ETH zürich

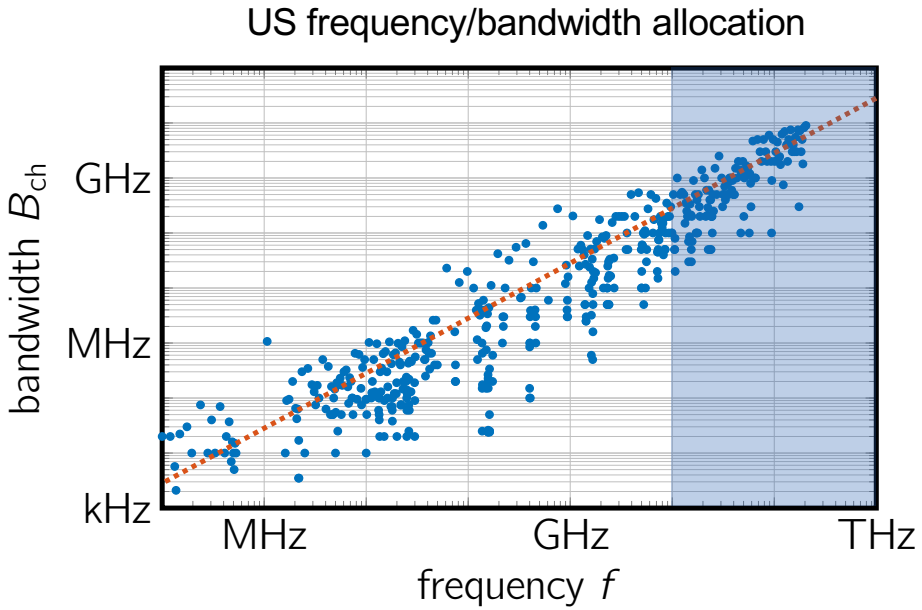
Integrated Information Processing (IIP) Group
Integrated Systems Laboratory

Algorithms, Architectures, and Hardware Implementations for Future MIMO Systems

Christoph Studer

Joint work with Oscar Castañeda, Seyed Hadi Mirfarshbafan, and Gian Marti

How to satisfy the never-ending need for higher data rates?



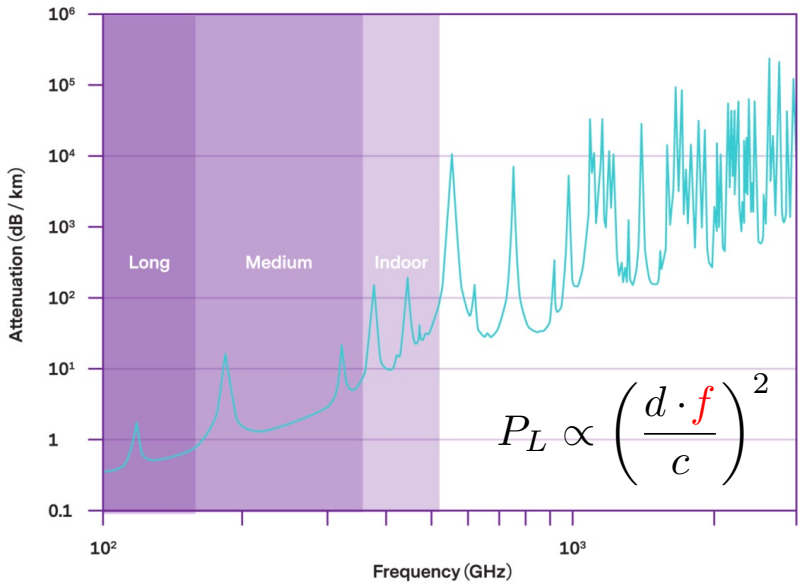
$$B_{\text{ch}} \approx 2\% \times f$$

Example: about 1 GHz contiguous bandwidth available at 50 GHz

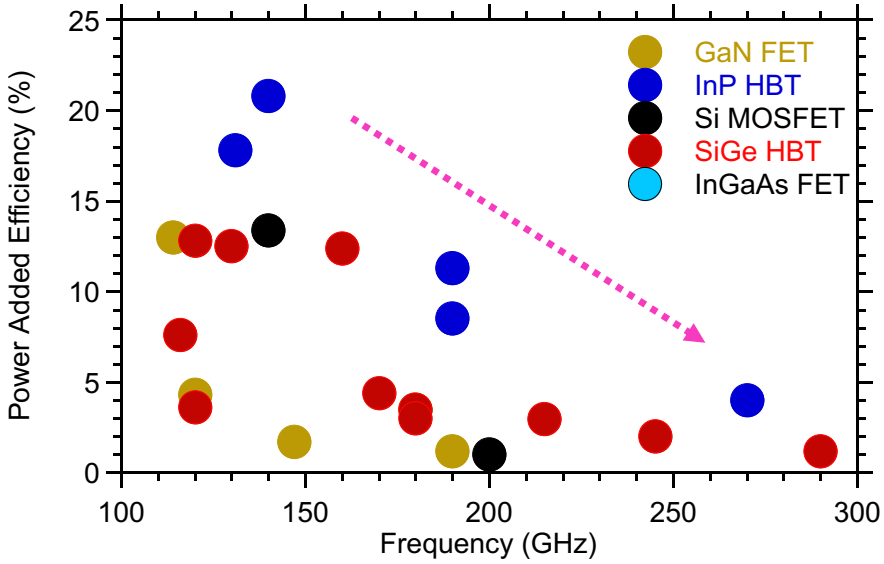
- Data rate = spectral efficiency × bandwidth B_{ch}
- More bandwidth available at higher carrier frequencies
- Possible solution: communication above 10 GHz (FR3, mmWave, or even THz)

Communication above 10 GHz is not easy!

Attenuation vs. carrier frequency



Transistor efficiency



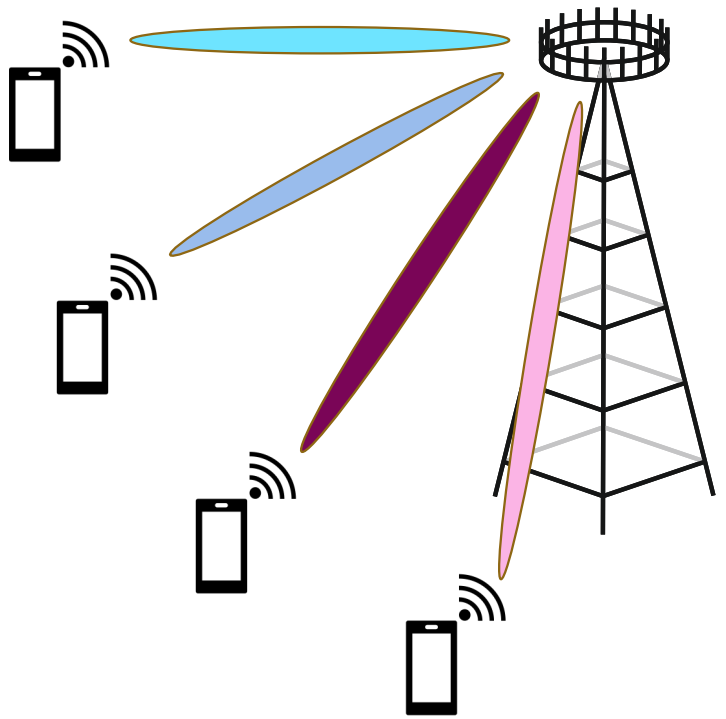
- Path loss P_L increases at higher frequencies
- Transistor efficiency decreases at higher frequencies

Source: SRC ComSenTer

Massive multiuser MIMO to the rescue

Equip basestation (BS) with hundreds of antennas [1]

- Many antennas at BS enable high array gain, fine-grained beamforming, and spatial multiplexing
- Array-gain compensates for high path loss at higher carrier frequencies
- Fine-grained beamforming and spatial multiplexing enable multiuser (MU) MIMO communication



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

The obvious challenges of massive MU-MIMO beyond 10 GHz

GHz of bandwidth:

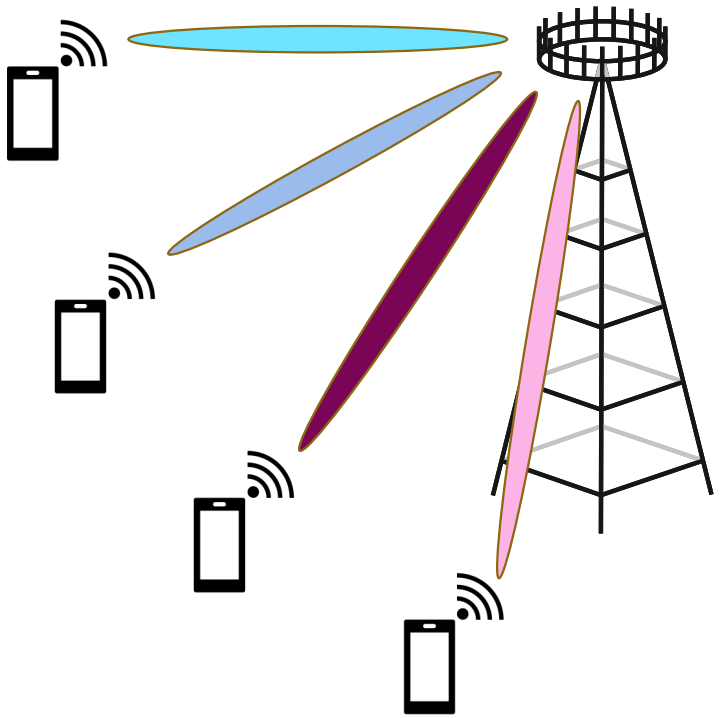
- Linearity requires power-hungry RF circuitry
- Signal processing of high data-rate signals

Hundreds of antennas:

- Expensive and power-hungry RF circuitry
- Signal processing of high-dimensional signals

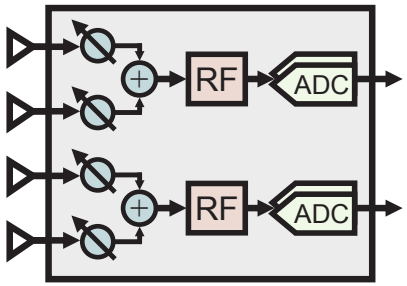
Complexity increases with bandwidth and number of antennas:

- Extreme requirements on baseband processing engines
- Extreme interconnect and chip-I/O data rates



We will focus on all-digital massive multiuser MIMO

Hybrid beamformer



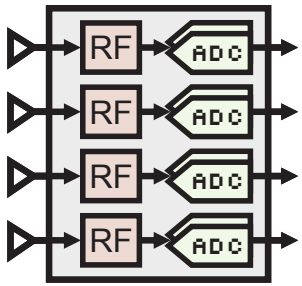
Pros:

- Energy efficient?

Cons:

- "Inflexible": limited spatial multiplexing
- Difficult to design, test, and calibrate

All-digital beamformer



Pros:

- Maximum flexibility
- Simplifies synchronization, impairment compensation, channel estimation, equalization, precoding, etc.
- Cheaper testing and technology migration

Cons:

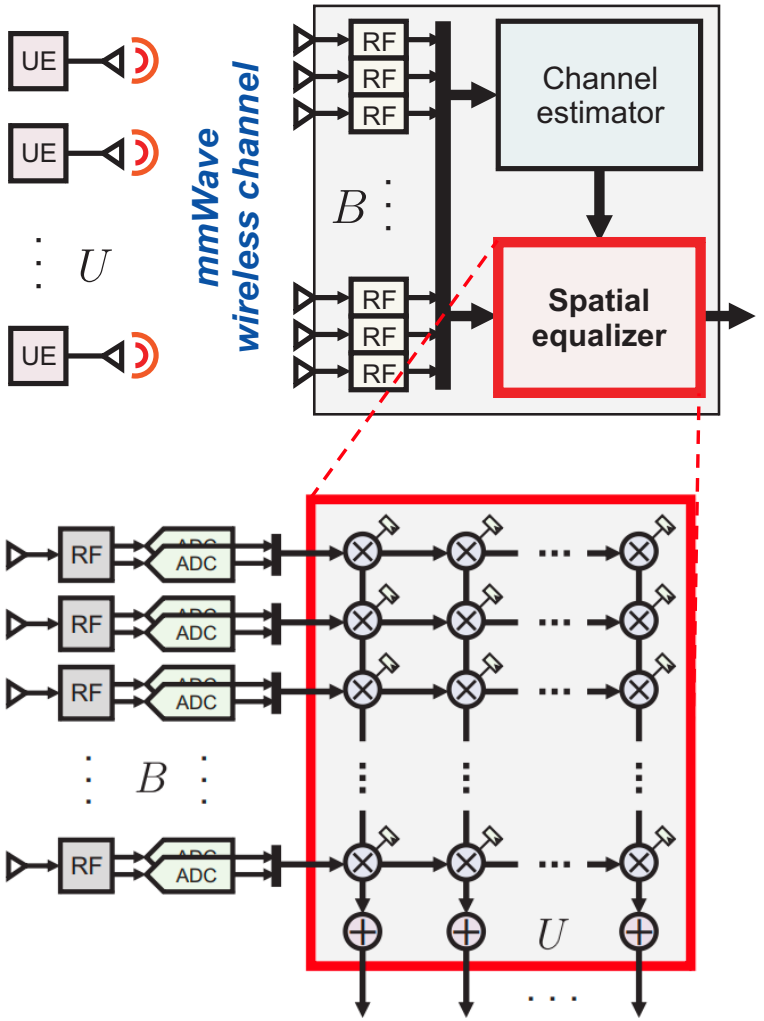
- Higher power consumption?
- High data rates from and to ADCs/DACs

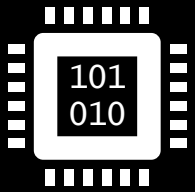
similar RF front-end power with low-resolution data converters [1]

[1] Roth *et al.*, "A comparison of hybrid and digital beamforming with low-resolution ADCs", IEEE JSTSP, Jun. 2018

Even “simple” tasks become challenging

- Uplink channel model: $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{n}$
- Spatial equalization: $\hat{\boldsymbol{s}} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{y}$
- Consider a massive MU-MIMO system with 16 UEs and 256 BS antennas supporting 1 GHz bandwidth
  - Spatial equalization requires 2B matrix-vector products per second (MVP/s) → 32 trillion multiplications per second…
- Consider 7-bit ADCs and a spatial equalization matrix $\boldsymbol{W}^{\mathrm{H}}$ with 10-bit entries (10-bit real, 10-bit imaginary)
  - A conventional application-specific integrated circuit (ASIC) requires more than 27mm^2 and 28W in 28nm CMOS [1]



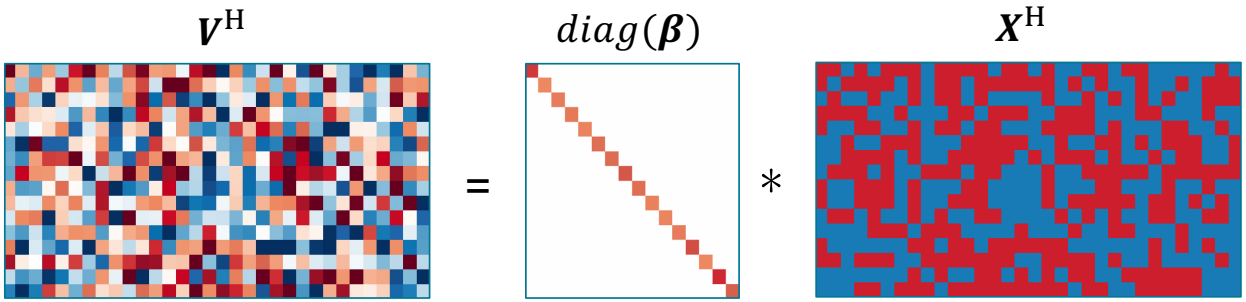


Part I: Low-resolution baseband processing

Only reducing ADC resolution is not enough!

- Multiplication of m-bit and n-bit number in hardware [1]:
  - Area = O(n*m)
  - Delay = O(log(max{m,n}))
  - Power is roughly proportional to area

- Idea: reduce precision of spatial equalization matrices

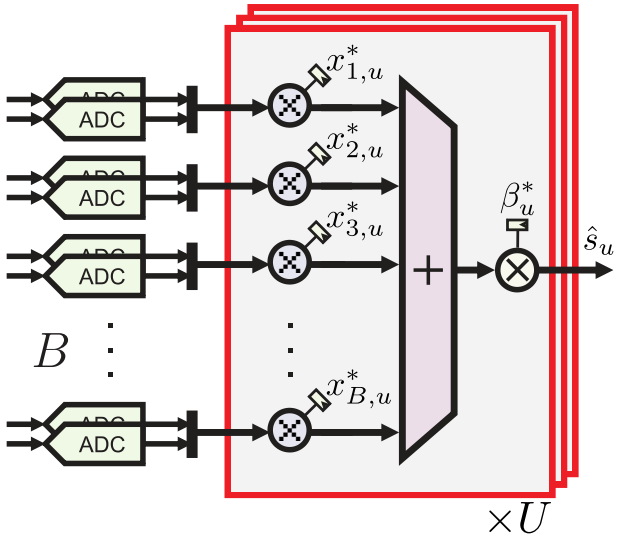


[1] Zimmermann, “Computer Arithmetic: Principles, Architectures, and VLSI Design,” Tech. Rep., Integrated Systems Laboratory, ETH Zurich, 1999

