A Review on Precoding Techniques for mm-Wave Massive MIMO Wireless Systems

Islam Osama [1,2], Mohamed Rihan [1,3], Mohamed Elhefnawy [2], Sami Eldolil [1] and Hend Abd El-Azem Malhat [1]

[1] Department of Electronics and Communication Engineering Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt
[2] Department of Electrical Engineering, Faculty of Engineering, October 6 University, Giza, Egypt
[3] School of Mathematics and Statistics, Shenzhen University, Shenzhen, China.

Abstract: The growing demands for high data rate wireless connectivity shed lights on the fact that appropriate spectrum regions need to be investigated so that the expected future needs will be satisfied. With this in mind, the research community has shown considerable interest in millimeter-wave (mm-wave) communication. Generally, hybrid transceivers combining the analog phase shifter and the RF chains with digital signal processing (DSP) systems are used for MIMO communication in the fifth generation (5G) wireless networks. This paper presents a survey for different precoding or beamforming techniques that have been proposed in the literature. These beamforming techniques are mainly classified based on their hardware structure into analog and digital beamforming. To reduce the hardware complexity and power consumption, the hybrid precoding techniques that combine analog and digital beamforming can be implemented for mm-wave massive MIMO wireless systems. The performance of the most common hybrid precoding algorithms has been investigated in this paper.

Keywords: mm-wave, massive MIMO, hybrid precoding, fully and partially connected hybrid precoder, transmission optimization.

1. Introduction

Millimeter radio waves have recently received much attention in academic and industrial research due to their wide spread use in military and limited space communication applications. This interest is due to the use of their spectrum frequency features, such as improved transmission capacity to several gigabits per second. For example, this technology can allow a transmission speed of up to 1 Gbps at 60 GHz because of its high transmission steering, which reduces the interference between connections [1]. Some studies predicted that the flow of data from smartphones, tablets, laptops, and many devices dealing with wireless data, would increase significantly in addition to the increase in the number of mobile devices, and this will require new technologies to meet such needs [2, 3]. Actually, this is due to the problem of the microwave spectrum scarcity, which extends to about 600 MHz, distributed in the frequency band, (700 MHz - 2600 MHz) and is currently in full use, researchers try to find a solution to this problem by trending to millimeter-wave frequencies which allow access to tens of spectral space provided in addition to the microwave spectrum [4, 5]. Millimeter communication technologies are of great importance in developing future generation's systems of wireless and cellular communications. Unfortunately, some challenges hinder the use of these technologies, such as:
The millimeter-wave signals are badly affected by a higher loss in propagation compared to other signals that operate at lower frequencies [1, 6, 7]. Figure 1 shows an increase in the loss as the distance between the transmitter and the receiver is increased or by increasing the frequency in case of an indoor environment. This means that increasing the transmission

frequencies of the current cellular systems to 30 GHz or higher frequencies will lead to an increase in path loss [8]. The low frequency signals can pass more smoothly through obstructions, unlike milli-meter signals that cannot penetrate most solid materials well, due to their short wavelength, which makes their ability to penetrate obstacles is minimal [9]. Also, the loss due to rain, snowfall, fog, water vapor during atmospheric closures, and oxygen absorption properties of (mm-wave) are among the factors hindering the spread of these waves [1, 10].
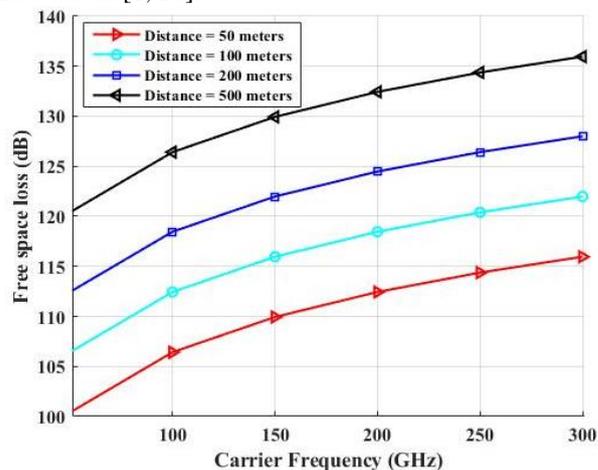

Figure 1. The effect of frequency on the free-space loss.

The user's movement will make important changes in the state of the channel, as with the movement of the users, the channel state will be varied due to the change in the distance between the TX to the RX. Assuming that the transmission process takes place through a line of sight between the BS and the user, then the capacity can be calculated, and it can be noted that this capacity will decrease significantly when the distance between the BS and the user is increased [1]. The Doppler effect on the wireless channel also based on the carrier frequency of the wave and the speed of its traffic. Assuming that the communication environment is full of obstacles, then the signal will scatter, and if the used antennas are omnidirectional, the movement at speed between 3 Km/h and 350 Km/h can make a maximum Doppler deviation of 10 Hz to 20 KHz, if the operating frequency in the range from 3-60 GHz [10].
The design of cost-effective and energy-efficient devices for use in high-frequency (mm-wave) communication systems with large bandwidths faces several technical challenges, including the design of circuits, elements, antennas, and phase shifters. It is known that the current radio equipment operating at these high frequencies is of low efficiency, as the efficiency of the current power amplifiers is estimated at less than 10%

at millimeter frequencies. Also, the high transmit power needed to overcome the loss of these systems and their large frequency bandwidth will cause non-linear distortions in the output of the power amplifiers, in addition to causing a phase noise [1, 11]. Also, the feeding design for multiple antenna arrays appropriately is another challenge in front of this technology, as each element of the antenna is related to its radio chain (RF chain). Therefore the total cost and the total energy consumption of the transmitter and receiver will increase according to the massive number of antennas [6]. To illustrate this idea, an example of measuring the power consumption levels for designing an analog-to-digital converter for 16 antennas, and we find that it consumes more than 250 mW. This means the need to use effective radio power amplifiers, and this combination is required when using phased array antenna systems [8]. On the other side, it is well understood that the cost of implementing analog to digital and digital to analog converters that are able to support the transmission of several megabytes per second can be a prohibitive cost by using the current techniques of millimeter-wave systems [7]. The remaining part of this paper is organized as follows: The analog and fully digital beamforming or precoding techniques are discussed in Section 2. Section 3 presents the hybrid precoding techniques; Section 4 discusses the Challenges of hybrid precoding and future research direction. Finally, Section 5 provides the conclusions.

2. The Analog and Fully Digital Beamforming

Antenna technologies play a critical role in optimizing network capacity. It all began with segmented antennas with a beam angle of 60 deg or 120 deg and acted as a single cell. The capacity in GSM can be tripled by using 120 deg antennas. Adaptive antenna arrays use narrow beams to maximize spatial multiplexing. Smart antennas are adaptive antenna arrays with different smart direction of arrival (DoA) estimations. Smart antennas can generate a beam that is specific to the user. The array system's complexity can be reduced by using optional feedback. The antenna array radiation pattern can be configured by using a precoding or beamforming techniques. It can be used in all antenna array systems, including MIMO systems [12]. Beamforming is generally considered as a spatial filtration technique which takes advantages of the spatial properties of the signals from multiple antenna array elements. For example, the phase and the magnitude of the signals from each antenna elements are processed to achieve the beamforming by adding these signals in a constructive or deconstructive way. The beamforming is used at the BS and the user's terminal to increase the received SNR [13]. According to the position of the phase shifters along the path of the signal, the precoding in the desired direction can be implemented in multiple ways. The precoding or beamforming structures can be categorized into three types: analog beamforming, digital precoding and hybrid precoding.

2.1   Analog Beamforming

The analog beamforming is implemented by phase-shifting at the RF stage or the local oscillator (LO) stage. Analog beamforming controls the initial signal phases to achieve maximum antenna arrays gain and effective SNRs.

2.1.1   Beam Steering with A Single Data Stream

Figure 2 shows the analog beamforming with a single-data stream (mm-Wave) and a massive MIMO system. Where a data stream is transmitted to a user with N_r antennas and the BS has N_t antennas and one RF chain.
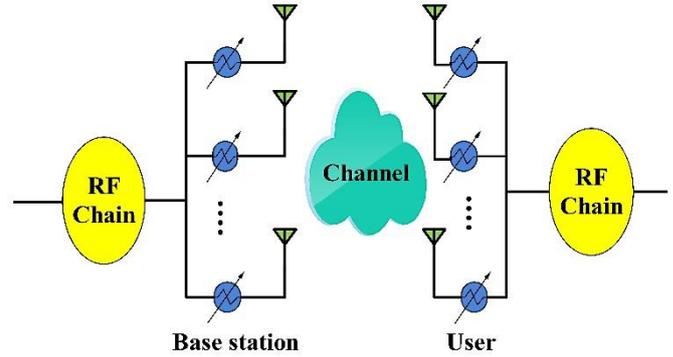

Figure 2. The structure of analog beamforming for a single-data stream mm-Wave massive MIMO system.

