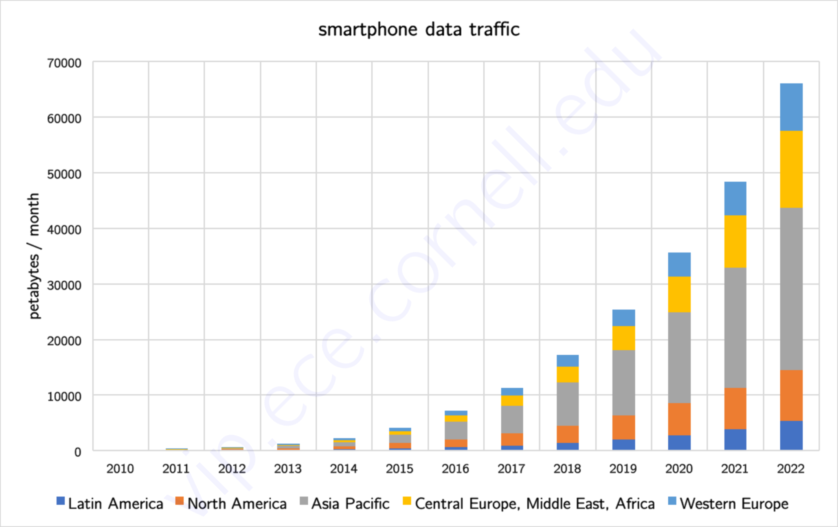
Reliable Communication in Massive MIMO with Low-Precision Converters

Christoph Studer

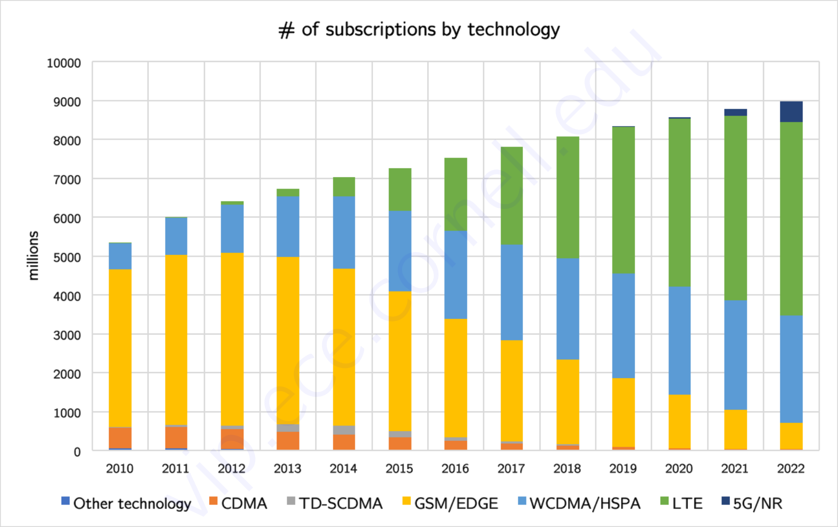
VLSI Information Processing Group
Cornell University

Smartphone traffic evolution needs technology revolution



Source: Ericsson, June 2017

Fifth-generation (5G) may come to rescue



Source: Ericsson, June 2017

5G has a wide range of requirements

data rates

traffic volume density

energy efficiency

reliability

Massive MIMO may provide solutions to all these

data rates

traffic volume density

energy efficiency

reliability

Multiple-input multiple-output (MIMO) principles

- ✓ Multipath propagation offers “spatial bandwidth”
- ✓ MIMO with spatial multiplexing improves throughput, coverage, and range at **no expense in transmit power**
- ✓ MIMO technology enjoys widespread use in many standards

Conventional small-scale point-to-point or multi-user (MU) MIMO systems already **reach their limits** in terms of **system throughput**

Massive MIMO*: anticipated solution for 5G

- Equip the basestation (BS) with hundreds or **thousands of antennas** B
- Serve tens of users U in the **same time-frequency resource**
- Large BS antenna array enables **high array gain** and **fine-grained beamforming**

*Other terms for the same technology: very-large MIMO, full-dimension MIMO, mega MIMO, hyper MIMO, extreme MIMO, large-scale antenna systems, etc.

Promised gains of massive MIMO (in theory)

- ✓ Improved spectral efficiency, coverage, and range
  - ➜ $10\times$ capacity increase over small-scale MIMO
  - ➜ $100\times$ increased radiated efficiency
- ✓ Fading can be mitigated substantially $\rightarrow$ "channel hardening"
- ✓ Significant cost and energy savings in analog RF circuitry
- ✓ Robust to RF and hardware impairments
- ✓ Simple baseband algorithms achieve optimal performance

Short "history" of massive MIMO

- 2010: Conceived by Tom Marzetta (Nokia Bell Labs) [1]
- 2012: First testbed for 64×15 massive MIMO system [2]
- 2013: Samsung achieves > 1 Gb/s with 64 BS antennas [3]
- 2016: ZTE releases first pre-5G BS with 64 antennas [4]
- 2017: Sprint & Ericsson field tests with 64 antennas [5]

Google Scholar search for "Massive MIMO" yields 13,300 results...