Finite-alphabet MMSE equalization (FAME)

- Define a finite-alphabet equalization matrix: $\boldsymbol{V}^{\mathrm{H}} = \mathrm{diag}(\boldsymbol{\beta}^*)\, \boldsymbol{X}^{\mathrm{H}}$ [1]
  - Low-resolution matrix $\boldsymbol{X}^{\mathrm{H}} \in \mathfrak{X}^{U \times B}$ with, e.g., $\mathfrak{X} = \{\pm 1 \pm \mathrm{i}\}$
  - Post-equalization scaling vector $\boldsymbol{\beta} \in \mathbb{C}^U$

- Per-UE spatial equalization can be implemented as $\hat{s}_u = \beta_u^* \, (\boldsymbol{x}_u^{\mathrm{H}} \, \boldsymbol{y})$
  - Inner products $\boldsymbol{x}_u^{\mathrm{H}} \, \boldsymbol{y}$ implemented with low-complexity hardware
  - Only scaling with β_u is carried out at higher precision

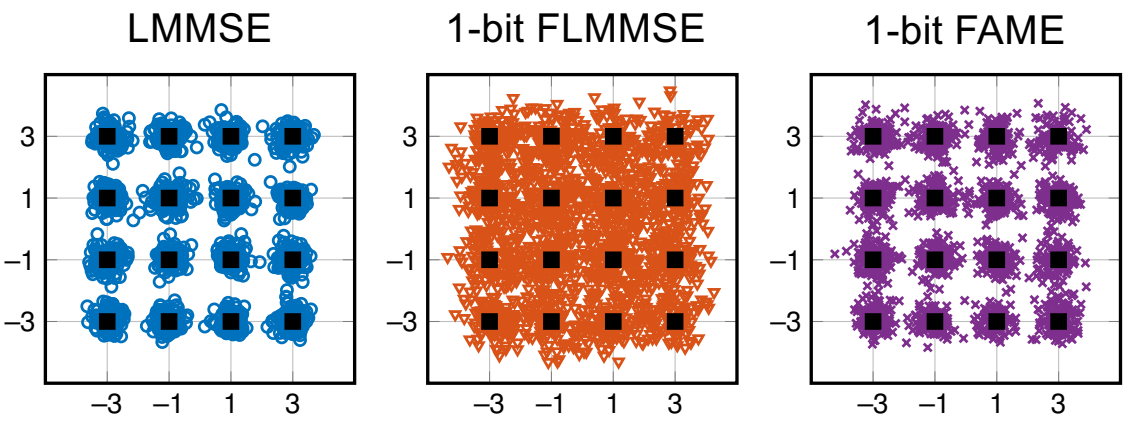


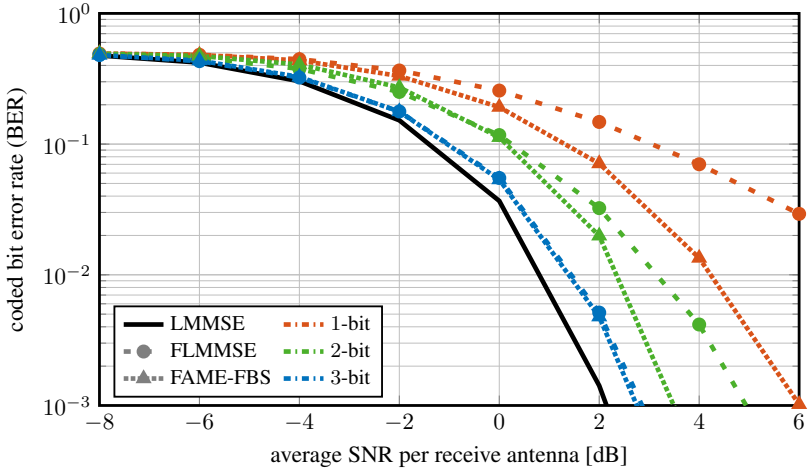
- Goal: find vector $\boldsymbol{\beta}$ and matrix $\boldsymbol{X}$ that minimize post-equalization MSE

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

[1] Castañeda, Jacobsson, Durisi, Goldstein, CS, "Finite-alphabet MMSE equalization for all-digital massive MU-MIMO mmWave communication," IEEE J-SAC, 2020

FAME is feasible for high-dimensional problems



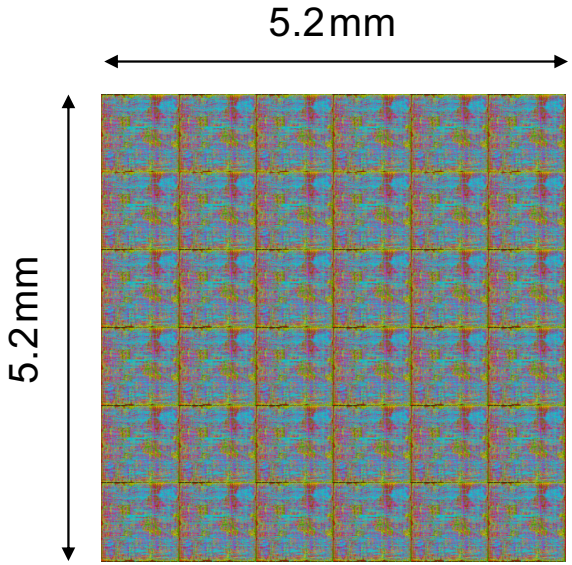


- Simulation setup: 16-QAM, 256 BS antennas, 16 UEs, non-LoS, convolutional code
- 3-bit FAME achieves near-infinite-precision performance!