The analog beamforming vector at the transmitter is defined by **p,** this vector size is $N_t \times 1$ and **w** refer to the analog combining vector at the receiver with a size of $N_r \times 1$. Our goal is to obtain **p** and **w** to optimize the SNR that can be seen as

$$(\mathbf{w}^{opt}, \mathbf{p}^{opt}) = argmax \; \|\mathbf{w}^H \mathbf{H} \, \mathbf{p}\|^2 \qquad (1)$$

$$Subject \; to: \; \mathbf{w}_i = \frac{1}{\sqrt{N_r}} e^{j\phi_i}, \forall i,$$

$$\mathbf{p}_l = \frac{1}{\sqrt{N_r}} e^{j\phi_l}, \forall l,$$

Where the Hermitian transpose is denoted $[\cdot]^H$. The SVD of the channel is obtained as $\mathbf{H} = \mathbf{U\Sigma V}^H$. Then, it is known that the optimum solutions to (1) should be $\mathbf{w}^{opt} = \mathbf{U}(:, 1)$ and $\mathbf{p}^{opt} = \mathbf{V}(:, 1)$. One available way for solving this unconstraint problem is to establish the solutions **p** and **w** to achieve the amplitude-constraint as near as possible to optimum $\mathbf{p}^{opt}$ and $\mathbf{w}^{opt}$ unconstraint solutions. A theorem shows that this design can be significantly simplified for (mm-Wave) systems with massive antenna array elements. This theorem states that the channel's description is converged to its SVD, resulting in the optimal array response vector. So, the beamforming and combining vectors in (mm-Wave) systems with large N_t and N_r can form the array response vectors, which approaches the ideal solutions [14].

2.1.2   Beam Training with A Single Data Stream

The channel matrix can not be directly observed by the BS or the user for (mm-Wave) systems with just one RF chain; instead, subspace sampling limitations would occur. These limitations mean that conventional algorithms for the channel estimation can not be used explicitly, and thus the channel matrix's elements can be practically evaluated. Without the full CSI, The beam training will represent a challenging task. The right precoding and combiner are selected from the predefined codebooks during the beam training. The beam steering model provides the predefined codebooks which can be described as:

$$\mathbf{p} \in \mathcal{F} = \{\mathbf{a_t}(\phi_1^{-t}, \theta_1^{-t}), \mathbf{a_t}(\phi_2^{-t}, \theta_2^{-t}), …, \mathbf{a_t}(\phi_{|F|}^{-t}, \theta_{|F|}^{-t})\}, \qquad (2)$$

$$\mathbf{w} \in \mathcal{W} = \{\mathbf{a_r}(\phi_1^{-r}, \theta_1^{-r}), \mathbf{a_r}(\phi_2^{-r}, \theta_2^{-r}), …, \mathbf{a_r}(\phi_{|W|}^{-r}, \theta_{|W|}^{-r})\}. \qquad (3)$$

The azimuth angle of departure is quantified and defined by $\phi_i^{-t}, (\theta_i^{-t})$, and the quantified azimuth angle of arrival is denoted by $\phi_l^{-r}, (\theta_l^{-r})$. Searching all conceivable $|\mathcal{F}|$, $|\mathcal{W}|$ pairs of beamforming and combining vectors will depend on

SNR optimization, the optimization of the efficient SNR is considered as the most efficient and straight forward beam training scheme. In mm-Wave systems, however, the possibly massive antenna array elements and the significant requirements of beamforming gain would require vast sizes of the codebook $|\mathcal{F}|$ and $|\mathcal{W}|$ which can lead to an inaccessible overhead for the comprehensive search. The problem can be solved by using a scheme of a hierarchical beam training. Then a sequence of codebooks $\mathcal{F}_1$, $\mathcal{F}_1,\ldots,\mathcal{F}_K$ ($\mathcal{W}_1,\mathcal{W}_1,..,\mathcal{W}_K$) with the increasing resolution are established as shown in figure 3"(a)". Next, the beams at the first level are trained by transmitting a training data which is divided into three steps:

a) The training data is sent by the BS to the user. The potential beamforming vector is selected from the codebook $\mathcal{F}_1$ and the user can define the best combination vector figure 3 "(b)".

b) User swaps its role with the BS by establishing a similar method of the best beamforming vector figure 3 "(c)".

c) They obtain each other's index of the chosen beamforming vector figure 3 "(d)".

This procedure is replicated with a higher-resolution codebook on the selected beam to take into consideration the final level (highest-resolution codebook $\mathcal{F}_K$). In contrast to the exhaustive search, this hierarchical beam training will effectively decrease the overhead. It should be pointed out that we are able to use shorter training sequences because of additional array gain each time we pass to the next step [14, 15].
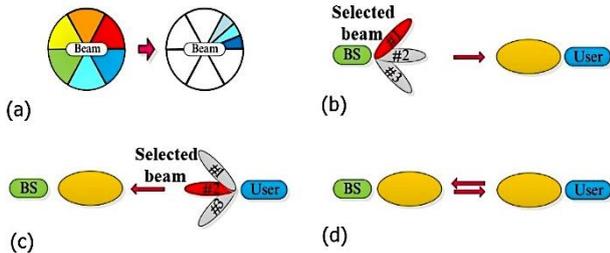

Figure 3. The beam training scheme. "(a)" Multi-level code-book, "(b)" Beam sweep at BS, "(c)" Beam sweep at the user, and "(d)" Feed-back phase [14].

2.1.3 Analog Beamforming with Multiple Data Streams

RF phase-shifting network is the most commonly used beamforming structure. Figure 4 shows a standard passive RF phase-shifting network consisting of a uniform linear array (ULA) which has Q -elements and the separation **d** (between its elements) with N $_{RF}$ channels.

The antenna array's elements are connected with the transmitter and receiver for each data stream through a duplexer or switch and phase shifter so that each beam can be controlled. The phase shifters number which are required can be calculated as the product of the antenna array elements and a number of beams [16].

ϕ_{nm} is assumed to be the phase shift between the m-th element of the antenna array and the n-th phase shifter. To obtain a beam with an off-broadside angle θ_n for λ_0 wavelength, the required phase ϕ_{nm} can be obtained by equation (4)

$$\phi_{nm} = \frac{2\pi d \sin\theta_n}{\lambda_0} \ , \forall m = 1, 2,\ldots, \text{Q}. \tag{4}$$

The narrower beam can be obtained as the distance between antenna elements **d** is increased. On the other hand, the grating lobe will be appeared in the radiation pattern when **d** is getting larger, which limits the scanning range. As shown in figure 5, the system structure for the LO phase shifting network, the direction of the beam can be adjusted by implementing the phase shifts at LO. So, for N beams which are generated by Q antenna array's elements, the N×Q mixers would be needed [17-19].

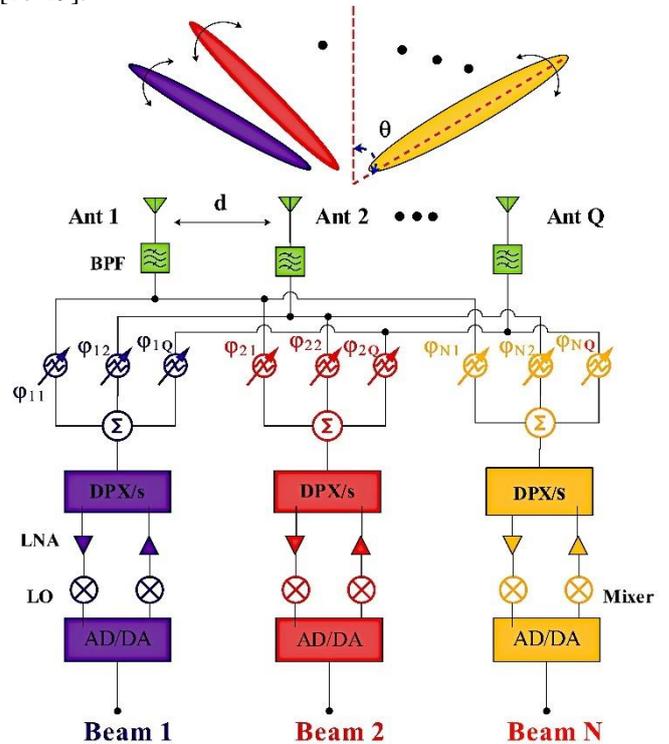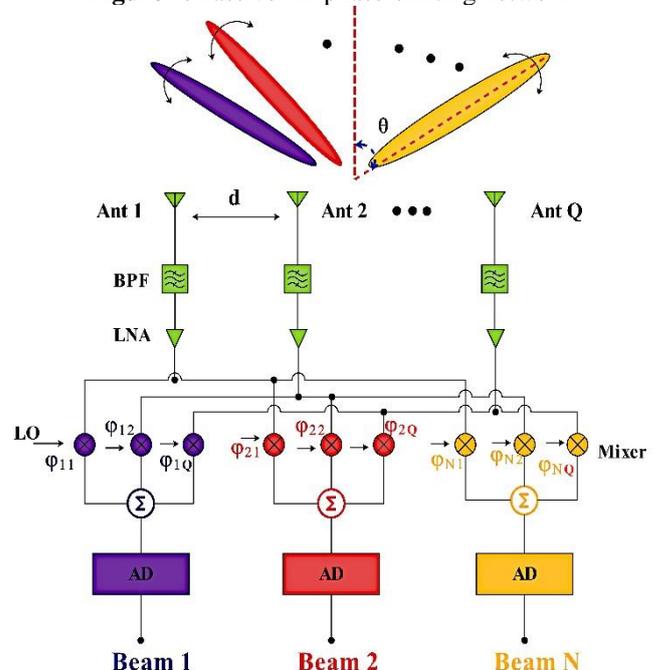**Figure 4.** Passive RF phase-shifting network

