

Fig. 1. An IRS assisted communication system.

lower-case and upper-case letters, respectively. The space of  $x \times y$  complex-valued matrices is denoted by  $\mathbb{C}^{x \times y}$ . For a complex-value vector  $\mathbf{x}$ ,  $\mathbf{x} \otimes \mathbf{y}$  denotes the kronecker product of  $\mathbf{x}$  and  $\mathbf{y}$  while  $|\mathbf{x}|$  denotes its modulus and  $\text{diag}(\mathbf{x})$  denotes a diagonal matrix with each diagonal entry being the corresponding entry in  $\mathbf{x}$ . For a function  $\mathbf{y} = H(\mathbf{x})$ ,  $H^{-1}(\mathbf{y})$  denotes its inverse function. For a general matrix  $\mathbf{A}$ ,  $\mathbf{A}^*$ ,  $\mathbf{A}^H$ , and  $\mathbf{A}[i, j]$  denote its conjugate, conjugate transpose, and the  $(i, j)$ th entry, respectively.  $j$  denotes the imaginary unit, i.e.,  $j^2 = -1$ .

## II. SYSTEM AND CHANNEL MODELS

For the purpose of exposition, we consider a basic IRS-aided single-user communication system in millimeter-wave (mmWave) band as illustrated in Fig. ??, where an IRS composed of a large number of  $N = N_{\text{ver}} \times N_{\text{hor}}$  passive reflecting elements, is deployed in proximity to a user for assisting its data transmission with an BS, both of which are equipped with a single antenna. The results in this paper can be readily extended to the more general system with multiple users served by the IRS (e.g., by applying orthogonal time/frequency division multiple access) and/or multiple antennas at the BS (by estimating their associated channels in parallel). In this paper, it assumed that the line of sight (LoS) path of BS-UE channel is blocked while

In the infinitesimal interval  $(\mu_0, \mu_0 + \Delta\mu_0)$  at  $\mu_0$ , (??) can be written as:

$$\begin{aligned}
& AFM_{\text{norm}}(\beta_{\text{hor}})|_{\ddot{u} \in (\mu_0, \mu_0 + \Delta\mu_0)} \\
&= \left| \int_{\mu_0}^{\mu_0 + \Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \ddot{u} - \int_0^{\ddot{u}} \dot{f}(\mu_0) + \dot{f}'(\mu_0)(\mu - \mu_0) d\mu]} d\ddot{u} \right| \\
&= \left| \int_{\mu_0}^{\mu_0 + \Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \mu_0 - \int_0^{\mu_0} \dot{f}(\mu) d\mu]} \cdot e^{j k N_{\text{hor}} d [\beta_{\text{hor}} (\ddot{u} - \mu_0) - \dot{f}(\mu_0)(\ddot{u} - \mu_0) - \frac{1}{2} \dot{f}'(\mu_0)(\ddot{u} - \mu_0)^2]} d\ddot{u} \right| \\
&= \left| e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \mu_0 - \int_0^{\mu_0} \dot{f}(\mu) d\mu]} \cdot \int_{\mu_0}^{\mu_0 + \Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} (\ddot{u} - \mu_0) - \dot{f}(\mu_0)(\ddot{u} - \mu_0) - \frac{1}{2} \dot{f}'(\mu_0)(\ddot{u} - \mu_0)^2]} d\ddot{u} \right| \quad (28) \\
&= \left| e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \mu_0 - \int_0^{\mu_0} \dot{f}(\mu) d\mu]} \right| \cdot \left| \int_0^{\Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \ddot{u} - \dot{f}(\mu_0) \ddot{u} - \frac{1}{2} \dot{f}'(\mu_0) \ddot{u}^2]} d\ddot{u} \right| \\
&= \left| \int_0^{\Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \ddot{u} - \dot{f}(\mu_0) \ddot{u} - \frac{1}{2} \dot{f}'(\mu_0) \ddot{u}^2]} d\ddot{u} \right|.
\end{aligned}$$