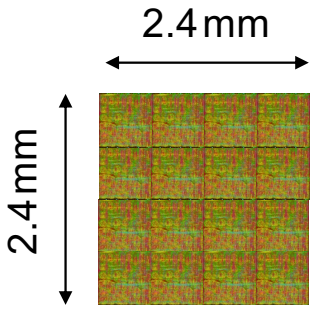
ETH zürich

Reducing resolution results in smaller circuit area and lower power

10 bit MVP ASIC

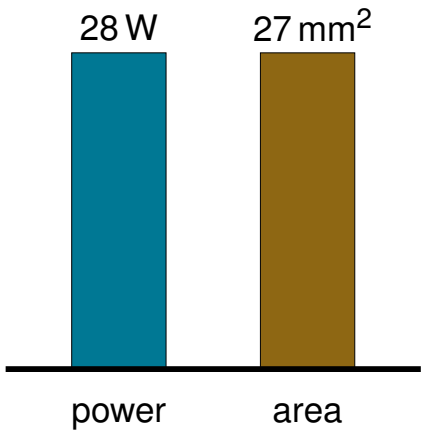


1 bit MVP ASIC

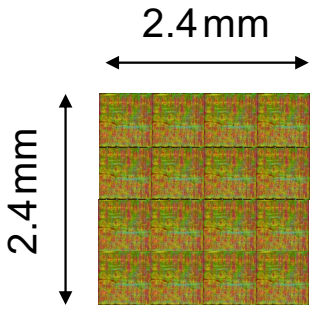


Reducing resolution results in smaller circuit area and lower power

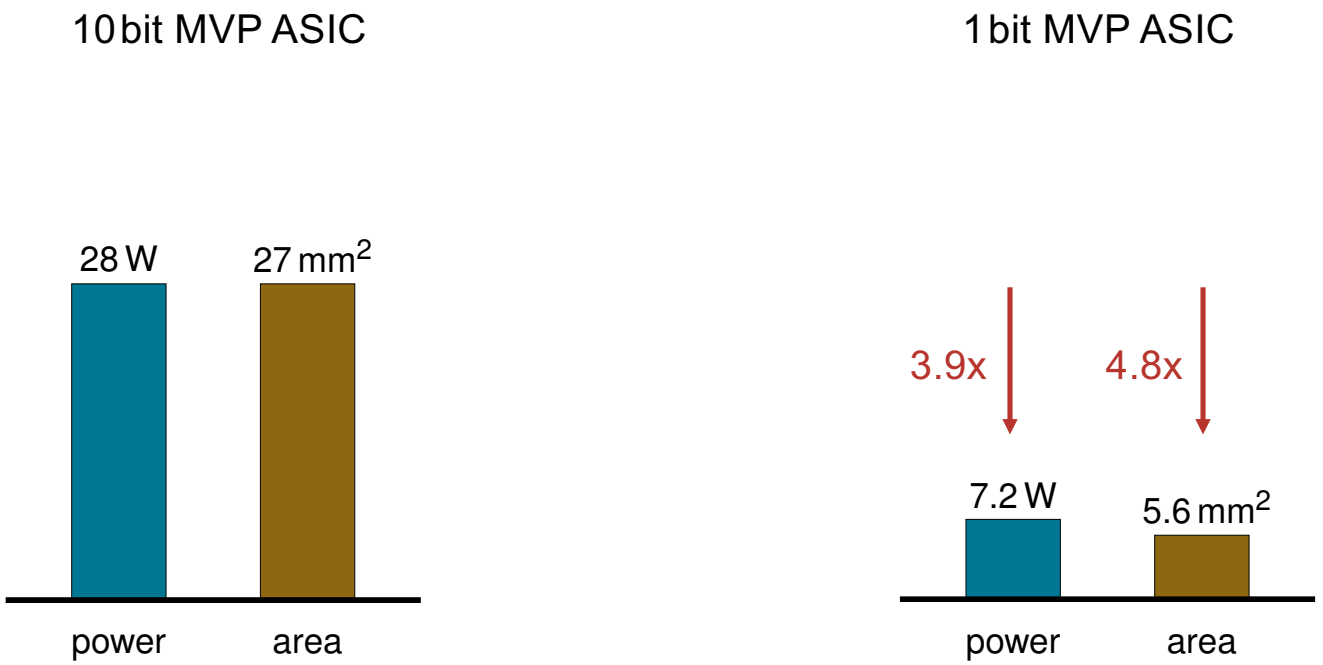
10 bit MVP ASIC



1 bit MVP ASIC

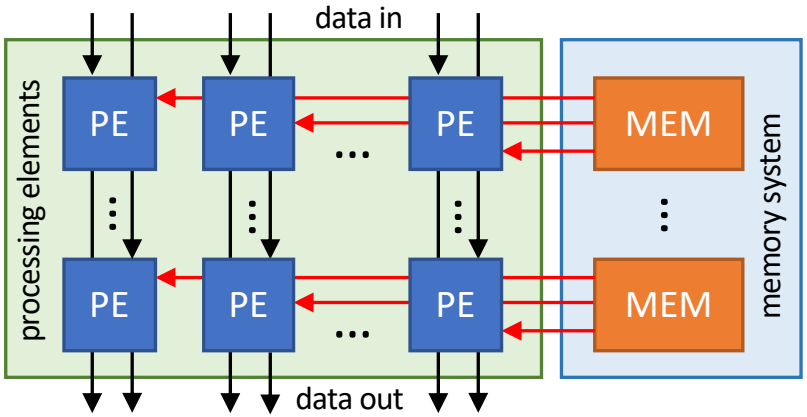


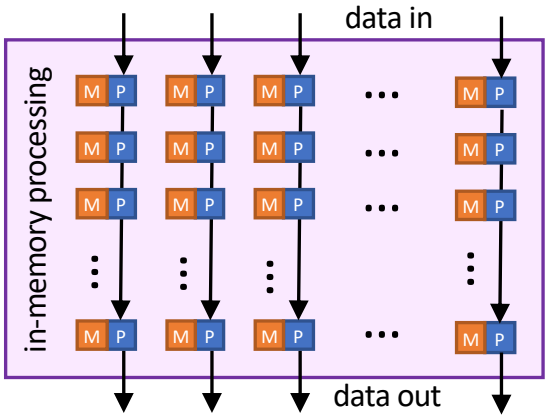
Reducing resolution results in smaller circuit area and lower power



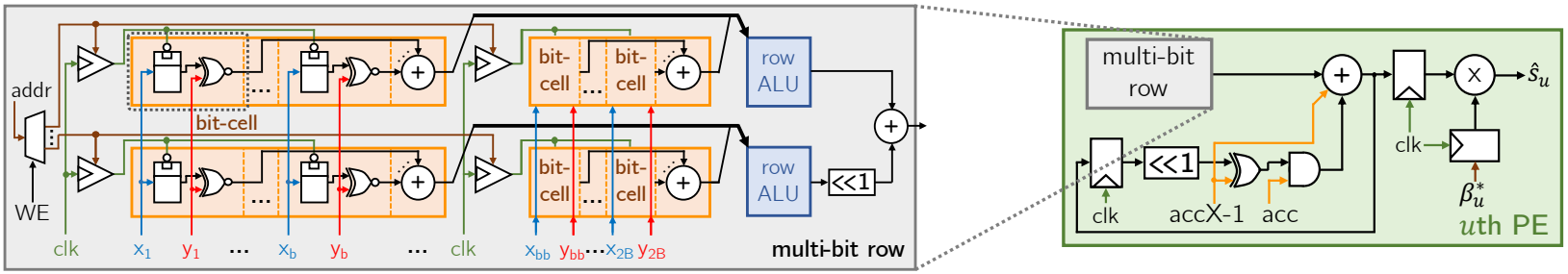
We can do much better!

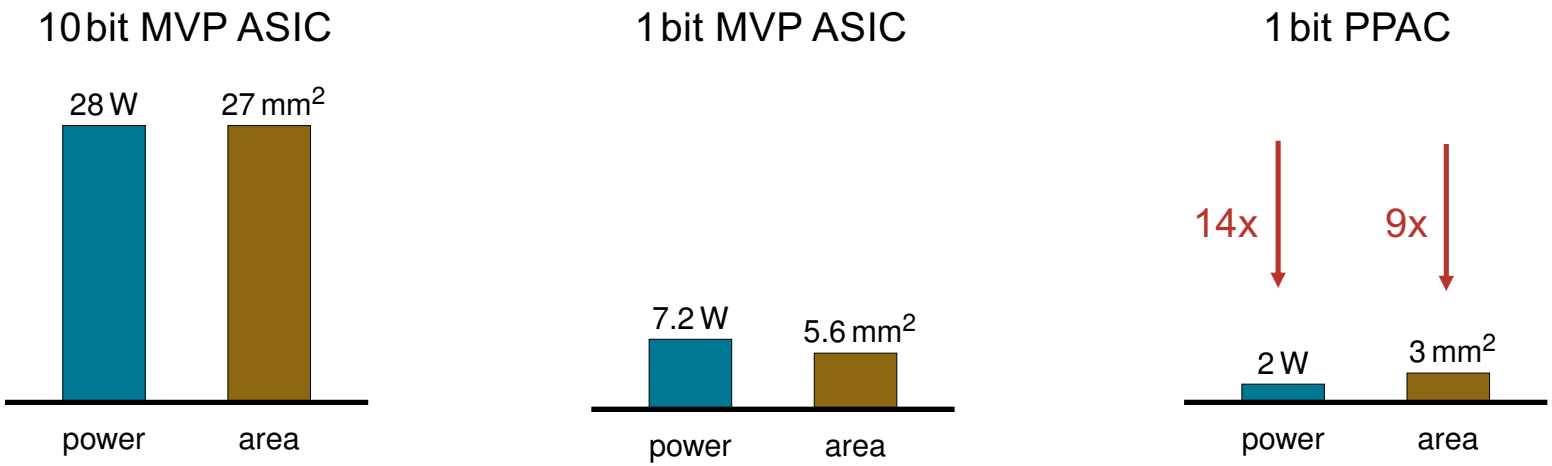
- Conventional MVP engines move data from memories to processing elements (PEs)
- Coefficient movement itself results in significant power dissipation
- Low-resolution computations enable in-memory computing which minimizes coefficient movement





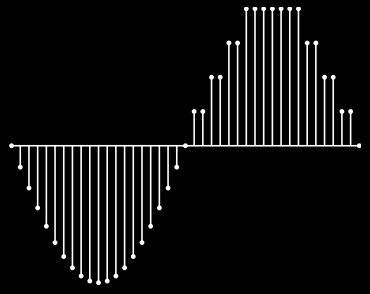
PPAC: parallel processor in associative content-addressable memory [1]





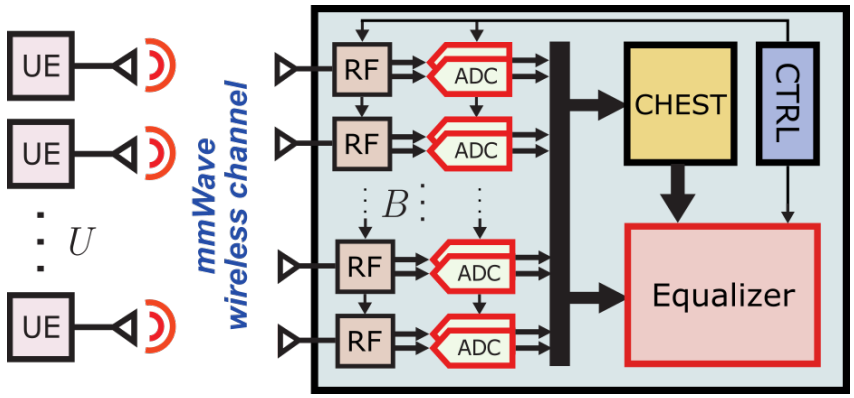
- In-memory computing with bit-serial computations further reduces area and power!