Figure 5. System architecture for LO phase shifting network

2.2 Digital Beamforming / Precoding

At mm-Wave frequencies with broad bandwidths and massive antenna array elements, full-digital processing is difficult to be realized because of the precoding / combining processing. Figure 6 shows that the dedicated RF chain per antenna in the transceiver [20].
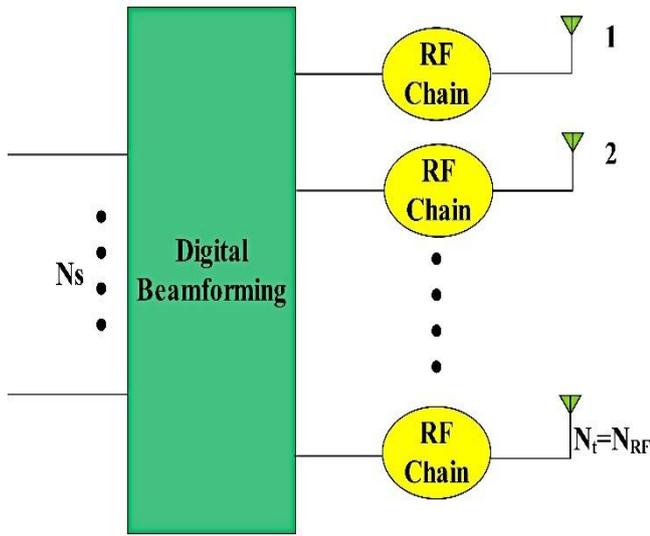
Figure 6. Fully digital beamforming architecture from

Digital precoding offers a robust approach for producing multiple independently controlled beams. It can provide robustness to avoid the failure of one or more antenna elements. Figure 7 shows the system architecture of Q-elements ULA, which can produce N beams by implementing the digital precoding. Behind each element of the transceiver, there is a power amplifier (PA), low noise amplifier (LNA), mixer, analog to digital converter (ADC), and digital to analog converter (DAC).   However, in contrast to the analog techniques, there are no attenuators or phase shifters. The RF signal is received by the antenna array, then it is filtered, amplified by LNA, down-converted by mixer, after that it is converted from analog to digital via A/D. The received complex baseband signal from the antenna array's elements can be written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots \dots, \mathbf{x}_Q \end{bmatrix}, \quad \forall m = 1, \dots \dots, Q \qquad (5)$$

In digital domain, this complex baseband signal which includes the in-phase and quadrature components is multiplied by a weighting matrix. The weighting matrix can be defined in (6).

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_N \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{N1} & \cdots & \mathbf{W}_{NQ} \end{bmatrix} \qquad (6)$$

Each element of the weighting matrix can be obtained by $W_{nm} = a_{nm} e^{j\phi_{nm}}$ with $m=1,..,Q$ and $n=1,..N$, and the n-th beam coming from m-th element is multiplied by a complex weight W_{nm}. The amplitude tapering is controlled by the coefficient a_{nm} and the phase delay for each antenna element is determined by ϕ_{nm}. These values may be modified according to various factor's frequencies. After the weighting process, the output of the $\boldsymbol{n}^{th}$ beam could be obtained by (7)

$$\mathbf{Y}_n(\theta) = \mathbf{W}_n^H \mathbf{X} \qquad (7)$$

The weighting vector could be represented by Eq. (8) if the signal power is brought out of the Eq. (7) as

$$\mathbf{W}_N = \left[1, e^{j\frac{2\pi d}{\lambda} \sin \theta_n}, \dots, e^{j(M-1)\frac{2\pi d}{\lambda} \sin \theta_n} \right]^T \qquad (8)$$

For the transmission mode, the signal will be converted from the digital domain into analog after the weighting process. Then this signal will be upconverted and radiated from the antenna array. Multiple beams independently controlled can be synthesized via the digital domain by creating a suitable weighting matrix. Each element has its own RF transceiver

chain in fully active digital beamforming. At mm-wave frequencies, the A/D and D/A in each RF chain will consume a high amount of power. The reduction in the number of required RF chains will result in reducing the consumed power and required hardware complexity. To ease the processing and power demands, the antenna array's elements are divided into subarrays as shown in figure 8.

2.2.1   Single-User Digital Precoding

In this section, simple linear-digital precoding will be focused on single-user primarily. Figure 9 shows the digital precoding when a mm-wave massive MIMO wireless system is assumed to have a single user.

For a transmission mode, the number of antenna array's elements at the BS is assumed to be N_t while the number of data streams is considered to be N_s. The number of antenna elements at the user side is assumed to be N_r, where $N_s < N_t$. The BS applies an $N_t \times N_r$ digital precoder **L** using its N_t^{RF} chains and the transmitted signal can be expressed as $\mathbf{x} = \mathbf{Ls}$, where **s** is $N_r \times 1$ refers to the original signal with normalized power. **L** and **s** satisfy the constraint of the total transmitted power. The received signal **y** is the $N_r \times 1$ can be expressed as:

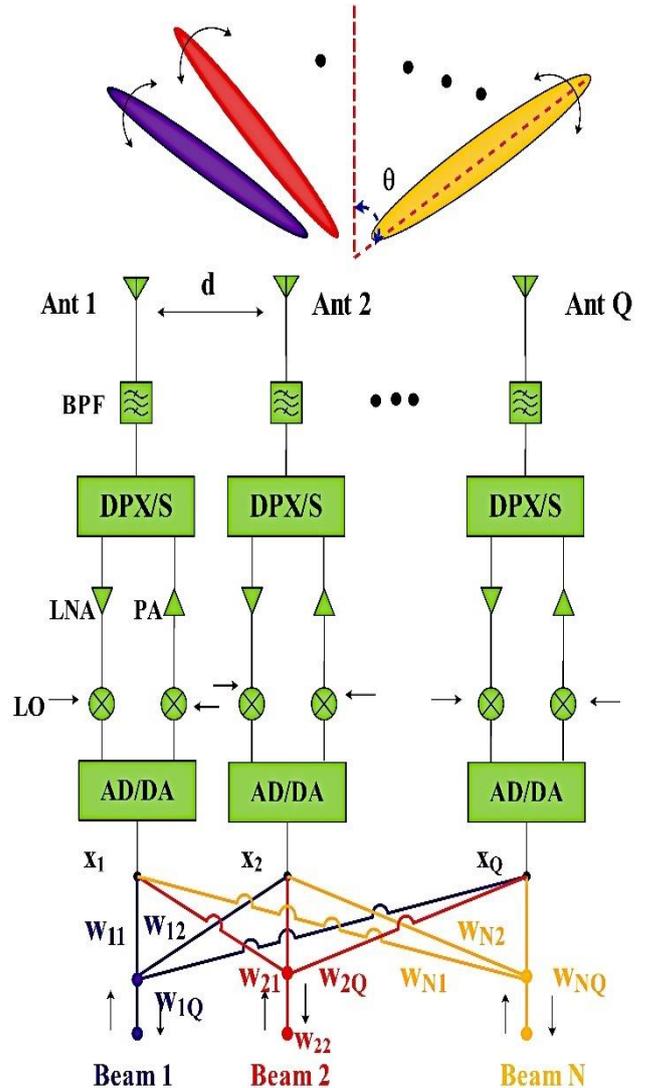$$\mathbf{y} = \sqrt{\rho} \ \mathbf{HLs} + \mathbf{n}, \qquad (9)$$

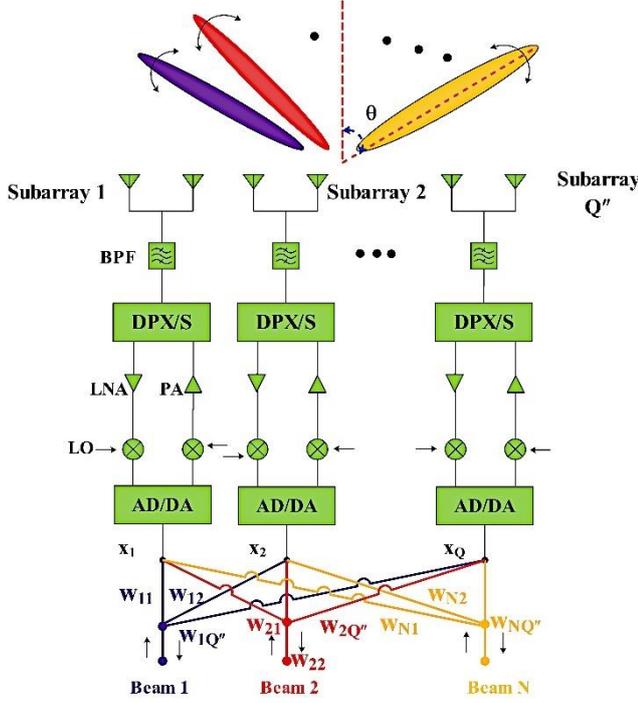**Figure 7.** System structure for Digital Beamforming