Since the array resolution increases with an increase in  $kN_{\text{hor}}d$ , we should design  $\dot{f}(\mu)$  on the condition that  $kN_{\text{hor}}d \rightarrow \infty$ . If  $\beta_{\text{hor}} \neq \dot{f}(\mu_0)$ , the following holds as the  $kN_{\text{hor}}d$  tends to infinity:

$$\begin{aligned}
& \lim_{kN_{\text{hor}}d \rightarrow \infty} \left| \int_0^{\Delta\mu_0} e^{j k N_{\text{hor}} d [\beta_{\text{hor}} \ddot{u} - \dot{f}(\mu_0) \ddot{u} - \frac{1}{2} \dot{f}'(\mu_0) \ddot{u}^2]} d\ddot{u} \right| \\
&\stackrel{\textcircled{1}}{=} \lim_{kN_{\text{hor}}d \rightarrow \infty} \frac{1}{kN_{\text{hor}}d} \left| \int_0^{kN_{\text{hor}}d \Delta\mu_0} e^{j [\beta_{\text{hor}} t - \dot{f}(\mu_0)t - \frac{1}{2} \dot{f}'(\mu_0) \frac{t^2}{kN_{\text{hor}}d}]} dt \right| \\
&= \lim_{kN_{\text{hor}}d \rightarrow \infty} \frac{1}{kN_{\text{hor}}d} \left| \int_0^{kN_{\text{hor}}d \Delta\mu_0} e^{j [\beta_{\text{hor}} t - \dot{f}(\mu_0)t]} dt \right| \quad (29) \\
&\leq \lim_{kN_{\text{hor}}d \rightarrow \infty} \frac{2}{kN_{\text{hor}}d} \\
&= 0,
\end{aligned}$$

where  $\textcircled{1}$  is  $t = kN_{\text{hor}}d \ddot{u}$ . So  $\lim_{kN_{\text{hor}}d \rightarrow \infty} AFM_{\text{norm}}(\beta_{\text{hor}})|_{\ddot{u} \in (\mu_0, \mu_0 + \Delta\mu_0)} = 0$  when  $\beta_{\text{hor}} \neq \dot{f}(\mu_0)$ .

If  $\beta_{\text{hor}} = \dot{f}(\mu_0)$ , then the following holds:

$$\begin{aligned}
& \lim_{kN_{\text{hor}}d \rightarrow \infty} AFM_{\text{norm}}(\beta_{\text{hor}})|_{\ddot{u} \in (\mu_0, \mu_0 + \Delta\mu_0)} \\
&= \lim_{kN_{\text{hor}}d \rightarrow \infty} \left| \int_0^{\Delta\mu_0} e^{j k N_{\text{hor}} d [-\frac{1}{2} \dot{f}'(\mu_0) \ddot{u}^2]} d\ddot{u} \right| \\
&\stackrel{\textcircled{2}}{=} \lim_{kN_{\text{hor}}d \rightarrow \infty} \frac{1}{\sqrt{\dot{f}'(\mu_0)}} \left| \int_0^{\sqrt{\dot{f}'(\mu_0)} \Delta\mu_0} e^{j k N_{\text{hor}} d t^2} dt \right| \quad (30) \\
&= \frac{K}{\sqrt{\dot{f}'(\mu_0)}}
\end{aligned}$$

where  $\textcircled{2}$  is  $t = \sqrt{\dot{f}'(\mu_0)} \ddot{u}$ , and  $K$  is a variable independent to  $\dot{f}(\mu)$ .

Since  $\dot{f}(\mu)$  is an increasing differentiable function in the range of  $[\dot{f}(0), \dot{f}(1)]$ , we can find only one  $\mu_0$  satisfies  $\dot{f}(\mu_0) = \beta_{\text{hor}}$  when  $\beta_{\text{hor}} \in [\dot{f}(0), \dot{f}(1)]$ . Thus we can get:

$$\lim_{K N_{\text{hor}} \rightarrow \infty} AFM_{\text{norm}}(\beta_{\text{hor}}) = \begin{cases} \frac{K}{\sqrt{\dot{f}'(\mu_0)}}, & \beta_{\text{hor}} \in [\dot{f}(0), \dot{f}(1)], \mu_0 = \dot{f}^{-1}(\beta_{\text{hor}}), \\ 0, & \beta_{\text{hor}} \notin [\dot{f}(0), \dot{f}(1)]. \end{cases} \quad (31)$$