[1] Castañeda, Bobett, Gallyas-Sanhueza, CS, "PPAC: A Versatile In-Memory Accelerator for Matrix-Vector-Product-Like Operations," IEEE ASAP, 2019



Part II: Resolution-adaptive BS architecture

Further improving energy-efficiency of all-digital equalization

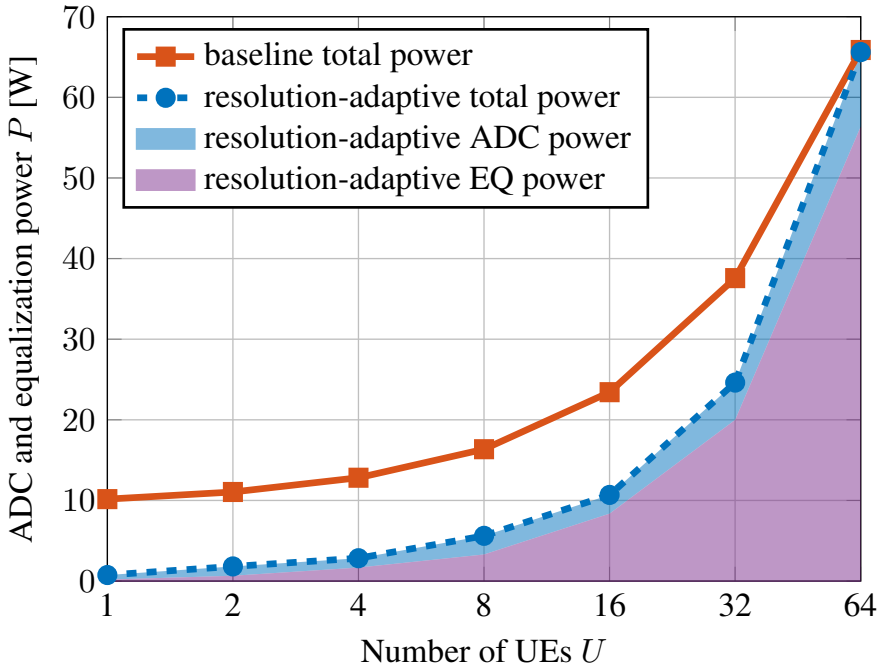


- Idea: adapt
  - ADC resolution q
  - finite-alphabet equalization resolution k, and
  - number of active BS antennas B'

  to number of active UEs U, modulation scheme, and channel propagation conditions

Resolution-adaptive architectures enable significant power savings [1]

- Baseline design:
  - Adapts P_{EQ} to number of active UEs U
  - Resolutions $q = 7$ and $k = 6$ chosen for 0.1dB worst-case SNR loss
  - $B' = B = 256$
- Resolution adaptive design adapts (q, k, B')

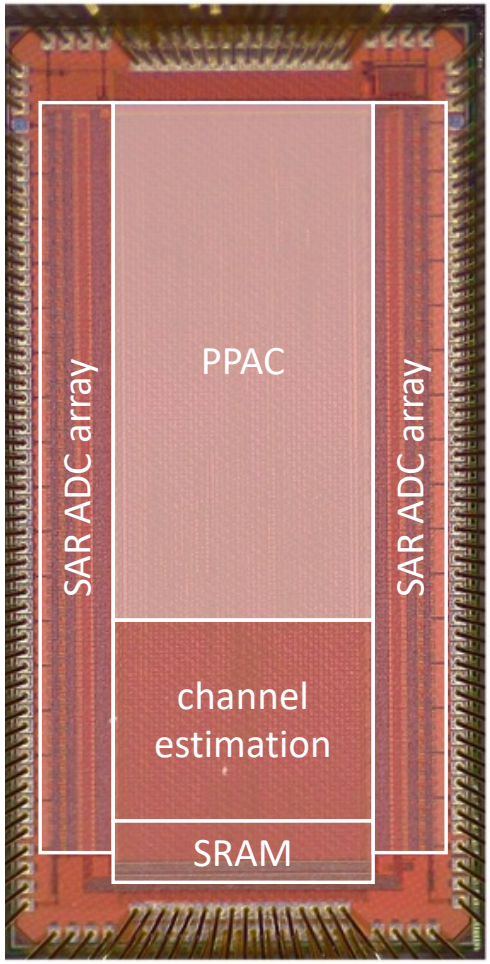


- Equalizer dominates total power P for $U > 2$
- Up to $22\times$ power reduction for $U = 1$

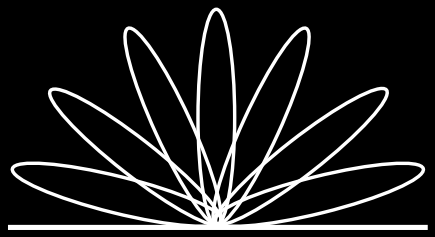
[1] Castañeda *et al.*, Resolution-adaptive all-digital spatial equalization for mmWave massive MU-MIMO," IEEE SPAWC, 2021

Putting the pieces together

- 8 mm^2 mixed-signal ASIC in TSMC 65nm [1]
  - Supports 32 BS antennas and up to 16 UEs
  - 32x2 (I/Q) SAR ADCs
    - 4x time-interleaved
    - 3b or 6b programmable resolution
  - Finite-alphabet spatial equalizer
    - PPAC-based in-memory processing architecture
    - 1b to 4b programmable resolution
- High throughput (up to 20 Gb/s)
- Record energy efficiency (as low as 14 pJ/b)



[1] Castañeda *et al.*, "A res.-ad. 8mm^2 32-ant. all-digital spatial eq. for mmWave massive MU-MIMO", IEEE ESSCIRC 2021



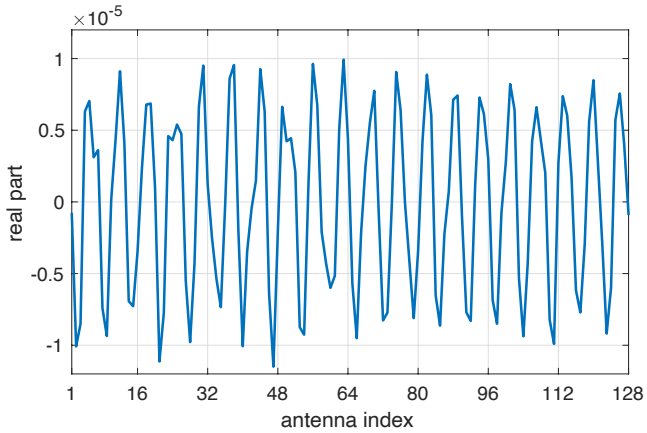
Part III: Beamspace processing

Channels become sparser at higher carrier frequencies

- Wave propagation becomes predominantly directional
- Antenna-domain UE channel vector at a BS with a uniform linear antenna array:

$$\boldsymbol{h} = \sum_{l=0}^{L-1} \alpha_l \boldsymbol{a}(\phi_l) \quad \text{with} \quad \boldsymbol{a}(\phi) = \left[e^{j0\phi}, e^{j1\phi}, \ldots, e^{j(B-1)\phi}\right]^T$$

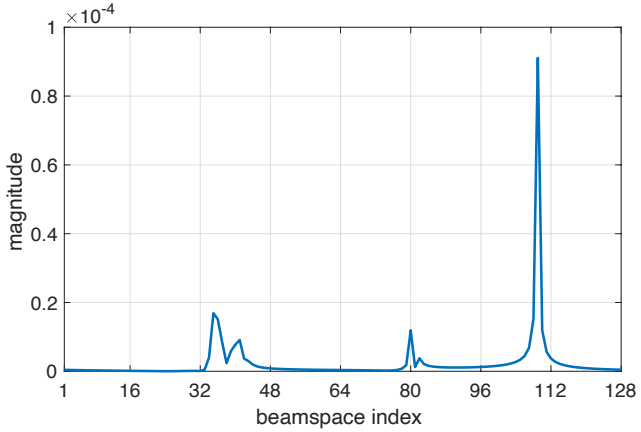
real part of $\boldsymbol{h}$ in antenna domain



