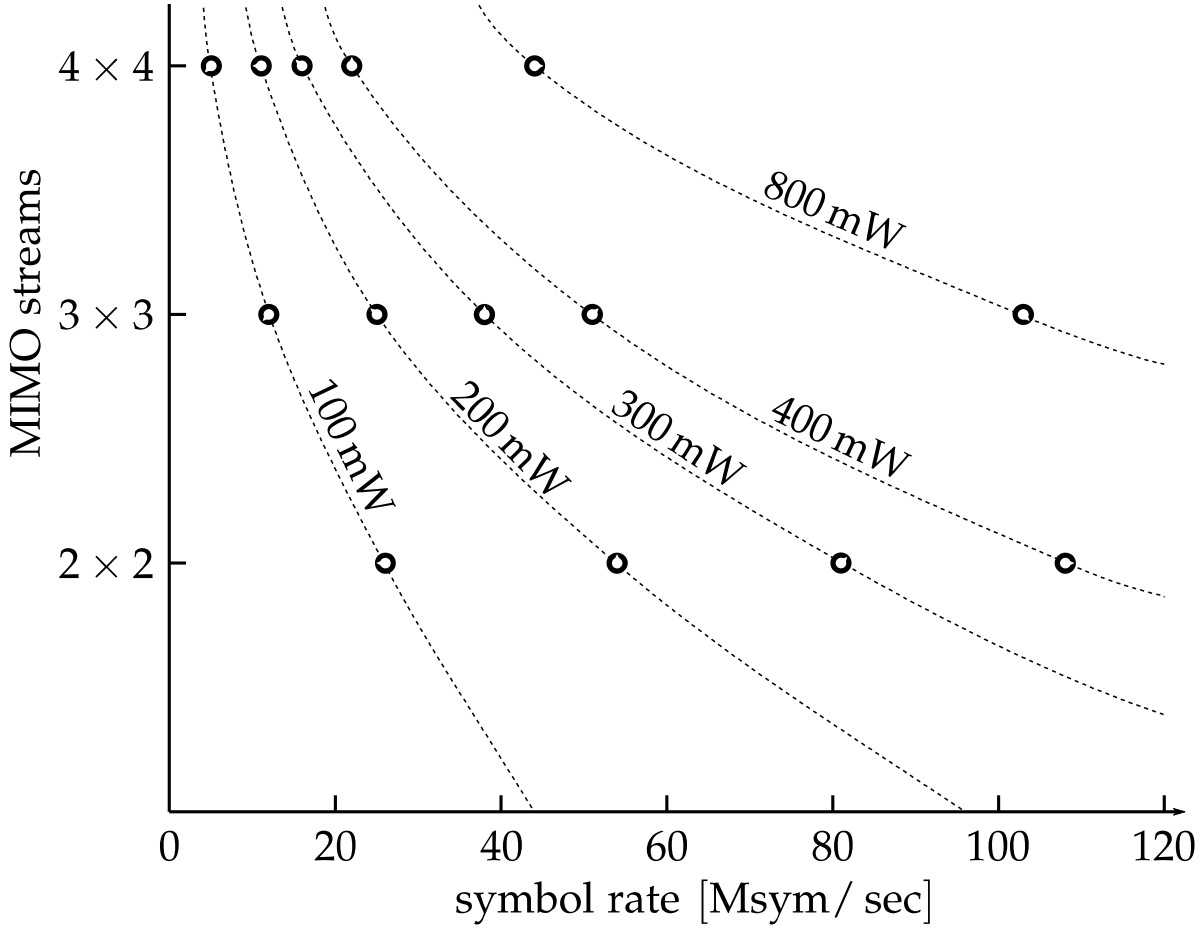
What symbol rate can be served given a power constraint?

