Scaling the "mostly digital" paradigm to mmWave massive MIMO

Upamanyu Madhow
ECE Department
University of California, Santa Barbara

Invited presentation, SPCOM 2022

Funded by NSF, DARPA/SRC JUMP

The mmWave → THz frontier

- "Unlimited" bandwidth
  - mmWave: Licensed (28 GHz), unlicensed (60 GHz)
  - Towards THz (100+ GHz), regulation TBD
- Tiny wavelengths → miniaturized antenna arrays
- Unique propagation characteristics
- Silicon RFICs, low-cost packaging

Didn’t you already say all this in 2016?

JTG/IEEE ITSoc Summer School 2016

Home | About | Speakers | Program | Accommodation | Registration | Contact | FAQs | Sponsors

2016 Joint Telematics Group/IEEE Information Theory Society Summer School
on Signal Processing, Communications and Networks.
IISc Bangalore, June 27 - July 01, 2016.

Venue: Golden Jubilee Hall, Department of ECE

Millimeter-Wave Systems: Theory, Systems, and Algorithms

mmWave represents the next frontier in wireless communication, providing "effectively unlimited" spectrum for short to medium range networks, due to the huge amounts of available spectrum and the aggressive spatial reuse enabled by highly directive links. Similarly, short-range mmWave radar is a key enabler for sensing applications such as gesture recognition and vehicular situational awareness. mmWave systems differ fundamentally from existing wireless systems because of the order of magnitude smaller carrier wavelength, and the order of magnitude higher available bandwidth. In this mini-course, we discuss some of these key differences and their design consequences, ranging from hardware/signal processing co-design to network protocols.

Lecture Titles (Milimeter-Wave Systems)

1. Overview: mmWave characteristics, concept systems, technical challenges
2. The mm wave channel: diversity, multiplexing, blockage
3. Steering large arrays: compressive and scan-based approaches
4. Fundamentals of compressive estimation: theory and algorithms
5. Networking with highly directional links
6. Networking (contd.)
7. Signal Processing at high bandwidths
8. Short-range mmWave radar

Today's agenda

- A brief summary of what we knew 5-6 years ago
- What's happened in mmWave@UCSB since then

mmWave at UCSB (2005-2017): a sampling

Directional Networking

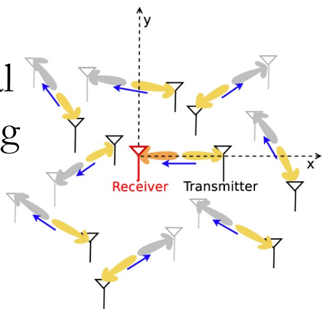
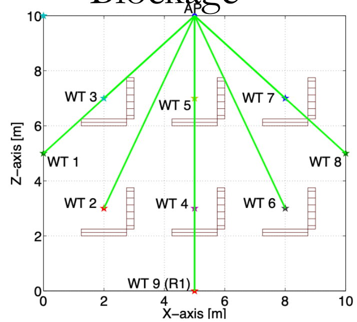


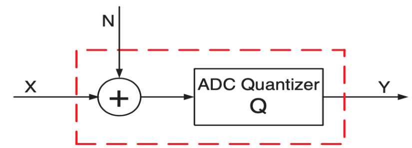
Fig. 2. Network model for interference analysis.

Blockage



(b) Office

ADC Fundamentals



NSF

QCOM, Samsung, Nokia
FB, Google

mmWave Picocells
Modeling, protocols, capacity

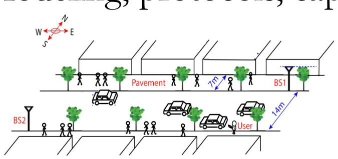
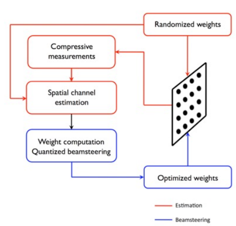
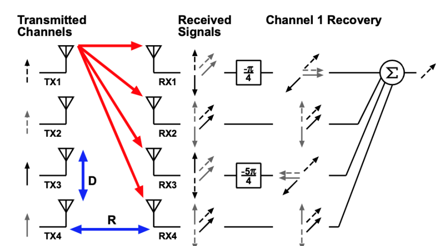


Fig. 1: Picocellular network deployed along an urban canyon

Compressive Estimation
Fundamentals, Algorithms, Demo



LoS MIMO
Fundamentals & Demo



mmWave Mesh Backhaul
Routing & Resource Alloc

Short-range mmWave Imaging
New models, Proof of Concept

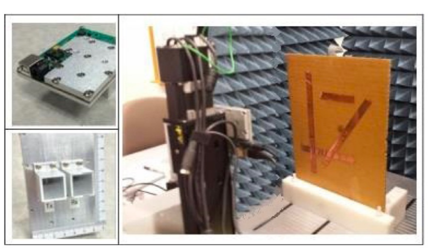


Fig. 5. Experimental data collection using 60 GHz quasi-monostatic radar system.

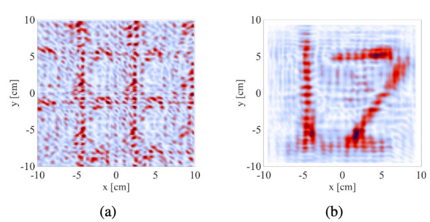
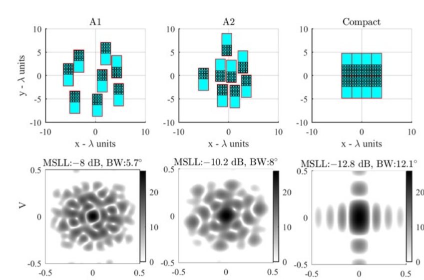


Fig. 8. Sparse array III (a) Point MF (b) Patch MF (1cm × 1cm).

mmWave Sensing
Sensor Geometries, Algorithms



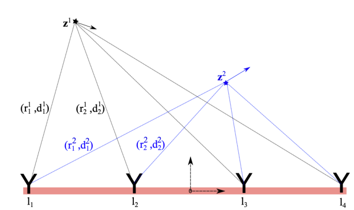


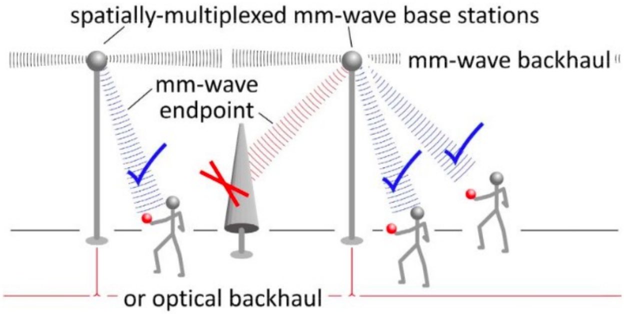
Fig. 1. 2D System model with a linear array of radar sensors placed on x-coordinates, $[l_1, l_2, l_3, l_4]$. The kinematic states z_1, z_2 of two targets are to be estimated using the unordered range and doppler observations from the sensors.

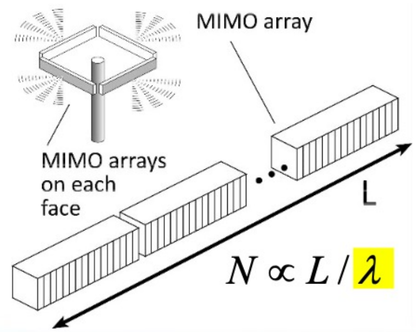
What we knew ~5 years ago

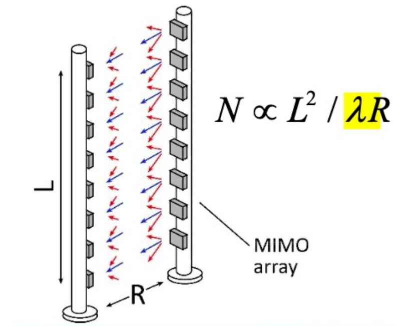
- Sweet spot is at short ranges
  - In-room indoors, ~100 meters outdoors
- Simple models for sparse channels are effective
- Blockage is not a killer: simulations and experiments
- Compressive estimation for efficient channel estimation & tracking
  - New super-resolution algorithms, experimental demonstrations
- LoS MIMO has huge potential: theory and prototyping
- Short-range sensing needs new models and algorithms
  - Patch models for extended objects (theory and experiments)
  - Exploiting geometric constraints

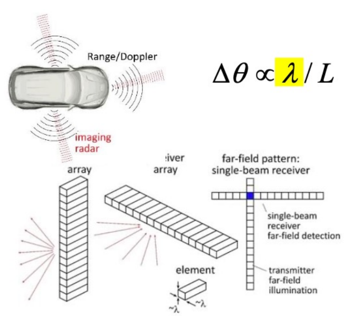
Industry was talking about 5G. What next for academia?

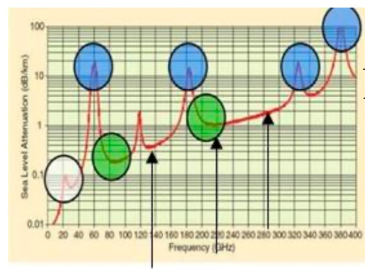
ComSenTer Vision
Communications & Sensing @ Terahertz











Band choices avoiding oxygen absorption peaks
(140, 210, 280 GHz)

Can mmWave hardware be scaled to these bands?

Massive increase in #RF chains

Low-cost packaging

Silicon whenever possible, augmented by III/V

How can system designs enable/exploit hardware scale?