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE T-WCOM, 2010

[2] C. Shepard, H. Yu, N. Anand, L. E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: practical many-antenna base stations," ACM MobiCom, 2012

[3] H.Benn, "Vision and key features for 5th generation (5G) cellular," Samsung R&D Institute UK, 2014

[4] "ZTE Pre5G massive MIMO base station sets record for capacity," ZTE Press Center, 2016

[5] "Sprint and Ericsson conduct first U.S. field tests for 2.5 GHz massive MIMO," Sprint Press Release, 2017

Practical challenges

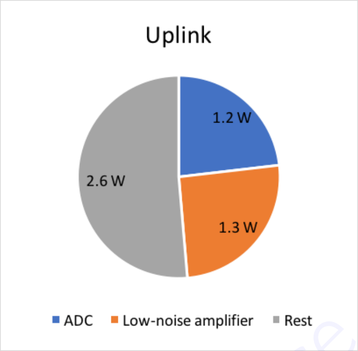
Practical challenges

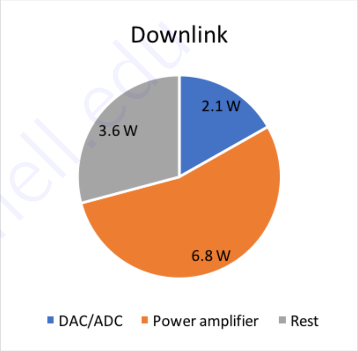
- ✗ The presence of hundreds or thousands of high-quality RF chains causes **excessive system costs** and **power consumption**

- ✗ High-precision ADCs/DACs cause **high amount of raw baseband data** that must be transported and processed

- ✗ The large amount of **data must be processed at high rates and low latency** and often within a single computing fabric

Power breakdown of a single, high-quality RF chain





Analog circuit power of a single RF chain in a picocell BS in Watt [1]

Data converters & **amplifiers** consume large portion of BS power

[1] C. Desset et al., "Flexible Power Modeling of LTE Base Stations," IEEE WCNC, 2012

We will show that massive MIMO enables reliable communication with low-precision data converters

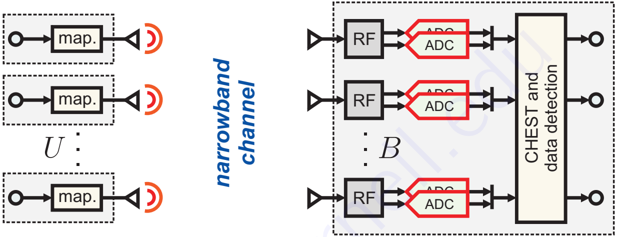
Why should we use low-resolution ADCs/DACs at BS?

- Lower resolution → **lower power consumption**
  - Power of ADCs/DACs scales exponentially with bits
  - Massive MIMO requires a large number of ADCs/DACs
- Lower resolution → **reduced hardware costs**
  - Remaining RF circuitry (**amplifiers**, filters, etc.) needs to operate at precision "just above" the quantization noise floor*
  - Extreme case of 1-bit data converters enables the use of high-efficiency, low-power, and nonlinear RF circuitry
- Lower resolution → **less data transported from/to BBU**
  - Example: 128 antenna BS and 10-bit ADCs/DACs operating at 80 MS/s produces more than **200 Gb/s** of raw baseband data

*terms and conditions apply

Uplink: users $\rightarrow$ basestation

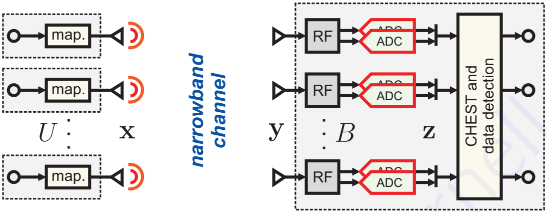
Quantized massive MIMO uplink



We consider infinite-precision DACs at the user equipments (UEs) and low-precision ADCs at the basestation (BS) side

- Is reliable communication with low-precision ADCs possible?
- How many quantization bits are required?
- Do we need complicated/complex baseband algorithms?