Note: IDD RX power consumption averaged over SNR range (3 to 33 dB)



Using Iterative Receivers in Mobile Devices

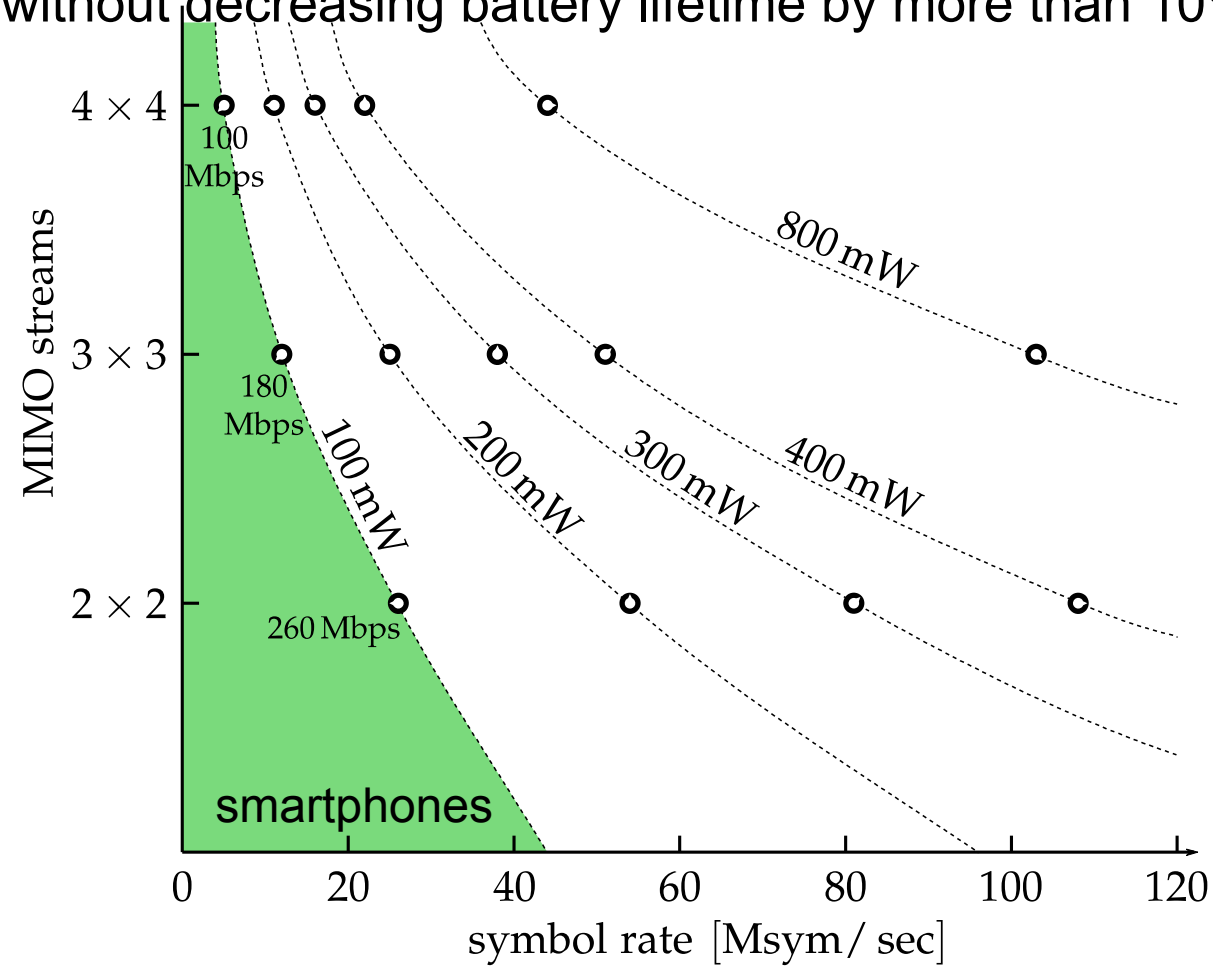
When using iterative detection, what symbol rate can be served without decreasing battery lifetime by more than 10%?



