Key Technologies for Beamforming in Millimeter Wave Communication System

Guanyu Qian¹, Weikun Liu¹,Dengfeng Hao¹ and Wei Li*¹

¹Smart Shine Microelectronics Technology Co. Ltd, Beijing, China;
*Corresponding author’s e-mail: liuweikun@sgitg.sgcc.com.cn

Abstract. Millimeter wave (mmWave) is one of the key technologies of 5G. It can solve the problem of insufficient spectrum resources. However, mmWave links are easily affected by the environment, such as the user equipment (UE) mobility, link blocking, etc. Moreover, the path loss and atmospheric absorption will also cause a great loss to the channel quality. Dealing with these challenges, beamforming is an effective solution. This paper tracks the progress of beamforming in mmWave communication system, including hybrid beamforming structure, mmWave channel state information (CSI) acquisition, and beam management methods during UE movement. We explore the applicability of different beamforming technologies, and discuss the future challenges in this domain.

1. Introduction
With the increasing demand for high data throughput in communication systems, mmWave communication has become the key technology for 5G[1]-[3]. By providing wide bandwidth resources, mmWave can overcome the problem of insufficient bandwidth resources in the low-frequency band[4]. In the mmWave frequency band, the electromagnetic wave will suffer a series of losses from path loss, shadow, atmospheric absorption, and so on. However, the high directional beamforming using large-scale antenna arrays can concentrate the signal energy in a certain direction, which can make up for the loss to a certain extent. For large-scale antenna arrays mmWave system, full digital beamforming requires one dedicated radio frequency (RF) chain per antenna element, which has unaffordable hardware and power consumption cost, so most mmWave systems use hybrid beamforming[5]. Meanwhile, the gain obtained by beamforming is closely related to the acquisition of channel state information (CSI). At present, the main methods to obtain channel information are beam training and channel estimation. High directional links need to align the beam of transmitter and receiver to achieve the desired system performance, especially in mobile scenarios, which require a set of operations called beam management. During movement, the main operations of beam management include beam tracking and beam handover.

This paper gives a comprehensive overview of beamforming in mmWave communication. It mainly includes the description of mmWave beamforming structure, CSI acquisition, beam measurement, and mobility management in beam space. The structure of this paper is shown in figure 1. As the attention of this paper is the survey of mmWave beamforming, this review generally doesn’t include the specific details of beamforming. But the basic concepts and the overview of beamforming technology can make readers clearly understand the recent development of beamforming.
2. Hybrid beamforming

In the communication system, large-scale antenna arrays mmWave system is realized by hybrid beamforming. It separates the beamformer into two parts, and one is low-dimensional baseband digital, the other is high-dimensional analog implemented with phase shifters, which reduces power consumption and guarantees the beamforming gain. In this part, according to the system model and structure of hybrid beamforming, we track the development of hybrid beamforming in millimeter wave communication.

2.1. System model

In a single-user MIMO system, the hybrid beamforming transmitter has $N_t$ antennas, $N_s$ independent data streams, and $N_R'_{RF}$ RF links, $N_s \leq N_R'_{RF} \leq N_t$. The receiver has $N_r$ antennas and $N_R$ RF links. At the transmitter, behind the RF link is a network composed of switches and phase shifters, which extends $N_R'_{RF}$ digital outputs to $N_t$ precoded analog signals on the antenna port. Similarly, at the receiver, there are the phase shifter network and switch network behind the antenna port. For simplicity, we assume $N_R'_{RF} = N_R'_{RF} = N_t$ and $H$ is the $N_t \times N_r$ channel matrix, and the transmitted symbol $\mathbf{s}$ is the vector of $N_s \times 1$ and satisfied $E[\mathbf{s}\mathbf{s}^H] = N_s^{-1}\mathbf{I}_{N_s}$, then the received analog signal at the receiver is:

$$\mathbf{y} = \sqrt{p}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{z}$$ (1)

Here $\sqrt{p}$ is the average received power, $\mathbf{F}_{BB} \in \mathbb{C}^{N_s \times N_R'_{RF}}$ is the digital precoding matrix, $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_R'_{RF}}$ is the analog precoder, and the set $\mathbf{F}$ is determined by the specific hardware scheme used. $\mathbf{z}$ is the noise vector obeying complex Gaussian distribution $CN(\mathbf{0},\sigma^2\mathbf{I}_{N_s})$. On the receiver, the analog combiner maps $N_r$ inputs to $N_R'_{RF}$ RF links, and then uses the digital combiner to process them in baseband. The final received signal is

$$\mathbf{r} = \sqrt{p}\mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{W}_{BB}^*\mathbf{W}_{RF}^*\mathbf{z}$$ (2)
$W_{BB} \in \mathbb{W}^{N_t \times N_r}$ is the digital combination matrix of the receiver and $W_{RF} \in \mathbb{W}^{N_t \times N_{RF}}$ is the analog combination matrix. We hope to design $W_{BB}, W_{RF}, F_R, F_{BB}$ to minimize the estimation error of $s$ from $r$ [6], [7].