System model details

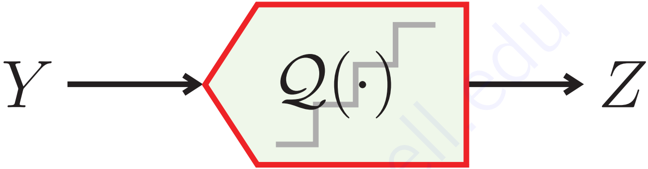


(Narrowband) channel model:

$$\mathbf{y} = \mathcal{Q}(\mathbf{H}\mathbf{x} + \mathbf{n})$$

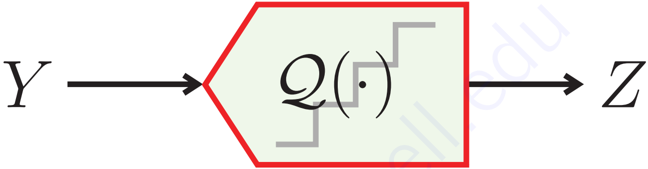
- $\mathbf{y} \in \mathcal{Y}^B$ receive signal at BS; $\mathcal{Y}$ quantization alphabet
- $\mathcal{Q}(\cdot)$ describes the joint operation of the $2B$ ADCs at the BS
- $\mathbf{H} \in \mathbb{C}^{B \times U}$ MIMO channel matrix
- $\mathbf{x} \in \mathcal{O}^U$ transmitted information symbols (e.g., QPSK)
- $\mathbf{n} \in \mathbb{C}^B$ noise; i.i.d. circularly symmetric Gaussian, variance N_0

How can we deal with quantization errors? Model 1



- Assume that input Y is a zero-mean Gaussian random variable
- Simple model: $Z = \mathcal{Q}(Y) = Y + Q$
  - Quantization error Q is statistically dependent on input Y
  - An exact **analysis** with this approximate model is **difficult**

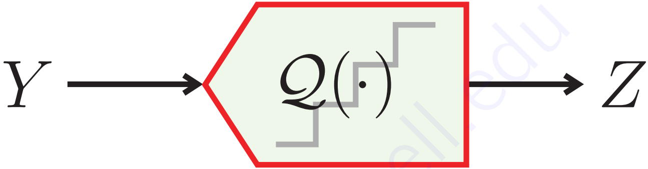
How can we deal with quantization errors? Model 2



- Assume that input Y is a zero-mean Gaussian random variable
- Model input-output relation statistically [1]
  - Probability distribution $p(Z \mid Y)$ has a known form
  - Exact model but a theoretical **analysis** is **difficult**

[1] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," IEEE SP-L, 2010

How can we deal with quantization errors? Model 3



- Assume that input Y is a zero-mean Gaussian random variable
- Bussgang's theorem [2]: $Z = \mathcal{Q}(Y) = gY + E$
  - Quantization error E is **uncorrelated** with input Y
  - This decomposition is exact $\rightarrow$ **theoretical analysis possible**

[1] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," IEEE SP-L, 2010

[2] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," MIT Research Laboratory of Electronics, technical report, **1952**

Consider linear channel estimation and detection

- **Bussgang's theorem linearizes the system model**:

$$\mathbf{y} = \mathcal{Q}(\mathbf{H}\mathbf{x} + \mathbf{n}) = \mathbf{G}\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{n} + \mathbf{e}$$

  where $\mathbf{G}$ is a diagonal matrix that depends on the ADC and error $\mathbf{e}$ is uncorrelated with $\mathbf{x}$

- Using Bussgang's theorem, we derive a linear channel estimator:

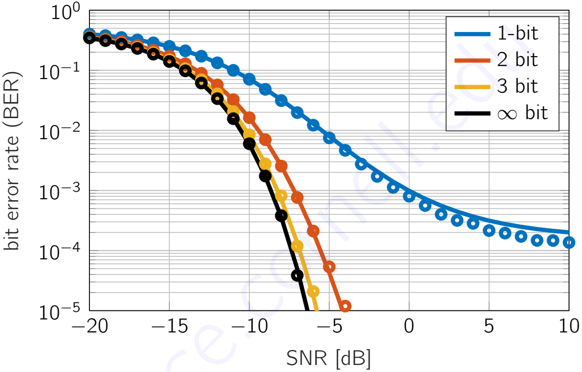
$$\hat{\mathbf{H}} = \frac{g \sum_{t=1}^{P} \mathbf{y}_t \mathbf{x}_t^H}{g^2 P \cdot SNR + g^2 + (1 - g^2)(U \cdot SNR + 1)}$$

  P = number of pilots; g = Bussgang gain that depends on ADC

- Zero-forcing (ZF) equalization: $\hat{\mathbf{x}} = (\hat{\mathbf{H}})^{\dagger}\mathbf{y}$

Do such simple receive algorithms work for coarse quantization?

Uncoded BER vs. SNR: ZF with QPSK



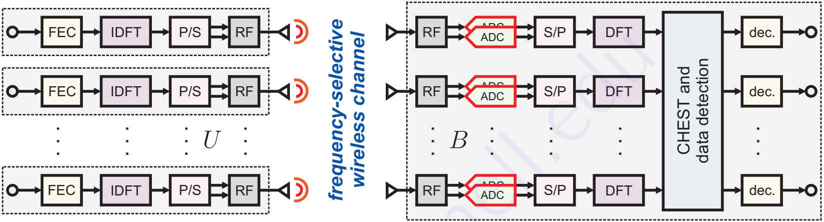
$B = 200$ antennas, $U = 10$ users, $P = 10$ pilots, Rayleigh fading

- Markers correspond to simulation results; solid lines correspond to Bussgang-based approximations

[1] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and CS, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," IEEE T-WC, 2017

Are these results still valid for realistic wideband massive MIMO-OFDM systems?