spatial discrete Fourier transform

$$\widehat{\boldsymbol{h}} = \boldsymbol{F}\boldsymbol{h}$$

magnitude of $\widehat{\boldsymbol{h}}$ in beamspace domain



Baseband processing in beamspace can exploit channel sparsity

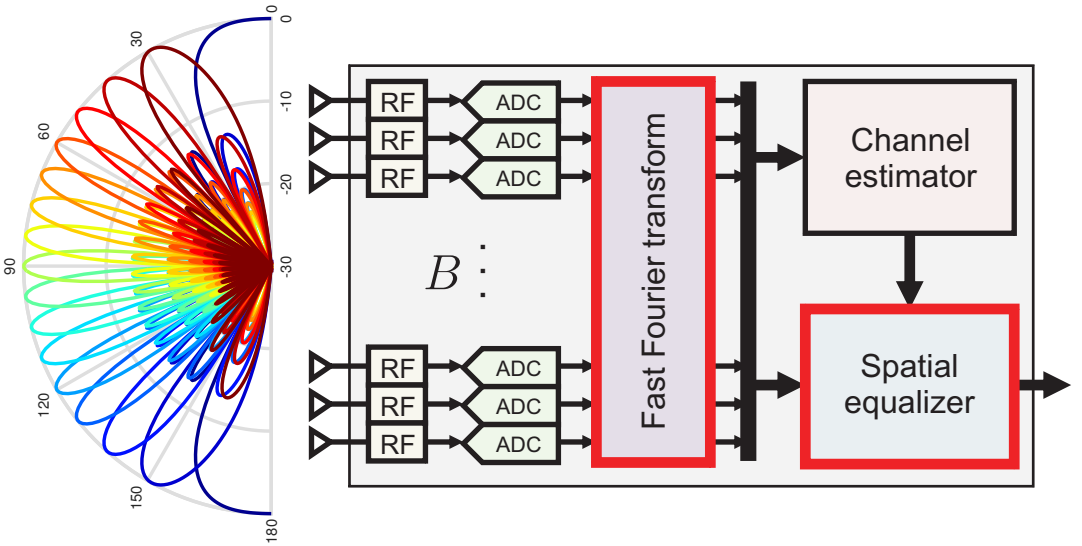
- Antenna-domain uplink system model:

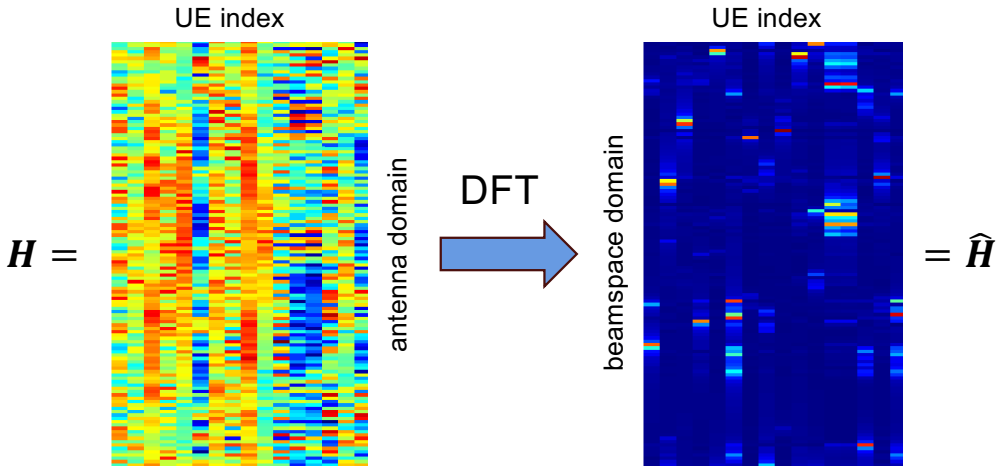
$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{n}$$

- Beamspace-domain uplink system model:

$$\hat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y} = \boldsymbol{F}\boldsymbol{H}\boldsymbol{s} + \boldsymbol{F}\boldsymbol{n},$$

where $\boldsymbol{F}$ is a $B \times B$ DFT matrix



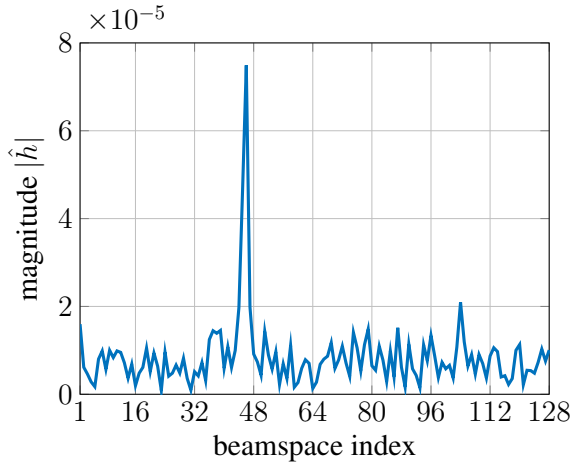


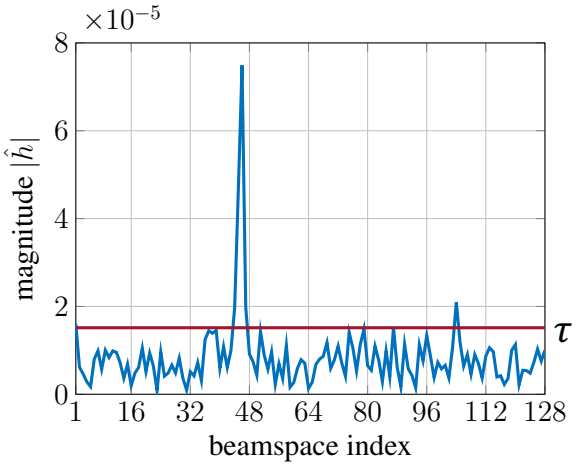
$\boldsymbol{H} =$ $= \hat{\boldsymbol{H}}$

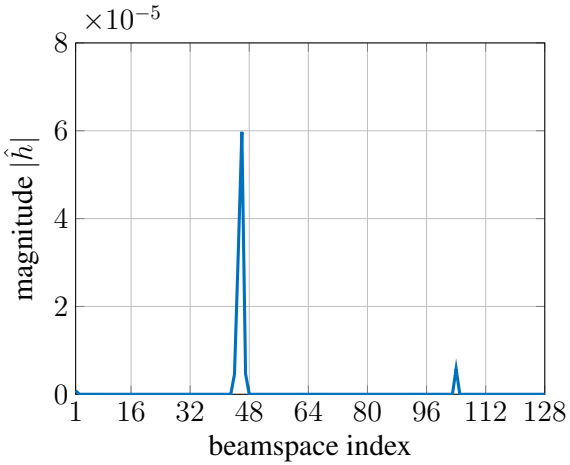
Channel matrices $\hat{\boldsymbol{H}}$ are sparse in beamspace domain: $\hat{\boldsymbol{H}} = \boldsymbol{F}\boldsymbol{H}$

ETH zürich

Sparsity-exploiting channel-vector denoising



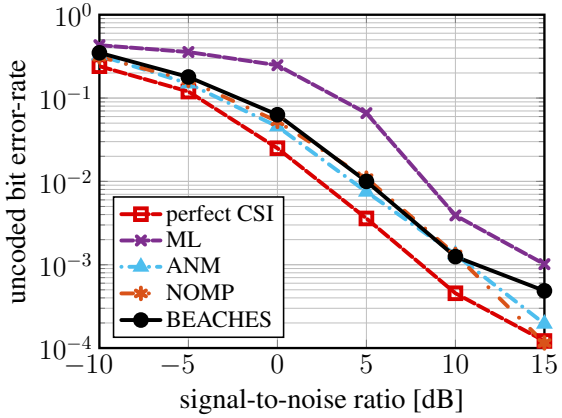




- Channel estimates $\tilde{\boldsymbol{h}} = \hat{\boldsymbol{h}} + \boldsymbol{e}$ are noisy with $e_b \sim CN(0, E_0)$
- We can use soft-shrinkage to denoise the sparse channel estimates
- Soft shrinkage requires a thresholding parameter τ that minimizes:

$$MSE(\tilde{\boldsymbol{h}}, \tau) = \mathbb{E}_{\boldsymbol{n}}\left[\left\|\eta(\tilde{\boldsymbol{h}}, \tau) - \hat{\boldsymbol{h}}\right\|_2^2\right]$$

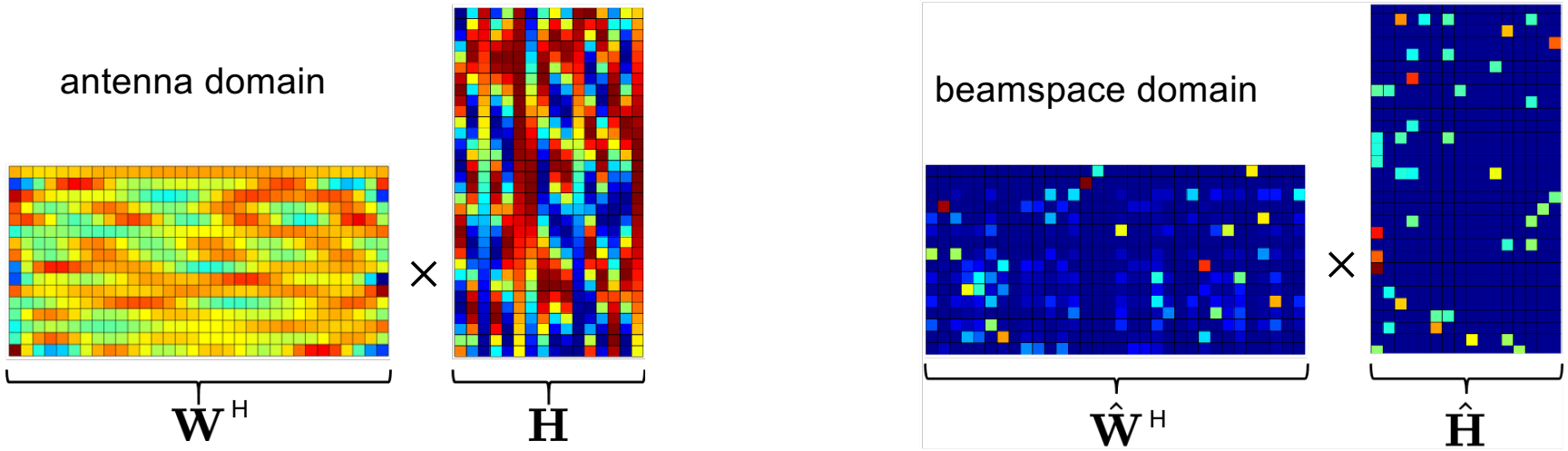
- We propose BEACHES, which minimizes Stein's unbiased risk estimate (SURE) at $O(B\log(B))$ complexity [1]