Assumption: Mobile device in WiFi internet browsing scenario

Using Iterative Receivers in Mobile Devices

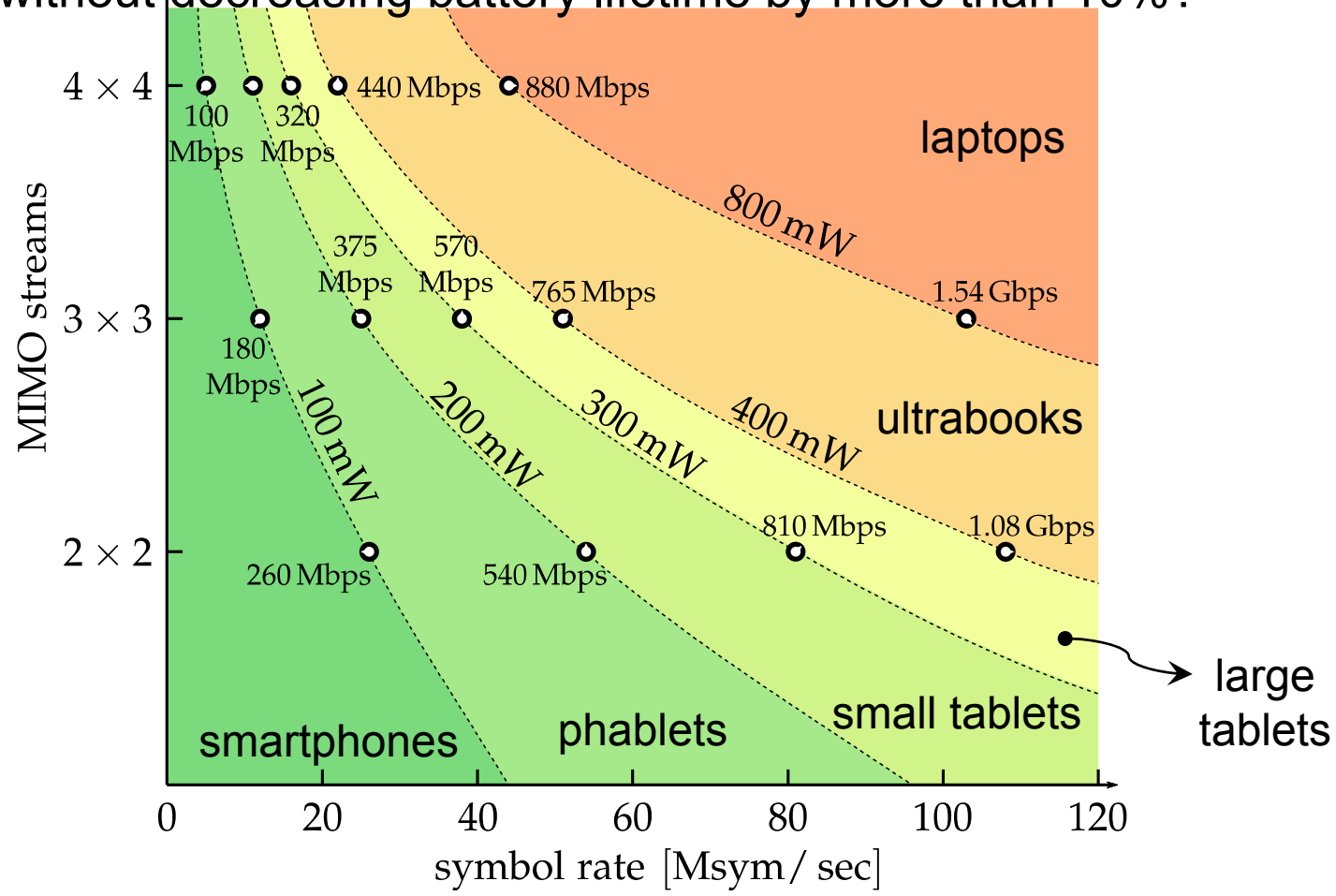
When using iterative detection, what symbol rate can be served without decreasing battery lifetime by more than 10%?