Full-fledged massive MIMO-OFDM system model [1]



- We consider quantized channel estimation and data detection
- We compare two methods:
  - Exact model (model quantization statistically)
  - Approximate model (treat as unorrelated noise)

➜ How many bits are required for reliable uplink transmission?

[1] CS and G. Durisi, "Quantized massive MIMO-OFDM uplink," IEEE T-WCOM, 2016

Two methods: exact & complex vs. simple & suboptimal

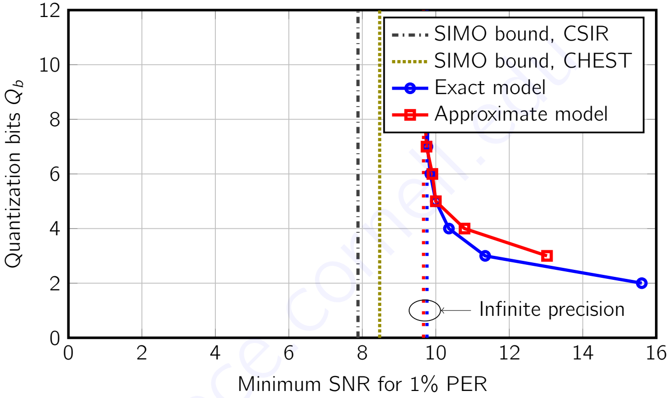
- Exact MMSE equalizer for the quantized system requires the solution to a large convex optimization problem:

$$\begin{cases} \underset{\tilde{s}_w, w \in \Omega_{\text{data}}}{\text{minimize}} & -\sum_{b=1}^{B} \log p(\mathbf{q}_b \,|\, \mathbf{F}^H \mathbf{z}_b) + \sum_{w \in \Omega_{\text{data}}} E_s^{-1} \|\mathbf{s}_w\|_2^2 \\ \text{subject to} & \{\mathbf{z}_b\}_{b=1}^{B} = \mathcal{T}\{\widehat{\mathbf{H}}_w \mathbf{s}_w\}_{w=1}^{W} \\ & \mathbf{s}_w = \mathbf{t}_w,\, w \in \Omega_{\text{pilot}} \end{cases}$$

- To minimize complexity, we can alternatively **use conventional MIMO-OFDM receivers** that ignore the quantizer altogether

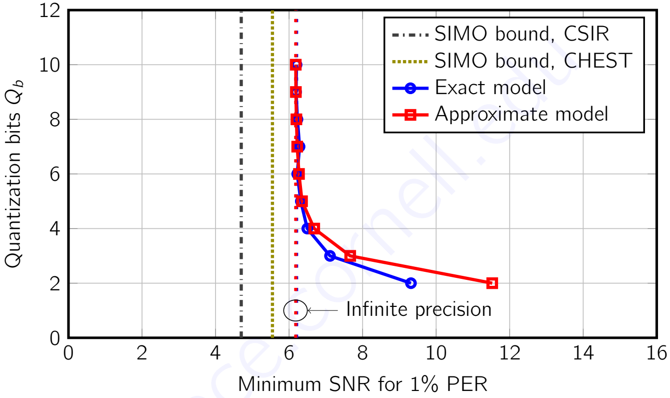
- The same two approaches exist for channel estimation

Performance of quantized massive MU-MIMO-OFDM



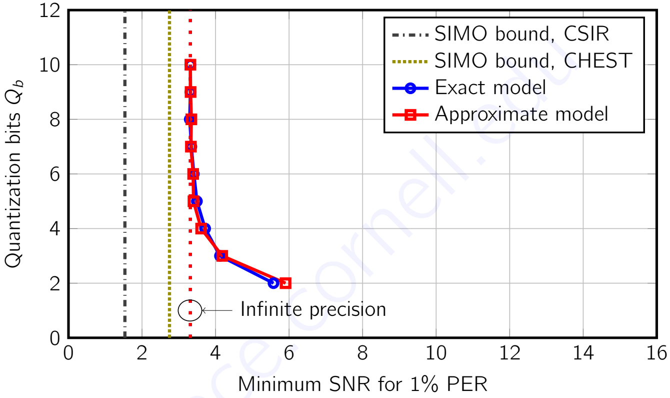
32 × 8 MU-MIMO-OFDM system, 128 subcarriers, 16-QAM, rate-5/6 convolutional code, Rayleigh fading, standard pilot-based training

Performance of quantized massive MU-MIMO-OFDM



64 × 8 MU-MIMO-OFDM system, 128 subcarriers, 16-QAM,
rate-5/6 convolutional code, Rayleigh fading, standard pilot-based training