Hardware-signal processing co-design

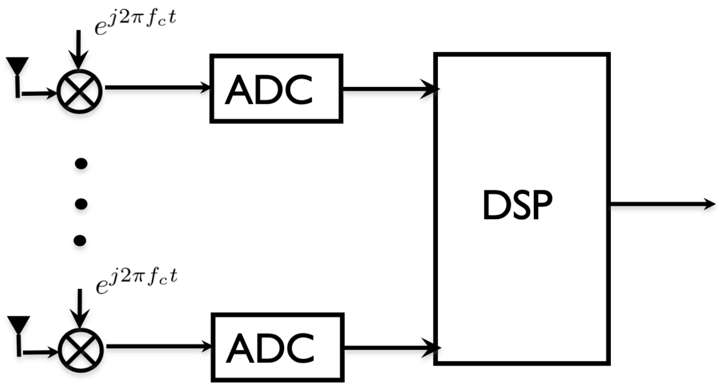
ComSenTer (2018-2022)
funded by DARPA/SRC JUMP program

$27M over 5 years

Prof. Mark Rodwell, Director

UCSB-led coalition of faculty from 10 universities

The "mostly digital" paradigm

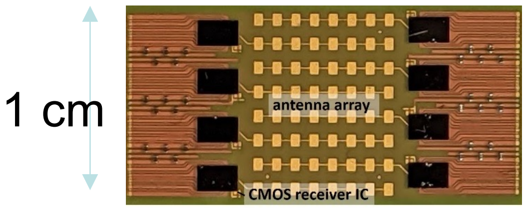


- Allows us to leverage Moore's law
- "Everything" is done in DSP
- Key to the success of cellular and WiFi

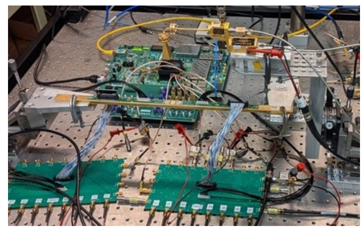
- **Can this scale to 100s of antennas at mmWave?**
  - RFIC designers: YES WE CAN (in low-cost CMOS)
- **Can this scale to bandwidths of ~10 GHz?**
  - Needs substantial innovations in DSP algorithms and architectures

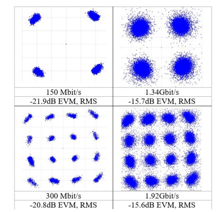
ComSenTer CMOS digital beamforming hardware at 140 GHz

Gen 1 CMOS 8 elt MU-MIMO RX



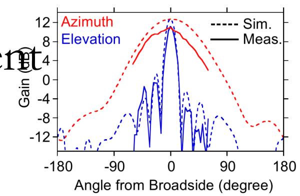
FPGA platform (RX)

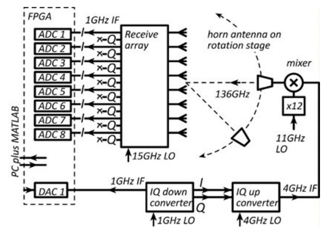




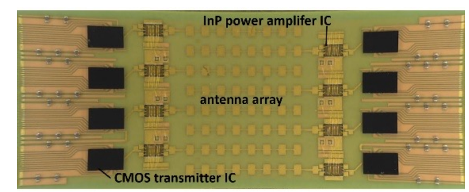
- Demonstrated MIMO beamforming
- Demonstrated 1.9Gb/s 16QAM transmission
- Baseband connector limits data rate
- Low assembly yield: ~60% working channels on most modules

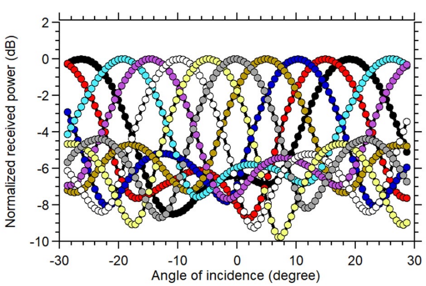
Single element pattern



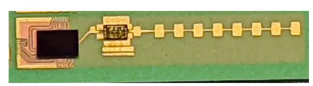


Analogous results for Gen 1 CMOS 8-elt MIMO TX





Multiple beams



CMOS + InP (increased power)

Improved packaging design for Gen 2 modules

Pushing the limits of the "mostly digital" paradigm

- With Moore's law winding down, does "mostly digital" still make sense?
  - YES: parallel & distributed architectures allow continued scaling of compute, compute is much more energy efficient than communication
- **But can we truly scale?**
  - **Need to rework canonical MIMO system designs**
- Massive MU-MIMO: the most obvious way to push boundaries
  - All-digital ➔ #users scales with #antennas
  - Bottlenecks: RF impairments (nonlinearities, phase noise), ADC precision, DSP complexity
- LoS MIMO
  - Can we exploit the DSP magic to enable flexible deployment of "wireless fiber" mesh networks?

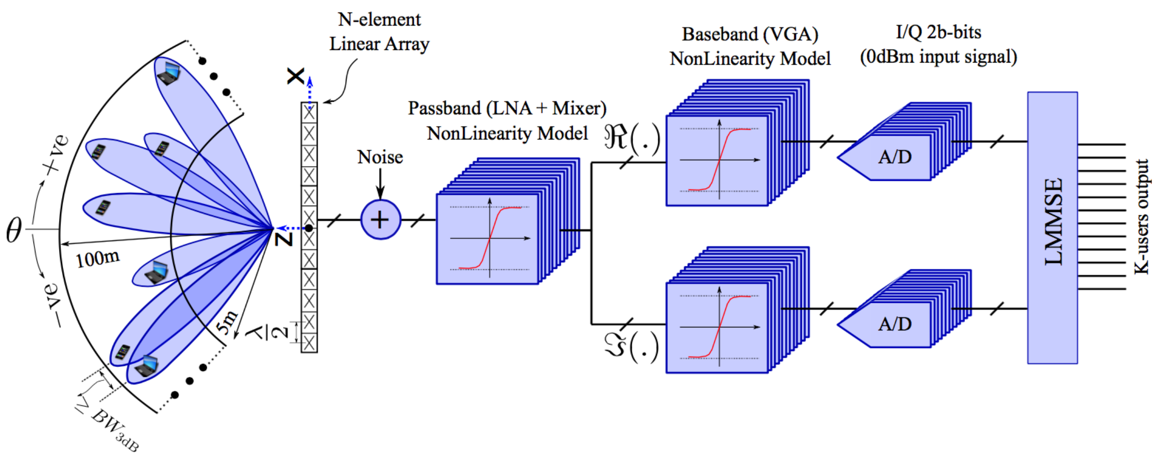
Massive MU-MIMO

Mohammed Abdelghany

Maryam E. Rasekh

12

Concept System: Tbps Massive MIMO @140GHz



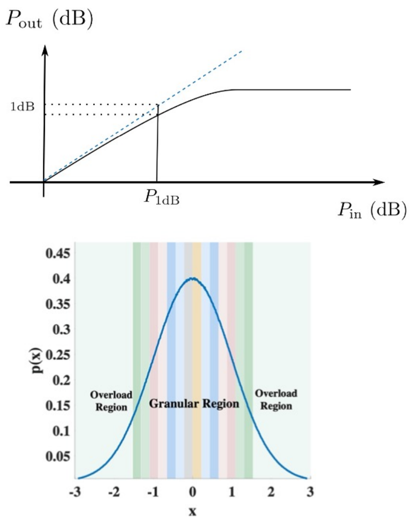
140 GHz
Picocellular
Uplink
(10 Gbps/user,
100 simultaneous
users)

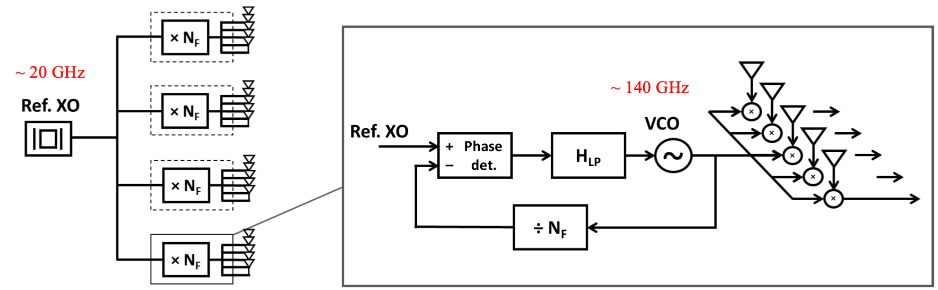
Key bottlenecks for all-digital architectures

- Need one RF chain for each antenna. Can we relax the specs enough that CMOS works?
- Phase noise is high at millimeter wave and THz. Don't things get worse as we scale to a large number of antennas?
- ADC cost, power consumption and availability is limited as we scale up bandwidth
- Multiuser detection is needed, but classic architectures do not scale

Hardware/signal processing co-design is crucial

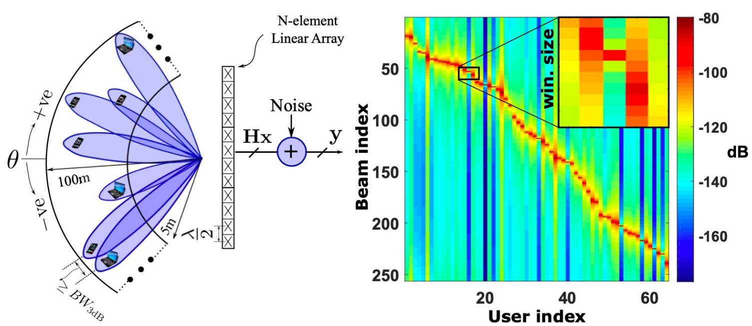
Example System Insights from ComSenTer





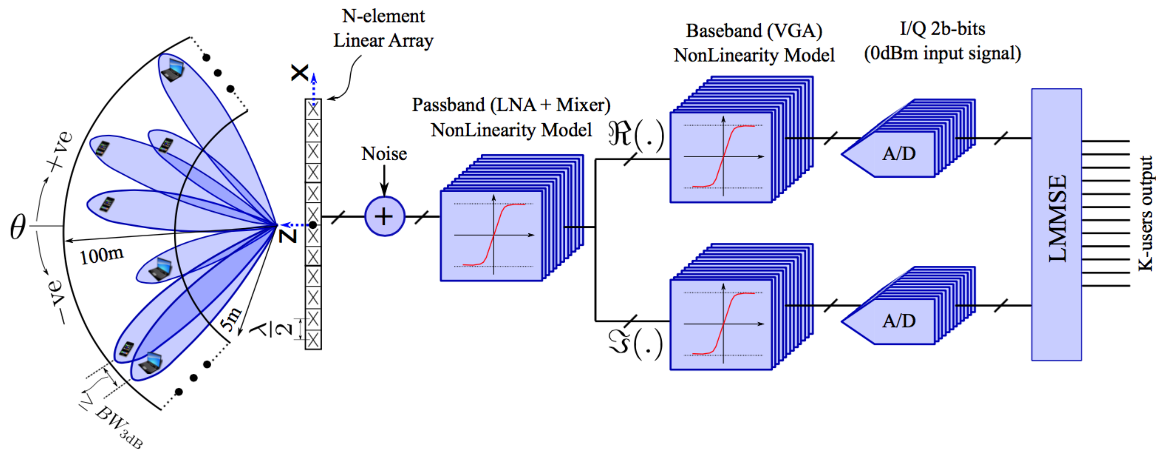
Scale can be attained with tiling
(phase noise specs can be relaxed)

Severe nonlinearities can be tolerated with scale
(hardware specs can be relaxed)



Channel sparsity can be exploited by going to beamspace
(significant reduction in signal processing complexity)

Beamspace Local LMMSE



M. Abdelghany, U. Madhow, A. Tolli, *Beamspace Local MMSE: an efficient digital backend for mmWave massive MIMO*, SPAWC 2019.