Assumption: Mobile device in WiFi internet browsing scenario

Using Iterative Receivers in Mobile Devices

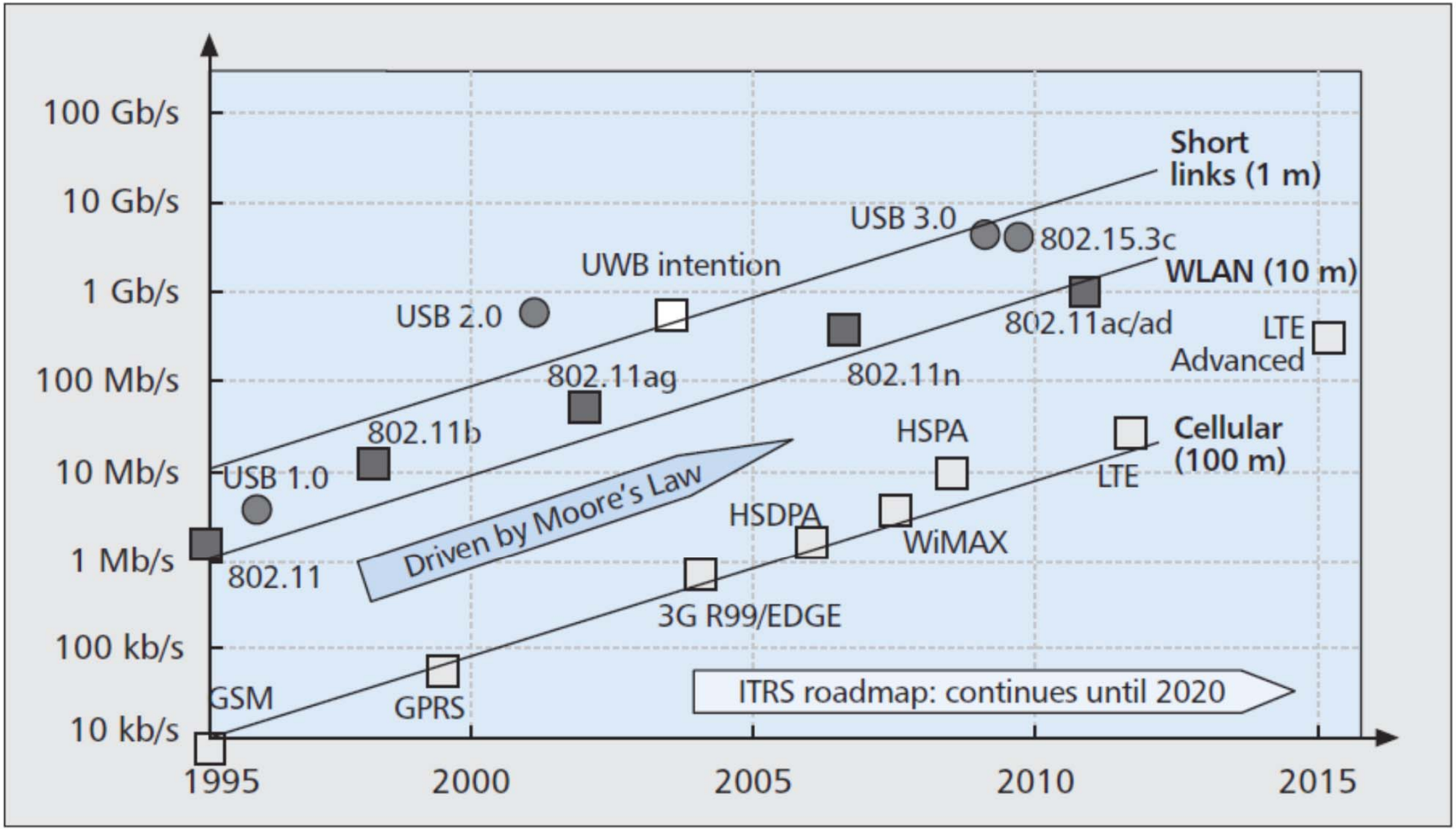
When using iterative detection, what symbol rate can be served without decreasing battery lifetime by more than 10%?



Assumption: Mobile device in WiFi internet browsing scenario

- This was for 65nm technology, now we are at 28nm
- What about future systems and technologies?

