Hybrid Precoding in Millimeter-Wave Massive MIMO Systems

Ph.D. Comprehensive Examination

Pranav Jha
Student ID: 40081750

Supervisor: Dr. Wei-Ping Zhu

Université Concordia University
GINA CODY SCHOOL OF ENGINEERING AND COMPUTER SCIENCE

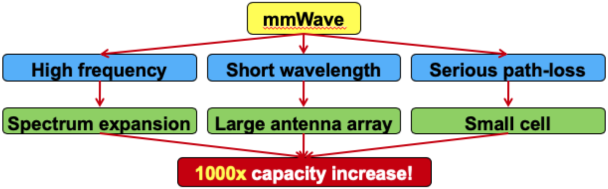
Department of Electrical and Computer Engineering

March 12, 2023

Outline

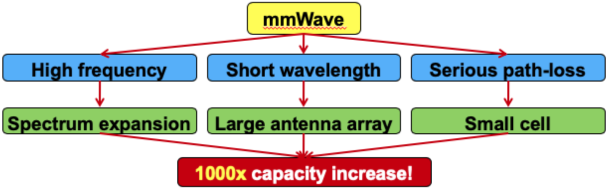
Mm-Wave Massive MIMO

Advantages



Mm-Wave Massive MIMO

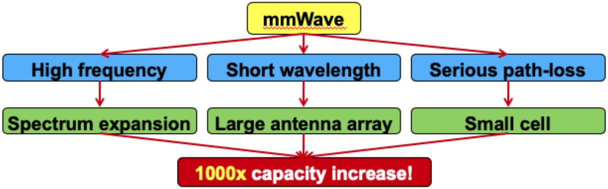
Advantages



- High frequency (30-300 GHz)
  - Larger bandwidth: 20 MHz $\rightarrow$ 2 GHz

Mm-Wave Massive MIMO

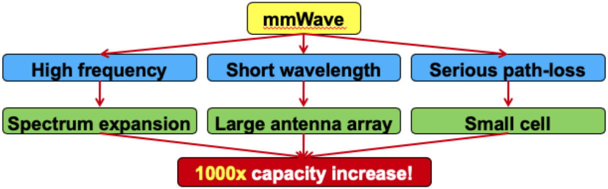
Advantages



- High frequency (30-300 GHz)
  - Larger bandwidth: 20 MHz $\rightarrow$ 2 GHz
- Short wavelength (1-10 mm)
  - Enable large antenna array (massive MIMO): At 70 GHz
    $\rightarrow$ Maximum #Antennas - 1024 and 64 (BSs and UEs)
    $\rightarrow$ Maximum #RF chains - 32 and 8 (BSs and UEs)

Mm-Wave Massive MIMO

Advantages



- High frequency (30-300 GHz)
  - Larger bandwidth: 20 MHz $\rightarrow$ 2 GHz
- Short wavelength (1-10 mm)
  - Enable large antenna array (massive MIMO): At 70 GHz
    $\rightarrow$ Maximum #Antennas - 1024 and 64 (BSs and UEs)
    $\rightarrow$ Maximum #RF chains - 32 and 8 (BSs and UEs)
- Serious path-loss and blockage
  - Massive MIMO provides sufficient gains to compensate the serious path-loss by using precoding
  - Avoid multi-cell interference, more appropriate for small cell

Precoding for Mm-Wave Massive MIMO

Traditional Precoding

- Preformed in digital domain with optimized performance
- One RF chain is required to support one transmit antenna
- Impractical in energy consumption for mm-Wave massive MIMO systems
  - 250 mW per RF chain, and 16 W for 64 antennas[1]

[1] P. V. Amadori and C. Masouros, "Low RF-Complexity Millimeter-Wave Beamspace-MIMO Systems by Beam Selection," in IEEE Transactions on Communications, vol. 63, no. 6, pp. 2212-2223, June 2015 [1]

Precoding for Mm-Wave Massive MIMO

Traditional Precoding

- Preformed in digital domain with optimized performance
- One RF chain is required to support one transmit antenna
- Impractical in energy consumption for mm-Wave massive MIMO systems
  - 250 mW per RF chain, and 16 W for 64 antennas[1]

Hybrid Analog and Digital Precoding

- Actual degree of freedom (i.e., number of users) is much smaller than number of antennas
- Divide digital precoding with large size into:
  - Digital precoding with small size
  - Analog precoding with large size (realized by phase shifter, PS)
- Significantly reduced number of RF chains
- Power-efficient, low complexity, without obvious performance loss

[1] P. V. Amadori and C. Masouros, "Low RF-Complexity Millimeter-Wave Beamspace-MIMO Systems by Beam Selection," in IEEE Transactions on Communications, vol. 63, no. 6, pp. 2212-2223, June 2015 [1]

Hybrid Precoding Architectures

Fully-connected Architecture

- RF chain is fully connected to all antennas
  - Large number of PSs (N^2M)
  - Near-optimal but energy-intensive
- Spatially sparse precoding [2]
- Codebook-based hybrid precoding [3]

[2] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014

[3] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," in IEEE Communications Magazine, vol. 52, no. 2, pp. 106-113, February 2014

