Received April 29, 2019, accepted May 17, 2019, date of publication May 22, 2019, date of current version June 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918281

Fast Beam Search for Massive MIMO Based on Mainlobe Overlapping State of Training Beam

WEIXIA ZOU[1,2], MINGYANG CUI[1], AND CHAO GUO[1]
[1]Key Laboratory of Universal Wireless Communications, MOE, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China

Corresponding author: Weixia Zou (zwx0218@bupt.edu.cn)

This work was supported in part by the NSFC under Grant 61571055, in part by the SKL MMW under Grant K201815, and in part by the Important National Science and Technology Specific Projects under Grant 2017ZX03001028.

ABSTRACT To reduce beam search complexity without decreasing feedback efficiency for millimeter-wave communication in the LoS environment, this paper proposes a fast beam search scheme based on joint judgment, which requires a masterly design of the training beams and let their mainlobe overlap according to certain rules. As a consequence, its state utilization efficiency has been improved to 100% while keeping feedback efficiency is still 100%. In this paper, through theoretical analysis, we find that to encode the index of every subinterval A_q using the Gray mapping can decrease the possibility and the impact of misestimating, compared with using a binary index, which is adopted. The transceiver emits the training beam in turn and then jointly determines optimal communication beam pair according to the relationship between the receiving power and the threshold. The simulation results show that our proposed scheme is more efficient and its search complexity has been further decreased while its FE remains 100%. Especially, this method has more obvious advantages in multi-user simultaneously beam search scenarios.

INDEX TERMS Fast beam search, millimeter-wave, Gray mapping, state utilization efficiency.

I. INTRODUCTION

To meet the rapid growth of high-speed wireless transmissions and mitigate severe spectrum resource scarce in the low frequency bands, millimeter wave (mmWave) communication has become a research hotspot [1], [2]. One merit promised by high frequency bands is the integration of more antennas. The large-scale antenna array not only greatly improve the spectrum utilization, but also obtain the considerable array gain, which can be beneficial to remedy the bad link quality aroused by tremendous propagation attenuation in mmWave bands [3]–[5]. However, with the number of array antennas increases, acquisition of the channel state information (CSI) will become a major obstacle, as it requires lots of feedback which is unaffordable in term of the time-consumption and energy efficiency. Moreover, the channel maybe already change when the transmitter obtains channel information, that is, the channel has been aging (aging channel) [6], [7]. Therefore, in order to ensure the timeliness of acquired channel information, it is necessary to design a more efficient and faster channel estimation scheme, especially for super-large scale antenna array.

In principle, the beamforming technique can be divided into two types, namely adaptive beamforming [8]–[11] and switching beamforming [12]–[16].Adaptive beamforming can automatically start to search process of optimal communication beam according to the change of the channel. Specifically, the authors of [10], [11] propose training designs to reduce training overhead via interleaving training and feedback. However, the high computational complexity and the high demand for hardware design, to some extent, may limit its practical application. In contrast, switching beamforming, which probes on a predefined codebook and searches the best antenna weight vector, seems to be more practical for low complexity scenarios [11]. The acquisition of CSI is actually a beam search process for beam switching system. Owing to both the low complexity and cost, the current standards on mmWave wireless local area network (WLAN) adopt this beam steering structure. In the emerging 5G mmWave mobile communication and future high speed mmWave WLAN, it will use the hybrid analog and digital beamforming technology [17], [18], which also involves the fast beam search process.

The associate editor coordinating the review of this manuscript and approving it for publication was Wenchi Cheng.

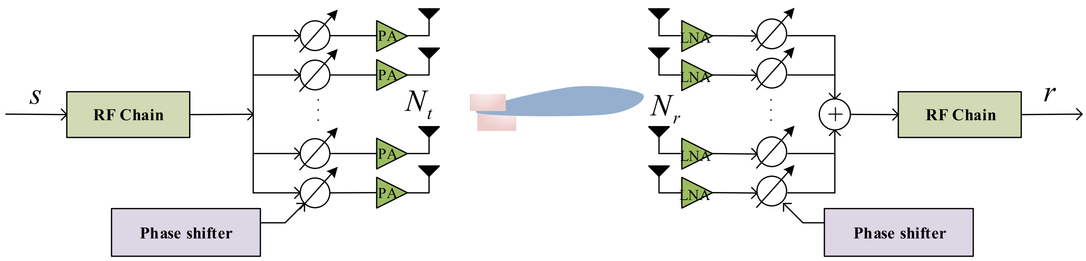

FIGURE 1. Block diagrams of the system.

At present, IEEE 802.15.3c [12] standard recommends a two-level search scheme, i.e., the sector-level following by the refined beam-level, and the search complexity can be proportional to antenna number square. The search complexity in 802.11ad [13] is reduced to a linear relationship with the antenna number. The dichotomy search scheme has been proposed in [14], whereby the complexity can be reduced to a logarithmic relationship with the antenna number (The search complexity refers to the time slots spent in the beam searching). Reference [15] considers a binary-search like (BSL) scheme, which further decreases the search complexity and improves the search efficiency. The authors of [16] propose the channel estimation scheme using sidelobe overlapping state, and the complexity is further reduced, but its feedback amount is a little big. Furthermore, in the existing beamforming system, the number of communication beams is usually 2 times that of antennas and is the integer power of 2. However, the number of communication beams generated in [16] is odd, and it is not a power relation between K and N in [16]. The amount of estimation stage needs to be rounded up, namely $\lceil \log_K N \rceil$, which causes the estimated sub-region does not exactly coincide with the region covered by the actual communication beam. So, its applicability is limited.

For these reasons, this paper proposes a beam search scheme from a new perspective to further reduce the search complexity without increasing the feedback amount. The contribution of the paper is summarized as follows.

- The concept of state utilization efficiency (SUE) and that of information representation efficiency (IRE) are first proposed, which is used to explain the reason of high search complexity of these methods mentioned above from the perspective of information theory. In addition, the concept of feedback efficiency (FE) is introduced to evaluate the performance.
- We discover that the number of training beams depends on the amount of minimum feedback required in order to obtain 100% SUE by theoretical analysis. Furthermore, the range covered by the training beam mainlobe can be determined by the index of subintervals, which is encoded respectively according to the number of communication beam and Gray mapping codes. Lastly, we discover that the index of subintervals can be coded with Gray mapping to reduce the mainlobe number of the training beam as much as possible.
- The joint judgment based on mainlobe overlapping state of all the training beams for beam search is proposed for the first time.

The remainder of this paper is organized as follows. In section II, we not only define SUE, IRE and FE, but also analyze that of the existing search methods. In section III, an efficient beam search method based on the joint judgment is proposed. Performance analysis is given in section IV. The conclusions are drawn lastly in section V.

II. ANALYSIS OF THE EXISTING BEAM SEARCH METHODS

A. PROBLEM DESCRIPTION

Without loss of generality, we consider an mmWave communication system with half-wavelength uniform linear arrays (ULAs) of N_t and N_r elements equipped at the transmitter and receiver, respectively, as shown in Fig. 1. In order to obtain better communication quality, it is necessary to search the optimal beam-pair to communicate. The transmitter needs to estimate the angle of departure (AoD) and the receiver similarly needs to estimate the angle of arrival (AoA) of the signal. The AoD or AoA is usually estimated based on the feedback by transmitting different training beams.

The objective of beam searching is to select the optimal transmit and receive weight vectors in Fig. 1 of transceiver to optimize the link quality. In this paper, we assume that $N_t = N_r$, the number of beams is Q and $Q = 2N_t$. The Q beams (which are used for communication, so referred to as communication beams in this paper) cover the entire AoD range $(-90°, 90°)$ at both the transmitter and receiver. It means that the Q communication beams actually divide the entire AoD range into Q AoD intervals, and the intervals are denoted by $A_q (q = 0, 1, 2, \ldots, Q-1)$, as shown in Fig. 2. Let AF denote the power of received signals in different directions, which is defined as

$$AF(\theta) = \left| \mathbf{w}^H \mathbf{a}(\theta) \right| = \left| \sum_{n=0}^{P-1} w_n e^{-j\pi n \sin\theta} \right|, \tag{1}$$

where P is the number of antennas, $\mathbf{w} = [w_0, w_1, \ldots, w_{P-1}]^T$ denotes the beam weight vector and $\mathbf{a}(\theta) = [1, e^{-j\pi \sin\theta}, \cdots, e^{-j\pi(P-1)\sin\theta}]$ represents the array response vector.