Data Rates in Wireless Communication



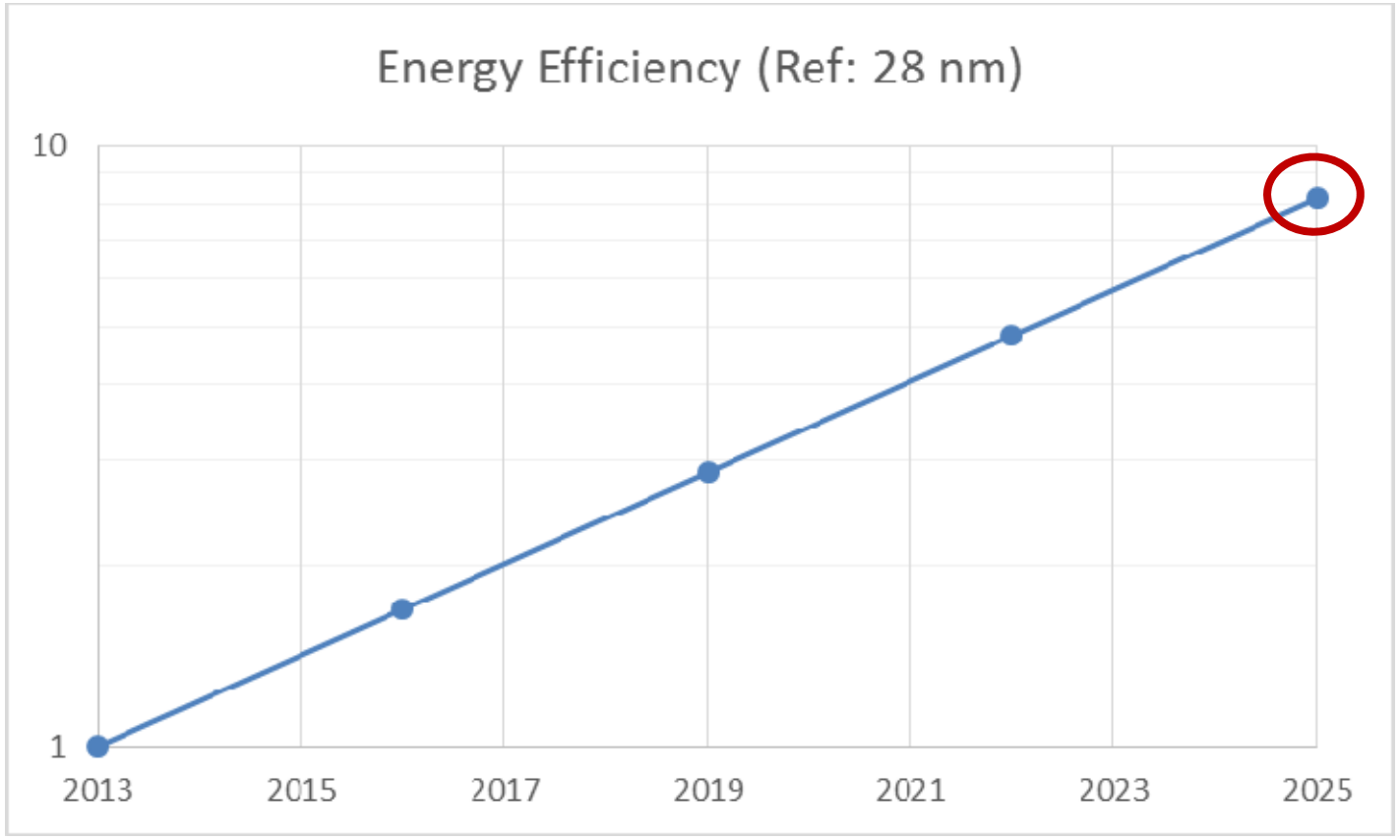
Source: Fehske, A.; Fettweis, G.; Malmodin, J.; Biczók,G.: The Global Footprint of Mobile Communications: The Ecological and Economic Perspective, IEEE Communication Magazine, August 2011, pp. 55-62.