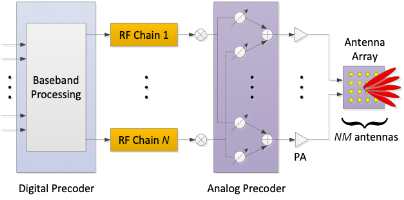
Hybrid Precoding Architectures

Fully-connected Architecture

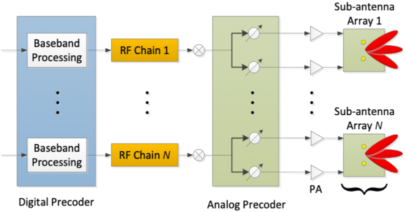
- RF chain is fully connected to all antennas
  - Large number of PSs (N^2M)
  - Near-optimal but energy-intensive
- Spatially sparse precoding [2]
- Codebook-based hybrid precoding [3]

Sub-connected Architecture

- RF chain is partially connected to a subset of antennas
  - Smaller number of PSs (NM)
  - More energy-efficient



(a)



(b)

2 3

[2] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014

[3] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," in IEEE Communications Magazine, vol. 52, no. 2, pp. 106-113, February 2014

Problem Statement

Hybrid Precoder Design

- Coupling between analog and baseband counterpart → Nonlinear
- The analog precoder rely on a phase-shifters network, which imposes a constant modulus constraint → Non-convex
- Quantization of the phase shifters → Combinatorial

Problem Statement

Hybrid Precoder Design

- Coupling between analog and baseband counterpart $\rightarrow$ Nonlinear
- The analog precoder rely on a phase-shifters network, which imposes a constant modulus constraint $\rightarrow$ Non-convex
- Quantization of the phase shifters $\rightarrow$ Combinatorial

Traditional Methods

- Based on SVD $\rightarrow$ complicated bit-allocation
- Based on GMD $\rightarrow$ brings great challenge in addressing non-convex constraints and exploiting sparsity statistics
- High computational complexity
- Traditional low-complexity schemes realization cost $\rightarrow$ Hybrid precoding performance degradation

Problem Statement

Hybrid Precoder Design

- Coupling between analog and baseband counterpart → Nonlinear
- The analog precoder rely on a phase-shifters network, which imposes a constant modulus constraint → Non-convex
- Quantization of the phase shifters → Combinatorial

Traditional Methods

- Based on SVD → complicated bit-allocation
- Based on GMD → brings great challenge in addressing non-convex constraints and exploiting sparsity statistics
- High computational complexity
- Traditional low-complexity schemes realization cost → Hybrid precoding performance degradation

Aim

- To develop a low-complex energy-efficient solution for the hybrid precoder design problem
- To consider the application of deep learning (DL) to develop optimal hybrid precoder

Problem and Methodology I

Problem Formulation [4]

- System Model

$$\mathbf{y} = \sqrt{\rho}\mathbf{HADs} + \mathbf{n} = \sqrt{\rho}\mathbf{HPs} + \mathbf{n}$$

- Total achievable rate

$$R = \log_2\left(\left|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{HPP}^H\mathbf{H}^H\right|\right)$$

[4] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath,"Energy-efficient hybrid analog and digital precoding for mm-wave mimo systems with large antenna arrays", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016 [4]

Problem and Methodology I

Problem Formulation [4]

- System Model

$$\mathbf{y} = \sqrt{\rho}\mathbf{HADs} + \mathbf{n} = \sqrt{\rho}\mathbf{HPs} + \mathbf{n}$$

- Total achievable rate

$$R = \log_2\left(\left|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{HPP}^H\mathbf{H}^H\right|\right)$$

- Target
  - Jointly design **A** and **D** to maximize the achievable rate

[4] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath,"Energy-efficient hybrid analog and digital precoding for mm-wave mimo systems with large antenna arrays", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016 [4]

Problem and Methodology I

Problem Formulation [4]

- System Model

$$\mathbf{y} = \sqrt{\rho}\mathbf{HADs} + \mathbf{n} = \sqrt{\rho}\mathbf{HPs} + \mathbf{n}$$

- Total achievable rate

$$R = \log_2\left(\left|\mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{HPP}^H\mathbf{H}^H\right|\right)$$

- Target
  - Jointly design **A** and **D** to maximize the achievable rate
- Three non-convex constraints
  - Structure constraint: $\mathbf{P} = \mathbf{AD} = \mathrm{diag}\{\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_N\} \cdot \mathrm{diag}\{d_1, \ldots, d_N\}$
  - Amplitude constraint: The amplitude of non-zero elements of the analog precoding matrix **A** is fixed to $1/\sqrt{M}$
  - Power constraint: $\|\mathbf{P}\|_F \leq N$

[4] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath,"Energy-efficient hybrid analog and digital precoding for mm-wave mimo systems with large antenna arrays", *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016 [4]

SIC-based Hybrid Precoding

- The total rate R can be decomposed as

$$R = \sum_{n=1}^{N} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H\mathbf{H}^H\mathbf{T}_{n-1}^{-1}\mathbf{H}\mathbf{p}_n\right)$$

where $\mathbf{p}_n$ be the N-th column of $\mathbf{P}$, $\mathbf{T}_n = \mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H}\mathbf{P}_n\mathbf{P}_n^H\mathbf{H}^H$ and $\mathbf{T}_0 = \mathbf{I}_N$