Typical channel: one path is dominant

$$\mathbf{y}_k = \mathbf{h}_k x_k$$

$$\mathbf{h}_k = \underbrace{\left(\frac{\lambda}{4\pi R_k}\right)^2}_{A_k^2} \left[\exp(-jn\, \underbrace{2\pi \frac{d \sin(\theta_k)}{\lambda}}_{\text{spatial freq. } \Omega_k}) \right]_{n=0}^{N-1}$$

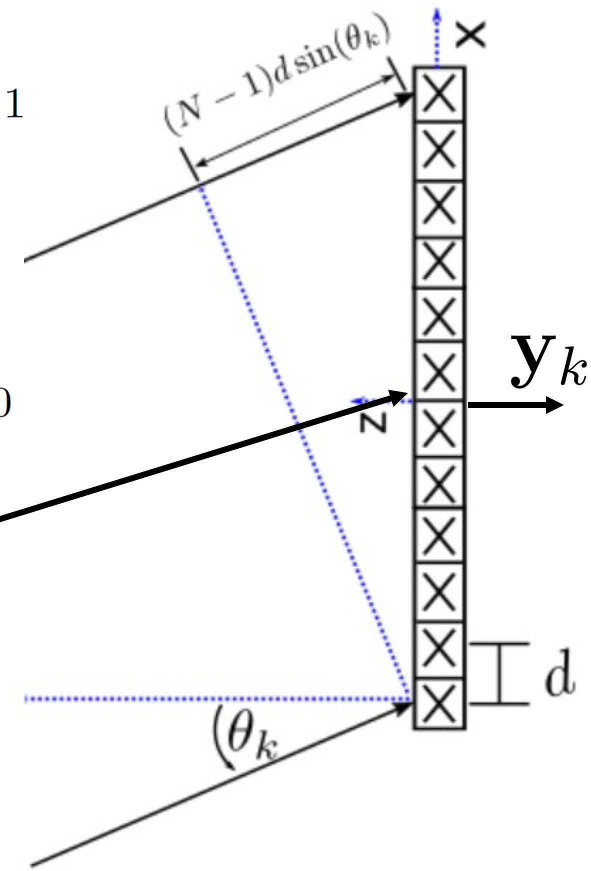
Path loss

Phase progression

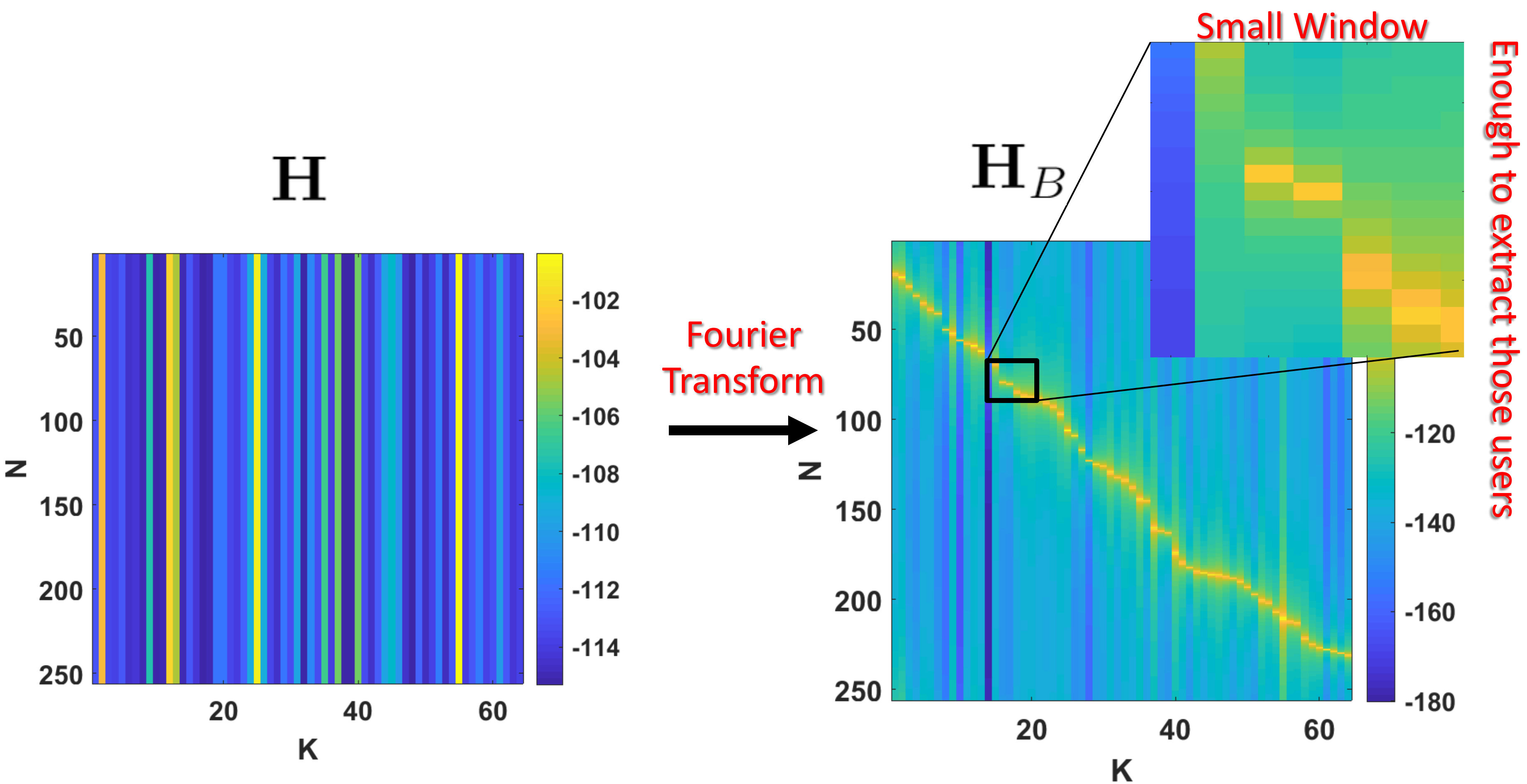
User k

x_k

R_k



Channel Matrix Sparsity in Beamspace



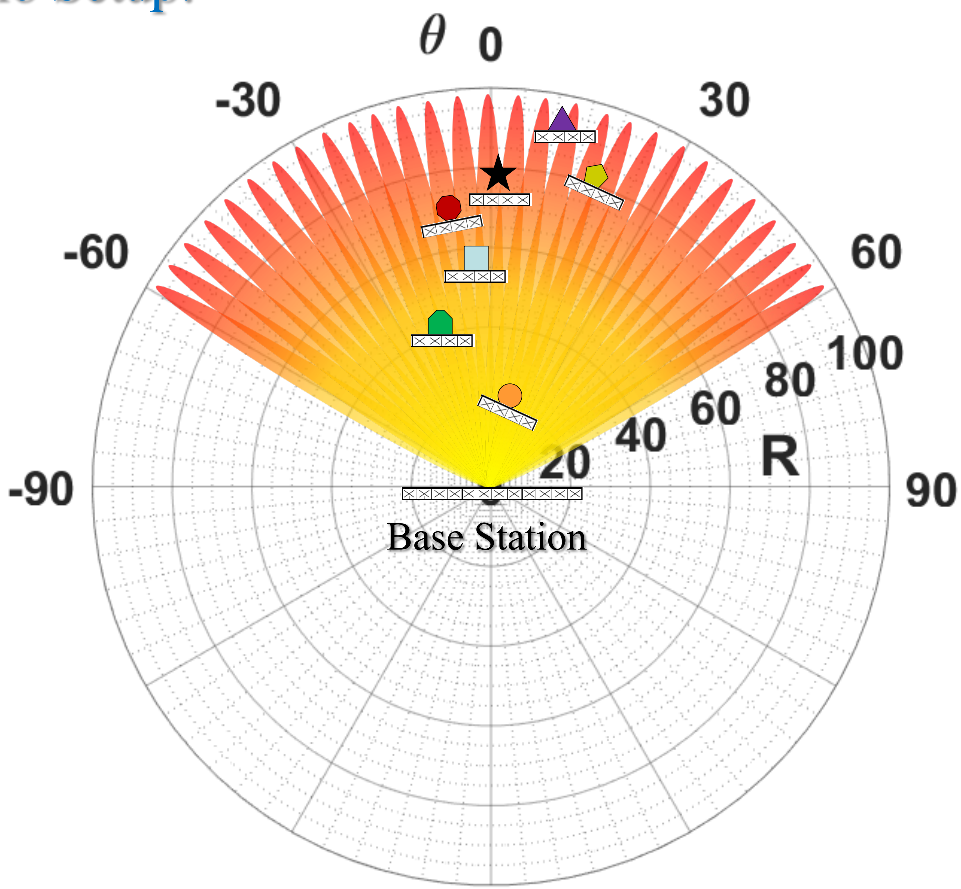
Most of the energy for a user can be captured in ~3 FFT bins (even as we scale #antennas)

Examples of beamspace processing

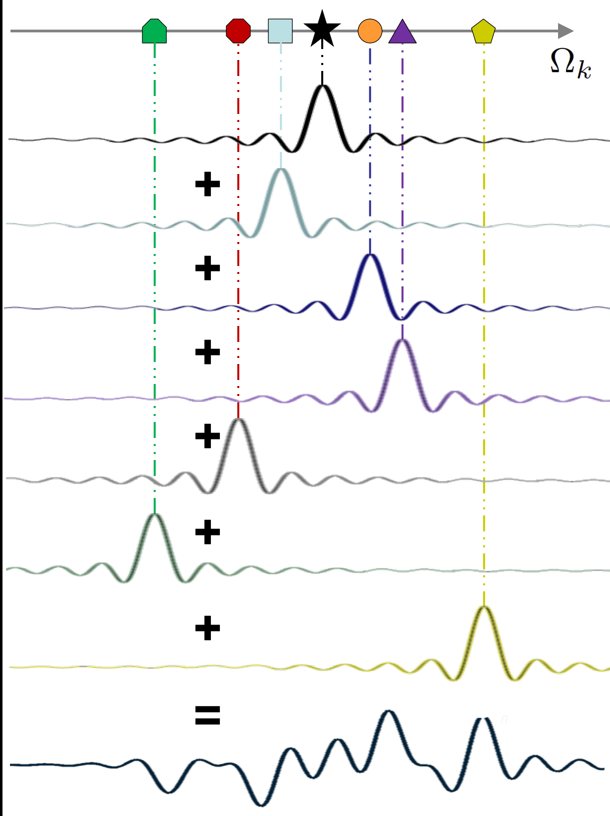
- Local LMMSE detection (SPAWC 2019)
  - Significantly reduces complexity at low load factors
- Nonlinear interference cancellation (SPAWC 2020)
  - SIC on top of LMMSE helps push load factors higher
- Wideband space-time interference suppression (Globecom 2019)
  - Space-time FFT instead of true time delay
- Downlink precoding (Asilomar 2019)
  - Applying uplink-downlink duality

