Soft-Output Finite Alphabet Equalization for mmWave Massive MIMO

Oscar Castañeda

Joint work with Sven Jacobsson, Giuseppe Durisi,
Tom Goldstein, and Christoph Studer

CORNELL TECH

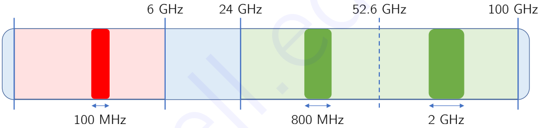
CHALMERS UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF MARYLAND

Future wireless systems will use various technologies

Millimeter-wave (mmWave) [1]

- More **bandwidth**



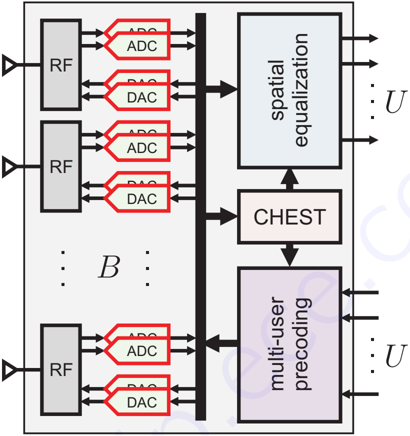
Massive MU-MIMO [2]

- Equip the basestation (BS) with **hundreds or thousands of antennas** B
- **High array gain** compensates for mmWave path-loss
- **Fine-grained beamforming** allows to serve tens of user equipments (UEs) U

[1] A. L. Swindlehurst, E. Ayaoglu, P. Heydari, and F. Capolino, "Millimeter-wave Massive MIMO: The Next Wireless revolution?," IEEE Commun. Mag., Sep. 2014

[2] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," IEEE T-WCOM, Nov. 2010

All-digital basestations are desirable in practice



All digital: Each antenna has a pair of ADCs/DACs

- Maximum flexibility for beam- and null-forming
- Simplified synchronization, channel estimation, beam finding and tracking, equalization, and precoding
- Minimal radio-frequency (RF) circuit design effort
- Inexpensive testing and technology migration

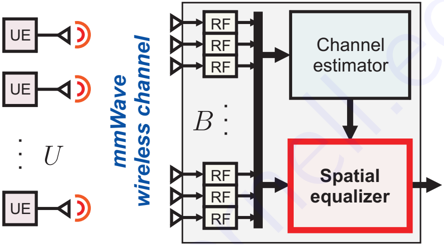
Past research focused on low-resolution ADCs/DACs

All-digital mmWave massive MU-MIMO BSs may entail excessive **interconnect**, **system costs**, and **power consumption**

- Lower resolution → **lower power consumption**
  - Power of ADCs/DACs scales exponentially with number of bits
- Lower resolution → **lower hardware complexity**
  - 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 → **lower raw data rates** from/to converters

Our focus: Not ADCs/DACs, but **low-resolution baseband processing**

We study the massive MU-MIMO OFDM uplink



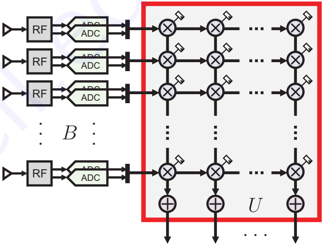
Per-subcarrier uplink channel model: $\mathbf{y} = \mathbf{Hs} + \mathbf{n}$

- $\mathbf{y} \in \mathbb{C}^B$ BS-receive signals; $\mathbf{y} = [y_1, \ldots, y_B]^T$
- $\mathbf{H} \in \mathbb{C}^{B \times U}$ (known) MIMO channel matrix; tall: $B \gg U$
- $\mathbf{s} \in \mathcal{S}^U$ UE-transmit vector; expected power E_s
- $\mathbf{n} \in \mathbb{C}^B$ noise; i.i.d. zero-mean Gaussian with variance N_0

Goal: **Recover** transmit data $\mathbf{s}$ given knowledge of $\mathbf{H}$ and $\mathbf{y}$

Baseband processing: An implementation bottleneck!