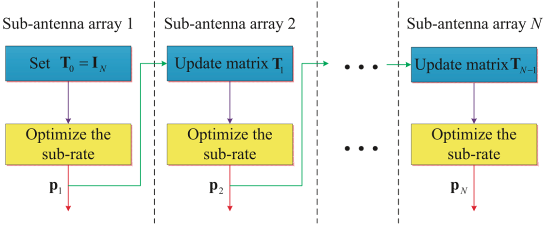
SIC-based Hybrid Precoding

- The total rate R can be decomposed as

$$R = \sum_{n=1}^{N} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H\mathbf{H}^H\mathbf{T}_{n-1}^{-1}\mathbf{H}\mathbf{p}_n\right)$$

where $\mathbf{p}_n$ be the N-th column of $\mathbf{P}$, $\mathbf{T}_n = \mathbf{I}_K + \frac{\rho}{N\sigma^2}\mathbf{H}\mathbf{P}_n\mathbf{P}_n^H\mathbf{H}^H$ and $\mathbf{T}_0 = \mathbf{I}_N$

- SIC-based hybrid precoding
  - Total rate → sub-rate of sub-antenna array
  - Optimize the sub-rate of each sub-antenna array one by one by exploiting the concept of SIC for multi-user detection



Solution to the sub-rate optimization problem

- Optimize achievable rate of the nth sub-antenna array

$$\mathbf{p}_n^{\text{opt}} = \underset{\mathbf{p}_n \in \mathscr{F}}{\arg\max} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H \mathbf{G}_{n-1}\mathbf{p}_n\right)$$

where $G_{n-1} = \mathbf{H}^H \mathbf{T}_{n-1}^{-1}\mathbf{H}$, $\mathscr{F}$ is the set of all feasible vectors which satisfy all the three constraints

Solution to the sub-rate optimization problem

- Optimize achievable rate of the nth sub-antenna array

$$\mathbf{p}_n^{\text{opt}} = \arg\max_{\mathbf{p}_n \in \mathscr{F}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H \mathbf{G}_{n-1}\mathbf{p}_n\right)$$

where $G_{n-1} = \mathbf{H}^H \mathbf{T}_{n-1}^{-1}\mathbf{H}$, $\mathscr{F}$ is the set of all feasible vectors which satisfy all the three constraints

- Considering non-zero elements, it is equivalent to a simplified problem as

$$\bar{\mathbf{p}}_n^{\text{opt}} = \arg\max_{\bar{\mathbf{p}}_n \in \bar{\mathscr{F}}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\bar{\mathbf{p}}_n^H \bar{\mathbf{G}}_{n-1}\bar{\mathbf{p}}_n\right)$$

where $\bar{\mathbf{G}}_{n-1} = \mathbf{R}\mathbf{G}_{n-1}\mathbf{R}^H$, $\mathbf{R} = [\mathbf{0}_{M\times M(n-1)}\ \mathbf{I}_M\ \mathbf{0}_{M\times M(N-n)}]$ is the corresponding selection matrix

Solution to the sub-rate optimization problem

- Optimize achievable rate of the nth sub-antenna array

$$\mathbf{p}_n^{\text{opt}} = \arg\max_{\mathbf{p}_n \in \mathscr{F}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H \mathbf{G}_{n-1}\mathbf{p}_n\right)$$

  where $G_{n-1} = \mathbf{H}^H \mathbf{T}_{n-1}^{-1}\mathbf{H}$, $\mathscr{F}$ is the set of all feasible vectors which satisfy all the three constraints

- Considering non-zero elements, it is equivalent to a simplified problem as

$$\bar{\mathbf{p}}_n^{\text{opt}} = \arg\max_{\bar{\mathbf{p}}_n \in \bar{\mathscr{F}}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\bar{\mathbf{p}}_n^H \bar{\mathbf{G}}_{n-1}\bar{\mathbf{p}}_n\right)$$

  where $\bar{\mathbf{G}}_{n-1} = \mathbf{R}\mathbf{G}_{n-1}\mathbf{R}^H$, $\mathbf{R} = \left[\mathbf{0}_{M\times M(n-1)}\ \mathbf{I}_M\ \mathbf{0}_{M\times M(N-n)}\right]$ is the corresponding selection matrix

- Simplify the optimization problem $\bar{\mathbf{p}}_n^{\text{opt}} = \arg\min_{\bar{\mathbf{p}}_n \in \bar{\mathscr{F}}} \|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2$, where $\mathbf{v}_1$ is the first right singular vector of $\bar{\mathbf{G}}_{n-1}$

Solution to the sub-rate optimization problem

- Optimize achievable rate of the nth sub-antenna array

$$\mathbf{p}_n^{\mathrm{opt}} = \arg\max_{\mathbf{p}_n \in \mathscr{F}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\mathbf{p}_n^H \mathbf{G}_{n-1}\mathbf{p}_n\right)$$

  where $G_{n-1} = \mathbf{H}^H \mathbf{T}_{n-1}^{-1}\mathbf{H}$, $\mathscr{F}$ is the set of all feasible vectors which satisfy all the three constraints

- Considering non-zero elements, it is equivalent to a simplified problem as

$$\bar{\mathbf{p}}_n^{\mathrm{opt}} = \arg\max_{\bar{\mathbf{p}}_n \in \bar{\mathscr{F}}} \log_2\left(1 + \frac{\rho}{N\sigma^2}\bar{\mathbf{p}}_n^H \bar{\mathbf{G}}_{n-1}\bar{\mathbf{p}}_n\right)$$

  where $\bar{\mathbf{G}}_{n-1} = \mathbf{R}\mathbf{G}_{n-1}\mathbf{R}^H$, $\mathbf{R} = \left[\mathbf{0}_{M\times M(n-1)}\ \mathbf{I}_M\ \mathbf{0}_{M\times M(N-n)}\right]$ is the corresponding selection matrix