Figure 8. System architecture for digital precoding

where the channel matrix $\mathbf{H}$ is $N_r \times N_t$ and it is assumed to have a normalized power, the average received power is defined by ρ, and $\mathbf{n}$ is a noise vector. Matched filter (MF) precoding technique can be used for the simplest linear-digital precoding [14] , which can be written as:

$$\mathbf{L} = \sqrt{\frac{N_r}{tr\,(\mathbf{FF}^H)}}\ \mathbf{F}\ ,\ \ \mathbf{F} = \mathbf{H}^H \tag{10}$$
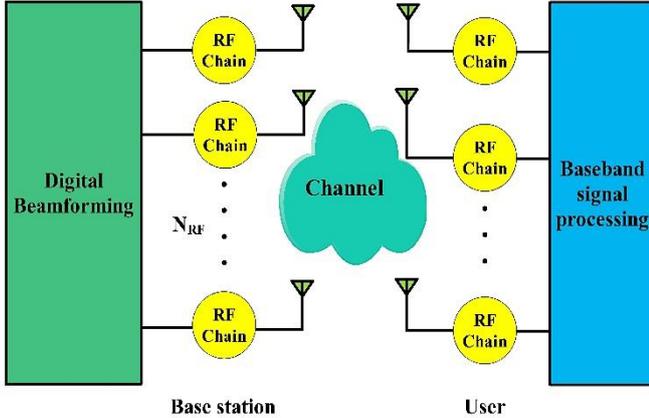

Figure 9. Structure of fully digital precoding for a mmWave massive MIMO wireless system with a single user

The received SNR at the user's terminal can be maximized by MF precoding; however, the interferences between different data streams are generally severe. The recognized ZF precoding is proposed for this reason. The digital precoding matrix $\mathbf{L}$ can be given by:

$$\mathbf{L} = \sqrt{\frac{N_r}{tr\,(\mathbf{FF}^H)}}\ \mathbf{F}\ ,\ \mathbf{F} = \mathbf{H}^H(\mathbf{HH}^H)^{-1} \tag{11}$$

ZF precoding can increase noise power which results in a certain loss of performance compared to optimum channel capacity [14].

2.2.2 Multi-User Digital Precoding

Figure 10 indicates the mm-Wave massive MIMO wireless system with multi-user fully digital precoding.
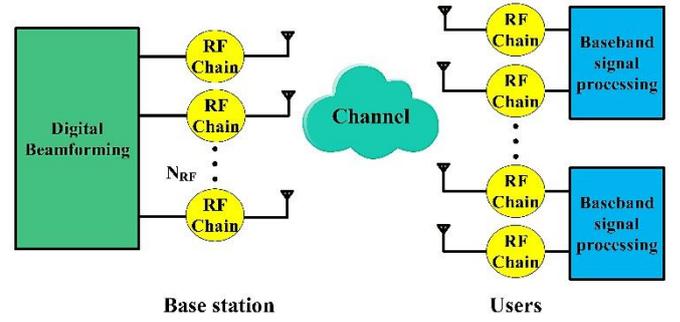

Figure 10. Structure of multi-user fully digital precoding for mmWave massive MIMO wireless system

In multi-user situation with a single antenna at each terminal, interference from other signals cannot be minimized. ZF precoding is commonly used to solve multi-user interference. Due to the simplicity of the ZF precoding technique, it is widely used; however, the Block Diagonalization method (BD) is more acceptable if terminals are equipped with multiple antennas [21, 22]. It is known that the BS with N_t antennas serve K terminals in the communication systems. N_r refers to a number of antennas for each terminal. The received signal $\mathbf{y}_i \in \mathbb{C}^{N_r \times 1}$, for the i-th terminal can be represented as:

$$\mathbf{y}_i = \mathbf{H}_i \sum_{k=1}^{K} \mathbf{P}_k^{BD}\,\mathbf{x}_k + \mathbf{n}_i \tag{12}$$

Where $\mathbf{H}_i \in \mathbb{C}^{N_r \times N_t}$ refers to the channel matrix, $\mathbf{P}_k^{BD} \in \mathbb{C}^{N_t \times N_r}$ refers to the BD precoding matrix, $\mathbf{x}_k \in \mathbb{C}^{N_r \times 1}$ refers to the source signal and $\mathbf{n}_i \in \mathbb{C}^{N_r \times 1}$ refers to the noise. The channel matrix of all terminals is represented as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_1 & \cdots & \mathbf{H}_1 \\ \mathbf{H}_2 & \mathbf{H}_2 & \cdots & \mathbf{H}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_K & \mathbf{H}_K & \cdots & \mathbf{H}_K \end{bmatrix} \tag{13}$$

So, the received signals for all terminals can be given by:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_1 & \cdots & \mathbf{H}_1 \\ \mathbf{H}_2 & \mathbf{H}_2 & \cdots & \mathbf{H}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_K & \mathbf{H}_K & \cdots & \mathbf{H}_K \end{bmatrix} \begin{bmatrix} \mathbf{P}_1^{BD}\mathbf{x}_1 \\ \mathbf{P}_2^{BD}\mathbf{x}_2 \\ \vdots \\ \mathbf{P}_K^{BD}\mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_K \end{bmatrix} \tag{14}$$

$$= \begin{bmatrix} \mathbf{H}_1\mathbf{P}_1^{BD} & \mathbf{H}_1\mathbf{P}_2^{BD} & \cdots & \mathbf{H}_1\mathbf{P}_K^{BD} \\ \mathbf{H}_2\mathbf{P}_1^{BD} & \mathbf{H}_2\mathbf{P}_2^{BD} & \cdots & \mathbf{H}_2\mathbf{P}_K^{BD} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_K\mathbf{P}_1^{BD} & \mathbf{H}_K\mathbf{P}_2^{BD} & \cdots & \mathbf{H}_K\mathbf{P}_K^{BD} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_K \end{bmatrix} \tag{15}$$

The $\mathbf{H}^{-i}$ matrix is constructed excluding the i-th terminal channel matrix.

$$\mathbf{H}^{-i} = [\mathbf{H}_1^H \quad \cdots \quad \mathbf{H}_{i-1}^H \quad \mathbf{H}_{i+1}^H \quad \cdots \quad \mathbf{H}_K^H]^H \tag{16}$$

Where $\mathbf{H}^{-i} \in \mathbb{C}^{(N_t-N_r)\times N_t}$ and $N_t = K\,N_r$. To overcome the multi-user interference, BD method is used to make the off-diagonal term $\mathbf{H}_i\mathbf{P}_j^{BD}$ equals to $\mathbf{0}_{N_r \times N_r}$, the received signals for all terminals can be expressed as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1\mathbf{P}_1^{BD} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_2\mathbf{P}_2^{BD} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{H}_K\mathbf{P}_K^{BD} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_K \end{bmatrix} \tag{17}$$

The design of (BD) precoding matrix, it is assumed the rank of $\mathbf{H}^{-i}$ is $N_{rank} = N_t - N_r$. The SVD of $\mathbf{H}^{-i}$ can be written as:

$$\mathbf{H}^{-i} = \mathbf{U}_i \sum_i \mathbf{V}_i^H = \mathbf{U}_i \sum_i \left[\mathbf{V}_i^{(1)} \quad \mathbf{V}_i^{(2)} \right]^H \qquad (18)$$

Where $\mathbf{U}_i$ refers to $N_{rank} \times N_{rank}$ unitary matrix, $\mathbf{V}_i$ refers to $N_t \times N_t$ unitary matrix and $\sum_i$ refer to $N_{rank} \times N_t$ diagonal matrix of singular values. $\mathbf{V}_i$ is split into two parts $\mathbf{V}_i^{(1)} \in \mathbb{C}^{N_t \times N_{rank}}$ consists of the N_{rank} non-zero singular vectors and $\mathbf{V}_i^{(2)} \in \mathbb{C}^{N_t \times N_r}$ consists of the N_r zero singular vectors. Multiplying $\mathbf{H}^{-i}$ with $\mathbf{V}_i^{(2)}$, the result will be

$$\mathbf{H}^{-i} \mathbf{V}_i^{(2)} = \mathbf{U}_i \left[\sum_i^{(1)} \quad \mathbf{0} \right] \begin{bmatrix} \mathbf{V}_i^{(1)H} \\ \mathbf{V}_i^{(2)H} \end{bmatrix} \mathbf{V}_i^{(2)}$$
$$= \mathbf{U}_i \sum_i^{(1)} \mathbf{V}_i^{(1)H} \mathbf{V}_i^{(2)}$$
$$= \mathbf{U}_i \sum_i^{(1)} \mathbf{0} = \mathbf{0} \qquad (19)$$

where $\sum_i^{(1)}$ refer to the first partition of dimension $N_{rank} \times N_{rank}$. Thus, $\mathbf{V}_i^{(2)}$ forms an orthogonal subspace for the null space of $\mathbf{H}^{-i}$ and $\mathbf{P}_i^{BD} = \mathbf{V}_i^{(2)}$ can be used for precoding the signal of the i-th terminal with BD constraint. The computational complexity of BD precoding is due to the SVD process that increases the computation complexity with the number of terminals and the system dimensions [21].

3. Hybrid Precoding

The main advantages of the hybrid precoding signal processing can be concluded as follows:

- The role of hybrid precoding for enabling (mm-Wave) massive MIMO communications is as important as the role of (mm-Wave) massive MIMO for enabling 5G wireless networks with 1000× capacity improvement. The hybrid precoding can reduce the high cost and hardware complexity, and also it can reduce inter-users interference.
- The antenna arrays with large number of elements, which realized by hybrid precoding in (mm-Wave) massive MIMO, potentially reduce the downlink and uplink transmit powers, so that the hybrid precoding can provide a high Energy Efficiency (EE) [20, 23-26].