Local LMMSE for Multiuser MIMO

Scenario Setup:

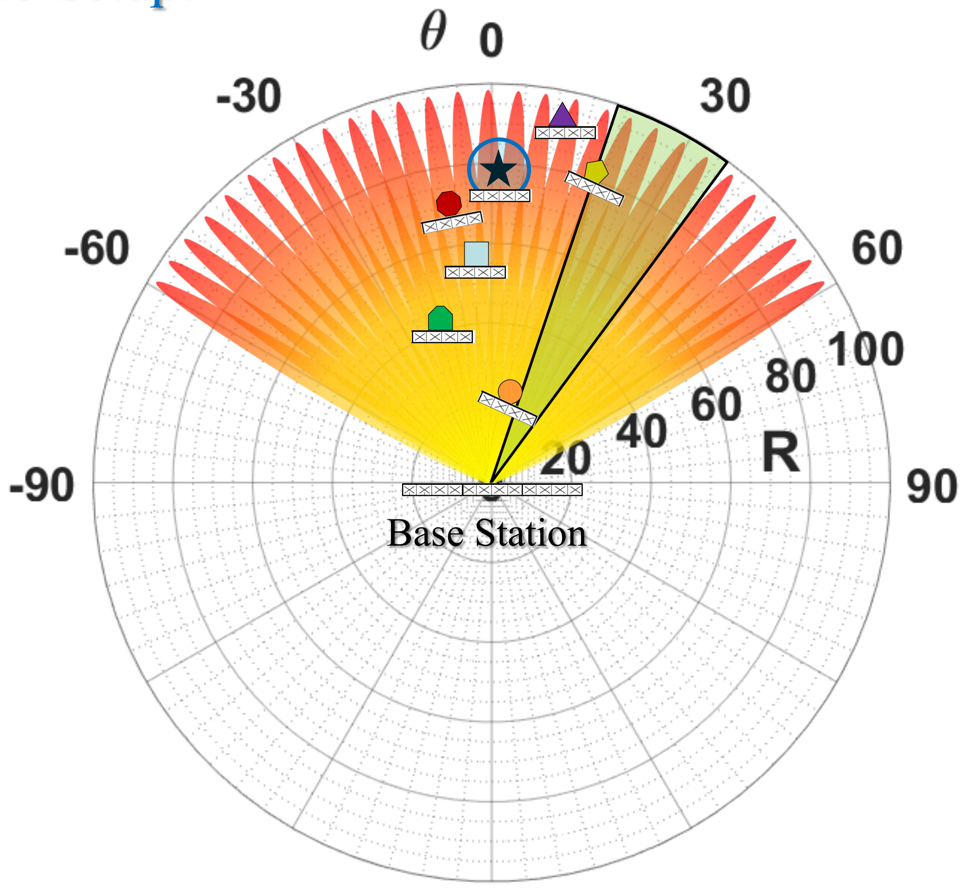


Received Signal in Beamspace:

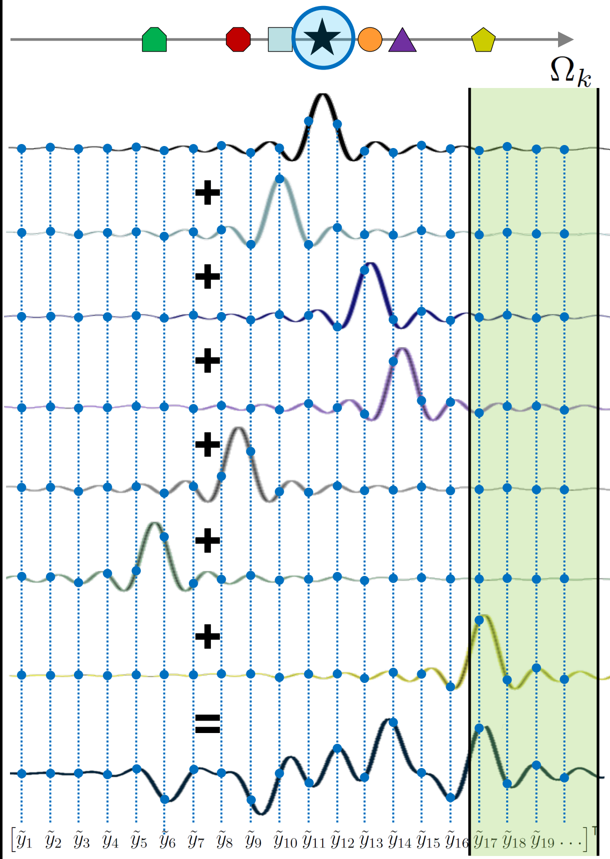


Local LMMSE: Multiuser

Scenario Setup:

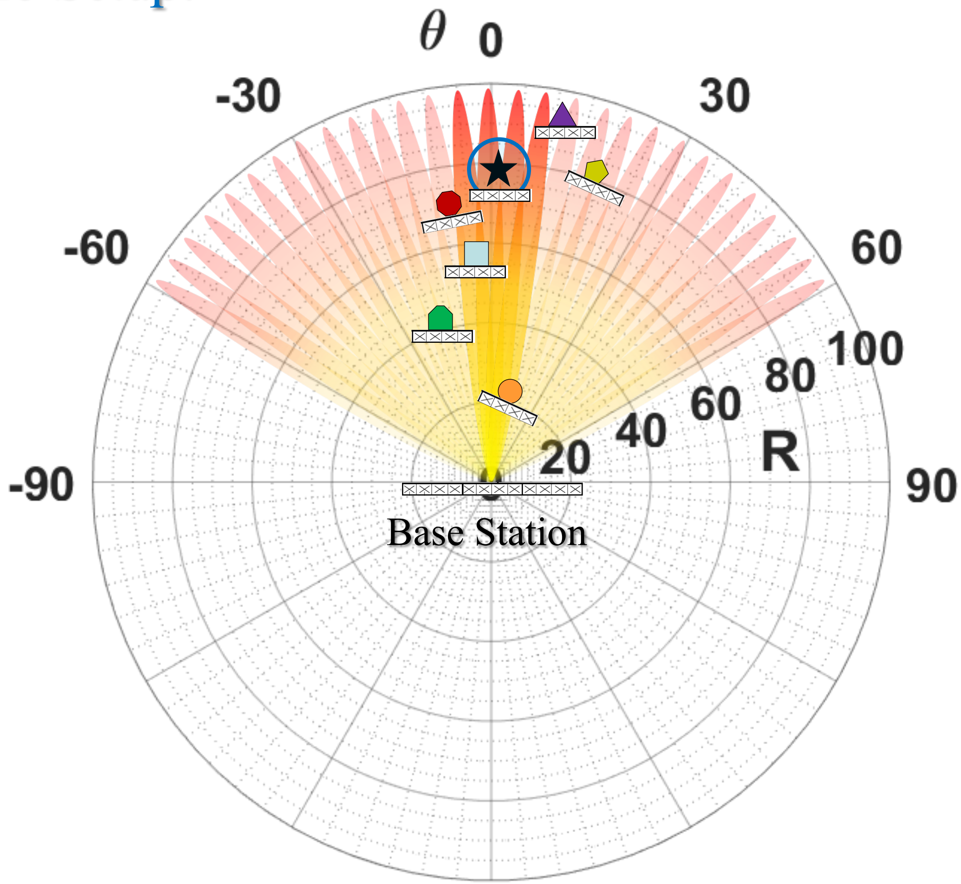


Received Signal in Beamspace:

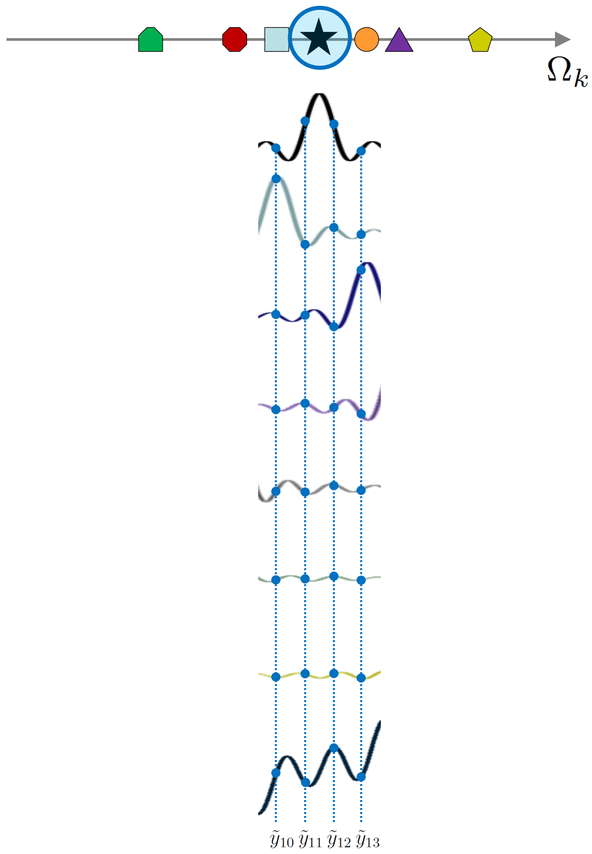


Local LMMSE: Multiuser

Scenario Setup:



Received Signal in Beamspace:



Local LMMSE: Multiuser

System Equation After Windowing:

$$\underbrace{\begin{bmatrix} \tilde{y}_{10} \\ \tilde{y}_{11} \\ \tilde{y}_{12} \\ \tilde{y}_{13} \end{bmatrix}}_{\tilde{\mathbf{y}}_{10:13}} = \underbrace{\begin{bmatrix} \overbrace{\tilde{H}_{10,1}}^{\tilde{\mathbf{u}}_{10:13}} & \tilde{H}_{10,2} & \cdots & \tilde{H}_{10,K} \\ \tilde{H}_{11,1} & \tilde{H}_{11,2} & \cdots & \tilde{H}_{11,K} \\ \tilde{H}_{12,1} & \tilde{H}_{12,2} & \cdots & \tilde{H}_{12,K} \\ \tilde{H}_{13,1} & \tilde{H}_{13,2} & \cdots & \tilde{H}_{13,K} \end{bmatrix}}_{\tilde{\mathbf{H}}_{10:13}} \mathbf{x} + \underbrace{\begin{bmatrix} \tilde{n}_{10} \\ \tilde{n}_{11} \\ \tilde{n}_{12} \\ \tilde{n}_{13} \end{bmatrix}}_{\tilde{\mathbf{n}}_{10:13}}$$