When the AoD interval estimated by the transmitter belongs to some interval A_q, the beam located in this interval

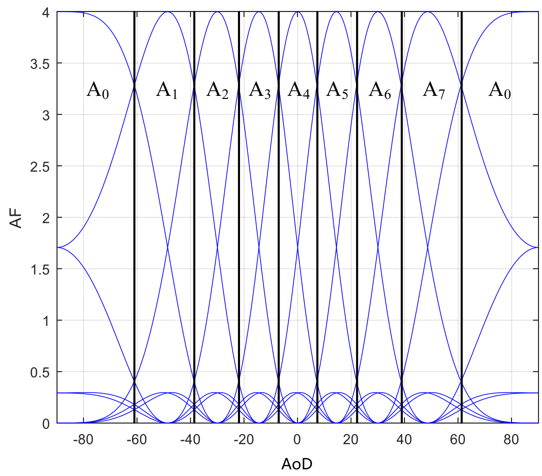

FIGURE 2. Communication beams with $Q = 8$ and corresponding interval A_q.

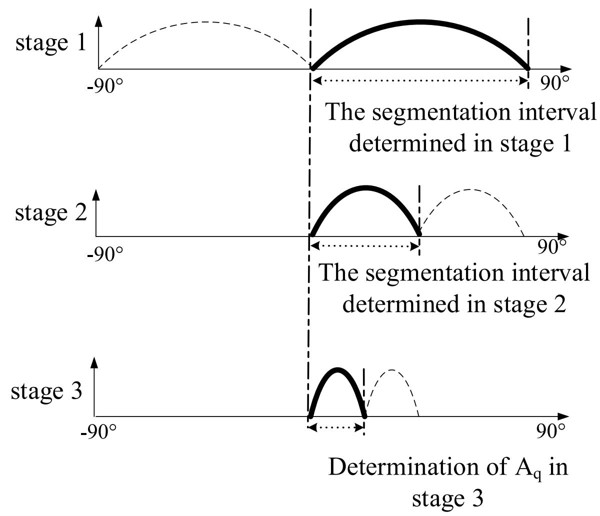

FIGURE 3. Dichotomy search process.

is chosen as the optimal communication beam, and the corresponding beam index q is denoted as $q(opt)$. That is, the optimal interval is $A_{q(opt)}$. In present beam switching systems, AoD estimation is usually performed in conjunction with the optimal beam search. The search mechanism and characteristics of several existing methods are described below.

B. BRIEF INTRODUCTION TO THE EXISTING BEAM SEARCH METHODS

For the traversal search method, training signals are transmitted through the Q different communication beams respectively. The receiver detects and compares all the received signals and feeds back the beam index with the strongest signal strength to the transmitter. It can be seen that, feedback times is 1, the feedback amount is $\log_2 Q$ bits, and the times of transmitting training signals are Q. Here, training beams are the same as communication beams.

The dichotomy method in [14] needs to design beams for fast search (referred to as the training beam in this paper). The mainlobe of the training beam at the beginning is the widest, and it will become narrower as the search process continues until it becomes the same wide as communication beam at the last search stage. The beam width at each stage is approximately twice the width of the beam at the next stage. That is, the search range at each stage is divided into two subintervals, which are covered respectively by two training beams. The transmitter sequentially transmits the training signal through the two beams and the receiver feedbacks the index of the beam with the higher signal strength to transmitter. That is to say, its corresponding subinterval is judged as $A_{q(opt)}$ at this stage. Then this subinterval is used as the beam search range of the next stage. Through this cyclic operation, the granularity of the AoD is refined step by step until a sufficiently fine $A_{q(opt)}$ is found, whose corresponding beam is determined as the optimal communication beam. As shown in Fig. 3, the dichotomy search process is divided into $\log_2 Q$ phases. At each stage, the transmitter needs to transmit 2 training signals, with 1 bit binary number to express the beam index, so the amount of feedback is 1 bit. After $\log_2 Q$ stages, the total feedback amount is $\log_2 Q$ bits, and the times of transmitting training signals are $2 \times \log_2 Q$.

The overall search process in [16] is also carried out in stages. Similar to the dichotomy method, the AoD interval determined at each stage is used as the search range at the next stage, which is divided into K ($K+1$ is an integer power of 2) subintervals. In theory, this algorithm requires $\log_K Q$ cycles, and the number of bits required for K subintervals to be expressed in binary at each stage is $\log_2 K$ bits. Therefore, the theoretical total amount of feedback is $\log_2 K \times \log_K Q = \log_2 Q$ bits. As previously analyzed, because the relation of Q and K is not a power relation, the amount of estimation stage needs to be rounded up. Furthermore, amount of actual training beam at each stage is $\log_2(K+1)$, so the actual times of transmitting training signals are $\log_2(K+1) \times \lceil \log_K Q \rceil$ and the actual feedback amount is $\log_2(K+1) \times \lceil \log_K Q \rceil$, which is bigger than $\log_2 Q$ bits. We take $K = 3$ and $Q = 8$ as a specific example to illustrate the search process of [16]. This algorithm requires $\lceil \log_3 8 \rceil = 2$ stages, and the number of training beam required in each stage is $\log_2(3+1) = 2$. Therefore, the actual feedback amount is $\lceil \log_3 8 \rceil \times \log_2 (3+1) = 4$, which is bigger than $\log_2 8 = 3$.

According to the above analysis, the communication beam is the same as the training beam in the traversal search, the training beam only at the last stage is the same as the communication beam in the dichotomy method. So the optimal beam final estimated must be actual communication beam for these two methods. However, the training beam is not always the same as the communication beam in [16], so the estimated optimal beam should be different from the actual communication beam. Furthermore, for Q communication beams, the minimum amount of feedback required is $\log_2 Q$ bits after the receiver selects the optimal communication beam. Whether the traversal search or the dichotomy search,

TABLE 1. The state of $S_{beam}^{(1)}$ for traversal search.

Bit sequence number	1	2	...	$q(opt)$	...	$Q-1$	Q
State	0	0	...	1	0	0	0

their actual total feedback bit is both $\log_2 Q$ bits, but total feedback bit in [16] is bigger than $\log_2 Q$ bits. This is another drawback of [16].

Thus, under the premise that the feedback amount is not greater than $\log_2 Q$ bits, how to design a beam search scheme to further reduce the complexity is considered in this paper. To analyze essential problems, we will define the concept of SUE, IRE and FE from new perspective below.

C. INTRODUCTION AND ANALYSIS OF SUE, IRE AND FE

1) DEFINITION OF SUE AND IRE

If the beam search process is divided into N_{level} stages, the amount of training beam in the $n_{th}(n = 1, 2, \ldots, N_{level})$ stage is $N_{beam}^{(n)}$, which can divide the search range of this stage into $N_{area}^{(n)}$ subintervals. Then in the search procedure of this stage, one of $N_{area}^{(n)}$ subintervals is chosen as $A_{q(opt)}$ based on $N_{beam}^{(n)}$ training beams. Here we use $N_{beam}^{(n)}$ bits binary number, denoted as $S_{beam}^{(n)}$, to represent the selected subinterval at this stage, which is one of $N_{area}^{(n)}$. Each bit of $S_{beam}^{(n)}$ is '1' or '0', depending on whether the receiver is located in the mainlobe range of corresponding training beam. If in, then the bit is '1', otherwise the bit is '0'.

In the n_{th} stage, although $S_{beam}^{(n)}$ has $2^{N_{beam}^{(n)}}$ states, only $N_{area}^{(n)}$ states are used. We introduce the concept of SUE to describe the state utilization efficiency. SUE in the n_{th} stage can be defined as $\eta^{(n)}$ =(number of used states) / (number of all states), that is

$$\eta^{(n)} = \frac{N_{area}^{(n)}}{2^{N_{beam}^{(n)}}}, \quad n = 1, \ldots, N_{level}. \tag{2}$$

For N_{level} stages, the total efficiency is

$$\eta = \frac{\sum_{n=1}^{N_{level}} N_{area}^{(n)}}{\sum_{n=1}^{N_{level}} 2^{N_{beam}^{(n)}}}. \tag{3}$$

If the search method for each stage is the same, then we can get

$$\eta = \frac{N_{level} \times N_{area}^{(n)}}{N_{level} \times 2^{N_{beam}^{(n)}}} = \eta^{(n)}, \quad n = 1, \ldots, N_{level}. \tag{4}$$

Similarly, although the information expressed by $S_{beam}^{(n)}$ are $N_{beam}^{(n)}$ bits, $N_{area}^{(n)}$ subintervals only need $\log_2 N_{area}^{(n)}$ bits when expressed in binary, which is also the feedback amount required.