The hybrid precoding can be a solution for improving the (EE), especially with mm-Wave massive MIMO systems. Digital precoding provides improved performance, but it requires high cost, power consumption and hardware complexity. Analog beamforming, on the other hand, is an easy and low cost technique with less versatility [27]. The Hybrid precoding architecture can offer sharp beams [28]. High directionality can be obtained by using large size antenna arrays of individual elements with small size [29]. The hybrid precoding architecture as in figure 11, offers a compromise between the hardware-complexity and the system-performance [29].

The benefits of mm-Wave massive MIMO wireless system are achieved by adopting the hybrid beamforming architecture [30]. Figure 12 shows that, in a fully-connected hybrid precoding system, all antennas link to each RF chain, so that each of the N_{RF} digital signals pass through different phase shifters. This architecture consists of analog beamforming and digital precoding, whereby each digital signal will achieve the maximum beamforming gain. Nevertheless, the fully-connected system is relatively complex, i.e., the number of

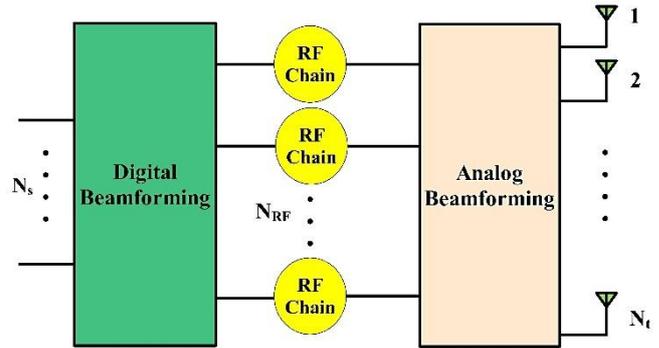antennas times the number of RF chains equal to the total number of phase shifters in this structure [31].

Figure 11. Hybrid precoding

Although the fully-connected hybrid precoding is relatively complex, its structure is used in many research works to enhance the EE and the spectral efficiency (SE). The authors in [32] propose energy-efficient hybrid precoding and it investigates the comparison between energy and cost efficiency for 5G wireless communication. Despite the high beamforming gain per transceiver of the fully-connected precoding architecture, all this causes high complexity. The partially-connected hybrid precoding structure decreases the complexity, as seen in figure 13. In a partially-connected structure, each RF-chain is linked to a set of antennas with phase shifters per N_{RF} elements. The total number of the phase shifters in this system can be shown by the number of antennas, which means that the analog beamformer's hardware complexity is reduced by a factor of the number of RF-chains [31, 33].
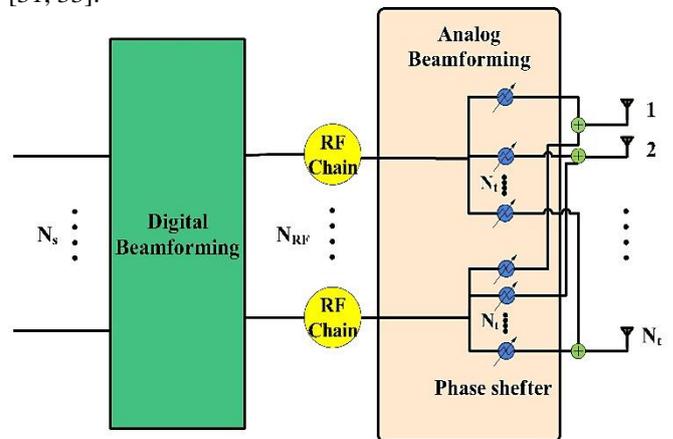
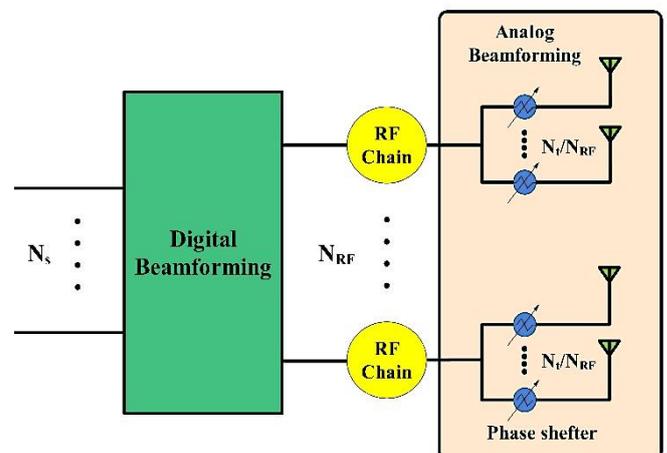**Figure 12.** Fully -connected hybrid precoding

Figure 13. Partially -connected hybrid precoding

At high-frequency (mm-Wave) bands, the transmission signal can be modified to overcome unfavorable channel properties [34]. The partially-connected system with multiple steering beams is realistic to the femtocell BS, which can move regularly and operate on batteries. From what has been mentioned above, we can conclude that, the fully-connected architecture will achieve optimal precoding gain. In contrast to the fully-connected system, the complexity of hardware design and power consumption of the partially-connected structure are substantially lower [31]. Spatially sparse hybrid precoding, manifold optimization-based hybrid precoding, successive interference cancellation-based hybrid precoding, and semidefinite relaxation hybrid precoding are the most common algorithms that are implemented to realize hybrid precoding. These hybrid precoding techniques are overviewed as follows:

3.1   Spatially-Sparse Hybrid Precoding

For the fully-connected structure, since the analog precoder $\mathbf{G}$ is achieved by the phase shifters, elements of $\mathbf{G}$ have different phases but equal amplitudes. Furthermore, by normalizing the digital beamformer $\mathbf{L}$ to satisfy $\|\mathbf{GL}\|_F^2 = N_s$, the total transmission power constraint is enforced. We want to design $(\mathbf{G}, \mathbf{L})$ in order to maximize the spectral efficiency R $(\mathbf{G}, \mathbf{L})$ over the (mm-Wave) channel. The corresponding spectral efficiency optimization problem can be presented with the design of $(\mathbf{G}, \mathbf{L})$ as:

$$(\mathbf{G}^{\text{opt}}, \mathbf{L}^{\text{opt}}) = argmax \, \text{R} \, (\mathbf{G}, \mathbf{L}), \qquad (20)$$

Subject to : $\mathbf{G} \in \mathcal{F}$,

$$\|\mathbf{GL}\|_F^2 = N_s,$$

All the feasible analog beamformers are included in $\mathcal{F}$ set, i.e., a set of $N_t \times N_t^{RF}$ matrices with equal-magnitude elements. An approximate solution of the above equation is proposed. Such approximation depends on converting achievable spectral efficiency into the distance between the optimal unconstrained precoder $\mathbf{P}_{\text{opt}}$ and the hybrid precoder $\mathbf{GL}$ [35]. The well-known orthogonal matching pursuit (OMP) can solve this problem. Algorithm 1 provides the precoder solution pseudocode obtained from OMP [35].

Algorithm 1 Orthogonal Matching Pursuit Algorithm

Input: $\mathbf{P}_{opt}$
1: $\mathbf{G}$ = Empty Matrix
2: $\mathbf{P}_{res} = \mathbf{P}_{opt}$
3: **for** $i \leq N_t^{RF}$ **do:**
4: $\psi = \mathbf{G}_t^H \mathbf{P}_{res}$
5: $k = \arg \max_{l=1,\dots,L} (\psi\psi^H)_{l,l}$
6: $\mathbf{G} = \left[\mathbf{G} | \mathbf{G}_t^{(k)} \right]$
7: $\mathbf{L} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{P}_{opt}$
8: $\mathbf{P}_{res} = \dfrac{\mathbf{P}_{opt}-\mathbf{GL}}{\|\mathbf{P}_{opt}-\mathbf{G\,L}\|_F}$
9: **end for**
10: $\mathbf{L} = \sqrt{N_s}\,\dfrac{\mathbf{L}}{\|\mathbf{GL}\|_F}$
Output: return $\mathbf{G}, \mathbf{L}$

---

what happens in the Algorithm 1 is that after the initial steps 1 and 2, this algorithm is started by obtaining the vector $\mathbf{a_t}$ (ϕ_l^t, θ_l^t), in step 5, the optimal precoder has the greatest projection along that vector. The chosen column vector is then appended in step 6, $\mathbf{a_t}(\phi_l^t, \theta_l^t)$ to the analog beamforming matrix $\mathbf{G}$. After finding the dominant vector, and the least-squares solution to $\mathbf{L}$ is determined in step 7, after that in step

8, the selected vector is removed, and this algorithm is proceeded to find the column with largest projection on the "residual precoding matrix" $\mathbf{P}_{res}$. The process is continued until the selection of all precoding vectors, and find the designed analog beamformer matrix $\mathbf{G}$ and the digital precoder matrix $\mathbf{L}$, which minimizes $\left\|\mathbf{P}_{opt} - \mathbf{GL}\right\|_F$. The constraint of power transmit $\|\mathbf{GL}\|_F^2 = N_s$ is satisfied in step 10 .

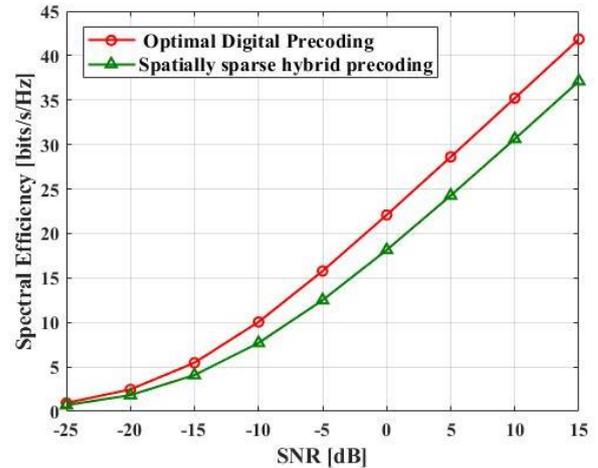**Figure 14.** Comparison of spectral efficiency in a 64×16 mm-Wave massive-MIMO wireless system with $N_t^{RF} = 4$.