- Spatial equalization:
  - Collect power from individual UEs
  - Suppress inter-UE interference
  - Map B data streams to U layers
- Spatial equalization must be **performed at ADC sampling rate**



- To minimize complexity and power, we focus on **linear** spatial equalization: $\bar{\mathbf{s}} = \mathbf{W}^H\mathbf{y}$

Linear spatial equalization of a single tap for a 256 BS array @ 1 GHz bandwidth **requires at least 27 mm^2 and 21 W** in 28nm CMOS [1]

- Area and power results when using 10b for $\mathbf{W}^H$ and 7b for $\mathbf{y}$

[1] O. Castañeda, S. Jacobsson, G. Durisi, T. Goldstein, and C. Studer, "High-Bandwidth Spatial Equalization for mmWave Massive MU-MIMO with Processing-In-Memory," to be presented at IEEE ISCAS 2020

What can we do? Finite-alphabet equalization

Only reducing precision of ADCs is **not enough**!

- Multiplication of m bit and n bit number in hardware [1]

$$Area = O(nm) \quad \text{and} \quad Delay = O(\log(\max\{m, n\}))$$

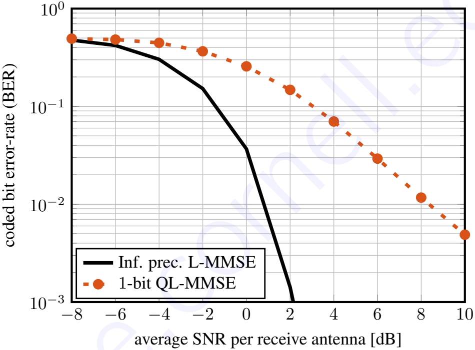
- Power consumption *roughly* proportional to area

Idea: Reduce precision of spatial equalization matrix $\mathbf{W}^H$

- Naïve approach: Compute L-MMSE $\mathbf{W}^H = (\mathbf{H}^H\mathbf{H} + \rho\mathbf{I}_U)^{-1}\mathbf{H}^H$ and quantize rows of $\mathbf{W}^H$ to $\{-\beta_u, +\beta_u\}$ with optimal β_u
- Multiplication with matrix that has 1-bit entries **only requires additions and subtractions**!

[1] R. Zimmermann, "Computer Arithmetic: Principles, Architectures, and VLSI Design," Technical Report, Integrated Systems Laboratory, ETH Zurich, 1999

Quantized L-MMSE bit error-rate performance



- B=256 BS antennas, U=16 UEs, 16-QAM, mmMAGIC NLoS, 60 GHz, R=3/4, OFDM, ±3 dB per-user power control

1-bit quantized L-MMSE results in **significant performance loss**

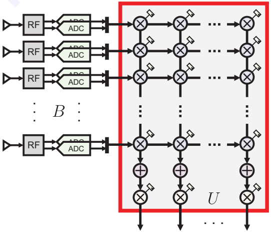
Let's do it right!

Main goal: Design matrices that **minimize MSE** and enable **hardware-efficient** VLSI implementations

Finite-alphabet matrix

$$\mathbf{V}^H = \text{diag}(\boldsymbol{\beta}^*)\mathbf{X}^H$$

- Low-resolution matrix $\mathbf{X}^H \in \mathcal{X}^{U\times B}$, e.g., $\mathcal{X} = \{\pm 1 \pm j\}$
- Post-equalization scaling $\boldsymbol{\beta}^* \in \mathbb{C}^U$



- **Hardware-friendly** per-UE *biased* spatial equalization

$$\bar{s}_u[k] = \beta_u^* \mathbf{x}_u^H \mathbf{y}, \quad u = 1, \ldots, U$$

  - Inner products $\mathbf{x}_u^H \mathbf{y}[k]$ can be implemented with simple hardware
  - Only scaling with β_u is carried out at higher precision