So we define IRE in the n_{th} stage as $\gamma^{(n)}$ = (the required feedback information amount when express $N_{area}^{(n)}$ in binary) / (the information amount represent by $S_{beam}^{(n)}$), that is

$$\gamma^{(n)} = \frac{\log_2 N_{area}^{(n)}}{N_{beam}^{(n)}}, \quad n = 1, \ldots, N_{level}. \tag{5}$$

For N_{level} stages, the total IRE is

$$\gamma = \frac{\sum_{n=1}^{N_{level}} \log_2 N_{area}^{(n)}}{\sum_{n=1}^{N_{level}} N_{beam}^{(n)}}. \tag{6}$$

Whatever the method, the total amount of feedback required is all $\log_2 Q$ bits for Q communication beams. So for each beam search method, IRE can be written as follows

$$\gamma = \frac{\log_2 Q}{\sum_{n=1}^{N_{level}} N_{beam}^{(n)}}. \tag{7}$$

According to (2) and (5), we can get the relationship between SUE and IRE as follows

$$\gamma^{(n)} = \frac{\log_2 \eta_{\text{numerator}}^{(n)}}{\log_2 \eta_{\text{denominator}}^{(n)}}. \tag{8}$$

2) SUE AND IRE OF BEAM SEARCH METHODS

The traversal search process has only one stage. Using Q beams, it divides the search range into Q subintervals, so $N_{level} = 1$, $N_{beam}^{(1)} = Q$, $N_{area}^{(1)} = Q$. That means that it uses one state of Q bits binary number to represent the optimal beam position information. For example, the receiver is located in the q_{th} A_q, which means that the q_{th} beam is the optimal beam, then the q_{th} bit of $S_{beam}^{(1)}$ is '1', while the other bits are all '0', as shown in Table 1. Although the Q bit binary number can represent up to 2^Q kinds of states, but the traversal search method only uses Q states, such as '00...01', ...,'10...00'. Therefore, its SUE and IRE respectively are

$$\eta = \frac{Q}{2^Q}, \tag{9a}$$

$$\gamma = \frac{\log_2 Q}{Q}. \tag{9b}$$

Generally, Q is a power of 2 and $Q \geq 4$, so $\eta(\gamma) < 50\%$.

The search process of dichotomy method is divided into $\log_2 Q$ stages, and the search range is divided into 2 subintervals by using 2 training beams at each stage, so $N_{level} = \log_2 Q$, $N_{beam}^{(n)} = 2$, $N_{area}^{(n)} = 2$. The useful states of $S_{beam}^{(n)}$ are two states, '01' and '10'. SUE and IRE respectively are

$$\eta = \eta^{(n)} = 2/2^2 = 50\%, \tag{10a}$$

$$\gamma = \frac{\log_2 Q}{2\log_2 Q} = \frac{1}{2} = 50\%. \tag{10b}$$

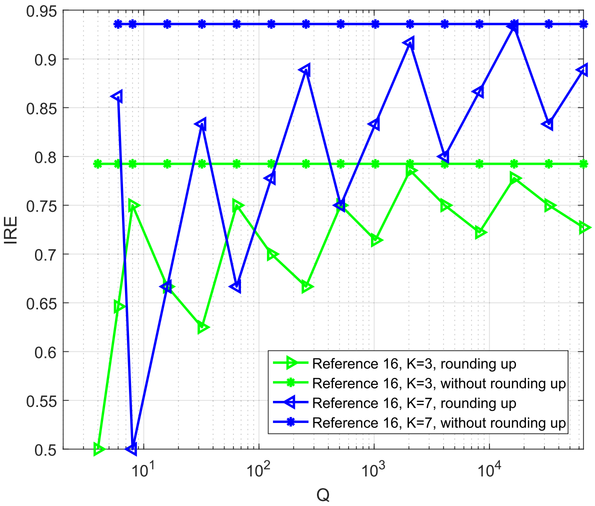

FIGURE 4. **Influence on IRE caused by rounding up.**

In [16], the amount of training beam transmitted in each stage is $\log_2(K+1)$, the search range is all divided into K subintervals, and the actual amount of stages is $\lceil \log_K Q \rceil$. So $N_{beam}^{(n)} = \log_2(K+1)$, $N_{area}^{(n)} = N_{area} = K$, the used states of $S_{beam}^{(n)}$ are K, not including the state of '00 . . . 00'. Its SUE and IRE respectively are

$$\eta = \frac{N_{area}^{(n)}}{2^{N_{beam}^{(n)}}} = \frac{K}{(K+1)} < 1, \tag{11a}$$

$$\gamma = \frac{\log_2 Q}{\log_2(K+1) \times \lceil \log_K Q \rceil}. \tag{11b}$$

When we get rid of the rounding up of the denominator in (11b), we can get the following equation

$$\bar{\gamma} = \frac{\log_2 Q}{\log_2(K+1) \times \log_K Q} = \frac{\log_2 K}{\log_2(K+1)}. \tag{12}$$

When Q is the integer power of K or the amount of estimation stage is not rounded up, (8) is also true for [16]. So the relation between IRE and SUE is always like (8) for each method. Next, we analyze the effect of rounding up operations on IRE in [16], and the results are shown in Fig. 4. Therefore, we can get the following conclusions.

- Rounding up causes a jitter on IRE of [16], whether $K = 3$ or $K = 7$. Let's take $K = 7$ as an example. When $Q = \{16, 32\}$, $\lceil \log_7 Q \rceil$ is always 2, so the IRE will increase with the increase of Q. When $Q = 64$, $\lceil \log_7 Q \rceil$ changes from 2 to 3, so the curve appears to fall. If the rounding up is removed, the IRE of [16] has always been a fixed value.
- When Q is small, the rounding up operation has a greater impact on IRE, resulting in a large jitter. As Q increases, the influence of $\lceil \log_K Q \rceil$ on the IRE is gradually reduced, so the volatility decreases. When Q is large enough, the IRE of [16] tends to be fixed, that is, the operation of the rounding up is removed.
- The smaller Q is, the stronger the jitter is, furthermore, it is more obvious for $K = 7$. Therefore, the IRE of $K = 3$ intersects the IRE of $K = 7$ when Q is small. But as Q increases, the two curve will gradually separate and there will be no crossover when Q is equal or greater than 512. Moreover, the IRE is higher for the larger K.

Rounding up not only causes a jitter, but increases the search complexity and feedback amount in some extent. When Q is large enough, the IRE of [16] tends to be fixed, it can achieve the values that are close to not rounded up. For the convenience of analysis, we will analyze the performance of [16] by removing the rounding up.

3) DEFINITION OF FE

We already know $\log_2 Q$ is minimum amount of feedback required for Q communication beams, and the denominator of (11b) is the actually feedback amount, so IRE of [16] also expresses physical significance of the feedback efficiency in finite state feedback. For further analysis, we define FE as $\lambda =$ (minimum amount of feedback) / (the actually feedback amount). The feedback amount is both $\log_2 Q$ bits for the traversal search and the dichotomy search, so their FE is both 100%. Obviously, FE is the same as IRE for [16], namely,

$$\lambda = \bar{\gamma} = \frac{\log_2 Q}{\log_2(K+1)\log_K Q} = \frac{\log_2 K}{\log_2(K+1)}. \tag{13}$$

4) ANALYSIS OF SUE, IRE AND FE

We analyze the IRE, SUE and FE of different beam search algorithms, and the results are shown in Fig. 5. Q should be respectively less than 4 and 8 for $K = 3$ and $K = 7$ in [16].

According to Fig. 5(a) and Fig. 5(b), the following conclusions can be drawn.