The comparison between the performance of the spatially sparse hybrid precoding and the performance of fully-digital precoding is illustrated in figure 14. The spatially sparse hybrid precoding will provide a near-optimal efficiency [14, 36]. Simulation results demonstrate that the spatially sparse hybrid precoding is very near to the optimal unconstrained precoder in classical mm-wave massive MIMO systems, although considerably reducing the number of necessary RF chains.

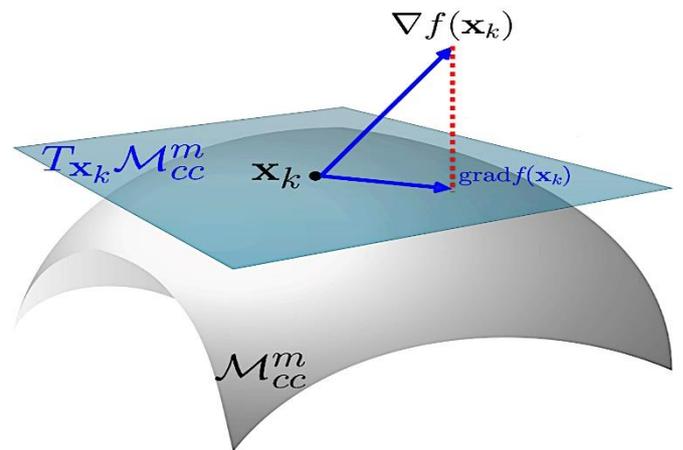### 3.2   Manifold Optimization (Mo) Based Hybrid Precoding

Figure 15. The tangent space of a Riemannian manifold [37]

In most applications, the topological manifolds can be categorized as a Riemannian manifold. Riemannian manifolds are rich in geometry and allow you to define cost function gradients as shown in figure 15 [37]. More specifically, the optimization of a Riemannian manifold is close to that of an Euclidean space where smooth restrictions are present. In Euclidean spaces, the conjugate gradient algorithm can obtain its equivalent on the specified Riemannian manifolds. The algorithm of conjugate gradient for analog precoding based on (MO) can be summarized in Algorithm 2 [37].

Algorithm 2 Conjugate Gradient Algorithm

Input: P_{opt} , L, $x_0 \in M_{cc}^m$

1: $\mathbf{d}_0 = -grad\ f(x_0)\ and\ k = 0$;

2: **Repeat**

3: Select Armijo backtracking line search step size α_k;

4: Define the next point x_{k+1} using $x_{k+1} = \text{Retr}_{x_k}(\alpha_k\ \mathbf{d}_k)$;

5: Determine Riemannian gradient $\mathbf{g}_{k+1} = \text{grad}\ f(x_{k+1})$

6: Find the vector transports g_k^+ and $\mathbf{d}_k^+$ of gradient g_k and conjugate direction $\mathbf{d}_k$ from x_k to x_{k+1};

7: Select Polak-Ribiere parameter β_{k+1} ;

8: Compute conjugate direction $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1}\ \mathbf{d}_k^+$;

9: $k \leftarrow k+1$;

10: **Till** a stopping criterion triggers

The direction with steepest descend can be determined by using the following equation.

$$\nabla f(x_k) = -2(\mathbf{L}^* \otimes \mathbf{I}_{N_t})\left[vec\left(\mathbf{P}_{opt}\right) - (\mathbf{L}^T \otimes \mathbf{I}_{N_t})x_k\right]\quad (21)$$

Where $\nabla f(x_k)$ is the Euclidean gradient at the point x_k. The orthogonal projection of $\nabla f(x_k)$ onto the tangent space can be defined as a tangent vector grade $f(x_k)$ which can be obtained by

$$\text{grad}f(x_k) = \text{Proj}_{x_k} \nabla f(x_k)\quad\quad\quad (22)$$

Algorithm 2 uses well recognized Armijo backtracking line search step and Polak-Ribiere parameter to ensure that the goal function in any iteration is not increased [38, 39]. After determining the step size, the destination is not on the manifold, so that mapping from tangent vector back to manifold itself is required and can be done by retraction $\text{Retr}_{x_k} (T_{x_k} \mathcal{M}_{cc}^m) \rightarrow \mathcal{M}_{cc}^m$ as shown in figure 16.
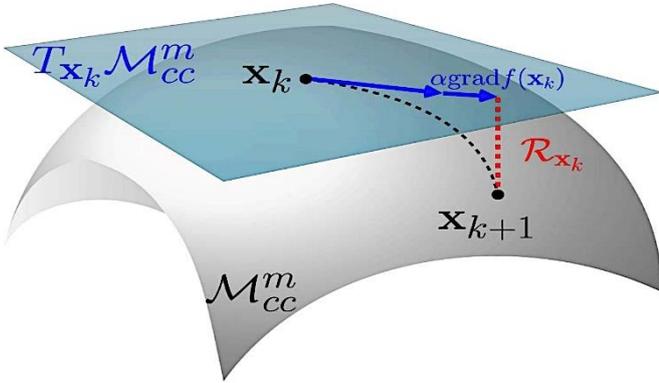

Figure 16. Mapping from tangent vector back to manifold itself [37]

Also, Algorithm 2 is ensured to correspond to the point where the objective function is zero (critical point). The hybrid precoder design with conjugate gradient algorithm for Analog Precoding is represented in the MO-AltMin algorithm through alternating minimization for the fully connected structure. The MO-AltMin Algorithm for the fully-connected structure can be done by the steps listed in Algorithm 3 [37].

Algorithm 3 The Manifold Optimization -Alternating Minimization Algorithm

---

Input: $\mathbf{P}_{opt}$

1: *Construct* $\mathbf{G}^{(0)}$ with random phase and set $k = 0$;

2: **Repeat**

3: Fix $\mathbf{G}^{(k)}$, and $\mathbf{L}^{(k)} = \mathbf{G}^{(k)\dagger}\ \mathbf{P}_{opt}$;

4: Optimize $\mathbf{G}^{(k+1)}$ using Conjugate Gradient algorithm when $\mathbf{L}^{(k)}$ is fixed;

5: $k \leftarrow k+1$;

6: **Till** a stopping criterion occurs ;

7: Normalize for the digital precoder using $\mathbf{L} = \frac{\sqrt{N_s}}{\|\mathbf{GL}\|_F}\ \mathbf{L}$.

The results of the simulation in figure 17 show an MO-AltMin algorithm with a near optimum efficiency. The complexity of the algorithm MO-AltMin is, however, very high. The analog precoder updates include the line search algorithm in each iteration, i.e., conjugate gradient algorithm, thus slowing down the entire solving process with the nested loops in the MO-AltMin algorithm. Although, the operation is highly complex, it is observed that the problem formulation of the hybrid precoder design under unit modulus constraints is directly solved by (MO)-based-MO-AltMin algorithm, which will increase spectral efficiency. So, the MO-AltMin algorithm will serve as the performance benchmark for spectral efficiency.
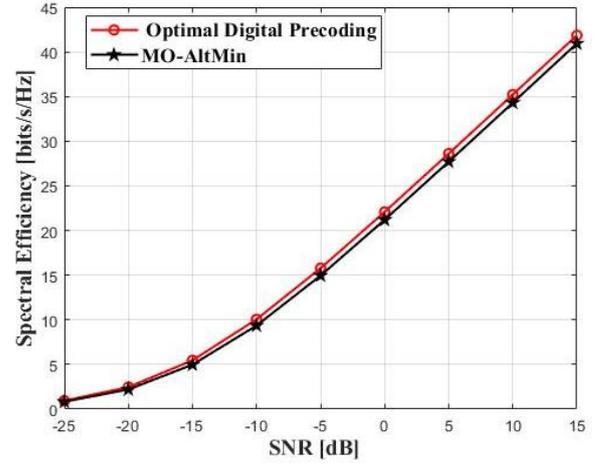

Figure 17. Spectral efficiency of the MO-AltMin algorithms given $N_{RF} = Ns = 4$

3.3 Successive Interference Cancelation (SIC)-Based Hybrid Precoding

In the SIC algorithm, a series of sum-rate of sub-antenna arrays can be individually optimized to obtain the total achievable rate optimization. The first sub-antenna array can be optimized and the matrix can be updated relying on the concept of SIC for multi-user signal detection, whose diagram is shown in Fig. 18. After that, the achievable sum rate of the second sub antenna array can be optimized by a similar process. This process is conducted until the last sub antenna array [14, 40].
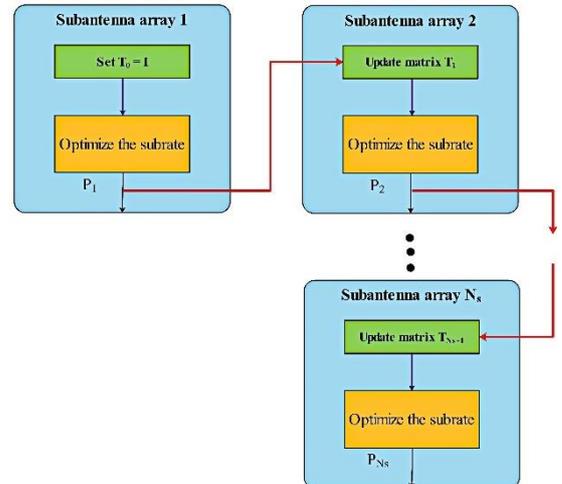