[1] Mirfarshbafan et al., "Beamspace channel estimation for massive MIMO mmWave systems: Algorithm and VLSI design," IEEE TCAS I, 2020

Sparsity-exploiting spatial equalization

- Not just channel matrices are sparse in beamspace…



- LMMSE equalization matrix is approximately sparse in beamspace as well!
- Sparsity of beamspace equalization matrix $\widehat{\boldsymbol{W}}^{\mathrm{H}}$ can reduce complexity of spatial equalization $\hat{\boldsymbol{s}} = \widehat{\boldsymbol{W}}^{\mathrm{H}}\hat{\boldsymbol{y}}$

SPADE: SParsity-Adaptive Equalization [1]

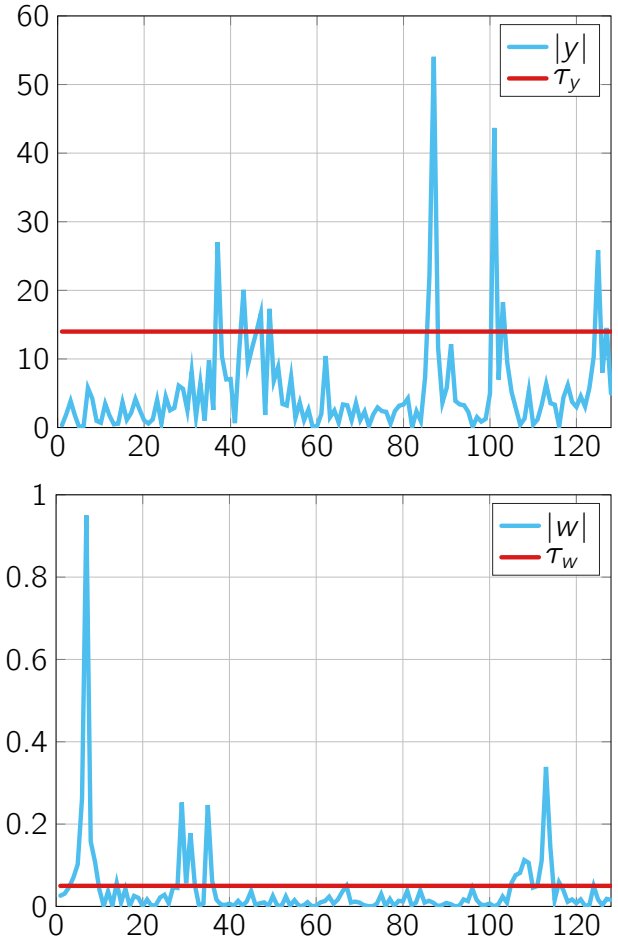
- Consider inner product between two real-valued vectors:

$$\langle \boldsymbol{w}, \boldsymbol{y} \rangle = \sum_{b=1}^{B} w_b y_b$$

- We can skip some multiplications if
  - $\boldsymbol{w}$ and $\boldsymbol{y}$ are both sparse
  - $\langle \boldsymbol{w}, \boldsymbol{y} \rangle \gg 0$, and
  - we do not need exact inner product

- All conditions met for beamspace equalization!

- Idea: skip individual multiplication if $|w_b| < \tau_w$ and $|y_b| < \tau_y$
  → power savings [1]

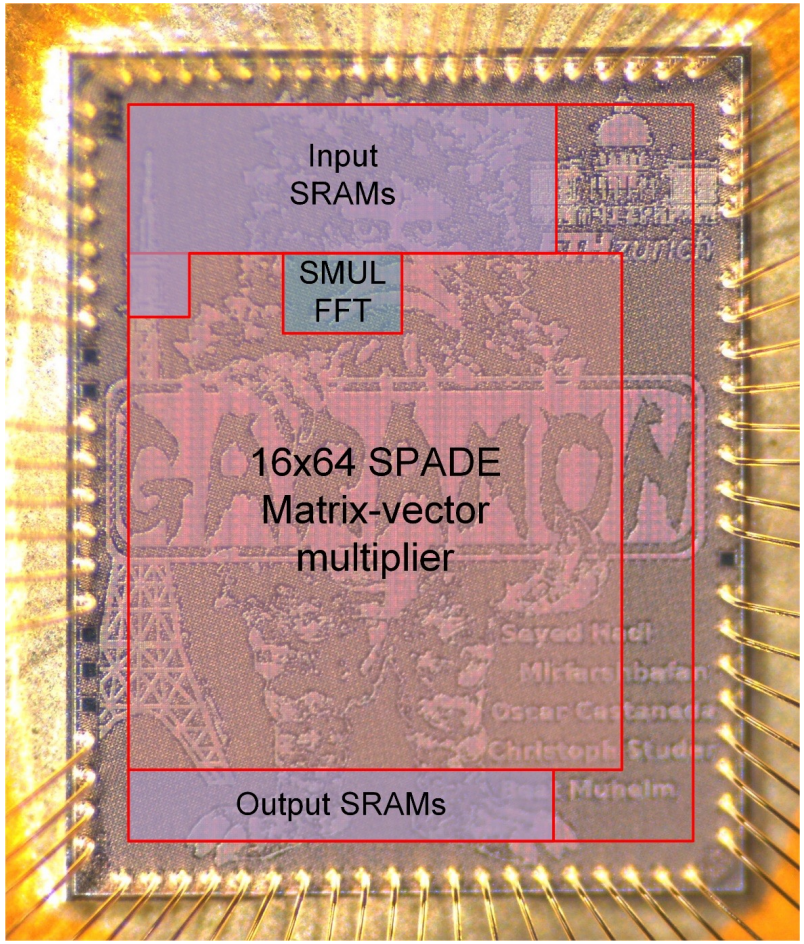


[1] Mirfarshbafan and CS, "A 46 Gbps 12 pJ/b Sparsity-Adaptive Beamspace Equalizer for mmWave Massive MIMO in 22FDX", TCAS II, 2024

Putting the pieces together

- 5 mm^2 ASIC in GlobalFoundries 22nm FDSOI [1]
  - Supports 64 BS antennas and up to 16 UEs
  - Supports antenna-domain and beamspace equalization
  - Contains SMUL-FFT beamspace transform [2]
  - Contains 64x16 SPADE array
- Record throughput (up to 57 Gb/s)
- Power consumption in LoS channels at 500 MHz:
  - 631 mW for antenna domain equalization
  - 387 mW for beamspace equalization with SPADE

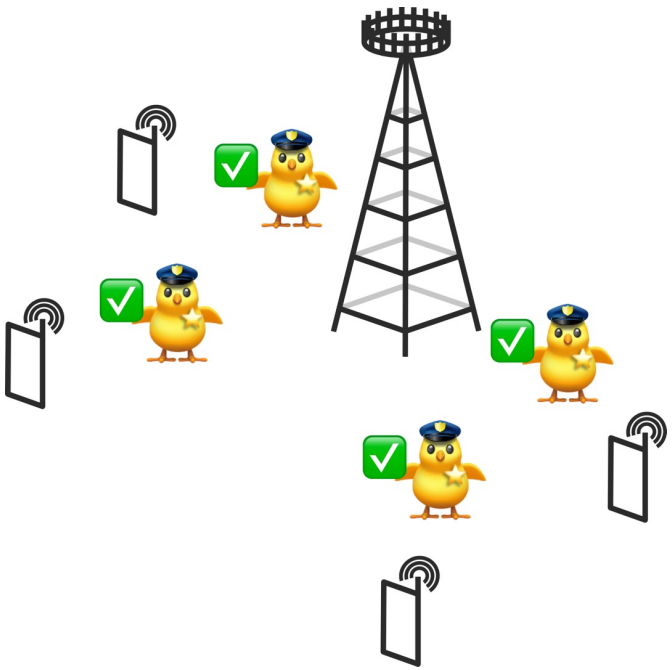
  → 38% power savings



[1] Mirfarshbafan and CS, "A 46 Gbps 12 pJ/b Sparsity-Adaptive Beamspace Equalizer for mmWave Massive MIMO in 22FDX", TCAS II, 2024
[2] Mirfarshbafan, Taner, CS, "SMUL-FFT: A streaming multiplierless fast Fourier transform," IEEE TCAS II, 2021