On-Chip Power Limitation

- Power consumption limit without forced cooling:
  order of magnitude is 1W
- Assume a data rate of 1 Gb/s, then the processing energy per bit is limited to 1W / 1 Gb/s = 1 nJ
- According to a NVIDIA study at 28nm:
  1-2 GOps/W ⇨ 1-2 Operations/nJ

Source: Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Processing Energy Efficiency



8.17
over
28 nm

On-Chip Power Limitation

- Power consumption limit without forced cooling:
  order of magnitude is 1W
- Assume a data rate of 1 Gb/s, then the processing energy per bit is limited to 1W / 1 Gb/s = 1 nJ
- According to a NVIDIA study at 28nm:
  1-2 GOps/W ⇨ 1-2 Ops/nJ or 1-2 Ops/bit
- Technology-based gain at 7 nm compared to 28nm : 8.17
  ⇨ 8-16 Ops/nJ or 8-16 Ops/bit
- Caveats
  - Above numbers do not consider analog and ADC/DAC
  - Consider 10+ Gb/s (numbers are for 2025 technology!)

Source: Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Project LP100: Wireless 100Gb/s and beyond

- Assuming 1 W for processing in a 100 Gb/s transmission (excluding transmit power)
  - Available energy per bit: 10pJ/bit

Research questions:

- How many bits/s within processing energy constraints?

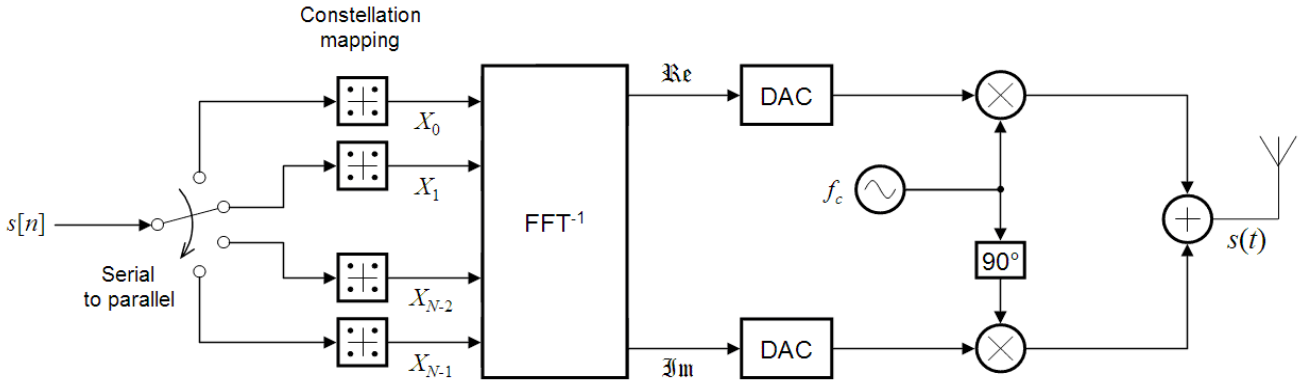
- Which system architecture (MIMO, modulation, coding, …) enables the maximum information bit rate?

    - 2x50 bit/symbol for 1 GHz bandwidth
  - Frequency ranges of interest around 60 GHz or beyond 100 GHz

    ⇨ several GHz of bandwidth available

Example: Modulation Schemes

- High complex modulation scheme OFDM not suitable



  - Extremely power hungry mixed signal processing:
    AD-DA conversion, power amplifier and low noise amplifier

- Alternative modulation scheme preferrable
  - Rectangular QAM
  - Circular QAM

Example: Power Efficient Processing

- Processing power can be reduced by lowering V_{DD}
- This leads to transient faults in case of same clock frequency
- When processing noisy signals the faults represent additional noise

  ⇨ Tradeoff between processing energy efficiency and communication performance

Summary

- In the past power efficiency only referred to maximizing bit/s for a given transmit power limit

- In the future processing power efficiency, i.e. bit/s for a given processing power limit, will be at least equally important

- Processing power efficiency is most critical for bad channel conditions and for very high data rates