Figure 18. SIC-based hybrid precoding

Figure 19 demonstrates that hybrid precoding performance based on SIC is less than that of the optimum digital precoding.
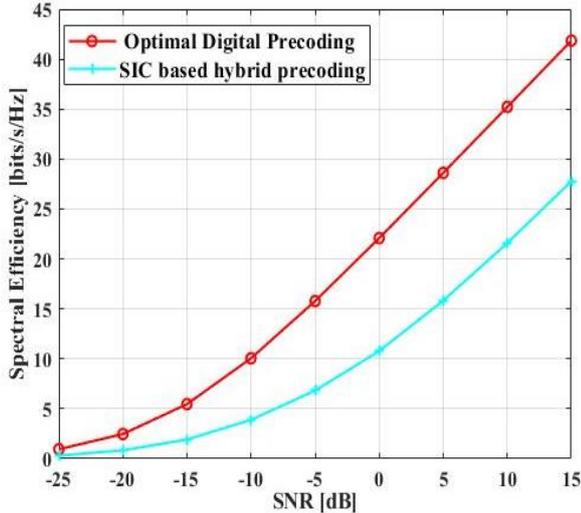

Figure 19. Comparison of spectral efficiency for an $N_t \times N_r$ = 64×16 ($N_t^{RF} = N_s = 4$) mm-Wave massive MIMO system

3.4   Semidefinite Relaxation (SDR)

The digital precoder design can be written as:

$$\underset{L}{minimize} \left\| \mathbf{P}_{opt} - \mathbf{GL} \right\|_F^2 \tag{23}$$

Subject to: $\|\mathbf{L}\|_F^2 = \frac{N_{RF}^t N_s}{N_t}$

This problem is a quadratic constraint quadratic programming (QCQP) non-convex problem, which means that it is NP-hard and not easy to solve this problem in polynomial time. To make this problem a convex programming problem that can be solved in polynomial time using the semidefinite programming methods, all of the matrices should be positive semidefinite.

$$minimize \quad \text{Tr}\,(\mathbf{CY})$$
$$\mathbf{Y} \in \mathbb{H}^n$$

$$Subject\ to \begin{cases} \text{Tr}\,(\boldsymbol{A}_1\mathbf{Y}) = \frac{N_{RF}^t N_s}{N_t} \\ \text{Tr}\,(\boldsymbol{A}_2\mathbf{Y}) = 1 \\ \mathbf{Y} \succeq 0 \quad rank(\mathbf{Y}) = 1 \end{cases} \tag{24}$$

Where $\mathbb{H}^n$ is complex Hermitian matrices, with $n = N_{RF}^t N_s + 1$ dimension. In addition, $y = [vec(\mathbf{L})\ t]^T$ with an auxiliary variable t, $\mathbf{Y} = yy^H$, $\mathbf{p} = vec(\mathbf{P}_{opt})$, and

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \mathbf{0}_{n-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} (\mathbf{I}_{N_s} \otimes \mathbf{G})^H(\mathbf{I}_{N_s} \otimes \mathbf{G}) & -(\mathbf{I}_{N_s} \otimes \mathbf{G})^H\mathbf{p} \\ -\mathbf{p}^H(\mathbf{I}_{N_s} \otimes \mathbf{G}) & \mathbf{pp}^H \end{bmatrix}.$$

Thus this problem without the rank-one constraint decreases to a problem of semi-finite programming (SDP) and can be resolved by typical convex optimization algorithms from which we can discover the optimum solution to the problem of digital precoder design [38, 41, 42].

The design of the analog precoder is formulated as:

$$\underset{G}{minimize} \left\| \mathbf{P}_{opt} - \mathbf{GL} \right\|_F^2 \tag{25}$$

Subject to: $\mathbf{G} \in \mathcal{A}_p$

This is essentially a vector approximation problem using phase rotation, and there is a closed form formulation for nonzero elements in $\mathbf{G}$, given by

$$\arg\,(\mathbf{G}_{i.l}) = \arg\,(\mathbf{G})_{i.:}(\mathbf{L})_{l,:,}^H,$$
$$1 \le i \le N_t, l = \left\lceil i\frac{N_{RF}^t}{N_t} \right\rceil \tag{26}$$

Hence, a detailed description of the SDR-AltMin algorithm is given Algorithm 4 [38].

Algorithm 4 Semidefinite Relaxation Algorithm
Input: $\mathbf{P}_{opt}$
1: *Construct* $\mathbf{G}^{(0)}$ with random phase and set $k = 0$;
2: **repeat**
3: Fix $\mathbf{G}^{(k)}$, solving = $\mathbf{L}^{(k)}$ using Semidefinite relaxation (24)
4: Fix $\mathbf{L}^{(k)}$ and update $\mathbf{G}^{(k+1)}$ by; (26)
5: $k \leftarrow k+1$;
6: **till** a stopping criterion occurs ;

Fig 20 demonstrates that the SDR-AltMin algorithm is superior to the SIC algorithm. This is primarily due to the full use of the digital precoder by the SDR-AltMin algorithm, whereas the method based on SIC only utilizes the digital precoder in order to distribute the power to the data streams.
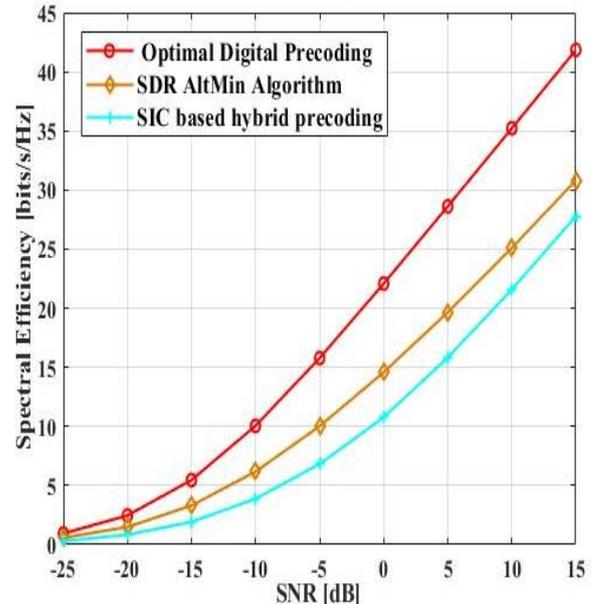

Figure 20. Spectral efficiency obtained by different precoding algorithms when $N_t^{RF} = N_s = 4$

3.5   Deep Learning-Based Hybrid Precoding

In this technique, the authors investigated low complexity hybrid precoding and also studied the design for (mm-Wave) MIMO systems using deep neural network, specifically CNN, that is why the proposed hybrid beamforming design is called HBDL, as an abbreviation to the hybrid beamforming scheme that based on deep learning. Specifically, they used two different CNNs to jointly optimize the precoding and combiner of a single user mm-Wave massive MIMO system. Additionally, they compared the performance of hybrid precoding based on deep learning (HBDL) to the other optimization-based hybrid beamforming techniques. The proposed deep learning-based design has been compared with legacy optimization-based hybrid beamforming techniques. It turns out that the deep learning-based approach presents an excellent performance with its legacy counterparts with much-reduced complexity as shown in figure 21 [43].
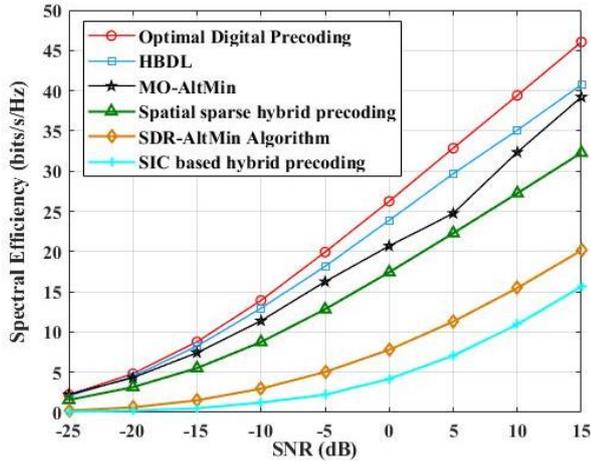
Figure 21. Spectral efficiency of the fully and partially-connected structures obtained by different precoding algorithms when $N_{RF}^t = N_{RF}^r = N_S = 4$, $N_R = N_T = 36$

4. Challenges Of Hybrid Precoding and Future Research Criteria

The hybrid precoding techniques have been investigated widely for the microwave and (mm Wave) frequencies, on the other hand, the Terahertz (THz) ultra-massive (UM) MIMO systems characteristics cause considerable challenges to the design of THz hybrid precoding.