- It can be seen from (9) that SUE and IRE of the traversal search are all reduced as the Q increases. Therefore, the trends of SUE and IRE are similar, which is also confirmed in (8).
- According to (10) and (11), both SUE and IRE of the dichotomy search and [16] can be converted into constants independent of Q.

For convenience, we focus on SUE in the next analysis.

The traversal search uses Q states of all 2^Q kinds of states, so its SUE is the lowest and is approaching 0 with the increase of Q; Dichotomy method uses 2 states of all 4 kinds of states at each stage, not using '00' and '11' states in each stage, so its SUE is always equal to 50% and still low. In [16], its SUE is improved and far greater than that of the traversal search and the dichotomy search. But the way of judgment in [16] determines that all-zero state is not used in every stage, so its SUE still has room for improvement.

In Fig. 5(c), for the FE of [16] neglects rounding up in (13), so the FE of [16] in this figure is bigger than the actual FE. Even so, it is still lower than the FE in the traversal search and the dichotomy search.

According to the above analysis, in order to obtain 100% SUE (IRE) and further reduce complexity without increasing

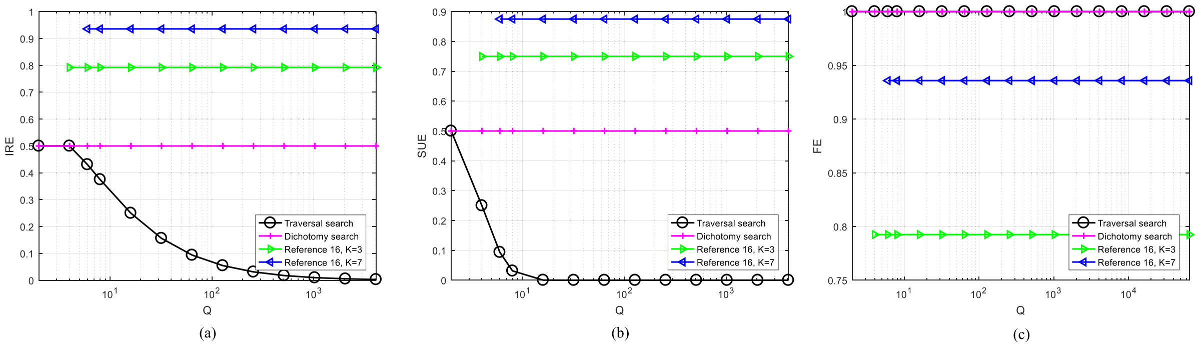

FIGURE 5. The IRE, SUE and FE of different methods.

the amount of feedback, we must use all the states at each stage. Based on this idea, this paper proposes a new method to estimate the channel AoD quickly for the LoS environment. The proposed scheme is only applicable to the LoS environment, and the beam search algorithm for the NLoS environment will be our future research direction.

III. A FAST BEAM SEARCH SCHEME BASED ON MAINLOBE OVERLAPPING STATE

In this section, the coverage area A_q of the communication beam is firstly encoded, and then the shape of the training beam is designed according to the codes. Finally, optimal beam-pair is found by utilizing mainlobe superposition state information of multiple beams based on joint judgment.

As mentioned above, the minimum feedback amount required by either method is all $\log_2 Q$ bits for Q communication beam, so not only to obtain 100% SUE, but also for 100% FE, the number of training beams required should be $\log_2 Q$. Then these training beams need to be designed cleverly and let them cover the entire AoD range $(-90°, 90°)$, which is bound to lead to the mainlobe of training beams overlay according to some regular. To illustrate the principle, we take eight communication beams ($Q = 8$) as an example.

A. A FAST SEARCH SCHEME BASED ON BITC

1) BINARY INDEX TO CODE A_Q (BITC)

According to the above analysis, the number of training beams required is 3 for eight communication beams, and the three training beams need to cover the range of eight communication beams. The indexes of the eight communication beam are '000', '001', '010', '011', '100', '101', '110', '111', respectively, and subintervals that they cover correspondingly are shown in Fig. 2. For the sake of description, the communication beam binary index is also used as the index of eight subintervals A_q, that means it is using binary index to code A_q, referred to as BITC.

2) DESIGN OF THE TRAINING BEAM BASED ON BITC

According to the definition of SUE and IRE, we know that three training beams correspond to 3 bits binary number $S_{beam}^{(3)}$, which happens to be consistent with the index number of subintervals. So we can design training beam according to the communication beam index to cover 8 subintervals.

Let training beam 1,2,3 respectively correspond to the leftmost bit, the middle bit and the rightmost bit of subinterval index. If the leftmost bit of subinterval index is 1, let training beam 1 cover this subinterval, otherwise training beam 1 doesn't cover this subinterval. So training beam 1 covers A_4, A_5, A_6, A_7. Similarly, training beam 2 covers A_2, A_3, A_6, A_7 and training beam 3 covers A_1, A_3, A_5, A_7. For eight communication beams, the training beam is designed as Fig. 6. In this paper, we define the coverage regions of the training beam as the mainlobe and the other areas as the sidelobes. Therefore, training beam 1, 2 and 3 have 1 mainlobe, 2 mainlobes and 4 mainlobes, respectively. Moreover, Fig. 6 is a schematic diagram in which the amplitudes are different only to distinguish different training beams, which are actually the same.

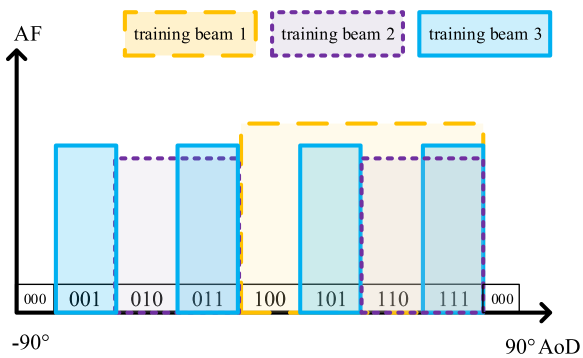

FIGURE 6. Design training beam based on 3-BITC.

3) FAST ESTIMATION OF AOD

During the AoD estimation, transmitter transmits the training beam 1, 2 and 3 in sequence and the receiver receives by an omnidirectional antenna. If the received power of only training beam 1 and 3 are bigger than the preset threshold (how to set the threshold will be given in IV-C), then the first and third bits of $A_{q(opt)}$ code are both set to be 1, while the second bit of $A_{q(opt)}$ code is set to be 0, that is, $A_{q(opt)}$ is '101'. The optimal beam index is 5. The receiver feeds back the $A_{q(opt)}$ code to the transmitter. Next, the receiver transmits three training beams in sequence to transmitter. And the transmitter uses the communication beam determined above, namely 5_{th} beam, to receive and feeds back the index of the optimal training beam to the receiver. Then the transmitter and receiver communicate with the selected optimal communication beam-pair.

4) DISADVANTAGES OF BITC

Because the edge of the mainlobe of the training beam actually generated is not steep enough, it will cause potential misjudgment. For example, in Fig. 6, if the AoD is located close to the right edge in region '001', and it is also close to the left edge of the training beam 2. If the gain of the training beam 2 here does not drop to a sufficiently low level, the receiver misjudges the second bit of the $A_{q(opt)}$ as '1', which come to decode the $A_{q(opt)}$ as '011'. Therefore, in order to reduce the misjudgment probability based on mainlobe overlapping, the edge number of the mainlobe of the training beam should be minimized, that is mainlobe number of each training beam should be minimized. So a new method to design training beam needs to be proposed.

B. A FAST SEARCH SCHEME BASED ON GMTC

1) GRAY MAPPING TO CODE A_Q (GMTC)

In order to make the training beam mainlobe as few as possible and find a better way to design a training beam, we first analyze the Hamming distance (HD) $d(x_i, x_j)$ of the communication beam index (or subinterval index), which represents amount of different characters corresponding to the positions between two equal-length strings. The HD of BITC is analyzed and shown as the following Fig. 7.