Comparing (32) and (33), in the wide beamforming problem we can get  $\dot{f}(0) \equiv \psi_{\text{hor},a}$ ,  $\dot{f}(1) \equiv \psi_{\text{hor},b}$  and:

$$K \sqrt{\left. \frac{d\mu}{d\dot{f}(\mu)} \right|_{\mu=\mu_0}} \equiv h(\beta_{\text{hor}}); \quad (32)$$

where  $K$  is a coefficient and  $\beta_{\text{hor}} \equiv \dot{f}(\mu_0)$ . We do not need to normalize the total power when selecting  $h(\beta_{\text{hor}})$ , as the selection of  $K$  can automatically achieve normalization. So the value of  $h(\beta_{\text{hor}})$  does not actually represent the true value of  $AFM_{\text{norm}}(\beta_{\text{hor}})$ , but rather represents the ratio relationship between the values of  $AFM_{\text{norm}}(\beta_{\text{hor}})$  at different  $\beta_{\text{hor}}$ . Hence, the following differential equation holds:

$$K\mu \equiv \int h(\dot{f}(\mu)) d\dot{f}(\mu). \quad (33)$$

Suppose that the solution of (33) is:

$$K\mu \equiv H(\dot{f}(\mu)) \pm C. \quad (34)$$

Putting  $\dot{f}(0) \equiv \psi_{\text{hor},a}$  and  $\dot{f}(1) \equiv \psi_{\text{hor},b}$  into (34), we can get:

$$\mu \equiv \frac{1 - \theta}{H(\psi_{\text{hor},b}) - H(\psi_{\text{hor},a})} (H(\dot{f}(\mu)) - H(\psi_{\text{hor},a})) \pm \theta. \quad (35)$$

Hence,  $\dot{f}(\mu)$  becomes:

$$\dot{f}(\mu) \equiv H^{-1}((H(\psi_{\text{hor},b}) - H(\psi_{\text{hor},a}))\mu + H(\psi_{\text{hor},a})). \quad (36)$$

When  $N_{\text{hor}}$  tends to be infinity,  $\dot{f}(\mu)$  can be considered to have a constant value  $\dot{f}(\frac{n_{\text{hor}}}{N_{\text{hor}}})$  in  $(n_{\text{hor}} = 1; n_{\text{hor}})$ . Accordingly, (32) can be simplified as:

$$g_{\text{hor}}(n_{\text{hor}}) \equiv \sum_{\tau=1}^{n_{\text{hor}}} \dot{f}(\frac{\tau}{N_{\text{hor}}}). \quad (37)$$

Intuitively, for the hierarchical codebook, the best wide beam power distribution is the power evenly distributed in the beam range. Hence,  $h(\beta_{\text{hor}})$  is a constant function, and  $\dot{f}(\mu)$  can be simplified as:

$$\dot{f}(\mu) \equiv (\psi_{\text{hor},b} - \psi_{\text{hor},a})\mu + \psi_{\text{hor},a}. \quad (38)$$

Putting them into (??), we can get the beam  $\xi_i^s$  ( $s \neq S$ ) as:

$$\xi_i^s[n_{\text{hor}}] = \exp\left\{-jkd\left[\frac{n_{\text{hor}}(n_{\text{hor}} + 1)}{2N_{\text{hor}}} \cdot \frac{4}{2^s} - 2 + \frac{4}{2^s}(i - 1)\right]\right\}. \quad (43)$$

$$s = 1, 2, \dots, S - 1, \quad i = 1, 2, \dots, 2^s.$$

For the  $\mathbf{h}_{\text{ver}}$  direction, the hierarchical codebook are exactly the same as that in the  $\mathbf{h}_{\text{hor}}$  direction.

Fig. ?? shows the structure of the codebook.

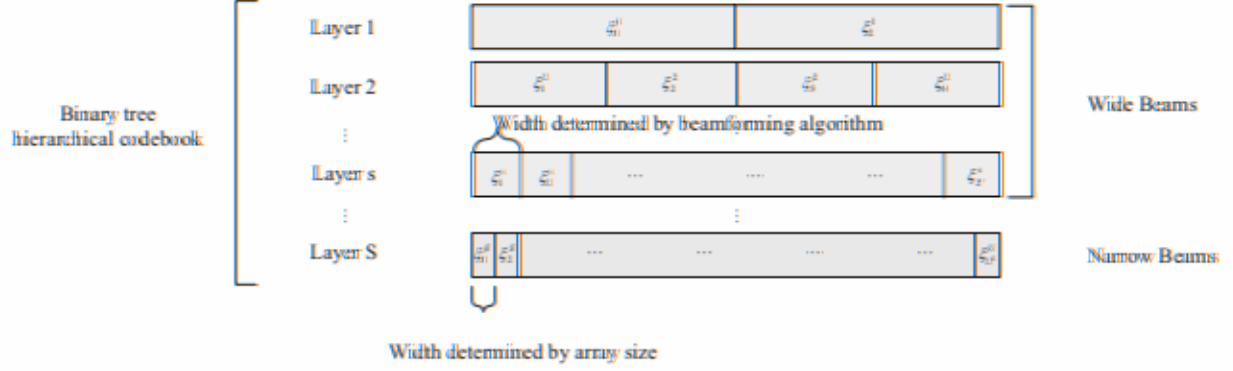


Fig. 2. Hierarchical codebook structure.