- Simplify the optimization problem $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \arg\min_{\bar{\mathbf{p}}_n \in \bar{\mathscr{F}}} \|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2$, where $\mathbf{v}_1$ is the first right singular vector of $\bar{\mathbf{G}}_{n-1}$
- Find a feasible precoding vector $\bar{\mathbf{p}}_n$, sufficiently close to $\mathbf{v}_1$, to maximize the achievable sub-rate

Design of Analog and Digital Precoder

- **Problem**
  - As we have $\bar{\mathbf{p}}_n = d_n \bar{\mathbf{a}}_n$, $\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2$ equals to

$$\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2 = \left(d_n - \mathrm{Re}\left(\mathbf{v}_1^H \bar{\mathbf{a}}_n\right)\right)^2 + \left(1 - \left[\mathrm{Re}\left(\mathbf{v}_1^H \bar{\mathbf{a}}_n\right)\right]^2\right)$$

Design of Analog and Digital Precoder

Problem

- As we have $\bar{\mathbf{p}}_n = d_n \bar{\mathbf{a}}_n$, $\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2$ equals to

$$\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2 = (d_n - \mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n))^2 + (1 - [\mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n)]^2)$$

Solution

- Analog precoder: $\bar{\mathbf{a}}_n^{\mathrm{opt}} = \frac{1}{\sqrt{M}} e^{j\,\mathrm{angle}(\mathbf{v}_1)}$
- Digital precoder: $d_n^{\mathrm{opt}} = \mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n) = \frac{1}{\sqrt{M}} \mathrm{Re}(\mathbf{v}_1^H e^{j\,\mathrm{angle}(\mathbf{v}_1)}) = \frac{\|\mathbf{v}_1\|_1}{\sqrt{M}}$
- Hybrid precoder: $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M} \|\mathbf{v}_1\|_1 e^{j\,\mathrm{angle}(\mathbf{v}_1)}$
- All the three constraints are satisfied

Design of Analog and Digital Precoder

Problem

- As we have $\bar{\mathbf{p}}_n = d_n \bar{\mathbf{a}}_n$, $\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2$ equals to

$$\|\mathbf{v}_1 - \bar{\mathbf{p}}_n\|_2^2 = (d_n - \mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n))^2 + (1 - [\mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n)]^2)$$

Solution

- Analog precoder: $\bar{\mathbf{a}}_n^{\mathrm{opt}} = \frac{1}{\sqrt{M}} e^{j\,\mathrm{angle}(\mathbf{v}_1)}$
- Digital precoder: $d_n^{\mathrm{opt}} = \mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n) = \frac{1}{\sqrt{M}} \mathrm{Re}(\mathbf{v}_1^H e^{j\,\mathrm{angle}(\mathbf{v}_1)}) = \frac{\|\mathbf{v}_1\|_1}{\sqrt{M}}$
- Hybrid precoder: $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M} \|\mathbf{v}_1\|_1 e^{j\,\mathrm{angle}(\mathbf{v}_1)}$
- All the three constraints are satisfied

Summary

- SVD of $\bar{\mathbf{G}}_{n-1}$ to obtain $\mathbf{v}_1$
- Compute $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M} \|\mathbf{v}_1\|_1 e^{j\,\mathrm{angle}(\mathbf{v}_1)}$ for the n-th sub-antenna array
- Update $\bar{\mathbf{G}}_n$ for the $(n+1)$-th sub-antenna array

Low-Complexity Solution

- Computation of $\mathbf{v}_1$
  - Only the first right singular vector of $\bar{\mathbf{G}}_{n-1}$ is required
  - Realized by power iteration algorithm with complexity $O(M^2)$

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

Low-Complexity Solution

- Computation of $\mathbf{v}_1$
  - Only the first right singular vector of $\bar{\mathbf{G}}_{n-1}$ is required
  - Realized by power iteration algorithm with complexity $O(M^2)$
- Acquire the optimal hybrid precoder
  - The complexity is only $O(M^2)$ to obtain $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M}\|\mathbf{v}_1\|_1 e^{j\mathrm{angle}(\mathbf{v}_1)}$

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

Low-Complexity Solution

- Computation of $\mathbf{v}_1$
  - Only the first right singular vector of $\bar{\mathbf{G}}_{n-1}$ is required
  - Realized by power iteration algorithm with complexity $O(M^2)$
- Acquire the optimal hybrid precoder
  - The complexity is only $O(M^2)$ to obtain $\bar{\mathbf{p}}_n^{\text{opt}} = \frac{1}{M}\|\mathbf{v}_1\|_1 e^{j\text{angle}(\mathbf{v}_1)}$
- Update $\bar{\mathbf{G}}_n$
  - The calculation can be simplified as

$$\bar{\mathbf{G}}_n \approx \bar{\mathbf{G}}_{n-1} - \frac{\frac{\rho}{N\sigma^2}\Sigma_1^2 \mathbf{v}_1 \mathbf{v}_1^H}{1 + \frac{\rho}{N\sigma^2}\Sigma_1}$$

    where Σ_1 is the largest singular value of $\bar{\mathbf{G}}_{n-1}$
  - Corresponding complexity is $O(M^2)$