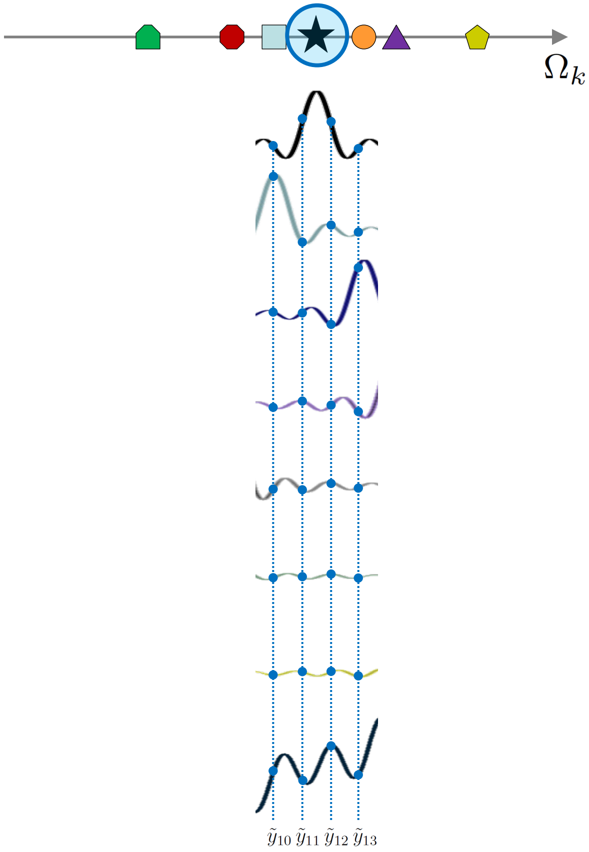
Performance Metric:

- Using optimal linear processing (LMMSE):

$$SINR \propto \frac{1}{MSE}$$

$$MSE_{10:13} = 1 - \tilde{\mathbf{u}}_{10:13}^H \overbrace{(\tilde{\mathbf{H}}_{10:13}\tilde{\mathbf{H}}_{10:13}^H + \sigma^2\mathbf{I})^{-1}}^{\mathbf{R}_{10:13}^{-1}} \tilde{\mathbf{u}}_{10:13}$$

Received Signal in Beamspace:



Local LMMSE: Multiuser

System Equation After Windowing:

$$\underbrace{\begin{bmatrix}\tilde{y}_{10}\\ \tilde{y}_{11}\\ \tilde{y}_{12}\\ \tilde{y}_{13}\end{bmatrix}}_{\tilde{\mathbf{y}}_{10:13}} = \underbrace{\begin{bmatrix}\overbrace{\tilde{H}_{10,1}}^{\tilde{\mathbf{u}}_{10:13}} & \tilde{H}_{10,2} & \cdots & \tilde{H}_{10,K}\\ \tilde{H}_{11,1} & \tilde{H}_{11,2} & \cdots & \tilde{H}_{11,K}\\ \tilde{H}_{12,1} & \tilde{H}_{12,2} & \cdots & \tilde{H}_{12,K}\\ \tilde{H}_{13,1} & \tilde{H}_{13,2} & \cdots & \tilde{H}_{13,K}\end{bmatrix}}_{\tilde{\mathbf{H}}_{10:13}} \mathbf{x} + \underbrace{\begin{bmatrix}\tilde{n}_{10}\\ \tilde{n}_{11}\\ \tilde{n}_{12}\\ \tilde{n}_{13}\end{bmatrix}}_{\tilde{\mathbf{n}}_{10:13}}$$

Performance Metric:

- Using optimal linear processing (LMMSE):

$$SINR \propto \frac{1}{MSE}$$

$$MSE_{10:13} = 1 - \tilde{\mathbf{u}}_{10:13}^H \overbrace{(\tilde{\mathbf{H}}_{10:13}\tilde{\mathbf{H}}_{10:13}^H + \sigma^2\mathbf{I})^{-1}}^{\mathbf{R}_{10:13}^{-1}} \tilde{\mathbf{u}}_{10:13}$$

Preprocessing:

1. Optimum Window Location:

$$I^* = \arg\min_I MSE_{I:I+W-1}$$

2. Local LMMSE Filter:

$$\mathbf{w}_1 = \mathbf{R}_{I^*:I^*+W-1}^{-1}\tilde{\mathbf{u}}_{I^*:I^*+W-1}$$

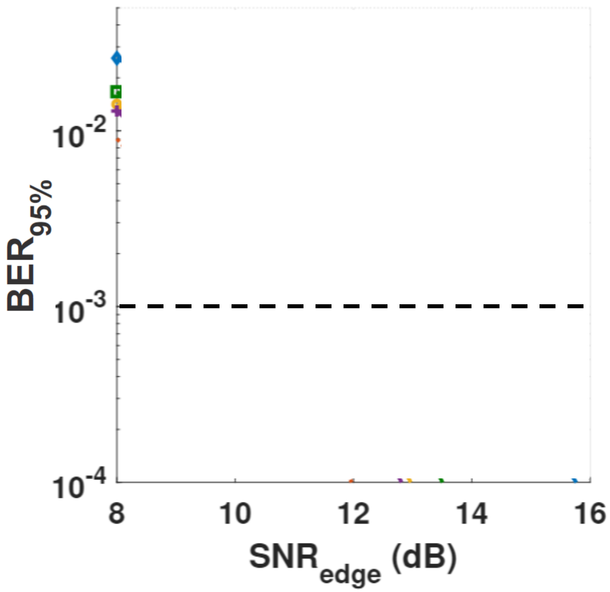
Beamformer:

$$\hat{x}_1 = \mathbf{w}_1^H \tilde{\mathbf{y}}_{I^*:I^*+W-1}$$

Local LMMSE: Multiuser

Performance Results:

- SNR_{edge}: The signal-to-noise ratio of the edge user
- $BER_{95\%}$: The BER achieved by at least 95% of the users



Preprocessing:

1. Optimum Window Location:

$$I^* = \arg\min_{I} MSE_{I:I+W-1}$$

2. Local LMMSE Filter:

$$\mathbf{w}_1 = \mathbf{R}^{-1}_{I^*:I^*+W-1}\tilde{\mathbf{u}}_{I^*:I^*+W-1}$$

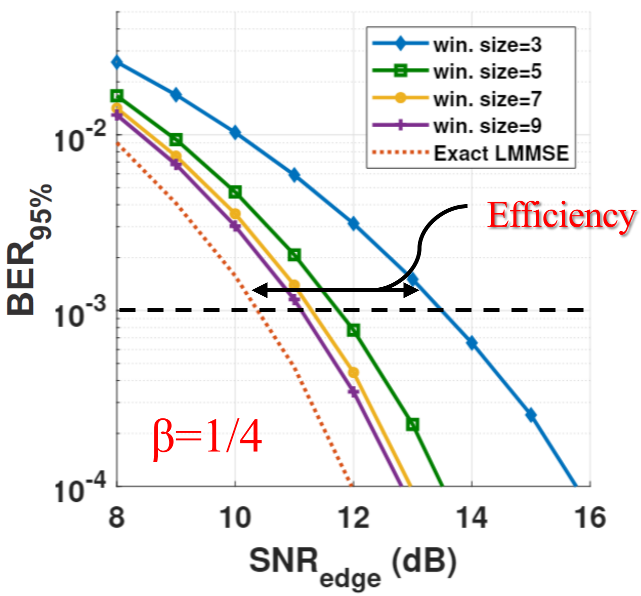
Beamformer:

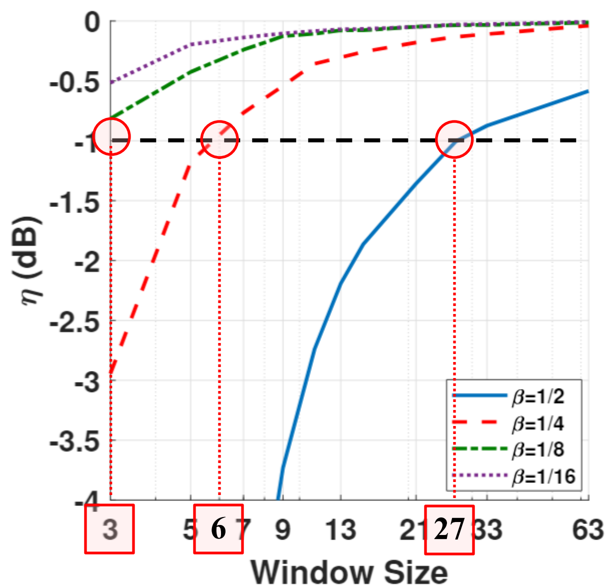
$$\hat{x}_1 = \mathbf{w}_1^H \tilde{\mathbf{y}}_{I^*:I^*+W-1}$$

Local LMMSE ➔ Parallelization, reduced complexity

Performance Results:

1. Window size (W) increases with the load factor (β)
2. Window size (W) does not scale with the number of elements (N)





$$\eta = \frac{SNR_{edge}(LMMSE)}{SNR_{edge}}$$

Preprocessing:

1. Optimum Window Location:

$$I^* = \arg\min_I MSE_{I:I+W-1}$$

2. Local LMMSE Filter:

$$\mathbf{w}_1 = \mathbf{R}^{-1}_{I^*:I^*+W-1}\tilde{\mathbf{u}}_{I^*:I^*+W-1}$$

Beamformer:

$$\hat{x}_1 = \mathbf{w}_1^H \tilde{\mathbf{y}}_{I^*:I^*+W-1}$$

Summary and Status

Promising first steps for multiuser MIMO for massive scale

Scale simplifies design of individual hardware components

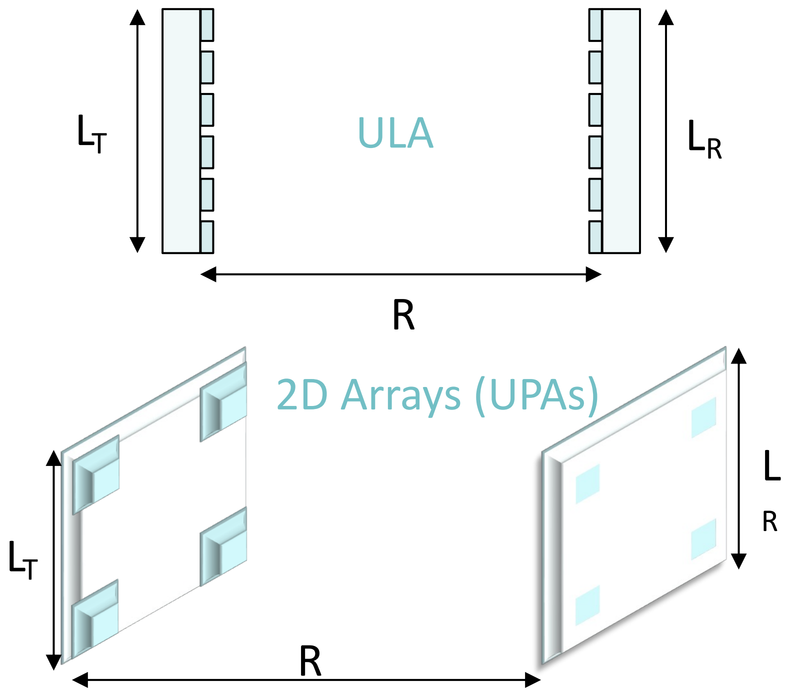
Scaling both antennas & bandwidth remains a huge DSP challenge!