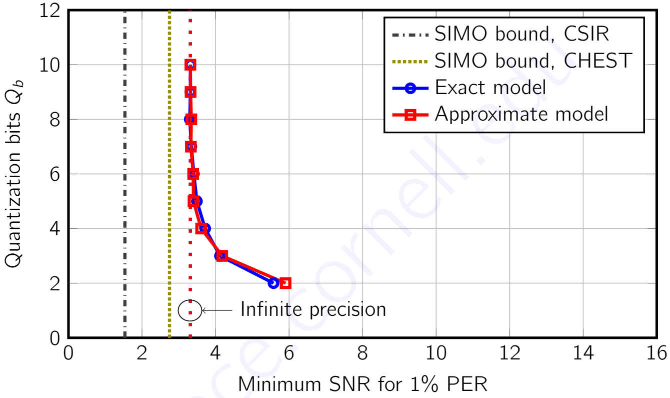
Performance of quantized massive MU-MIMO-OFDM



128 × 8 MU-MIMO-OFDM system, 128 subcarriers, 16-QAM,
rate-5/6 convolutional code, Rayleigh fading, standard pilot-based training

We can use traditional MIMO-OFDM receivers with 4-6 bit ADCs

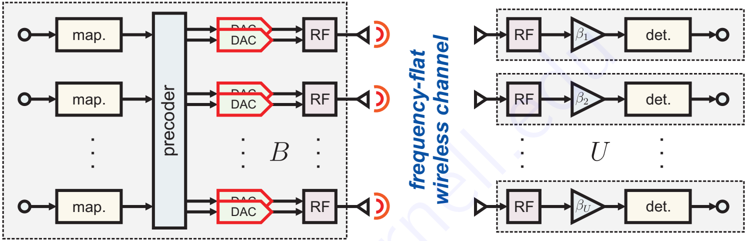
Performance of quantized massive MU-MIMO-OFDM



128 × 8 MU-MIMO-OFDM system, 128 subcarriers, 16-QAM,
rate-5/6 convolutional code, Rayleigh fading, standard pilot-based training

Downlink: basestation → users

Quantized massive MIMO downlink with low-res. DACs



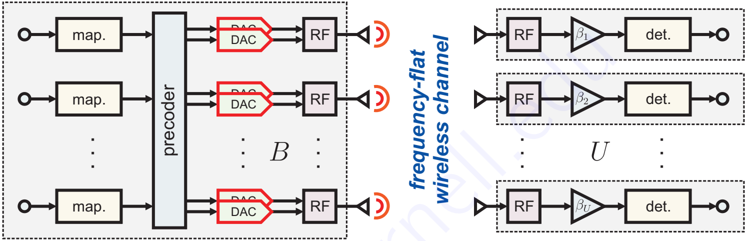
We consider low-precision DACs at the basestation (BS)
and infinite-precision ADCs at the UE side

➜ Is reliable communication with low-precision DACs possible?

➜ How many quantization bits are required?

➜ Do we need complicated/complex baseband algorithms?

Quantized massive MIMO downlink with low-res. DACs



(Narrowband) channel model: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$

- $\mathbf{y} \in \mathbb{C}^U$ receive signals at U users; $\mathbf{y} = [y_1, \ldots, y_U]^T$
- $\mathbf{H} \in \mathbb{C}^{U \times B}$ MIMO channel matrix
- $\mathbf{x}(\mathbf{s}, \mathbf{H}) \in \mathcal{X}^B$ transmitted vector; satisfies $\mathbb{E}\left[\|\mathbf{x}\|^2\right] \leq \rho$
- $\mathbf{s} \in \mathcal{O}^U$ are the information symbols (e.g., QPSK symbols)
- $\mathbf{n} \in \mathbb{C}^U$ noise; i.i.d. zero-mean Gaussian with variance N_0

The quantized precoding (QP) **problem**

- Optimal precoder finds transmit vector $\mathbf{x}$ and associated β that **minimizes the receive-side MSE** between $\hat{\mathbf{s}}$ and $\mathbf{s}$

$$MSE = \mathbb{E}_{\mathrm{n}}\left[\|\hat{\mathbf{s}} - \mathbf{s}\|_2^2\right] = \|\mathbf{s} - \beta\mathbf{H}\mathbf{x}\|_2^2 + \beta^2 U N_0$$