000	001	010	011	100	101	110	111	000
HD	1	2	1	3	1	2	1	3

FIGURE 7. **The HD of BITC.**

By comparing Fig. 6 and Fig. 7, it can be found that when the HD between the indexes of two adjacent subintervals in Fig. 7 is n, there are n beam edges between these corresponding two adjacent subintervals in Fig. 6. For example, the HD of the subintervals '001' and '010' in Fig. 7 is 2, then the edge number between the corresponding training beams in Fig. 6 is also 2. We define the sum of the edges of all training beams as the Total Edge (TE), half of which is actually the number of mainlobes, for one mainlobe corresponds to two edges. The number of mainlobes is 7 in Fig. 6 and the TE is 14, which equals to sum of HD (SHD) in Fig. 7.

According to the above analysis, the smaller the mainlobe number of training beams, the lower the probability of misjudgment. Therefore, the criterion to design training beam is to minimize TE or SHD. Obviously, to minimize SHD must minimize each HD. Because the indexes of 8 subintervals are different, the HD between the adjacent subinterval indexes is at least 1. That is to say that we need encode the index of each subinterval again and let the HD between the indexes of two adjacent subintervals always equals to 1. The codewords that satisfy this requirement are only Gray code. That is, 8 subintervals covered by 8 communication beams need to be encoded by Gray code. The minimum value of SHD is 8, namely TE is also 8, which means the mainlobe number of three training beams is 4 at least. (It should be noted that these 8 subintervals are actually connected end to end in space, so 8 subintervals have 8 boundaries). So at least one training beam has two mainlobes and the other two training beams each have one mainlobe, which is the best case.

There are many kinds of Gray code. For the 3-bit Gray code, when 8 codewords are connected end to end and still meet the Gray code rule, there are 12 kinds. Through analysis, it can be found that the training beams designed based on these Gray codes are essentially the same. Therefore, this paper uses the most representative typical Binary Gray Code as an example to design the training beams. Its HD is shown as the following Fig. 8.

000	001	011	010	110	111	101	100	000
HD	1	1	1	1	1	1	1	1

FIGURE 8. **The HD of GMTC.**

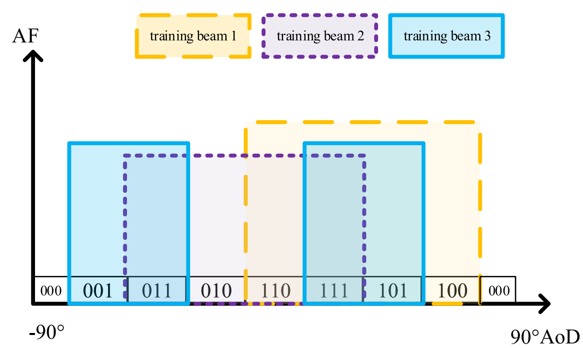

FIGURE 9. **Design training beam based on 3-GMTC.**

The proposed method is using Gray mapping to code A_q, and the corresponding training beam shape designed is shown in Fig. 9. There are four mainlobes indeed for three training beams and the third beam has two mainlobes, which is completely consistent with the above analysis. Eight communication beams need 3 bits to code, referred to as 3-GMTC. The method of estimating AoD for GMTC is the same as that of BITC.

For BITC, the training beam 2 and 3 has 2 and 4 mainlobes, respectively. In the scheme of GMTC, mainlobe numbers of training beams 2 and 3 are reduced by half. So Gray mapping A_q coding indeed reduces mainlobe numbers of training beams, thereby reducing the misjudgment probability.

2) ADVANTAGES OF GMTC

Compared with BITC, GMTC has two other advantages:

- Reduce the possibility of multiple bit misjudgment simultaneously.

According to the HD of BITC and GMTC, we can see that the probability of multi-bit misjudgment in GMTC is much lower than that of BITC. For example, in Fig. 6, the border between $A_q s$ numbered '001' and '010' is at not only the left edge of the training beam 2, but also the right edge of the training beam 3. The second bit and the third bit of the $A_{q(opt)}$ code may be misjudged at the same time when the device is located at this boundary. However, this is not the case in GMTC.Therefore, only one-bit misjudgment will occur, thus reducing the possibility of multi-bit misjudgment.

- Reduce the negative impact of misjudgment.

When an error occurs for GMTC, the misjudged A_q is actually adjacent to the $A_{q(opt)}$. It means that sub-optimal communication beam is found, which can still provide relatively significant gain. For example, in Fig. 9, when AoD is located close to the right edge of A_q coded '011'. If the receiving power of the training beam 3 does not reach the threshold, the receiver will judge the third bit of the A_q as '0', then the result of the wrong judgment is '010', which is right next to the correct subintervals '011'.

C. CASES FOR ARBITRARY Q VALUES

Next, we analyze the case of $Q = 4$ and $Q > 8$. Without loss of generality, in practical applications, the number of communication beams is usually an integer power of 2, that is, $Q = 2^M$ (M is positive integer).

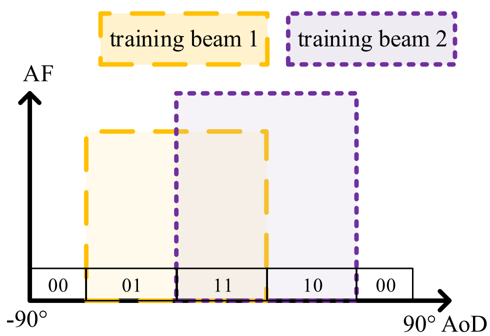

FIGURE 10. Design training beams based on 2-GMTC.

1) CASE FOR $Q = 4$

Using the above method, it is easy to design two training beams for $Q = 4$ in Fig. 10, referred to as 2-GMTC.

2) CASE FOR $Q > 8$

When $Q = 16$, if A_q is encoded according to the 4-bit Gray Code of the index of the communication beam, then four training beams can be designed, as shown in Fig. 11.

As discussed in section III-A 4), smaller mainlobe number of each training beam is, lower the misestimate probability of channel is. But the training beam 1 and 3 in Fig. 11 have respectively 2 and 3 mainlobes, so the misjudgment probability may be a little high. So for $Q = 16$, the search

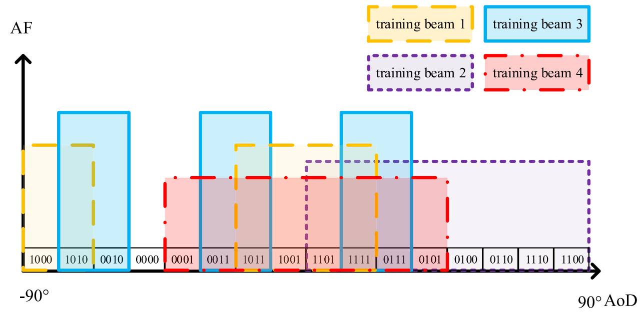

FIGURE 11. Design training beams based on 4-GMTC.

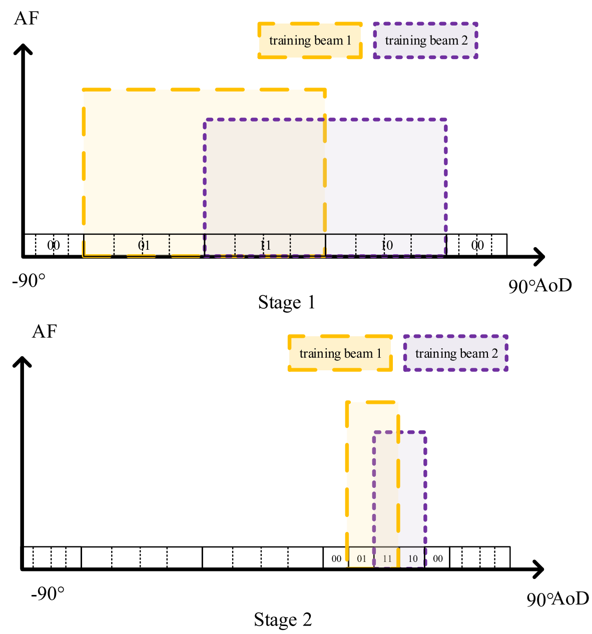

FIGURE 12. Search process carried out in two stages for $Q = 16$.

process is recommended to be conducted in stages and the search range at each stage is divided into four subintervals by two training beams, which is designed as shown in Fig. 12. Each training beams only has one mainlobe. Whether 4-bit Gray mapping (only one stage) or 2-bit Gray mapping (two stages), times of transmitting training signal is both 4 and the amount of feedback information is both 4 bits. The only difference is that the latter feeds back two times and the former only feeds back one time.