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

Low-Complexity Solution

- Computation of $\mathbf{v}_1$
  - Only the first right singular vector of $\bar{\mathbf{G}}_{n-1}$ is required
  - Realized by power iteration algorithm with complexity $O(M^2)$
- Acquire the optimal hybrid precoder
  - The complexity is only $O(M^2)$ to obtain $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M}\|\mathbf{v}_1\|_1 e^{j\mathrm{angle}(\mathbf{v}_1)}$
- Update $\bar{\mathbf{G}}_n$
  - The calculation can be simplified as

$$\bar{\mathbf{G}}_n \approx \bar{\mathbf{G}}_{n-1} - \frac{\frac{\rho}{N\sigma^2}\Sigma_1^2 \mathbf{v}_1 \mathbf{v}_1^H}{1 + \frac{\rho}{N\sigma^2}\Sigma_1}$$

    where Σ_1 is the largest singular value of $\bar{\mathbf{G}}_{n-1}$
  - Corresponding complexity is $O(M^2)$
- Total complexity $O(M^2(NS + K))$
  - Only 10% of SVD-based spatially sparse precoding[5]

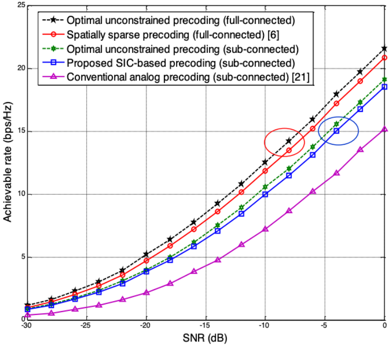
[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

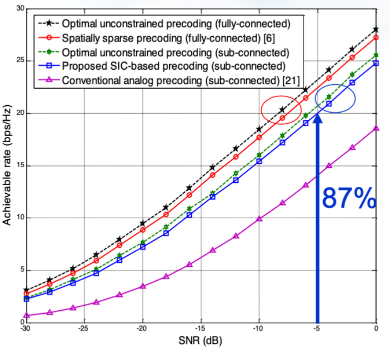
Achievable Rate

- Simulation Setup
  - Antenna: (1) $NM \times K = 64 \times 16$ (2) $NM \times K = 128 \times 32$
  - RF chains: (1) $N = 8$ (2) $N = 16$
  - Channel: Geometric Saleh-Valenzuela model

Achievable Rate

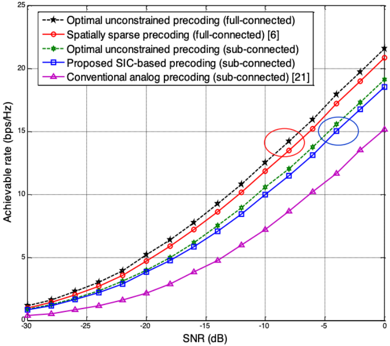
- Simulation Setup
  - Antenna: (1) $NM \times K = 64 \times 16$ (2) $NM \times K = 128 \times 32$
  - RF chains: (1) $N = 8$ (2) $N = 16$
  - Channel: Geometric Saleh-Valenzuela model

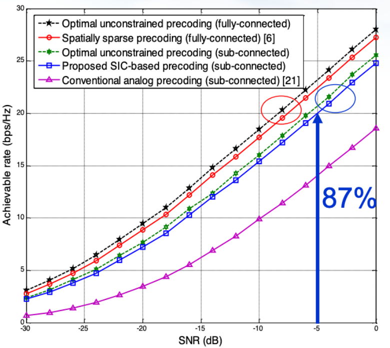




Achievable Rate

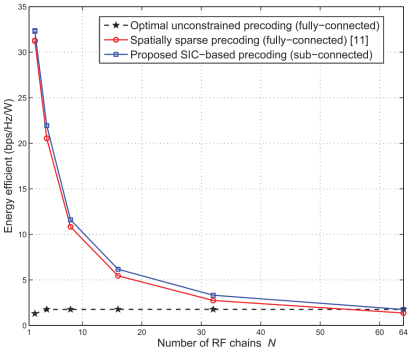
- Simulation Setup
  - Antenna: (1) $NM \times K = 64 \times 16$ (2) $NM \times K = 128 \times 32$
  - RF chains: (1) $N = 8$ (2) $N = 16$
  - Channel: Geometric Saleh-Valenzuela model





SIC-based hybrid precoding is near-optimal!

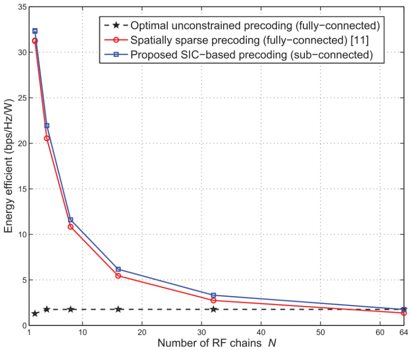
Energy Efficiency



- Both the SIC-based precoding and the spatially sparse precoding[6] can achieve higher energy-efficiency than the optimal unconstrained precoding.

[6] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

Energy Efficiency



- Both the SIC-based precoding and the spatially sparse precoding[6] can achieve higher energy-efficiency than the optimal unconstrained precoding.

SIC-based hybrid precoding is more energy-efficient!