2.2. The architecture of hybrid beamforming

The goal of all hybrid beamforming architectures is to reduce the complexity of hardware and signal processing, and to provide optimal performance, that is, near the performance of pure digital beamforming.

In the aspect of the hardware architecture of hybrid beamforming, the typical structures are fully connected and sub-connected. This paper only gives an overview of the structure of beamforming at the transmitter, and the structure of the receiver is similar. As shown in figure 2-(a), in a fully connected hybrid beamforming structure, each RF link is connected with $N_t$ phase shifters, so there are $N_{RF} \times N_t$ phase shifters in this structure. In figure 2-(b), there is the sub-connected hybrid beamforming structure, each RF link is connected with $N_t / N_{RF}$ phase shifters, so there are $N_t$ phase shifters. In terms of spectral efficiency (SE), the fully connected structure provides a gain of $N_{RF} \log_2 (N_{RF})$ over the sub-connected structure, but also has $N_{RF}$ times energy consumption[8].

![Figure 2. Hybrid beamforming architectures in mmWave massive MIMO systems.](image)

To reduce the signaling overhead and complexity of fully connected architecture, there is a full connection architecture based on the “virtual sector”[9], as shown in figure 3. It groups users in the analog beamformer and serves them together in the digital beamformer. In order to improve the spectral efficiency of the sub-link structure, a hybrid beamforming structure of dynamic sub-array is proposed [10]. As shown in figure 4, the selection of its dynamic sub-array is based on the channel covariance matrix, which can maximize the mutual information rate.

![Figure 3. The structure of full connection based on the "virtual sector".](image)

![Figure 4. The structure of dynamic sub-array.](image)
In figure 5, we compare the spectral efficiency of several hybrid beamforming structures mentioned above with a single user, and $N_r = 64$, $N_t = 4$, $N_{RF}^r = N_{RF}^t = 4$.

![Figure 5. The comparison of different precoder structures.](image)

3. CSI acquisition

In mmWave communication, the large-scale antenna array is used to generate a highly directional beam to compensate for severe path loss. However, the gain of beamforming depends on the accuracy of channel state information. The acquisition methods of CSI in mmWave Massive MIMO mainly include beam training and channel estimation. Beam training avoids the estimation of high-dimensional channel matrix by selecting the best beam, and channel estimation could use compressed sensing or neural network to estimate high-dimensional channel matrix to obtain more accurate CSI.

3.1. Beam training

Beam training is also called beam alignment. How to find the best beam efficiently is the main purpose of beam training. Codebook based beam training method is widely used. In this training method, a pair of codewords are selected at transmitter and receiver to generate the beam, and then measuring the reference signal receiving power (RSRP). The purpose of beam training is to find the largest RSRP beam, that is, the best codeword pair. The main beam training methods include beam sweeping[11], hierarchical beam training[12], [13], and machine learning based methods[15]:

- For beam sweeping, all beams need to be measured to find the best beam quality, but this exhaustive search method costs a huge resource load and time. To reduce the complexity of this process, the reference [11] designed an efficient beam sweeping algorithm, called 3-dimensional peak finding (3DPF), which can find the best beam in logarithmic time. It describes the beam sweeping process as a sparse problem, and uses a compressed sensing method to determine the minimum number of measurements required in this process. Based on beam partitioning and gradient ascending search, 3DPF collects the number of beam probes needed for CS analysis, and allows mmWave devices to discover each other directly. It can maintain a high beamforming gain and a low misdetection rate, and compared with the scanning approach in 802.11, it reduces the search time by 90%.
Hierarchical beam training uses hierarchical code to train the beam[12], [13]. The width and resolution of the beam represented by the code words in different layers are different. The hierarchical codebook is designed by using the adaptive design method[13]. The current layer codewords are designed according to the training results of the previous layer code words. The scheme can approach the performance of beam scanning.