Part IV: Jamming-resilient communication

One more challenge: reliability should not be an afterthought



- Existing wireless systems for civil use are highly vulnerable
- Communication infrastructure must be protected against jammers

- Challenges of all-digital architectures:
  - Jammers either saturate ADCs or drown signal in quantization noise [1]
  - Smart jammers can evade estimation of their spatial signature [2]

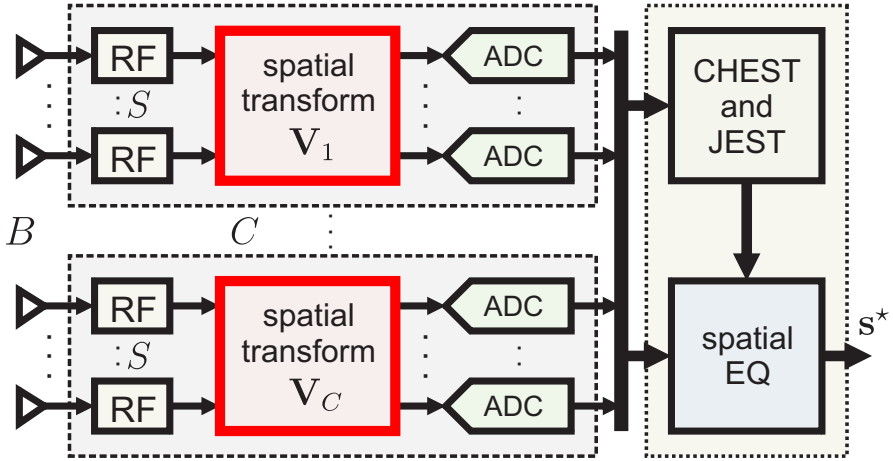
- Multi-antenna wireless systems can be protected
  - Before ADCs: spatial transforms
  - After ADCs: spatial nulling

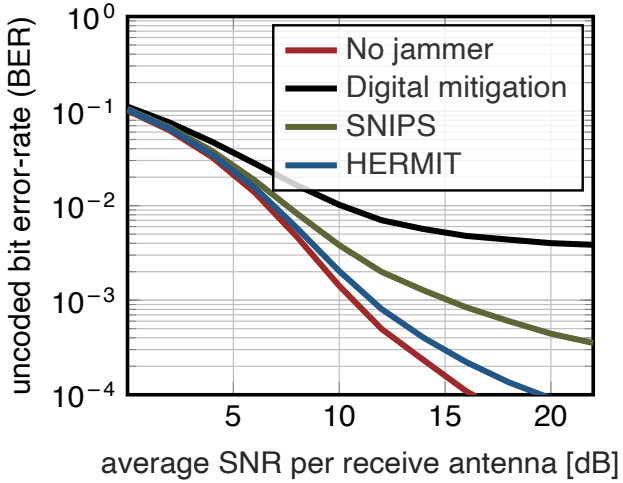
[1] Marti, Stutz-Tirri, CS, "Fundamental Limits for Jammer-Resilient Communication in Finite-Resolution MIMO," Asilomar 2024
[2] Marti, Kölle, CS, "Mitigating Spart Jammers in Multi-User MIMO," T-SP 2023

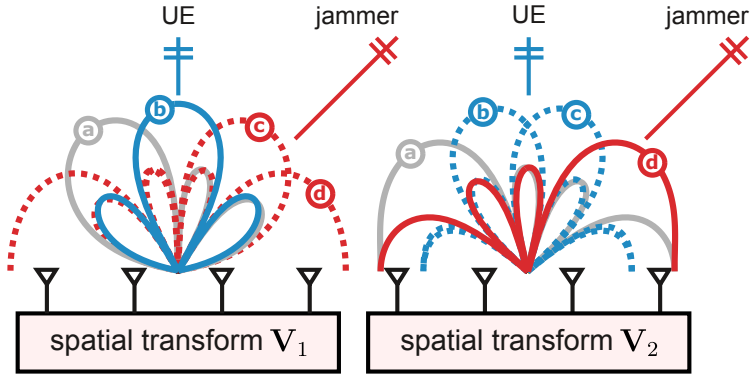
Dealing with jammers before the ADCs

- To avoid ADC saturation or drowning useful signal in quantization noise, jammers should be mitigated already before the ADCs

- SNIPS: nonadaptive spatial transform [1]
- HERMIT: adaptive, structured spatial transform [2]





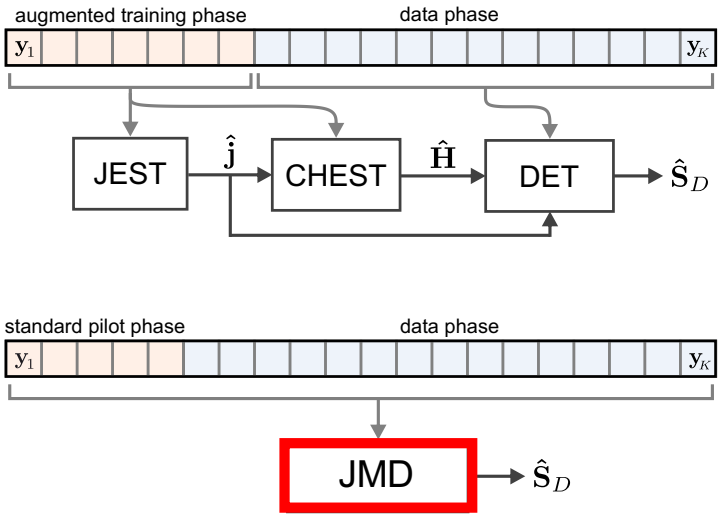


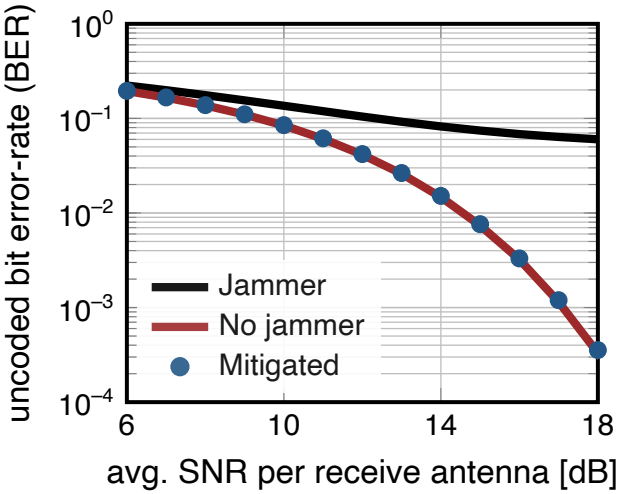
[1] Marti, Castañeda, CS, "Jammer mitigation via beam-slicing for low-resolution mmWave massive MU-MIMO," OJ-CAS 2021
[2] Marti, Castañeda, Jacobsson, Durisi, Goldstein, CS, "Hybrid jammer mitigation for all-digital mmWave massive MU-MIMO," Asilomar 2021

Dealing with jammers after the ADCs

- If jammers' spatial signature is known, then they can be removed by spatial nulling in the digital domain
  - Smart jammers can evade estimation of spatial signature

- Solution: Joint jammer mitigation and data detection (JMD) [1]





First jammer resilient data detector for multiuser MIMO [2]

← 22FDX ASIC of JMD

[1] Marti and CS, "Mitigating Smart Jammers in MU-MIMO via Joint Channel Estimation and Data Detection," ICC 2022
[2] Bucheli, Castañeda, Marti, CS, "A Jammer-Mitigating 267 Mb/s 3.78mm^2 583mW 32x8 Multi-User MIMO Receiver in 22FDX," VLSI 2024

Conclusions

Summary

- Wideband multi-antenna communication requires processing of high-dimensional signals at extreme rates
- All-digital architectures provide more flexibility and simplify baseband processing
- All-digital architectures are susceptible to jamming (hybrid architectures are too!)

- We have shown hat
  - efficient equalization is feasible using low-resolution techniques
  - beamspace processing can exploit channel sparsity for further power savings
  - systems can be made resilient to jamming before and after the ADCs

More information at **iip.ethz.ch**

Dr. Oscar Castañeda

Seyed Hadi Mirfarshbafan

Gian Marti