FAME: Finite-Alphabet MMSE Equalizer

- Goal: Find vector $\boldsymbol{\beta}$ and matrix $\mathbf{X}^H$ that minimize MSE

$$\{\boldsymbol{\beta}, \mathbf{X}^H\} = \underset{\tilde{\boldsymbol{\beta}} \in \mathbb{C}^U, \tilde{\mathbf{X}}^H \in \mathcal{X}^{U \times B}}{\arg\min} \mathbb{E}_{\mathbf{s},\mathbf{n}}\left[\|\text{diag}(\tilde{\boldsymbol{\beta}}^*)\tilde{\mathbf{X}}^H\mathbf{y} - \mathbf{s}\|_2^2\right]$$

- Problem can be formulated per UE $u = 1, 2, \ldots, U$

$$\{\beta_u, \mathbf{x}_u\} = \underset{\beta \in \mathbb{C}, \tilde{\mathbf{x}} \in \mathcal{X}^B}{\arg\min} \|\mathbf{e}_u - \mathbf{H}^H\tilde{\beta}\tilde{\mathbf{x}}\|^2 + \rho\|\tilde{\beta}\tilde{\mathbf{x}}\|^2, \quad \rho = N_0/E_s$$

- FAME can be solved in two steps:

$$\mathbf{x}_u = \underset{\tilde{\mathbf{x}} \in \mathcal{X}^B}{\arg\min} \frac{\|\mathbf{H}^H\tilde{\mathbf{x}}\|_2^2 + \rho\|\tilde{\mathbf{x}}\|_2^2}{|\mathbf{h}_u^H\tilde{\mathbf{x}}|^2} \qquad \beta_u(\mathbf{x}_u) = \frac{\mathbf{x}_u^H\mathbf{h}_u}{\|\mathbf{H}^H\mathbf{x}_u\|_2^2 + \rho\|\mathbf{x}_u\|_2^2}$$

Optimization problem is NP-hard: Exhaustive search for 1-bit with $B = 256$ requires 10^{154} evaluations of objective function

Solving FAME in practice [1]

We need fast, even if approximate, algorithms to solve:

$$\mathbf{x}_u = \underset{\tilde{\mathbf{x}} \in \mathcal{X}^B}{\arg\min} \frac{\|\mathbf{H}^H \tilde{\mathbf{x}}\|_2^2 + \rho \|\tilde{\mathbf{x}}\|_2^2}{|\mathbf{h}_u^H \tilde{\mathbf{x}}|^2}$$

Finite-alphabet L-MMSE (FL-MMSE)

- Quantize each L-MMSE matrix row using uniform-width bins

FAME via forward-backward splitting (FBS)

- Iterative procedure to approximately solve the FAME problem

FL-MMSE and FAME-FBS have both the
same complexity scaling $O(BU^2)$ as L-MMSE

[1] O. Castañeda, S. Jacobsson, G. Durisi, T. Goldstein, and C. Studer, "Finite-Alphabet MMSE Equalization for All-Digital mmWave Massive MU-MIMO," to appear in IEEE J-SAC

Unbiased finite-alphabet equalization

- Per-UE **biased** equalization with the L-MMSE $\mathbf{W}^H$

$$\bar{s}_u = \mathbf{w}_u^H \mathbf{y} = \mathbf{w}_u^H \mathbf{h}_u s_u + \mathbf{w}_u^H \tilde{\mathbf{n}}_u$$

where $\tilde{\mathbf{n}}_u = \sum_{i=1, i\neq u}^{U} \mathbf{h}_i s_i + \mathbf{n}$ is noise-plus-interference (NPI)