4.1 Hardware-Optimized Implementation

Increasing the frequency to the THz band leads to increasing hardware challenges. For instance, it is still difficult to generate the high-resolution phase shifter and RF chain. Hence, by considering low-resolution phase shifter and RF chain, the need for THz hybrid Precoding architectures and algorithms for practical implementation becomes a necessity. a further prospective for future research is to investigate the usage of lens array for achieving THz hybrid beamforming, i.e., the THz Wave is focused by using an electromagnetic lens and a matching antenna array on the focal surface of the lens, which in turn cancels the need for phase shifters and so the hardware complexity is lowered significantly. However, algorithms are needed for choosing beams from ultra-massive antennas, that increases the hardware complexity. so, there is a need for more further researching on hardware-optimized lens array architectures and algorithms.

4.2 Impact of Imperfect Channel State Information

The THz UM-MIMO channel is high-dimensional, as it contains exceptionally high antennas that provide difficulty in finding perfect channel state information (CSI). In addition, the decline in performance caused by poor CSI and beam distortion is substantial with an ultra-sharp beam created by THz UM-MIMO systems. Thus, for THz hybrid beamforming architecture, super-resolution channel and angle calculation algorithms need to be improved. On the other side, it is necessary to investigate THz hybrid precoding algorithms which alleviates imperfect CSI. The probabilistic method is an appealing concept for the development of the strong THz hybrid algorithm.

4.3 Hybrid Precoding Based on Deep Learning Algorithms

Despite considerable hybrid beamforming techniques have been researched, their computational complexity is often

dependent on number of antennas. For example, Computational complexity is very high for THz hybrid precoding with extremely large number of antennas that reach more than thousand antennas. DL algorithms have recently attracted considerable interest for the resolution of high-complexity, Computational time and resource allocation problems. According to that, paper [43] suggests using deep learning method to handle the precoding problem hybrid in mmWave frequency. The challenges must be considered carefully in case of using deep learning algorithms for THz hybrid precoding.

5. Conclusions

This paper surveys the state-of-the-art approaches of hybrid beamforming. The analog beamforming requires low hardware complexity, but it can provide a poor performance compared to the performance of the fully digital precoding. The fully digital beamforming, on the other hand, is difficult to be realized. The hybrid precoding or beamforming can improve the spectral efficiency of the mm-wave massive MIMO wireless communication system without high hardware requirements. The deep learning-based design has been compared with legacy optimization-based hybrid precoding techniques. It turns out that the deep learning-based approach presents an excellent performance compared to its legacy counterparts with much-reduced complexity and computational time, while the performance of the SIC-based is the worst amongst the hybrid precoding algorithms. Finally, this paper overviews challenges of implementing the hybrid precoding techniques with future THz wireless communication systems.

References

[1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. J. W. n. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," vol. 21, no. 8, pp. 2657-2676, 2015.

[2] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution," IEEE Communications Magazine, vol. 52, no. 9, pp. 56-62, 2014, doi: 10.1109/MCOM.2014.6894453.

[3] C. Dehos, J. L. González, A. D. Domenico, D. Kténas, and L. Dussopt, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems," IEEE Communications Magazine, vol. 52, no. 9, pp. 88-95, 2014, doi: 10.1109/MCOM.2014.6894457.

[4] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "Mimo for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both," IEEE Communications Magazine, vol. 52, no. 12, pp. 110-121, 2014.

[5] S. K and T. srinivasulu, "Design and Development of Novel Hybrid Precoder for Millimeter-Wave MIMO System," 2021, vol. 13, no. 3, 2021-12-25 2021, doi: 10.54039/ijcnis.v13i3.5096 J International Journal of Communication Networks and Information Security (IJCNIS).

[6] J. Karjalainen, M. Nekovee, H. Benn, W. Kim, J. Park, and H. Sungsoo, "Challenges and opportunities of mm-wave communication in 5G networks," in 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2-4 June 2014 2014, pp. 372-376.

[7] M. R. Akdeniz, Y. Liu, S. Rangan, and E. Erkip, "Millimeter wave picocellular system evaluation for urban deployments," in 2013 IEEE Globecom Workshops (GC Wkshps), 9-13 Dec. 2013 2013, pp. 105-110.

[8] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," Proceedings of the IEEE, vol. 102, no. 3, pp. 366-385, 2014.

[9] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60 GHz Indoor Networking through Flexible Beams: A Link-Level Profiling," presented at the Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Portland, Oregon, USA, 2015. [Online]. Available: https://doi.org/10.1145/2745844.2745858.

[10] Z. Pi and F. Khan, "A millimeter-wave massive MIMO system for next generation mobile broadband," in 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 4-7 Nov. 2012 2012, pp. 693-698.

[11] J. Chen, "Advanced architectures for efficient mm-Wave CMOS wireless transmitters," University of California at Berkeley, 2014.

[12] S. Schindler, "Introduction to MIMO. Application Note, Rhode & Schwarz.," 2009.

[13] S. Payami, "Hybrid beamforming for massive MIMO systems," Doctoral dissertation, University of Surrey, 2017.

[14] S. Mumtaz, J. Rodriguez, and L. Dai, mmWave Massive MIMO: A Paradigm for 5G. Elsevier Science, 2016.

[15] W. Junyi et al., "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," IEEE Journal on Selected Areas in Communications, vol. 27, no. 8, pp. 1390-1399, 2009.

[16] W. Hong et al., "Multibeam Antenna Technologies for 5G Wireless Communications," IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 6231-6249, 2017.

[17] P. Chen, W. Hong, H. Zhang, J. Chen, H. Tang, and Z. Chen, "Virtual Phase Shifter Array and Its Application on Ku Band Mobile Satellite Reception," IEEE Transactions on Antennas and Propagation, vol. 63, no. 4, pp. 1408-1416, 2015.

[18] G. Xiang, H. Hashemi, and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," IEEE Journal of Solid-State Circuits, vol. 39, no. 12, pp. 2311-2320, 2004.

[19] S. Jeon et al., "A Scalable 6-to-18 GHz Concurrent Dual-Band Quad-Beam Phased-Array Receiver in CMOS," IEEE Journal of Solid-State Circuits, vol. 43, no. 12, pp. 2660-2673, 2008.

[20] I. Ahmed et al., "A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3060-3097, 2018.

[21] Y. Chen, "Low Complexity Precoding Schemes for Massive MIMO Systems " Doctor of Philosophy School of Engineering Newcastle University June, 2019.

[22] D. Kaur, N. J. I. J. o. C. N. Kumar, and I. Security, "Capacity enhancement of multiuser wireless communication system through adaptive non-linear pre coding," vol. 10, no. 1, pp. 67-78, 2018.

[23] S. Kutty and D. Sen, "Beamforming for Millimeter Wave Communications: An Inclusive Survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 949-973, 2016.

[24] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the Number of RF Chains and Phase Shifters, and Scheduling Design With Hybrid Analog–Digital Beamforming," IEEE Transactions on Wireless Communications, vol. 15, no. 5, pp. 3311-3326, 2016.

[25] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, 2013.

[26] F. Rusek et al., "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40-60, 2013.

[27] S. Buzzi, I. C, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead," IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 697-709, 2016.

[28] D. Jiang and B. Natarajan, "Hybrid precoding with compressive sensing based limited feedback in massive MIMO systems," vol. 27, no. 12, pp. 1672-1678, 2016, doi: 10.1002/ett.3108.

[29] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and Analyzing Millimeter Wave Cellular Systems," IEEE Transactions on Communications, vol. 65, no. 1, pp. 403-430, 2017.

[30] W. Roh et al., "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.

[31] F. Sohrabi, "Hybrid Beamforming and One-Bit Precoding for Large-Scale Antenna Arrays," Doctoral dissertation, Electrical and Computer Engineering, 2018.

[32] R. Zi, X. Ge, J. Thompson, C. Wang, H. Wang, and T. Han, "Energy Efficiency Optimization of 5G Radio Frequency Chain Systems," IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 758-771, 2016.

[33] S. Salman, "Analysis of Antenna and RF Front-End Topologies for Multi-Beam Systems," Master of Science, Delft University of Technology, 2018.

[34] C. Kim, T. Kim, and J. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in 2013 IEEE Globecom Workshops (GC Wkshps), 9-13 Dec. 2013 2013, pp. 61-65.

[35] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. J. I. t. o. w. c. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," vol. 13, no. 3, pp. 1499-1513, 2014.

[36] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in 2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 17-20 June 2012 2012, pp. 100-104.

[37] P. A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds. Princeton University Press, 2009.

[38] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 485-500, 2016.

[39] D. P. Bertsekas, "Nonlinear Programming," Journal of the Operational Research Society, vol. 48, no. 3, pp. 334-334, 1997/03/01 1997, doi: 10.1057/palgrave.jors.2600425.

[40] Y. Liang, E. Y. Cheu, L. Bai, and G. Pan, "On the Relationship Between MMSE-SIC and BI-GDFE Receivers for Large Multiple-Input Multiple-Output Channels," IEEE Transactions on Signal Processing, vol. 56, no. 8, pp. 3627-3637, 2008.

[41] W. K. K. Ma, "Semidefinite relaxation of quadratic optimization problems and applications," IEEE Signal Processing Magazine, 1053(5888/10). 2010.

[42] S. Boyd, S. P. Boyd, and L. Vandenberghe, Convex optimization. Cambridge university press, 2004.

[43] I. Osama, M. Rihan, M. Elhefnawy, and S. Eldolil, "Deep Learning Based Hybrid Precoding Technique for Millimeter-Wave Massive MIMO Systems," in 2021 International Conference on Electronic Engineering (ICEEM), 3-4 July 2021 2021, pp. 1-7, doi: 10.1109/ICEEM52022.2021.9480386.