- The optimal quantized precoding (QP) problem is given by

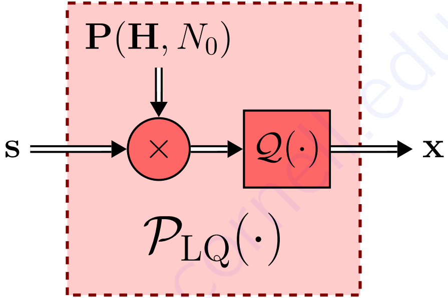
$$(\mathrm{QP}) \quad \begin{cases} \underset{\mathbf{x}\in\mathcal{X}^B,\, \beta\in\mathbb{R}}{\text{minimize}} & \|\mathbf{s} - \beta\mathbf{H}\mathbf{x}\|_2^2 + \beta^2 U N_0 \\ \text{subject to} & \|\mathbf{x}\|_2^2 \leq \rho \end{cases}$$

- **Problem is NP-hard**: Transmit vector $\mathbf{x} \in \mathcal{X}^B$ belongs to a finite lattice due to the finite-precision of DACs

  ✗ For 128 BS antennas with 1-bit DACs, an exhaustive search would evaluate the objective more than 10^{77} times...

We need more efficient, approximate algorithms!

Linear-quantized (LQ) precoding



- **Idea**: multiply the information vector $\mathbf{s} \in \mathcal{O}^U$ with a (linear) precoding matrix $\mathbf{P} \in \mathbb{C}^{B\times U}$ and quantize the result:

$$\mathbf{x} = \mathcal{P}_{\mathrm{LQ}}(\mathbf{s}) = \mathcal{Q}(\mathbf{P}\mathbf{s})$$

$\mathcal{Q}(\cdot)$ models the effect of the $2B$ DACs

Linear-quantized precoding can be analyzed [1]

- We can derive simple expressions for the signal-to-interference-noise-and-distortion ratio (SINDR) **using Bussgang's theorem**:

$$\mathrm{SINDR}_{\mathrm{ZF}} \approx \frac{g^2(B-U)/U}{(1-g^2)+N_0/\rho}$$

g depends on the DAC resolution; ρ is the transmit power

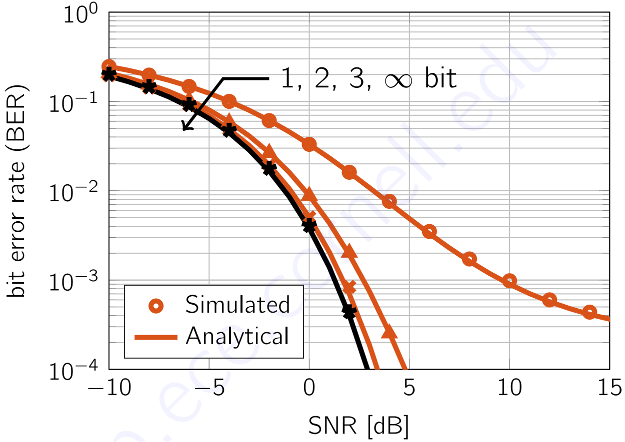
- The SINDR can be used to approximate

$$\mathrm{BER} \approx Q(\sqrt{\mathrm{SINDR}}) \qquad \text{(for QPSK inputs)}$$

$$R_{\mathrm{sum}} \approx U \log_2(1+\mathrm{SINDR}) \qquad \text{(for Gaussian inputs)}$$

[1] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and CS, "Quantized precoding for massive MU-MIMO," IEEE T-COM, 2017

Uncoded BER: simulations vs. analytical expressions

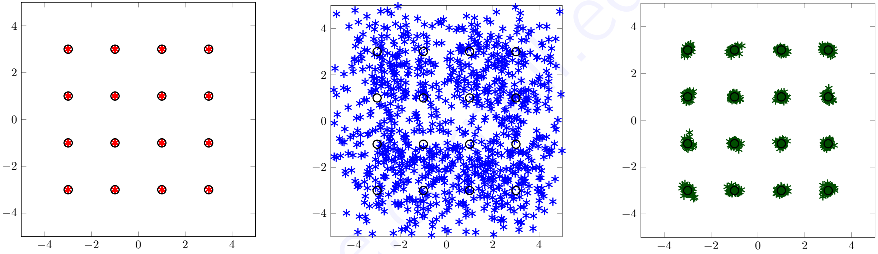


ZF precoding; QPSK signaling; $B = 128$, $U = 16$; Rayleigh fading

Do linear precoders achieve near-optimal performance?

No! Linear-quantized precoding is far from optimal

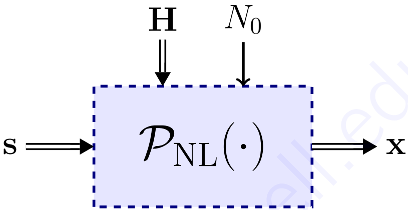
ZF, 10-bit DACs | ZF, 1-bit DACs | optimal, 1-bit DACs



16-QAM signaling, $B = 8$, $U = 2$, SNR $= \infty$; Rayleigh fading

Can we design precoders that achieve **near-optimal performance without resorting to an exhaustive search**?