In recent years, machine learning (ML) has been applied to solve different problems in physical layer communication[14]. In reference [15], it proposes an alignment method with partial beams using ML for multi-user mmWave massive MIMO systems. It simulates the channel environment according to mmWave channel model for off-line training, and then deploys online, using partial beams to predict the beam distribution vector. Then, according to the beam distribution vector, the beams of all users are aligned at the same time. Different from the existing method based on the hierarchical codebook, this method aligns beams for all users at the same time and saves the training time greatly. In addition, it doesn’t need prior knowledge to train the neural network, such as user location information, so it can significantly reduce the system overhead.

The performance of the above beam training methods are all near the beam sweeping methods, and the main difference is reflected in the computational complexity and overhead. The specific comparison is shown in table 1.

Table 1. Comparisons of different beam training schemes.

| Scheme                      | Computational complexity | Overhead |
|-----------------------------|--------------------------|----------|
| Beam sweeping[11]           | Low                      | Medium   |
| Hierarchical beam training[13] | Low                      | High     |
| ML based method[15]         | High                     | Low      |

3.2. Channel estimation
CSI information can be accurately obtained by channel estimation. In the mmWave scenario, to reduce the high complexity of large-scale antenna array, the main channel estimation methods are based on compressed sensing and machine learning:

- In the hybrid beamforming precoding mmWave channel, most non-line of sight (NLOS) paths are much weaker than the line of sight (LOS) paths due to the weak scattering and reflection ability of mmWave, so mmWave MIMO channel is sparse in the angular domain, and sparse channel estimation is modeled as a problem of compressed sensing. A sparse channel estimation method based on $\ell_{1/2}$ regularization is proposed[16], and in this method, an objective function is constructed, which is the weighted sum of $\ell_{1/2}$ regularization and error constraints. Then the gradient descent method is used for iterative optimization, and the weight parameters in the function are updated in each iteration, to realize the super-resolution channel estimation.

- Machine learning has the characteristics of offline training and online operation. It also uses a neural network to extract features, which has a unique advantage compared with traditional algorithms. Therefore, the channel parameters can be estimated by the neural network, including AOAs, AODs, gain, etc. A channel estimation method based on spatial frequency CNN (SFCNN) is proposed[17]. It is based on space and frequency correlation, and the channel matrix on adjacent subcarriers is input to CNN at the same time. Then, a spatial-frequency-temporal CNN (SFT CNN) method is proposed based on the temporal correlation in time-varying channels, which further improves the accuracy.
4. Beam management during movement

In the process of UE moving, it is necessary to adjust the beam to make the beam between the transmitter and the receiver align to ensure the gain of beamforming. The whole beam management process includes beam scanning, beam measurement, beam determination, and beam reporting:

- **Beam scanning**, the base station sends directional beams in sequence to scan the beam in a specific angle domain space.
- **Beam measurement**, the UE receives the beam and measures the beam quality. The measurement quantity can be equal to SNR and RSRP.
- **Beam determination**, the mmWave base station allocates one or several optimal service beams to the UE based on the feedback information of the UE.
- **Beam reporting**, UE selects one or several best beams from the measured beam quality and reports them to mmWave base station.

In figure 6, the flow chart of UE initial access in downlink channel, in which Synchronization Signal Block (SSB) is the reference signal in downlink beam management. When the UE determines the best beam, the base station needs to scan the beam again so that the UE can obtain the corresponding resources of the best beam, and send the Random Access Channel (RACH) preamble to the base station to make the base station know the best beam information selected by the UE.

![Figure 6. The process of beam management](image)

The operation of beam management in this process mainly includes beam tracking and handover.

4.1. Beam tracking

To solve the problem of beam alignment, a direct method is beam training mentioned above, but the overload is too high, and it is not suitable for fast changing environment, especially in the process of UE moving. By estimating AOAs and AODs to track the change of channel, the beam can be adjusted effectively. At present, the main beam tracking methods include Kalman filter, particle filter, and machine learning.

Kalman filter is the main scheme of beam tracking, including extended Kalman filter (EKF) and unscented Kalman filter (UKF), and the specific methods are introduced in detail in references [18] and [19]. The disadvantage of the Kalman filter is the accumulation of estimation errors. In reference [20], a beam tracking method based on adaptive beamwidth control is proposed, which uses the particle filter to estimate channel parameters. The main innovation is that the beamwidth can be changed adaptively with the estimation error. The method of machine learning can also be used for beam tracking. The neural network can be used to predict AOA and AOD under LoS conditions[21], which can achieve a compromise in spectrum efficiency and load overhead. In reference [22], the beam and transmission rate...
is selected by using the reinforcement learning method, and then the beam and transmission rate used in the next transmission period is selected according to the updated posterior distribution.