Rethinking LoS MIMO

Lalitha Giridhar

Ahmet Sezer

LoS MIMO is a natural concept for mmWave and THz



Number of spatial degrees of freedom
(based on information-theoretic considerations):

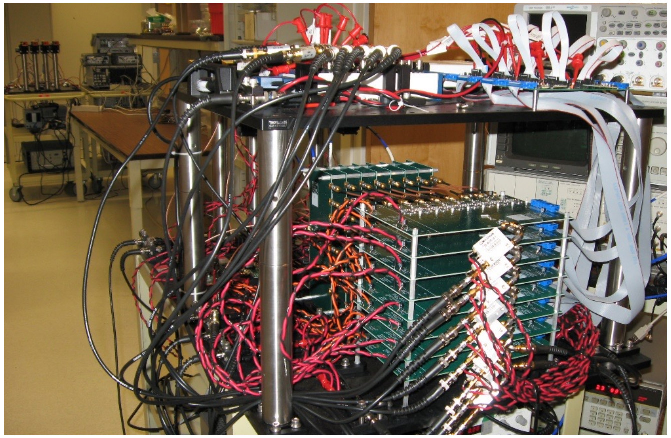
$$N \approx \frac{L_T L_R}{R\lambda} + 1$$

$$N \approx \left(\frac{L_T L_R}{R\lambda}\right)^2 + 1$$

Bandwidth also scales with carrier frequency
($fc \propto 1/\lambda$) where λ : wavelength

→ Data Rate $\propto f_c^3$

Significant progress in past decade



2 orders of magnitude in range & data rate

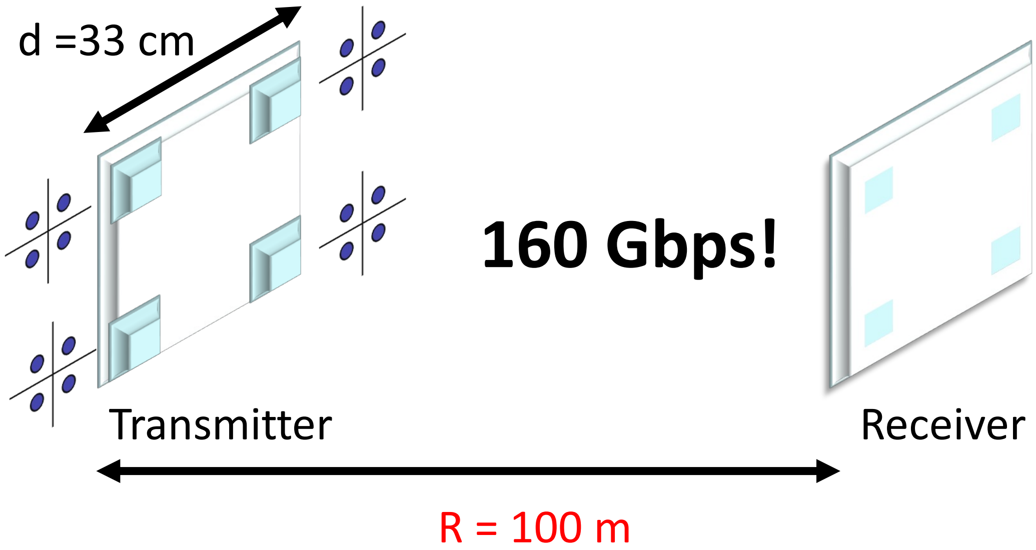
UCSB lab demo @ 60 GHz (2010)
4-fold spatial multiplexing
2.4 Gbps aggregate data rate
Range 10-40 meters

Ericsson prototype link in E-band (2019)
8-fold multiplexing: 4 spatial, 2X polarization
100 Gbps aggregate data rate
Range 1500 meters

Widespread deployment of LoS MIMO requires less bulky and expensive equipment

Short-range backhaul more interesting?

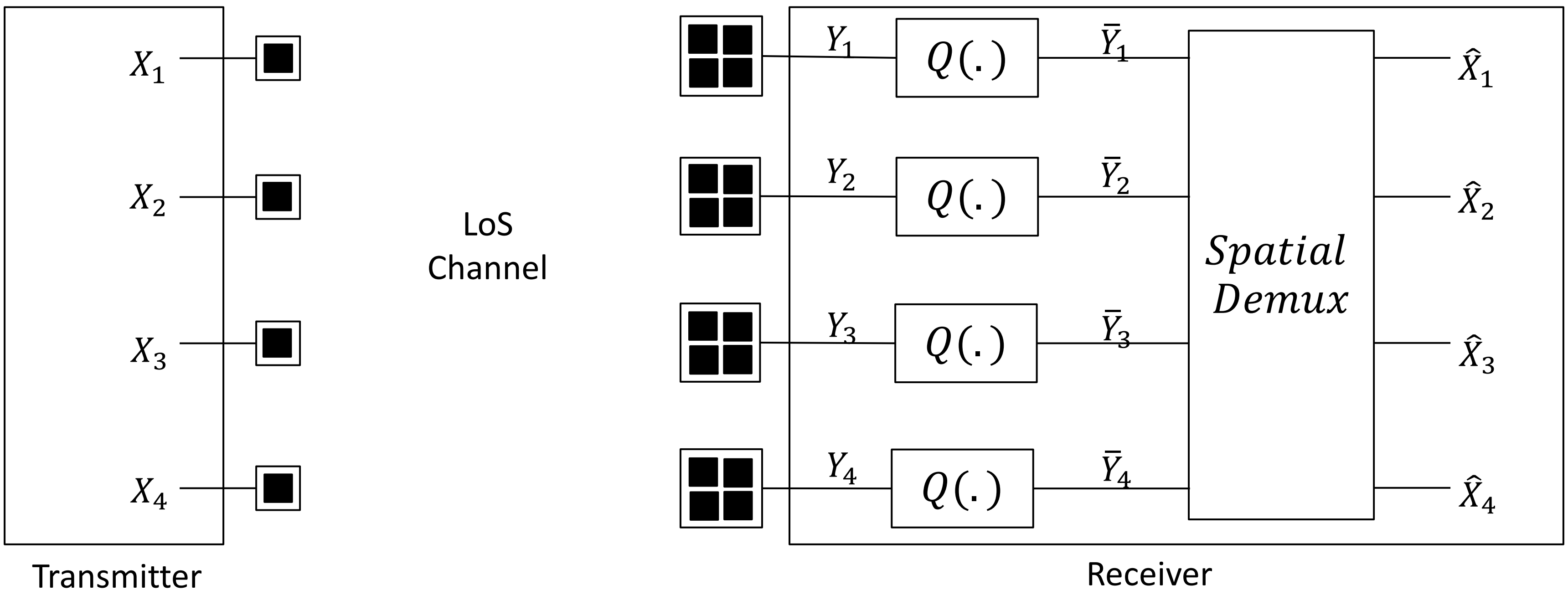
4 x 4 MIMO
130/140 GHz carrier frequency
40 Gbps per stream
Antenna spacing 33cm
(*lamppost-compatible*)



Reasonable form factor, but how about cost & power?

- CMOS or SiGe RFICs with required power are within reach
- DSP is key to economies of scale in baseband processing
- But ADC is a bottleneck at 10-20 GHz bandwidths
- Geometric misalignments result in channel dispersion

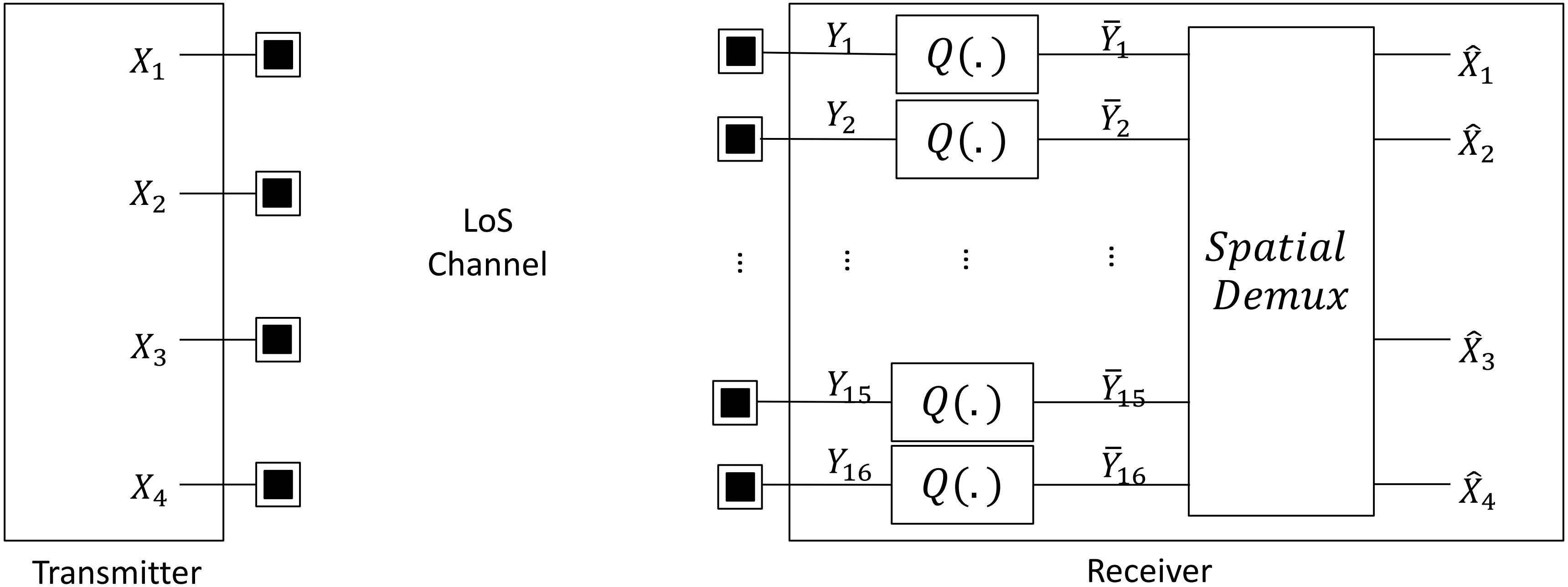
Standard Array of Subarrays Architecture