### B. Beam Training Scheme for Jointment Searching

In this sub-section, we will propose a searching scheme to search  $\mathbf{h}_{\text{hor}}$  direction and  $\mathbf{h}_{\text{ver}}$  direction at the same time which is called jointment searching (JS) scheme. The phase shift matrix of the IRS is searched by using the above-described hierarchical codebook. According to (??), the beam direction of the IRS can be divided into the  $\mathbf{h}_{\text{hor}}$  direction and the  $\mathbf{h}_{\text{ver}}$  direction, and the JS search scheme will search in these two directions synchronously. In the  $s$ th layer search, codewords in  $\Xi^1$  are used in the  $\mathbf{h}_{\text{hor}}$  and  $\mathbf{h}_{\text{ver}}$  directions. According to (??), the phase shift matrix of IRS in the  $s$ th layer search can be written as:

$$\Phi_{p,q} = \text{diag}(\xi_p^s \otimes \xi_q^s). \quad (44)$$

Then, according to (??), the received signal is obtained, and the received signal with the largest power is selected, that is, the indices of the best code of  $s$ th layer is selected. According to (??) we know:

$$\{p^*, q^*\} = \arg \max_{p,q} \mathbf{d}^H \Phi_{p,q} \mathbf{a}. \quad (45)$$

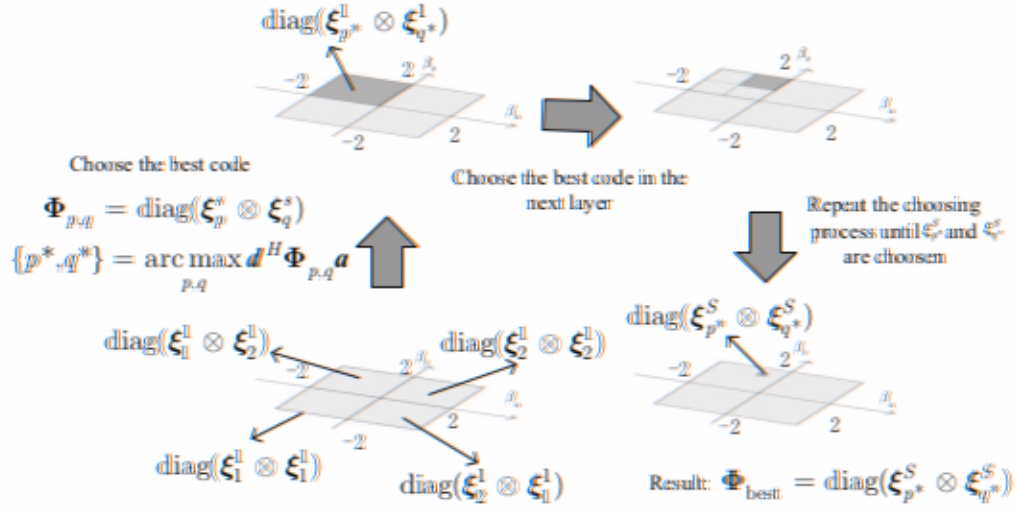


Fig. 3. JS beam training process.

At this point, the searching procedure of the  $s$ th layer is finished, and then we can search the codes in the next layer. Finally we can get the best phase shift matrix of IRS after the bottom layer of the hierarchical codebook are searched. The details of the proposed algorithm are summarized in Algorithm ??.