4.2. Beam handover
When there is a sudden interrupt between the UE and the base station, or when the UE moves to the edge of beam coverage, beam handover is needed. In the 3GPP standard[23], two thresholds are defined for cell handover, which are time-to-trigger (TTT) $\beta$ and hysteresis $\Delta$. When the difference of the RSRP between the adjacent cell and the current service cell exceeds $\Delta$ and lasts for $\beta$, the handover process will be triggered, as shown in figure 7.

![Figure 7. The diagram of the handover](image)

In the actual handover process, to achieve different purposes, such as reducing the handover overhead, improving the system performance, reducing the handover failure rate, and so on, there are different research concerns. Therefore, when to handover and how to select the target beam are important issues in the beam handover. Reference [24] proposes a method to adjust the user's handover frequency according to the user's mobile speed. In this paper, the UE with different mobile speeds is divided into multiple inter beam handover class (IBCH). According to different UE mobile levels, different handover frequencies and the number of adjacent beams can be calculated as a handover period, to determine the handover strategy. Although this method reduces the handover frequency, the utilization of spectrum resources is low because the adjacent beams serve a UE. In reference [25], it proposes a way to consider handover Strategy based on effective beam coverage probability (EBCP). EBCP is defined as the probability that UE is in the coverage area of the same beam at the next time point. The location of UE is obtained by GPS and the moving speed of UE is estimated. The Gauss Markov model is established for the moving mode of UE. The distribution of EBCP is obtained according to the geometric relationship and the moving mode of UE. Finally, the handover beam is selected for UE and the handover decision is made according to the judgment of hysteresis, TTT, and EBCP. In reference [26], a beam handover algorithm based on fuzzy logic is proposed. According to the geometric relationship between UE and beam, the algorithm formulates a fuzzy logic block, and identifies the target beam with the largest displacement of UE in the beam coverage by fuzzy logic block to reduce the handover frequency and improve the stability of the connection.

In vehicle networking, a beam switching algorithm based on machine learning is proposed[27]. It uses the continuous CSI data collected in the low-frequency link, then uses ML algorithm to predict the position of the vehicle to preactivate the mmWave remote radio unit (RRU) nearby, and use K-Nearest Neighbor (KNN) algorithm to directly determine the handover target for the handover vehicle according
to the historical handover data. This method uses the collected CSI to predict the vehicle location, so it does not need additional signaling overhead.

A switching method based on reinforcement learning is proposed[28]. When switching occurs, the optional BS is used as an action, and the ultimate goal is to select the action in the current state of UE that may produce the maximum reward in the long term as the switching target. The reward can be the beam quality or system throughput after handover. This method only needs to look up the reward table when switching online, and it has low time complexity and can find the best choice with long-term performance. But at the same time, this method is limited to the training scenario, so its practical value is not high.

5. Conclusion and future work
Beamforming technology enables signal energy to be concentrated in a certain direction, which helps mmWave link overcome its inherent difficulties, so it also greatly attracts research interest. In this paper, beamforming is reviewed from a global perspective, and the latest research is investigated. The structure of mmWave hybrid beamforming is analyzed and compared. It also summarizes the current two methods of channel state information acquisition, focusing on the latest algorithms. Considering the dynamic change of beam in mmWave mobile scene, we analyze the related beam management technology. Although some propositions have been put forward, it is important to further study the complexity, accuracy, and cost related issues of beamforming in the future.

References
[1] Rappaport et al, T.S. (2013) Millimeter wave mobile communications for 5G cellular: It will work! IEEE Access., 1: 335–349.
[2] Pi, Z., Khan, F. (2011) An introduction to millimeter-wave mobile broadband systems. IEEE Commun. Mag., 49: 101–107.
[3] Roh et al, W. (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results. IEEE Commun. Mag., 52: 106–113.
[4] Rangan, S., Rappaport, T.S., Erkip, E. (2014) Millimeter-wave cellular wireless networks: Potentials and challenges. Proc. IEEE., 102: 366–385.
[5] Roh, W., Seol, J., Park, J., Lee, B., Lee, J., Kim, Y., Cho, J., Cheun, K., Aryanfar, F. (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. IEEE Commun. Mag., 52: 106–113.
[6] Lin, T., Cong, J., Zhu, Y., Zhang, J., Ben Letaief, K. (2019) Hybrid Beamforming for Millimeter Wave Systems Using the MMSE Criterion. IEEE Transactions on Communications. 67: 3693-3708.
[7] Ioushua, S. S., Eldar, Y. C. (2014) A Family of Hybrid Analog–Digital Beamforming Methods for Massive MIMO Systems. IEEE Transactions on Signal Processing., 52: 106–113.
[8] Han, S., I, C.L., Xu, Z., Rowell, C. (2015) Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. IEEE Commun. Mag., 53: 186–194.
[9] Molisch et al, A.F. (2017) Hybrid Beamforming for Massive MIMO: A Survey. IEEE Communications Magazine., 55: 134-141.
[10] Park, S., Alkhateeb, A., Heath, R.W. (2017) Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems. IEEE Transactions on Wireless Communications., 16: 2907-2920.
[11] Aykin, I., Krunz, M. (2020) Efficient Beam Sweeping Algorithms and Initial Access Protocols for Millimeter-Wave Networks. IEEE Transactions on Wireless Communications., 19: 2504-2514.
[12] Noh, S.M., Zoltowski, D., Love, D.J. (2017) Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. IEEE Trans. Wireless Commun., 16: 5689–5701.
[13] Chen, K., Qi, C., Dobre, O. A., Li, G. (2019) Simultaneous Multi-user Beam Training Using Adaptive Hierarchical Codebook for mmWave Massive MIMO. In: 2019 IEEE Global Communications Conference (GLOBECOM). Hawaii. pp. 1-6.