[6] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014 [2]

Problem and Methodology II

Problem Formulation [7]

- System Model

$$\mathbf{x} = \mathbf{D}\mathbf{s}$$

$$\mathbf{y} = \mathbf{B}^H\mathbf{H}\mathbf{x} + \mathbf{B}^H\mathbf{n}$$

- Here, $\mathbf{D} = \mathbf{D}_A\mathbf{D}_D$ and $\mathbf{B} = \mathbf{B}_A\mathbf{B}_D$

[7] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive mimo for hybrid precoding", *IEEE Transactions on Vehicular Technology*, vol. 68,no. 3, pp. 3027–3032, 2019 [5]

Problem and Methodology II

Problem Formulation [7]

- System Model

$$\mathbf{x} = \mathbf{D}\mathbf{s}$$

$$\mathbf{y} = \mathbf{B}^H\mathbf{H}\mathbf{x} + \mathbf{B}^H\mathbf{n}$$

- Here, $\mathbf{D} = \mathbf{D}_A\mathbf{D}_D$ and $\mathbf{B} = \mathbf{B}_A\mathbf{B}_D$
- Transmit power constraint:

$$\mathrm{tr}\{\mathbf{D}\mathbf{D}^H\} \leq N_s$$

[7] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive mimo for hybrid precoding", *IEEE Transactions on Vehicular Technology*, vol. 68,no. 3, pp. 3027–3032, 2019 [5]

Problem and Methodology II

Problem Formulation [7]

- System Model

$$\mathbf{x} = \mathbf{D}\mathbf{s}$$
$$\mathbf{y} = \mathbf{B}^H\mathbf{H}\mathbf{x} + \mathbf{B}^H\mathbf{n}$$

- Here, $\mathbf{D} = \mathbf{D}_A\mathbf{D}_D$ and $\mathbf{B} = \mathbf{B}_A\mathbf{B}_D$
- Transmit power constraint:

$$\mathrm{tr}\{\mathbf{D}\mathbf{D}^H\} \leq N_s$$

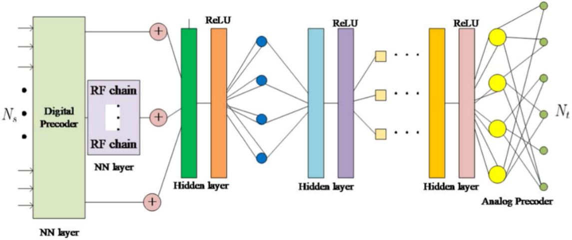
- Constraints on the elements of $\mathbf{D}_A$ and $\mathbf{B}_A$:

$$|\{\mathbf{D}_A\}_{i,j}| = \frac{1}{\sqrt{N_t}}, \quad |\{\mathbf{B}_A\}_{i,j}| = \frac{1}{\sqrt{N_r}}$$

[7] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive mimo for hybrid precoding", *IEEE Transactions on Vehicular Technology*, vol. 68,no. 3, pp. 3027–3032, 2019 [5]

Deep-Neural Network

DNN Framework



- Input Layer: 128 units
- Hidden Layers: 400 units and 256 units
- Noise Layer: 200 units
- Hidden Layers: 128 units and 64 units
- Output Layer: Deployed to generate output signals
- Input and hidden layers activation: ReLU
- Output layer activation:

$$f(\mathbf{s}) = \min(\max(\mathbf{s}, 0), N_s)$$

Learning Policy

- Decompose $\mathbf{H}$ using GMD as

$$\mathbf{H} = \mathbf{WQR}^H = [\mathbf{W}_1, \mathbf{W}_2] \begin{bmatrix} \mathbf{Q}_1 & * \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^H \\ \mathbf{R}_2^H \end{bmatrix}$$

- The largest N_s singular values: $q_{i,i} = (\delta_1, \delta_2, ..., \delta_{N_s})^{\frac{1}{N_s}} \in \bar{\mathbf{q}} \; \forall i$

Learning Policy

- Decompose $\mathbf{H}$ using GMD as

$$\mathbf{H} = \mathbf{W}\mathbf{Q}\mathbf{R}^H = [\mathbf{W}_1, \mathbf{W}_2] \begin{bmatrix} \mathbf{Q}_1 & * \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^H \\ \mathbf{R}_2^H \end{bmatrix}$$

- The largest N_s singular values: $q_{i,i} = (\delta_1, \delta_2, \ldots, \delta_{N_s})^{\frac{1}{N_s}} \in \bar{\mathbf{q}} \ \forall i$
- The received signal is given as

$$\begin{aligned} \mathbf{y} &= \mathbf{W}_1^H \mathbf{H} \mathbf{R}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \\ &= \mathbf{Q}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \end{aligned}$$

Learning Policy

- Decompose $\mathbf{H}$ using GMD as

$$\mathbf{H} = \mathbf{W}\mathbf{Q}\mathbf{R}^H = [\mathbf{W}_1, \mathbf{W}_2] \begin{bmatrix} \mathbf{Q}_1 & * \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^H \\ \mathbf{R}_2^H \end{bmatrix}$$

- The largest N_s singular values: $q_{i,i} = (\delta_1, \delta_2, ..., \delta_{N_s})^{\frac{1}{N_s}} \in \bar{\mathbf{q}}\ \forall i$
- The received signal is given as

$$\begin{aligned} \mathbf{y} &= \mathbf{W}_1^H \mathbf{H} \mathbf{R}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \\ &= \mathbf{Q}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \end{aligned}$$