Solution: Nonlinear (NL) precoding [1]



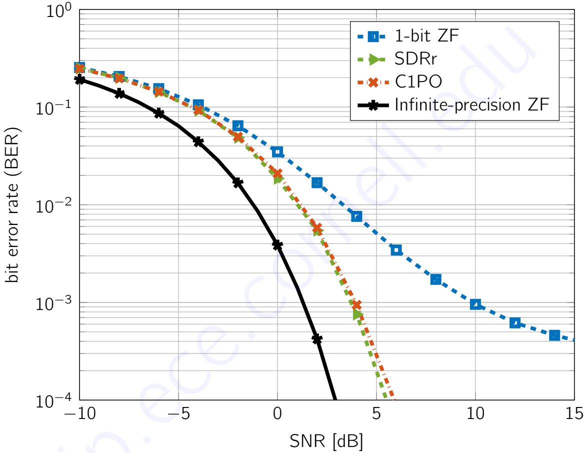
- We now return to the original QP problem:

$$(\text{QP}) \quad \begin{cases} \underset{\mathbf{x}\in\mathcal{X}^B,\, \beta\in\mathbb{R}}{\text{minimize}} & \|\mathbf{s}-\beta\mathbf{H}\mathbf{x}\|_2^2+\beta^2 U N_0 \\ \text{subject to} & \|\mathbf{x}\|_2^2 \leq P \end{cases}$$

- **Idea**: Relax the QP problem so that we can solve it more efficiently using **(non-)convex optimization** techniques:
  - SDR (semidefinite relaxation)
  - C1PO (biConvex 1-bit PrecOding)

[1] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and CS, "Quantized precoding for massive MU-MIMO," IEEE T-COM, 2017

Uncoded BER for NL precoders with 1-bit DACs



QPSK signaling; $B = 128$, $U = 16$; Rayleigh fading

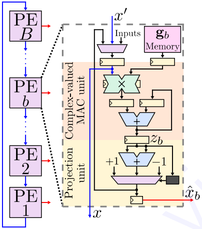
Non-linear precoders significantly outperform LQ precoders

"In theory, theory and practice are the same.
In practice, they are not." [A. Einstein]

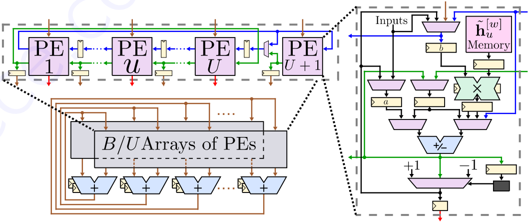
Non-linear precoding can be implemented in practice

- Algorithms that seem efficient can often not be implemented efficiently in very-large scale integration (VLSI) circuits
- Semidefinite relaxation is notoriously difficult to implement
- **CxPO were specifically designed and optimized for VLSI** [1]

C1PO



C2PO



[1] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit Massive MU-MIMO Precoding in VLSI," under review, IEEE JETCAS

FPGAs implementation results [1]

C2PO implementation on a Xilinx Virtex-7 XC7VX690T FPGA

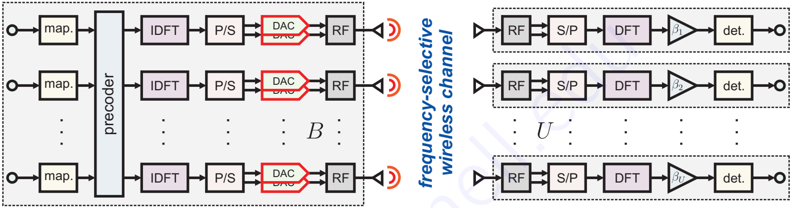
BS antennas B	64	128	256
Slices	6 519	12 690	24 748
LUTs	21 920	43 710	85 323
Flipflops	12 461	26 083	53 409
DSP48 units	272	544	1 088
Clock freq. [MHz]	206	208	193
Latency [clock cycles]	40	41	42
Mvectors/s	5.13	5.06	4.63

16-QAM in 128×16 system yields max. 324 Mb/s throughput

[1] O. Castañeda, S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "1-bit Massive MU-MIMO Precoding in VLSI," under review, IEEE JETCAS

Are these results still valid for
wideband massive MIMO-OFDM systems?

Full-fledged massive MIMO-OFDM system model [1]

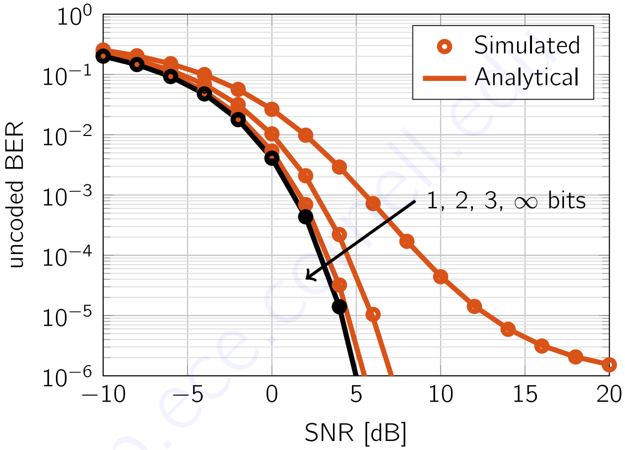


- **Downlink precoding is far more challenging** than uplink:
  - ➜ BS must avoid MU interference via precoding
  - ➜ Nonlinearity introduced by DACs causes intercarrier interference