When Q is other values, the following method is applied.

- When M is an integer multiple of 2 (that is $M = 2 \times M_2$, where M_2 is positive integer), the whole process is divided into M_2 stages and the search range at each stage is divided into 4 subintervals. The method 2-GMTC is used at each stage. After M_2 times cycles, the training signal will be transmitted $2 \times M_2$ times and amount of actual feedback is $2 \times M_2 = M$ bits.
- When M is an integer multiple of 3 (that is $M = 3 \times M_3$ where M_3 is positive integer), the whole process is divided into M_3 stages and the search range at each stage is divided into 8 subintervals. The method 3-GMTC is

carried out M_3 times. The training signal will eventually be transmitted $3 \times M_3$ times and amount of actual feedback is $3 \times M_3 = M$ bits.

- When $M = 2\times M_2+3\times M_3$, the whole process is divided into $M_2 + M_3$ stages. The method 2-GMTC is carried out M_2 times, the method 3-GMTC is carried out M_3 times. The training signal will eventually be transmitted $2 \times M_2 + 3 \times M_3$ times and amount of actual feedback is $2 \times M_2 + 3 \times M_3 = M$ bits.

Therefore, it can be seen that no matter what Q is, beam search scheme proposed in this paper is feasible. Furthermore, the amount of feedback is always $\log_2 Q = M$ bits, which is the same conclusion as before.

When the communication beam is greater than 8, the estimation of AoD needs to be carried out in stages. In every stage, the operation is similar. We take sixteen communication beams as an example. In the first stage, the transmitter transmits the wider training beam 1 and 2 in sequence. The receiver obtains the optimal interval of the first stage according to the above judgment method, for example, '10', and feeds the value back to the transmitter. In the second stage, the transmitter sends the finer training beam 1' and 2' in turn in the optimal interval determined in the first stage. The receiver judges that the optimal interval of the second stage, for example, '11', and feeds the value back to the transmitter. The transmitter determines the optimal interval number '1011', and then determines the optimal communication beams. The operation of the receiver is similar to that of the transmitter.

For the method proposed in this paper, it all needs to design training beams with one or multiple mainlobes, whose mainlobes are superimposed with each other. A feasible beam design method has been proposed in [19], [20], which has strong control on the beam shape and can generate arbitrary shape beam. In this paper, we will generate training beams with desired shape according to the method in [19], [20].

In order to facilitate the understanding of our algorithm, we use pseudo code to explain the whole process in detail for the case of $Q = 2^M (M = 2 \times M_2 + 3 \times M_3)$.

Algorithm 1 Fast Beam Search for Millimeter-wave Communication Based on Gray Mapping

Input: M_2, M_3, the initial AoD ranges of the transceiver in $[-90°, 90°]$.

for $m = 1$ to M_2 **do**

  2-GMTC is executed at transmitter according to the corresponding AoD ranges.

  The receiver determines the $A^{(t,m)}_{q(opt)}$ (that is the $A_{q(opt)}$ in the m_{th} cycle of the transmitter) according to the relationship between the received power and the threshold, and feeds the index of subinterval back to the transmitter.

  2-GMTC is executed at receiver according to the corresponding AoD ranges. And the transmitter receives the signal through the AoD range of $A^{(t,m)}_{q(opt)}$.

  The transmitter determines the $A^{(r,m)}_{q(opt)}$ (that is $A_{q(opt)}$ in the m_{th} cycle of the receiver) according to the relationship between the received power and the threshold, and feeds back the index to the receiver.

end for

Initialize the AoD ranges at the transceivers of the next stage as $A^{(t,M_2)}_{q(opt)}$ and $A^{(r,M_2)}_{q(opt)}$.

for $n = 1$ to M_3 **do**

  3-GMTC is executed at transmitter according to the corresponding AoD ranges.

  The receiver determines the $A^{(t,n)}_{q(opt)}$ (that is $A_{q(opt)}$ in the n_{th} cycle of the transmitter) according to the relationship between the received power and the threshold, and feeds back the index to the transmitter.

  3-GMTC is executed at receiver according to the corresponding AoD ranges. And the transmitter receives the signal through the AoD range of $A^{(t,n)}_{q(opt)}$.

  The transmitter determines the $A^{(r,n)}_{q(opt)}$ (that is $A_{q(opt)}$ in the n_{th} cycle of the receiver) according to the relationship between the received power and the threshold, and feeds back the index to the receiver.

end for

Output: $A^{(t,M_3)}_{q(opt)}$ and $A^{(r,M_3)}_{q(opt)}$.

IV. PERFORMANCE ANALYSIS

A. SUE AND FE OF THE PROPOSED METHOD

1) SUE OF THE PROPOSED METHOD

Using the definition of Section II-C, SUE of the proposed scheme is analyzed for any Q values.

- When M is an integer multiple of 2 (that is $M = 2 \times M_2$, where M_2 is positive integer), the whole process is divided into M_2 stages and the search range at each stage is divided into 4 subintervals. So $N_{level} = M_2$, $N^{(m)}_{beam} = 2$, $N^{(m)}_{area} = 4$, then SUE is $\eta = \eta^{(m)} = \frac{N^{(m)}_{area}}{2^{N^{(m)}_{beam}}} = \frac{4}{2^2} = 100\%, (m = 1, 2, \cdots, M_2)$. For $M_2 = 1$, it is the case of $Q = 4$.
- When M is an integer multiple of 3 (that is $M = 3 \times M_3$, where M_3 is positive integer), the whole process is divided into M_3 stages and the search range at each stage is divided into 8 subintervals. So $N_{level} = M_3$, $N^{(n)}_{beam} = 3$, $N^{(n)}_{area} = 8$, then SUE is $\eta = \eta^{(n)} = \frac{N^{(n)}_{area}}{2^{N^{(n)}_{beam}}} = \frac{8}{2^3} = 100\%, (n = 1, 2, \cdots, M_3)$. For $M_3 = 1$, it is the case of $Q = 8$.
- When $M = 2 \times M_2 + 3 \times M_3$, the whole process is divided into $M_2 + M_3$ stages, and $N_{level} = M_2 + M_3$. For the method 2-GMTC, $N^{(n)}_{beam} = 2, N^{(n)}_{area} = 4$, and for the method 3-GMTC, $N^{(m)}_{beam} = 3, N^{(m)}_{area} = 8$. So, $\eta = \frac{\sum_{m=1}^{M_2} N^{(m)}_{area} + \sum_{n=1}^{M_3} N^{(n)}_{area}}{\sum_{m=1}^{M_2} 2^{N^{(m)}_{beam}} + \sum_{n=1}^{M_3} 2^{N^{(n)}_{beam}}} = \frac{4M_2+8M_3}{M_2\times 2^2+M_3\times 2^3} = 100\%$. For $M_2 = 2, M_3 = 0$, it is the case of $Q = 16$.

Thus, it can be seen that SUE of the proposed method is indeed improved up to 100%.

2) FE OF THE PROPOSED METHOD

No matter what Q is, the amount of actual feedback is always $\log_2 Q$ bits for our proposed method, so its FE is always100%.

B. COMPLEXITY ANALYSIS

If the transmitter and receiver have the same number of antennas and communication beams, N_{sector} is sector number, N_{beam} is beam number in each sector, and the amount of communication beam is $Q = N_{sector} \times N_{beam}$. In this paper, we define the time period of a training sequence as a time slot to measure complexity.

The minimum time slot in 802.15.3c is

$$N_{3c} = N_{sector}^2 \times N_{beam}^2. \tag{14}$$

The minimum time slot in 802.11.ad is

$$N_{ad} = 2 \times N_{sector} \times N_{beam}. \tag{15}$$

The minimum time slot of the dichotomy method is

$$N_{2dic} = 2 \times 2 \times \log_2 Q. \tag{16}$$

The search process in [16] is similar to 802.15.3c, so the time slot is $\lceil \log_K Q \rceil \times \log_2^2(K+1)$. In fact, if [16] adopts the method similar to 802.11.ad, the complexity will be lower, which is also adopted in this paper. Therefore, the minimum time slot of [16] is

$$N_{[16]} = 2 \times \lceil \log_K Q \rceil \times \log_2(K+1). \tag{17}$$

For the scheme proposed in this paper, if Q is an integer power of 2 and greater than 8, then it can be carried out in stages. Without loss of generality, this paper assumes that the method 2-GMTC is carried out M2 times, the method 3-GMTC is carried out M3 times. Complexity of the method 2-GMTC is four and complexity of the method 3-GMTC is six. Therefore, the complexity of the proposed method is

$$N = 4M_2 + 6M_3 = 2 \times \log_2(2^M) = 2 \times \log_2 Q. \tag{18}$$

Obviously, the complexity of the proposed method is reduced to half complexity of the dichotomy method. The complexity comparison is shown in Fig. 13.