- The loss function is given as

$$loss = \|\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D\|_F = \sqrt{\Sigma_{i=1}^{\min\{N_t, N_s\}} \delta_i^2 (\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)}$$

where $\mathbf{R}_A$ and $\mathbf{R}_D$ are the GMD-based analog and digital precoder, respectively and $\delta_i(\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)$ implies the singular values of matrix $(\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)$

Learning Policy

- Decompose $\mathbf{H}$ using GMD as

$$\mathbf{H} = \mathbf{W}\mathbf{Q}\mathbf{R}^H = [\mathbf{W}_1, \mathbf{W}_2] \begin{bmatrix} \mathbf{Q}_1 & * \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1^H \\ \mathbf{R}_2^H \end{bmatrix}$$

- The largest N_s singular values: $q_{i,i} = (\delta_1, \delta_2, \ldots, \delta_{N_s})^{\frac{1}{N_s}} \in \bar{\mathbf{q}}\ \forall i$
- The received signal is given as

$$\begin{aligned} \mathbf{y} &= \mathbf{W}_1^H \mathbf{H} \mathbf{R}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \\ &= \mathbf{Q}_1 \mathbf{s} + \mathbf{W}_1^H \mathbf{n} \end{aligned}$$

- The loss function is given as

$$loss = \|\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D\|_F = \sqrt{\Sigma_{i=1}^{\min\{N_t, N_s\}} \delta_i^2 (\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)}$$

  where $\mathbf{R}_A$ and $\mathbf{R}_D$ are the GMD-based analog and digital precoder, respectively and $\delta_i(\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)$ implies the singular values of matrix $(\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)$
- These constraints need to be satisfied:

$$|\{\mathbf{R}_A\}_{i,j}| = \frac{1}{\sqrt{N_t}}, \quad \mathrm{tr}(\mathbf{R}_A \mathbf{R}_D \mathbf{R}_D^H \mathbf{R}_A^H) \leq N_s$$

Université Concordia University

Autoencoder

- To construct an autoencoder, the deep-neaural network (DNN) framework is employed as

$$\mathbf{R}_1 = f(\mathbf{R}_A \mathbf{R}_D; \Omega)$$

where Ω is the dataset of the samples and $f(\cdot)$ is the mapping relation.

Autoencoder

- To construct an autoencoder, the deep-neaural network (DNN) framework is employed as

$$\mathbf{R}_1 = f(\mathbf{R}_A \mathbf{R}_D; \Omega)$$

  where Ω is the dataset of the samples and $f(\cdot)$ is the mapping relation.

- Training Procedure:
  - $\mathbf{R}_A$, $\mathbf{R}_D$ - Empty matrices
  - The DNN is trained with the input data sequences
  - Update $\mathbf{R}_A$, $\mathbf{R}_D$
  - AoA θ_p^r, AoD θ_p^t - generated randomly
  - Bias between $\mathbf{R}_1$ and $\mathbf{R}_A \mathbf{R}_D$ - from the output layer
- The training set Ω is obtained

Stochastic Gradient Descent

- SGD algorithm with momentum to process the loss function

$$\mathbf{R}_A^{j+1} = \mathbf{R}_A^j + v \tag{1}$$

$$\mathbf{R}_D^{j+1} = \mathbf{R}_D^j + v \tag{2}$$

where v is the velocity for facilitating the gradient element and j represents the iteration

Stochastic Gradient Descent

- SGD algorithm with momentum to process the loss function

$$\mathbf{R}_A^{j+1} = \mathbf{R}_A^j + v \tag{1}$$

$$\mathbf{R}_D^{j+1} = \mathbf{R}_D^j + v \tag{2}$$

  where v is the velocity for facilitating the gradient element and j represents the iteration

- The update procedure of v can be given as

$$\begin{aligned} v &= \alpha v - \varepsilon g \\ &= \alpha v - \varepsilon \frac{1}{N} \bigtriangledown_{\mathbf{R}_A, \mathbf{R}_D} \sqrt{\Sigma_{i=1}^{\min\{N_t, N_s\}} \delta_i^2 (\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)}, \end{aligned} \tag{3}$$

  where α is the momentum parameter and ε is the learning rate. Also, g and N are the gradient element and the number of samples, respectively

Stochastic Gradient Descent

- SGD algorithm with momentum to process the loss function

$$\mathbf{R}_A^{j+1} = \mathbf{R}_A^{j} + v \tag{1}$$

$$\mathbf{R}_D^{j+1} = \mathbf{R}_D^{j} + v \tag{2}$$

where v is the velocity for facilitating the gradient element and j represents the iteration

- The update procedure of v can be given as

$$\begin{aligned} v &= \alpha v - \varepsilon g \\ &= \alpha v - \varepsilon \frac{1}{N} \bigtriangledown_{\mathbf{R}_A, \mathbf{R}_D} \sqrt{\Sigma_{i=1}^{\min\{N_t, N_s\}} \delta_i^2 (\mathbf{R}_1 - \mathbf{R}_A \mathbf{R}_D)}, \end{aligned} \tag{3}$$

where α is the momentum parameter and ε is the learning rate. Also, g and N are the gradient element and the number of samples, respectively

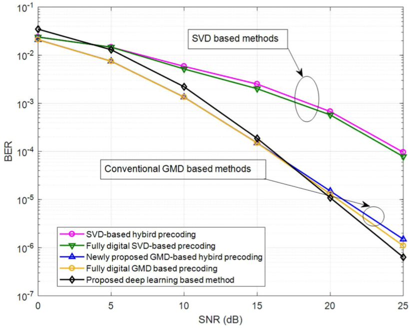
- DL-based scheme has the lowest complexity, where the number of multiplications and divisions are $O(N_s N_t^2)$ and $O(L^2)$, respectively