[14] Qin, Z., Ye, H., Li, G.Y., Juang, B.H. (2019) Deep learning in physical layer communications. IEEE Wireless Commun., 26: 93–99.

[15] Ma, W., Qi, C., Li, G.Y. (2020) Machine Learning for Beam Alignment in Millimeter Wave Massive MIMO. IEEE Wireless Communications Letters., 9: 875-878.

[16] Zhang, Z., Shi, W., Yuan, L., Gu, G. (2019) Regularization-Based Super-Resolution Sparse Channel Estimation for MmWave Massive MIMO Systems. IEEE Access., 7: 75837-75844.

[17] Dong, P., Zhang, H., Li, G.Y., Gaspar, I. S., NaderiAlizadeh, N. (2019) Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems. IEEE Journal of Selected Topics in Signal Processing., 13: 989-1000.

[18] Jayaprakasam, S., Ma, X., Choi, J.W., Kim, S. (2019) Robust Beam-Tracking for mmWave Mobile Communications. IEEE Communications Letters., 13: 989-1000.

[19] Song, H. -L., Ko, Y. -C. (2021) Robust and Low Complexity Beam Tracking With Monopulse Signal for UAV Communications. IEEE Transactions on Vehicular Technology., 70:3505-3513.

[20] Chung, H., Kang, J., Kim, H., Park, Y.M., Kim, S. (2021) Adaptive Beamwidth Control for mmWave Beam Tracking. IEEE Communications Letters., 25: 137-141.

[21] Wang, C., Chen, Y., Lu, Z., Wen, X., Wang, Z. Wang, L. (2021) FCBET: A Fast Consecutive Beam Tracking Scheme for MmWave Vehicular Communications. In: 2021 IEEE Wireless Communications and Networking Conference (WCNC). Nanjing. pp. 1-6.

[22] Aykin, I., Akgun, B., Feng, M., Krunz, M. (2020) MAMBA: A Multi-armed Bandit Framework for Beam Tracking in Millimeter wave Systems. In: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. Beijing. pp. 1469-1478.

[23] 3rd Generation Partnership Project (3GPP). (2020) Technical Specification (TS) 38.300: Technical Specification Group Radio Access Network. 3GPP. Sophia Antipolis.

[24] Kim, J. S., Lee, W. J., Chung, M.Y. (2016) A multiple beam management scheme on 5G mobile communication systems for supporting high mobility. In: 2016 International Conference on Information Networking (ICOIN). Kota Kinabalu. pp. 260-264.

[25] Chen, Y., Hsu, T., Wang, L. (2017) Improving Handover Performance in 5G mm-Wave HetNets. In: GLOBECOM 2017 - 2017 IEEE Global Communications Conference. Singapore. pp. 1-6.

[26] Kose, A., Foh, C. H., Lee, H., Dianati, M. (2020) Beam-centric Handover Decision in Dense 5G-mmWave Networks. In: 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. London. pp. 1-6.

[27] Yan et al, L. (2019) Machine Learning-Based Handovers for Sub-6 GHz and mmWave Integrated Vehicular Networks. IEEE Transactions on Wireless Communications., 18: 4873-4885.

[28] Yajnanarayana, V., Rydén, H., ‘Evizi, L. (2020) 5G Handover using Reinforcement Learning. In: 2020 IEEE 3rd 5G World Forum (5GWF). Bangalore. pp. 349-354.