- In general for the L-MMSE equalizer, $\mathbf{w}_u^H \mathbf{h}_u \neq 1$
- An **unbiased** estimate can be computed as

$$\hat{s}_u = \frac{\bar{s}_u}{\mathbf{w}_u^H \mathbf{h}_u} = s_u + \frac{\mathbf{w}_u^H \tilde{\mathbf{n}}_u}{\mathbf{w}_u^H \mathbf{h}_u}$$

Unbiased finite-alphabet equalization remains **hardware-friendly**:

$$\hat{s}_u = \frac{\beta_u^* \mathbf{x}_u^H \mathbf{y}}{\beta_u^* \mathbf{x}_u^H \mathbf{h}_u} = \frac{\mathbf{x}_u^H \mathbf{y}}{\mathbf{x}_u^H \mathbf{h}_u}$$

Generating soft-output for finite-alphabet equalization

- We find the NPI variance:

$$\nu_u^2 = \mathbb{E}_{\mathrm{s,n}}\left[|\hat{s}_u - s_u|^2\right] = E_s\left((\beta_u(\mathbf{x}_u)\mathbf{h}_u^H\mathbf{x}_u)^{-1} - 1\right)$$

- We compute LLR values by assuming that the residual error $\hat{s}_u - s_u$ is circularly-symmetric Gaussian with variance ν_u^2 [1]:

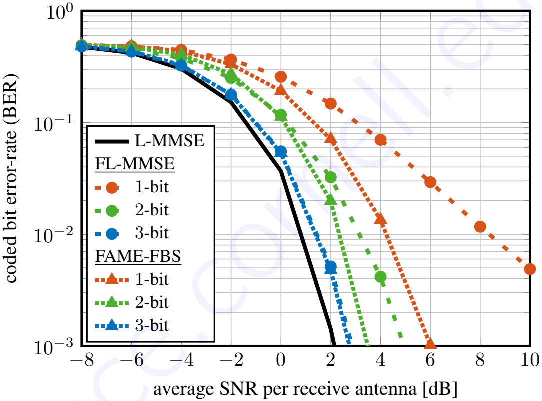
$$\Lambda_{u,q} = \log\left(\sum_{s\in\mathcal{S}_q^{(1)}} \exp\left(-\frac{|\hat{s}_u - s|^2}{\nu_u^2}\right)\right) - \log\left(\sum_{s\in\mathcal{S}_q^{(0)}} \exp\left(-\frac{|\hat{s}_u - s|^2}{\nu_u^2}\right)\right),$$

where $\mathcal{S}_q^{(1)}$ and $\mathcal{S}_q^{(0)}$ are the subsets of the constellation $\mathcal{S}$ in which the qth bit is 1 and 0, respectively.

Computing soft-outputs for finite-alphabet equalizers entails the **same complexity** as for traditional L-MMSE

[1] C. Studer, S. Fateh, and D. Seethaler, "ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation," IEEE J-SSC, Jul. 2011

FAME offers competitive error-rate performance



- B=256 BS antennas, U=16 UEs, 16-QAM, mmMAGIC NLoS, 60 GHz, R=3/4, OFDM, ±3 dB per-user power control
- Finite-alphabet equalization also supports multi-bit quantization

Summary and conclusions

Operation at extreme mmWave bandwidths results in
high silicon area and power consumption

Novel paradigm that offers **significant area and power savings**
compared to conventional baseband processing

- ✓ Finite-alphabet matrices enable the use of low-precision hardware while minimizing performance loss
- ✓ Approach also provides **unbiased estimates with soft-outputs**
- ✓ Hardware results for matrix-vector product demonstrate **savings in area and power of up to 5.8× and 3.9×, respectively** [1]

More information → vip.ece.cornell.edu

[1] O. Castañeda, S. Jacobsson, G. Durisi, T. Goldstein, and C. Studer, "Finite-Alphabet MMSE Equalization for All-Digital mmWave Massive MU-MIMO," to appear in IEEE J-SAC