The complexity of the scheme proposed in this paper is lower than that of all the existing search methods, including [16]. Compared with [16], firstly, the FE of the proposed method is 100%. Secondly, its search complexity is lower than that in [16]. Finally, since the method of [16] cannot completely cover the communication beam of integer power of 2, the estimated optimal subinterval is not covered by the actual communication beam. So various performances of our proposed method in this paper are all better than that of [16].

C. SEARCHING SUCCESS RATIO

Consider a room whose length, width and height are respectively 20m, 10m, and 2.4m. One of the corners is used as the coordinate origin. In the room, the transmitter is located at coordinates (0.5, 5, 1.3). The transmitter is equipped with one

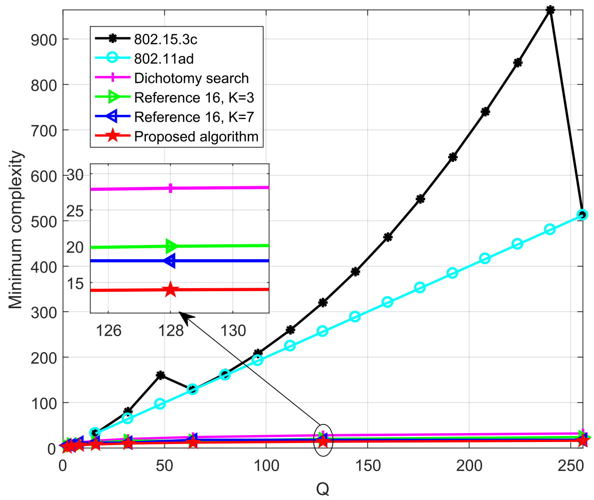

FIGURE 13. Complexity comparison of different methods.

dimensional linear antenna array consisting of four associated antenna elements, which are parallel to the Y axis. In the range of $1 \le x < 20$, $0.2 \le y < 10$, $0.2 \le z < 2.4$, we divide the entire space into 51205 check points at 0.2m intervals, which are actually the receiver's locations. The receiver is also equipped with four associated antenna elements.

The 16-phase codebook can be used to generate 8 communication beams to cover the whole angle space, which can be seen in Fig. 2. The desired training beam shape (that is rectangular signal, which can be referred to Fig. 9) and the actually generated beams of GMTC and BITC are shown in Fig. 14. We use the methods in [19], [20] to generate training beams. It can be seen that the generated beam 1 and the generated beam 2 of BITC are the same as the generated beam 1 and the generated beam 3 of GMTC, respectively. The generated beam 3 of BITC is divided into four equal-width regions with relatively poor shape.

Firstly, the transmitter transmits three training beams in sequence to a randomly selected check point. Then the receiver at this check point determines optimal communication beam by comparing the relative receiving power with the threshold and feeds back to the transmitter.

The threshold setting is most critical part of this method, so let's analyze it in detail first. We assume that the transmitting power of each training beam is all P_T. We first generate the desired beams, and then design the training beams. Finally, the generated beams are normalized to meet the total transmitting power of P_T. The power received by the receiver at each check point is denoted as $P_r^i (i=1, \cdots, 51205)$, and the corresponding path loss (PL) is PL_r^i. In the simulation, we use the classic PL model, which can be written as

$$\begin{aligned} PL_r^i(dB) &= P_T(dB) - P_r^i(dB) \\ &= PL_r^i(d_0) + 10n\lg(\frac{d_1}{d_0}) + X_\sigma(dB), \end{aligned} \tag{19}$$

where n represents the path loss index, and X_σ represents a normal random variable with a standard difference of σ. d_0 is the reference distance, and the d_1 indicates the distance

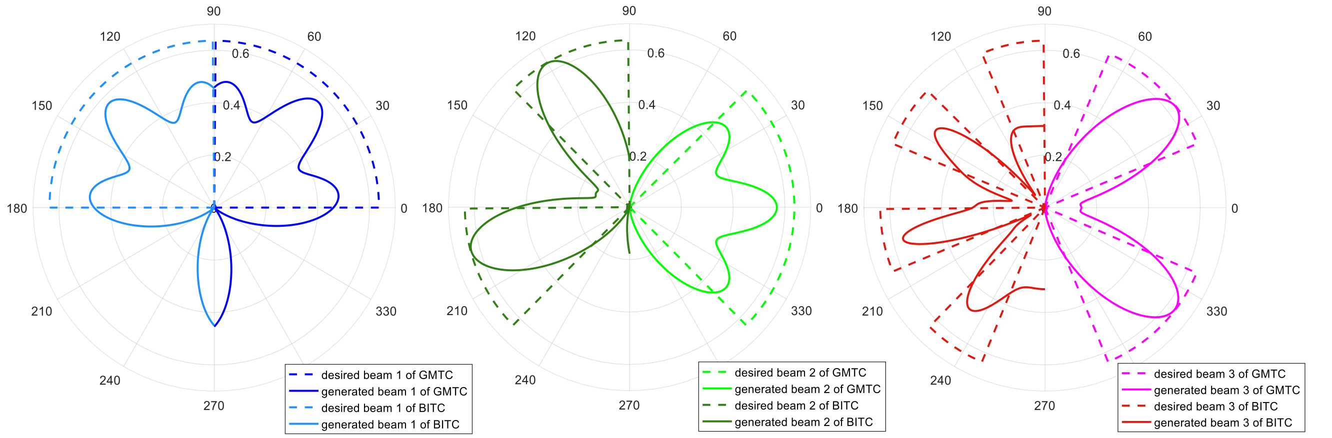

FIGURE 14. Beam shape of the desired beam and the actual generated beam.

between the transmitter and the receiver. $PL_r^i(d_0)$ indicates that the path loss at the reference distance, that is

$$PL_r^i(d_0) = 20\lg(4\pi d_0) - 20\lg(\lambda), \tag{20}$$

where λ is the wavelength.

Because the receiving power is not only related to the transmitting power, but also related to the path loss of each check point, we do the following setting for the convenience of comparison. Firstly, we introduce the concept of equivalent receiving power by adding the corresponding path loss to the receiving power of each check points, that is $\overline{P_r^i} = P_r^i + PL_r^i$, which is equivalent to placing all the check points on the surface of a sphere that is centered on the transmitter, thus the effect of distance on receiving power can be eliminated. Then, we define the relative receiving power $\beta = \frac{\overline{P_r^i}}{P_T}$ to eliminate the impact of transmit power. So we can use any value of the relative receiving power within some range as the threshold to determine whether it is within the mainlobe of training beam.

In this paper, we define the success ratio as finding the real optimal beam or suboptimal beam. Then search success ratio can be calculated by comparing the searched communication beam with the real optimal or suboptimal communication beam in 51205 different positions. Through plenty of simulations, we can get the relationship between the success ratio and the threshold, as shown in Fig. 15.

Through the analysis and comparison of Fig. 15, we can obtain the following conclusions:

- The green line and the black line indicate the success ratio of the generated beam based on the Gray code and based on Binary mapping, respectively. It can be seen that the search success ratio of the proposed method is close to 100%, much higher than that of BITC. According to the analysis in III-A, this is due to the excessive number of mainlobes and imperfect real shape of BITC training beams. This result will prompt us to design more efficient coding methods. Furthermore, the success ratio of BITC is more susceptible to the thresholds than GMTC.