Rayleigh-spaced “subarrays” attain the available spatial DoF

“Subarrays” can use RF beamforming (one RF chain per subarray)

All-digital architectures are more flexible



Recent developments in mmWave RFICs: one RF chain for each antenna

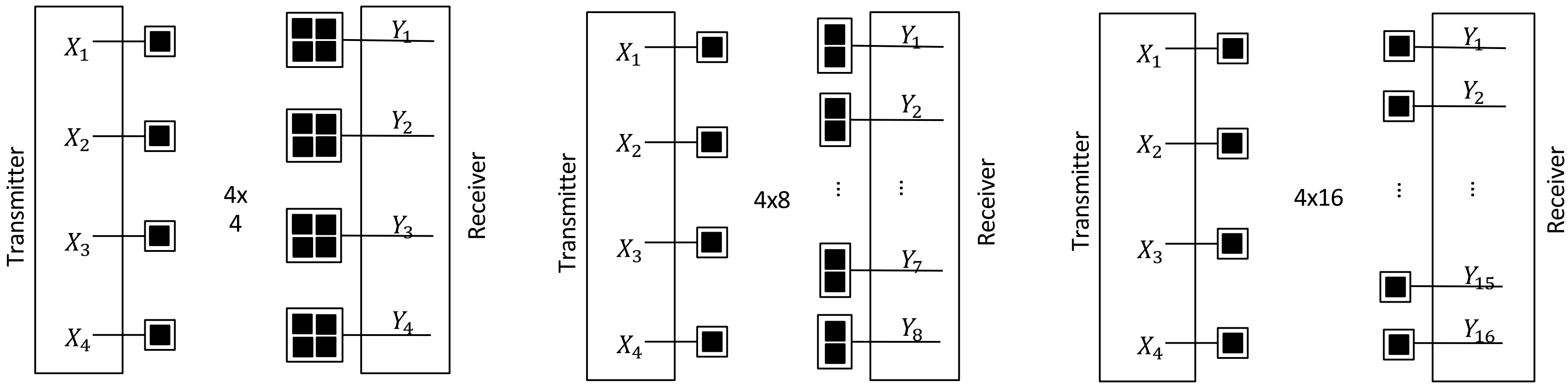
Can spread receive antennas evenly across the aperture

Spatial redundancy: a diversity of “looks” at the channel
When is this useful?

Does packaging antennas differently help?

Three example configurations

Configuration #1 (Benchmark) | Configuration #2 | Configuration #3



Each configuration has a total of 16 receive antennas.

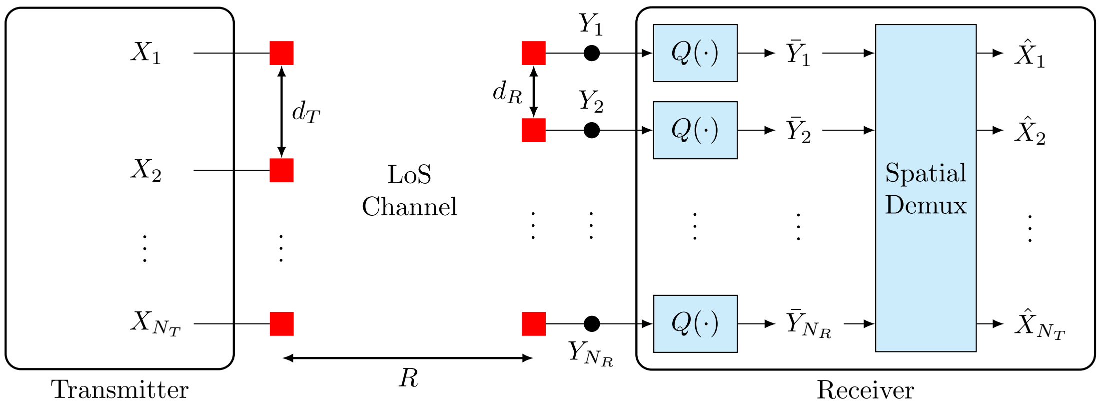
Receive units (subarrays) are spaced evenly across the same aperture

Each subarray employs RF beamforming

Does spreading antennas out help in the presence of severe quantization?

Mathematical model for LoS MIMO

No transmit precoding!



Received Signal

$$\mathbf{Y} = \mathbf{H}\,\mathbf{X} + \mathbf{N}$$

Channel between TX element *n* and RX unit *m*

$$\mathbf{H}(m,n) = e^{-j\Phi}\, e^{-j\theta_{m,n}}$$

Common phase

$$\Phi \sim \mathcal{U}(0, 2\pi)$$

"Cross-over" Phase

$$\theta_{m,n} \approx \frac{\pi((n-1)d_T - (m-1)d_R)^2}{\lambda R}$$

for $R \gg (N_T - 1)d_T = (N_R - 1)d_R$

Inter-antenna spacing of transmit elements

$$d_T = \sqrt{\frac{\lambda R}{N_T}}$$

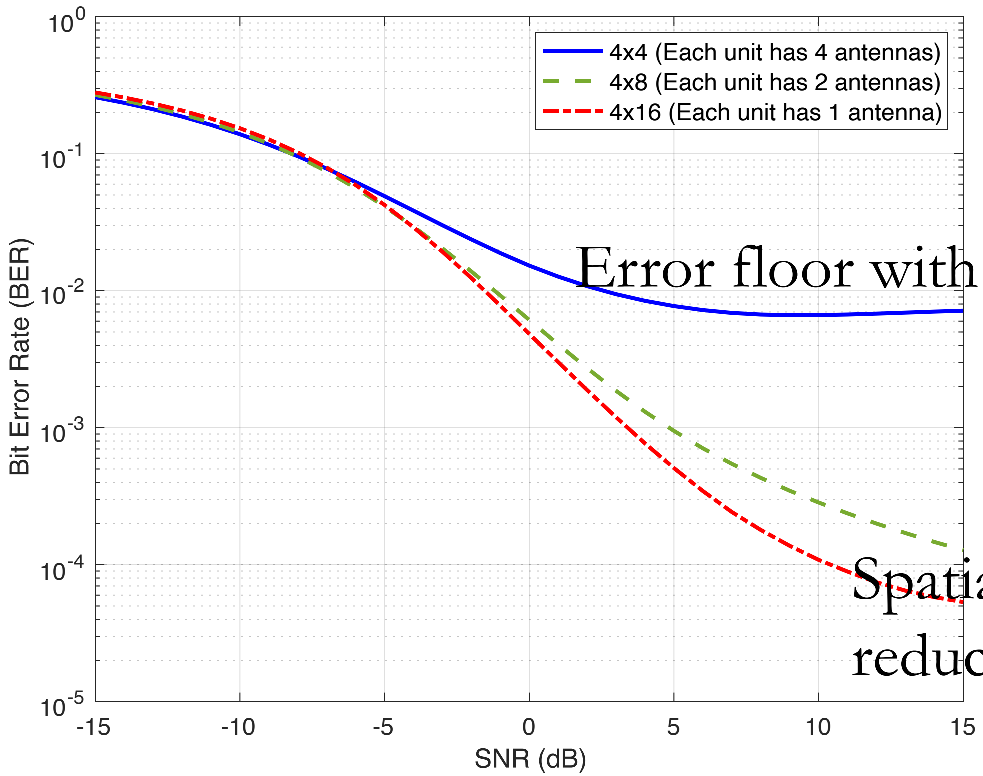
Rayleigh spaced (assuming same aperture for TX and RX)