Bussgang-based analysis can be extended to MIMO-OFDM systems and oversampling DACs [1]

[1] S. Jacobson, G. Durisi, M. Coldrey, and CS, "Linear Precoding with Low-Resolution DACs for Massive MU-MIMO-OFDM Downlink ," submitted to a journal, 2017

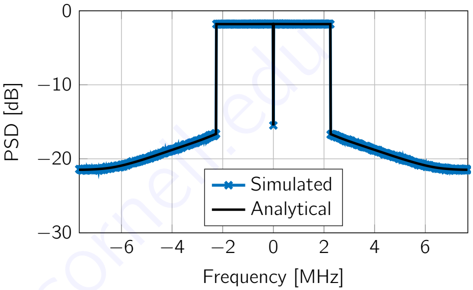
LQ precoding is possible* even with 1-bit DACs



Massive MU-MIMO-OFDM, 1024 subcarriers (300 occupied), 128 BS antennas, 16 users, ZF, QPSK, uncoded, 3.4× oversampling

*terms and conditions apply

Practical issue: out-of-band (OOB) interference [1]



- ✗ Conversion with low-precision DACs causes OOB interference
- ✗ Practical systems would require sharp analog filters to meet stringent OOB requirements

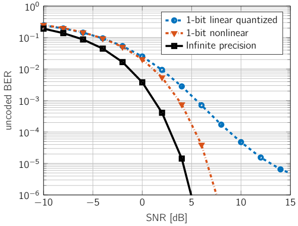
Spectrum regulations may prevent the use of 1-bit precoders

[1] S. Jacobson, G. Durisi, M. Coldrey, and CS, "Linear Precoding with Low-Resolution DACs for Massive MU-MIMO-OFDM Downlink," submitted to a journal, 2017

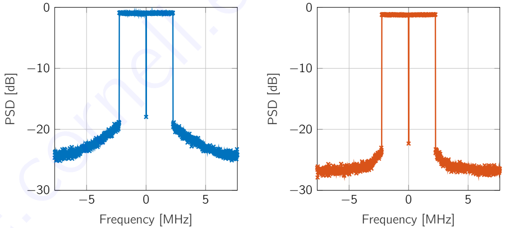
Solution: nonlinear precoders (again)

- Nonlinear precoders approximate optimal quantized precoder for MIMO-OFDM, e.g., using convex relaxation [1]

error-rate comparison



OOB interference comparison

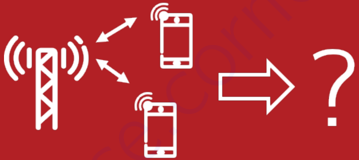


Massive MU-MIMO-OFDM, 4096 subcarriers (1200 occupied), 128 BS antennas, 16 users, QPSK, uncoded, 3.4× oversampling

Significant OOB suppression, better BER, but higher complexity

[1] S. Jacobson, G. Durisi, M. Coldrey, and CS, "Massive MU-MIMO-OFDM downlink with one-bit DACs and linear precoding," submitted to a journal, 2017

Conclusions and open problems



Summary

The use of high-quality RF chains at the BS would result in excessive system costs and power consumption

- ✓ Massive MU-MIMO enables reliable uplink and downlink communication with low-precision data converters
- ✓ Quantization is a nonlinear operation but its artifacts can be analyzed via Bussgang's theorem
- ✓ The uplink requires no changes; 4-to-6 bit are sufficient
- ✓ The downlink is significantly more challenging but feasible

Preliminary results show that nonlinear precoders can be implemented in VLSI and mitigate OOB interference

Open problems

- Uplink
  - Is robust timing, sampling rate, and frequency synchronization still possible with coarse quantization?
  - Can we still use digital time-domain filters after the ADCs?
- Downlink
  - We need new ideas of how to reduce OOB interference
  - We need efficient nonlinear precoders for massive MIMO-OFDM
- System design
  - Precision, power, and cost trade-offs between number of BS antennas and ADC/DAC quality are not well-understood
  - Do all these results still hold for mm-wave systems?

Thanks to my collaborators!

Oscar Castañeda
Cornell

Mikael Coldrey
Ericsson

Giuseppe Durisi
Chalmers

Tom Goldstein
UMD

Ulf Gustavsson
Ericsson

Sven Jacobsson
Ericsson, Chalmers,
& Cornell

Olav Tirkkonen
Aalto & Cornell

VIP Group
website

More information $\rightarrow$ vip.ece.cornell.edu