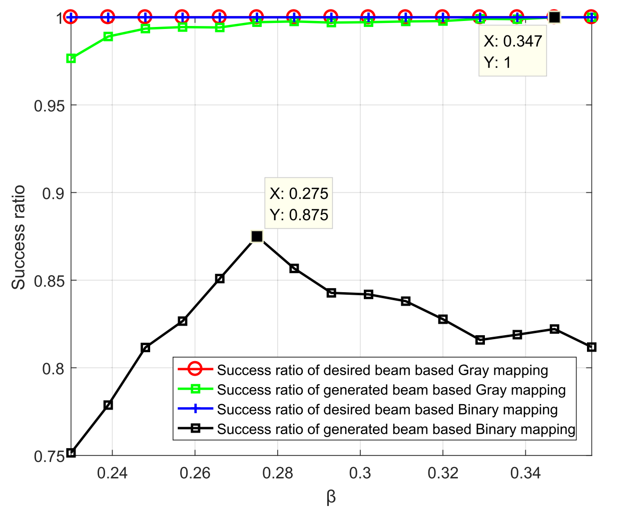

FIGURE 15. The relationship between the success ratio and the threshold.

- The success ratio of desired beam based on Gray mapping and Binary mapping are both 100%. This is because the desired beams are ideal, so the success ratio is not affected by the threshold. This motivates us to design training beams with better shape.
- For a specific scenario, we should first measure the parameters of the scene and perform numerous simulations to determine the optimal threshold range, so as to get the best possible performance.

D. EFFICIENCY OF MULTI-USER AOD ESTIMATION SIMULTANEOUSLY

For millimeter-wave system, point-to-multipoint star networks are a common topology, such as a base station communicating with multiple mobile stations in a cell, or AP (Access Point) communicating with multiple portable devices in a WLAN. In this scenario, if the center node uses beamforming technique, it needs to estimate the

TABLE 2. Dichotomy search process with 8 communication beams in multi-user scenario.

<table>
<tr><td rowspan="2">stage I</td><td>feedback information</td><td colspan="4">0</td><td colspan="4">1</td></tr>
<tr><td>determine the first bit of $A_{q(opt)}$</td><td colspan="4">0XX</td><td colspan="4">1XX</td></tr>
<tr><td rowspan="2">stage II</td><td>feedback information</td><td colspan="2">0</td><td colspan="2">1</td><td colspan="2">0</td><td colspan="2">1</td></tr>
<tr><td>determine the second bit of $A_{q(opt)}$</td><td colspan="2">00X</td><td colspan="2">01X</td><td colspan="2">10X</td><td colspan="2">11X</td></tr>
<tr><td rowspan="2">stage III</td><td>feedback information</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>determine the third bit of $A_{q(opt)}$</td><td>000</td><td>001</td><td>010</td><td>$\underbrace{011}_{U_1}$</td><td>100</td><td>$\underbrace{101}_{U_2}$</td><td>110</td><td>$\underbrace{111}_{U_3}$</td></tr>
</table>

corresponding AoD of each end node, and find the optimal communication beam pointing to each end node.

To illustrate the advantages of the proposed scheme in multi-user scenarios, we take 8 communication beams and three end nodes as an example, and assume optimal beam index of three users U_1, U_2, U_3 are numbered '011', '101' and '111'.

Firstly, for traversal search, the number of training signals is 8 and the feedback information is 9 bits.

Secondly, dichotomy search and the search scheme in [16] are carried out in stages. Compared with single-user, the complexity is greatly increased. Taking dichotomy search as an example, the search process needs three stages, as shown in Table 2. In the first stage, the training signals are transmitted twice. 3 users determine that their respective $A_{q(opt)}$ numbers are '0XX', '1XX', '1XX', and feed back '0','1','1' to the transmitter, respectively; In the second stage, the training signals are transmitted four times. The $A_{q(opt)}$ numbers of 3 users are '01X', '10X', '11X' respectively, and feedback information received at the transmitter is '101'; In the last stage, the training signals are transmitted six times. 3 users determine that their respective $A_{q(opt)}$ numbers are '011', '101', '111', and the feedback is '111'; Therefore, the number of training signals is 12 and the feedback is 9 bits. Through analysis, we can know that the minimum number of training signals required by dichotomy search is 8, that is, 3 users are located in three regions with the same first bit.

Finally, the proposed scheme has great advantages in multi-user scenarios. Compared with the single-user scenarios, the similarity is that the center node also only sends 3 training signals. The difference is that each end node all needs to feed back information. Therefore, the amount of feedback information increases from 3 ($\log_2 8 = 3$) bits to 9 ($3 \times \log_2 8 = 9$) bits. The number of training signals required in the whole process is still 3, so the search complexity does not increase. Obviously, with the constant communication beams, the number of training signals is always 3 as the number of users continues to increase, only the necessary total feedback information is increased. In summary, beam search complexity of the proposed method in multi-user scenario is lowest.

V. CONCLUSION

This paper defines SUE and IRE from a totally new perspective and discovers the reason that SUE of the existing beam search methods is so low. Then a fast beam search scheme based on the mainlobe overlapping of training beams is proposed. It can not only ensure high success ratio, but also improve the search speed without increasing FE. In particular, the greater the number of antennas is, the more complexity is reduced. Its advantage is also more obvious under the multi-user scenarios. One drawback of the proposed scheme is that it can only be used in LoS scenarios. Therefore, our future work will study a more appropriate method for NLoS scenarios.

REFERENCES

[1] 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 Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[2] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[3] T. S. Rappaport, G. R. Maccartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[4] R. W. Heath and N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[6] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.

[7] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1428–1442, Feb. 2017.

[8] M. Park and P. Gopalakrishnan, "Analysis on spatial reuse and interference in 60-GHz wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1443–1452, Oct. 2009.

[9] Z. Xiao, "Suboptimal Spatial Diversity Scheme for 60 GHz millimeter-wave WLAN," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1790–1793, Sep. 2013.

[10] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Interleaved training and training-based transmission design for hybrid massive antenna downlink," Oct. 2017, *arXiv: 1710.06949*. [Online]. Available: https://arxiv.org/abs/1710.06949

[11] E. Koyuncu, X. Zou, and H. Jafarkhani, "Interleaving channel estimation and limited feedback for point-to-point systems with a large number of transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6762–6774, Oct. 2018.

[12] J. P. Gilb, *Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANS) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension*, IEEE Standard 802.15.3c-2009, IEEE Computer Society, New York, NY, USA, 2009.

[13] C. Cordeiro, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications-Amendment 6: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard P802.11adTM/d0.1, IEEE, New York, NY, USA, 2010.

[14] W.-X. Zou, G.-L. Du, B. Li, and Z. Zhou, ''Step-wisely refinement based beam searching scheme for 60 GHz communications,'' *Wireless Pers. Commun.*, vol. 71, no. 4, pp. 2993–3010, 2013.

[15] T. He and Z. Xiao, ''Suboptimal beam search algorithm and codebook design for millimeter-wave communications,'' *Mobile Netw. Appl.*, vol. 20, no. 1, pp. 86–97, Feb. 2015.

[16] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, ''Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation,'' *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 601–616, Feb. 2017.

[17] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, ''Channel-reconstruction-based hybrid precoding for millimeter-wave multi-user MIMO systems,'' *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 383–398, May 2018.

[18] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, ''Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems,'' *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[19] W.-X. Zou, C. Guo, G.-L. Du, and X. Han, ''Arbitrary shaped beamforming codebook design for millimeter-wave communications,'' *Wireless Pers. Commun.*, vol. 85, no. 4, pp. 2773–2794, 2015.

[20] Z. Weixia, G. Chao, D. Guanglong, W. Zhenyu, and G. Ying, ''A new codebook design scheme for fast beam searching in millimeter-wave communications,'' *China Commun.*, vol. 11, no. 6, pp. 12–22, Jun. 2014.

WEIXIA ZOU received the B.S. degree in electric traction and drive control from Tongji University, in 1994, the M.S. degree in circuits and systems from Shandong University, in 2002, and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, in 2006. She is currently an Associate Professor with the Beijing University of Posts and Telecommunications, Beijing, China. Her current research interests include new technologies of short-range wireless communication and millimeter wave communication.

MINGYANG CUI was born in Henan, China, in 1993. He is currently pursuing the Ph.D. degree in telecommunications engineering with the Beijing University of Posts and Telecommunications (BUPT). His research interests include millimeter wave communication and performance optimization in massive MIMO systems.

CHAO GUO received the B.S. degree in communication engineering from the Hebei University of Technology, in 2011, and the master's degree in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2016.