Inter-antenna spacing of receive units

$$d_R = \frac{d_T(N_T - 1)}{N_R - 1}$$

Same aperture as TX, but possibly different spacing

Spatial redundancy alleviates impact of severe quantization

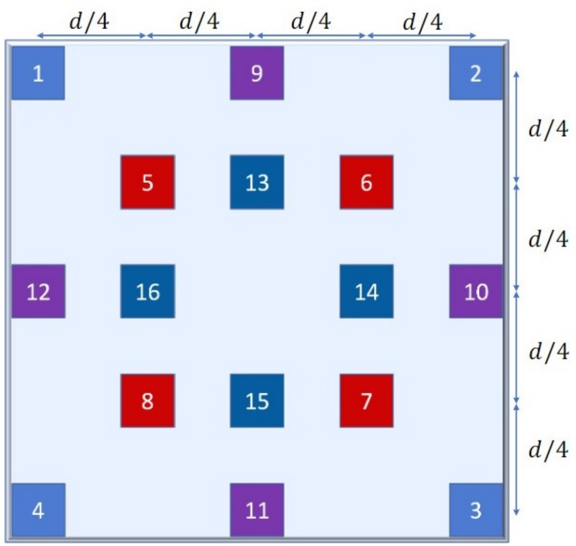


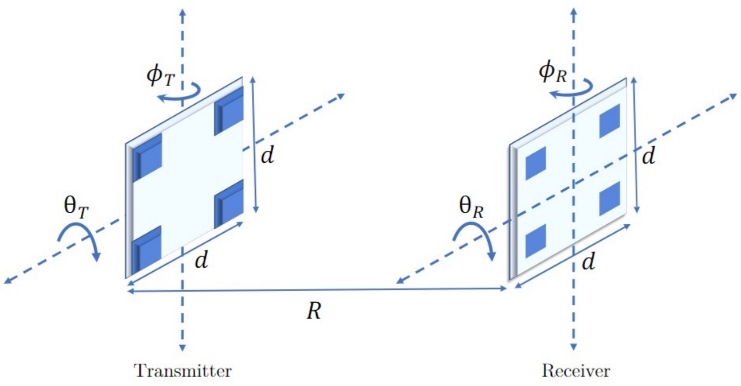
Error floor with array of subarrays

Spatial redundancy reduces error floor

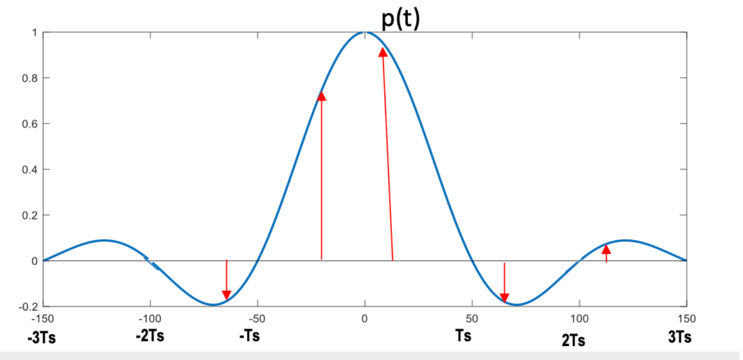
QPSK with 2-bit quantization

Spatial redundancy also helps with misalignment



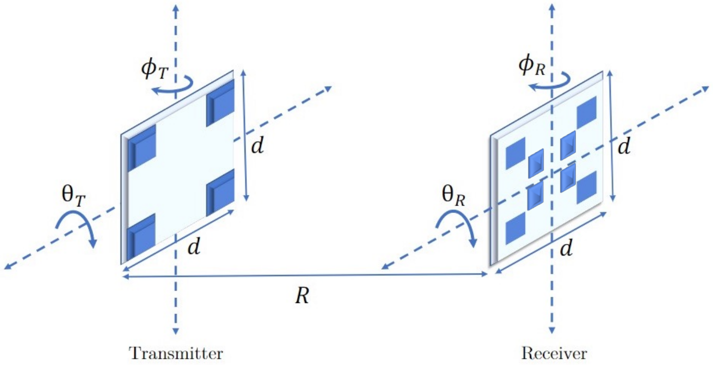


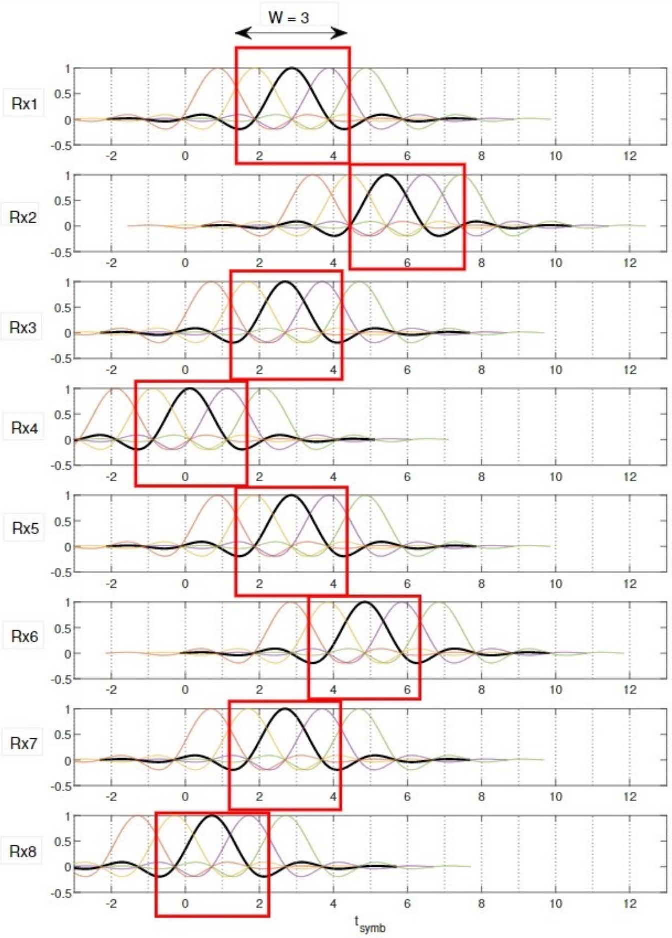
- QPSK modulation
- BW = 20 GHz for f_c = 130 GHz
- Symbol duration T = 50 ps and λ = 2.3 mm
- Transmit pulse - RC waveform with β = 0.25
- Symbol rate sampling assumed; T_S = T



Misalignment will be routine in mesh networks with LoS MIMO links

Adaptive windowing with spatial oversampling

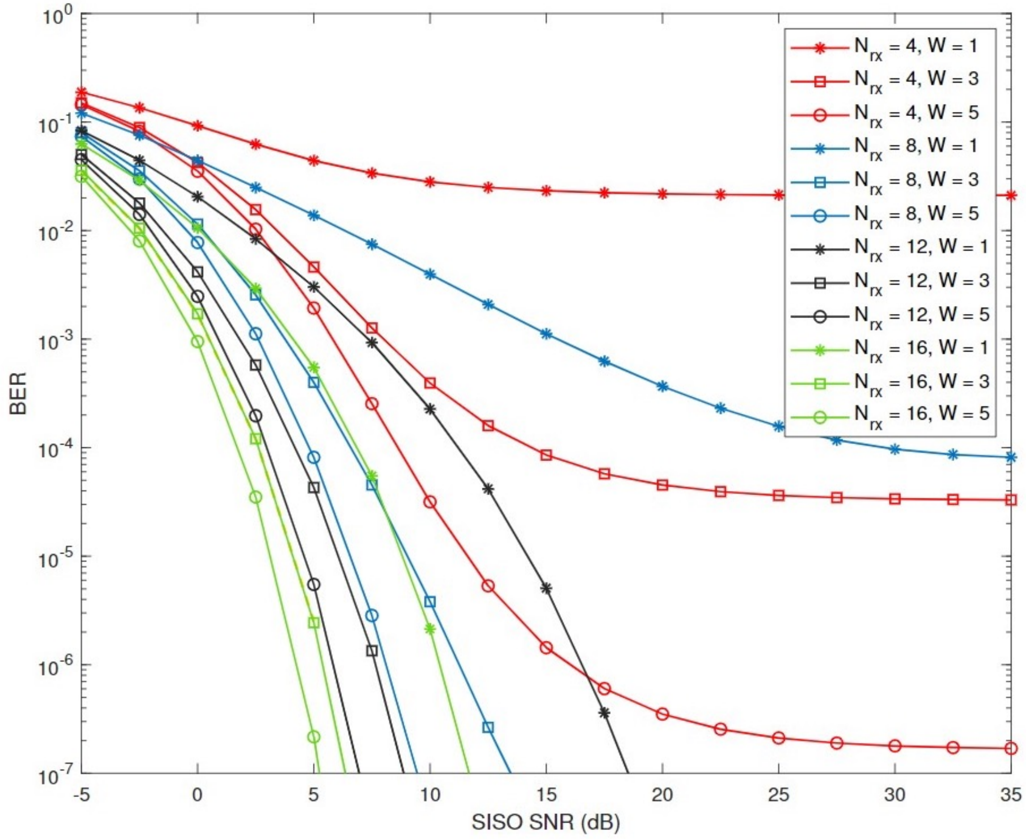




Misalignment example:

$$\theta_T = 3.67^\circ$$
$$\varphi_T = -4.30^\circ$$
$$\theta_R = 6.36^\circ$$
$$\varphi_R = 7.19^\circ$$

BER curves and dimension counting



W	Interference Vectors	Dimension of Signal Space				
		$N_{\mathrm{RX}}=4$	$N_{\mathrm{RX}}=6$	$N_{\mathrm{RX}}=8$	$N_{\mathrm{RX}}=12$	$N_{\mathrm{RX}}=16$
1	19	4	6	8	12	16
3	27	12	18	24	36	48
5	35	20	30	40	60	80

Error floors avoided when signal space dimension is bigger than # strong interference vectors

Summary and Status

One RF chain per antenna enables novel LoS MIMO configurations

Spatial redundancy increases resilience to impairments

Much work remains for enabling flexible deployment of "wireless fiber"

Parting Thoughts

- Pushing to higher carrier frequencies keeps opening up new intellectual challenges via hardware/signal processing entanglement
  - Hardware bottlenecks force system innovations
  - Hardware advances open up new system possibilities
  - Key ideas: antenna scaling, bandwidth scaling, sparsity, geometry
- Ambitious system specs today become industry focus ~10 yrs from now
  - The only legitimate barriers are physics and information theory fundamentals
- What next?
  - Cost- and energy-efficient hardware scaling: co-design of RFIC, packaging, ADC/DAC, DSP
  - Cost- and energy-efficient network scaling: virtualization, fronthaul, backhaul
  - (Sub-)THz joint communication and sensing infrastructure

Scaling “mostly digital:” selected pubs

Averaging out impairments via antenna scaling in mostly digital MU-MIMO

- Mohammed Abdelghany, Ali A. Farid, Maryam Eslami Rasekh, Upamanyu Madhow, Mark J. W. Rodwell, "A design framework for all-digital mmWave massive MIMO with per-antenna nonlinearities", TWC 2021.
- Maryam Eslami Rasekh, Mohammed Abdelghany, Upamanyu Madhow, Mark Rodwell, "Phase noise in modular millimeter wave massive MIMO", TWC 2021.

Beamspace multiuser MIMO

- Mohammed Abdelghany, Upamanyu Madhow, Antti Tölli, "Beamspace Local LMMSE: An Efficient Digital Backend for mmWave Massive MIMO", SPAWC 2019.
- Mohammed Abdelghany, Upamanyu Madhow, and Mark J. Rodwell, "An efficient digital backend for wideband single-carrier mmWave massive MIMO", Globecom 2019.
- Mohammed Abdelghany, Maryam Eslami Rasekh, Upamanyu Madhow , "Scalable Nonlinear Multiuser Detection for mmWave Massive MIMO", SPAWC 2020.

LoS MIMO: spatial redundancy, quantization constraints, misalignment

- Ahmet Dundar Sezer and Upamanyu Madhow, "All-Digital LoS MIMO with Low-Precision Analog-to-Digital Conversion", TWC 2022.
- Ahmet Dundar Sezer, Upamanyu Madhow, and Mark J. W. Rodwell, "Spatial Oversampling for Quantized LoS MIMO", Asilomar 2021.
- Lalitha Giridhar, Maryam Eslami Rasekh, Ahmet Dundar Sezer, and Upamanyu Madhow, "Adaptive Space-Time Equalization with Spatial Oversampling for Misaligned LoS MIMO", WCNC 2022.
- Ahmet Dundar Sezer and Upamanyu Madhow, "Spatially redundant, precision-constrained transmit precoding for mmWave LoS MIMO", SPAWC 2022.

For further exploration

- Wireless Communication and Sensornets Lab (WCSL) home page: https://wcsl.ece.ucsb.edu
- NSF Giganets project (2015-2020): https://wcsl.ece.ucsb.edu/giganets
  - Papers from interdisciplinary collaborations involving hardware, signal processing and systems
  - Lectures at IISc, Bangalore summer school, 2016
  - Tutorial at ACM SigComm 2017
- ComSenTer (2018-2022): https://comsenter.engr.ucsb.edu/
  - UCSB-led center funded by DARPA and SRC
  - Pushing the limits of mm-wave and THz comm and sensing: both hardware and signal processing