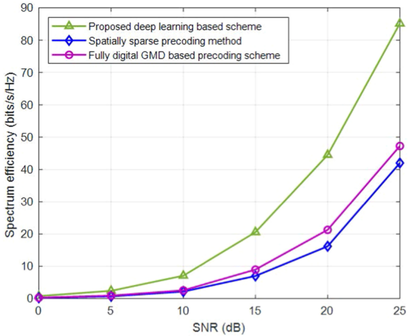
Numerical Results

- **Simulation Setup**
  - Channel: Geometric Saleh-Valenzuela (SV) model with P=3 at 28 GHz
  - Angles are generated randomly in $\{-\pi/2, \pi/2\}$
  - To generate channel measurements: Environment simulator
  - To construct and process the DNN framework: *Keras*
  - Learning rate 0.001, momentum of 0.85

Numerical Results

- **Simulation Setup**
  - Channel: Geometric Saleh-Valenzuela (SV) model with P=3 at 28 GHz
  - Angles are generated randomly in $\{-\pi/2, \pi/2\}$
  - To generate channel measurements: Environment simulator
  - To construct and process the DNN framework: *Keras*
  - Learning rate 0.001, momentum of 0.85





DL-based scheme is superior!

Critical Review and Future Work

- **Critical Points**
  - Single user scenario

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel
  - Further, SGD with momentum is good for computation in less time, but not reliable

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel
  - Further, SGD with momentum is good for computation in less time, but not reliable

- **Future Work**
  - Design of optimal hybrid precoders for multi-user systems
  - To address the high energy consumption problem, the usage of low-resolution PSs [6, 7] and low-resolution DACs/ADCs [8, 9, 10]

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel
  - Further, SGD with momentum is good for computation in less time, but not reliable

- **Future Work**
  - Design of optimal hybrid precoders for multi-user systems
  - To address the high energy consumption problem, the usage of low-resolution PSs [6, 7] and low-resolution DACs/ADCs [8, 9, 10]
  - An appropriate trade-off between energy efficiency and system sum-rate for a wideband mm-Wave channel is still an open research problem

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel
  - Further, SGD with momentum is good for computation in less time, but not reliable

- **Future Work**
  - Design of optimal hybrid precoders for multi-user systems
  - To address the high energy consumption problem, the usage of low-resolution PSs [6, 7] and low-resolution DACs/ADCs [8, 9, 10]
  - An appropriate trade-off between energy efficiency and system sum-rate for a wideband mm-Wave channel is still an open research problem
  - To solve the design of finding the best hybrid precoder the idea of cross-entropy optimization (CEO) in deep learning solutions [11]

Critical Review and Future Work

- **Critical Points**
  - Single user scenario
  - Infinite or high-resolution phase shifters (PSs)
  - The number of RF chains with high-resolution DACs/ADCs
  - The channel model as a narrowband mm-Wave channel
  - Further, SGD with momentum is good for computation in less time, but not reliable

- **Future Work**
  - Design of optimal hybrid precoders for multi-user systems
  - To address the high energy consumption problem, the usage of low-resolution PSs [6, 7] and low-resolution DACs/ADCs [8, 9, 10]
  - An appropriate trade-off between energy efficiency and system sum-rate for a wideband mm-Wave channel is still an open research problem
  - To solve the design of finding the best hybrid precoder the idea of cross-entropy optimization (CEO) in deep learning solutions [11]

It could be more practical and of great benefit to leverage deep learning (DL) in mm-Wave massive MIMO systems!

References I

- P. V. Amadori and C. Masouros, “Low rf-complexity millimeter-wave beamspace-mimo systems by beam selection,” *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2212–2223, 2015.
- O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave mimo systems,” *IEEE transactions on wireless communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, “Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results,” *IEEE communications magazine*, vol. 52, no. 2, pp. 106–113, 2014.
- X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, “Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.
- H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, “Deep-learning-based millimeter-wave massive mimo for hybrid precoding,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3027–3032, 2019.
- X. Gao, L. Dai, Y. Sun, S. Han, and I. Chih-Lin, “Machine learning inspired energy-efficient hybrid precoding for mmwave massive mimo systems,” *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2017.

References II

- S. Gao, Y. Dong, C. Chen, and Y. Jin, “Hierarchical beam selection in mmwave multiuser mimo systems with one-bit analog phase shifters,” *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, pp. 1–5, 2016.
- Y. Dong and L. Qiu, “Spectral efficiency of massive mimo systems with low-resolution adcs and mmse receiver,” *IEEE Communications Letters*, vol. 21, no. 8, pp. 1771–1774, 2017.
- J. Zhang, L. Dai, S. Sun, and Z. Wang, “On the spectral efficiency of massive mimo systems with low-resolution adcs,” *IEEE Communications Letters*, vol. 20, no. 5, pp. 842–845, 2016.
- C. Kong, C. Zhong, S. Jin, S. Yang, H. Lin, and Z. Zhang, “Full-duplex massive mimo relaying systems with low-resolution adcs,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5033–5047, 2017.
- R. Y. Rubinstein and D. P. Kroese, “Simulation and the monte carlo method,” *John Wiley & Sons*, vol. 10, 2016.

Thank You