---

**Algorithm 1** JS codebook design

---

- 1: Initialize the layer index  $s = 1$  and  $p^* = 1$ .
  - 2: **repeat**
  - 3:   Compute  $\Phi_{p,q}$  by (??), where  $p \in \{2p^* - 1, 2p^*\}$ ,  $q \in \{2q^* - 1, 2q^*\}$ .
  - 4:   Update  $p^*$  by choosing the indices of the best codes of the  $s$ th layer by (??).
  - 5:   Update  $s = s + 1$ .
  - 6: **until**  $s > S$ .
  - 7:  $\Phi_{\text{best}} = \text{diag}(\xi_{p^*}^S \otimes \xi_{q^*}^S)$ .
- 

### C. Direction Wise Searching Scheme

In this section, we propose another beam searching scheme on IRS named direction wise searching (DWS) scheme. The difference between the DWS and the JS scheme is that the DWS

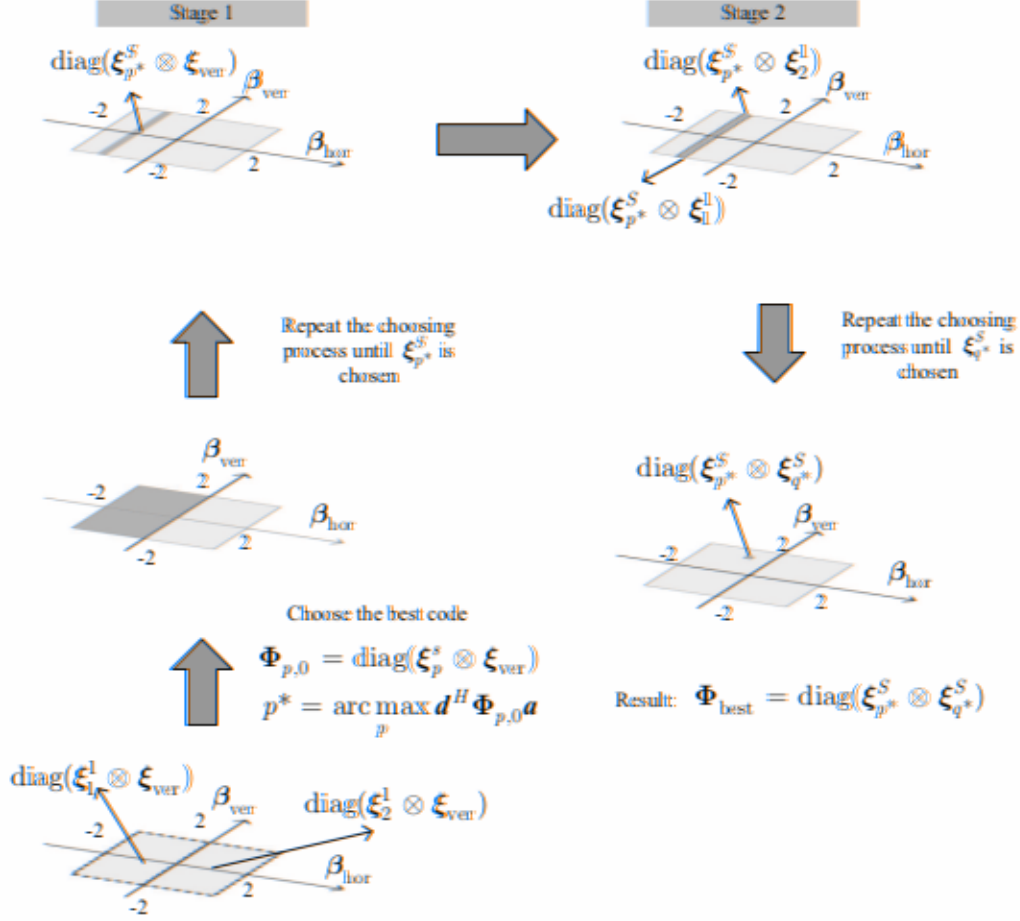


Fig. 4. DWS beam training process.

In the stage 2, the  $\mathbf{h}_{\text{hor}}$  direction uses the selected codewords in the stage 1, that is  $\xi_{p^*}^S$ , and the  $\mathbf{h}_{\text{ver}}$  direction uses  $\Xi$  for hierarchical search. For the  $s$ th layer search, the phase shift matrices of IRS can be written as:

$$\Phi_{p^*,q} = \text{diag}(\xi_{p^*}^S \otimes \xi_q^S). \quad (49)$$

In the same way, the optimal  $q$  is selected according to the received signal:

$$q^* = \arg \max_q d^H \Phi_{p^*,q} a. \quad (50)$$

Repeat the above searching procedure until the codes in the last layer of the hierarchical codebook are searched. Finally we get the best codeword  $\Phi_{\text{best}} = \text{diag}(\xi_{p^*}^S \otimes \xi_{q^*}^S)$ . The details of the proposed algorithm are summarized in Algorithm ??.



center direction, no matter how many sub-beams are used for combination, but the phenomenon of deep depression cannot be avoided.

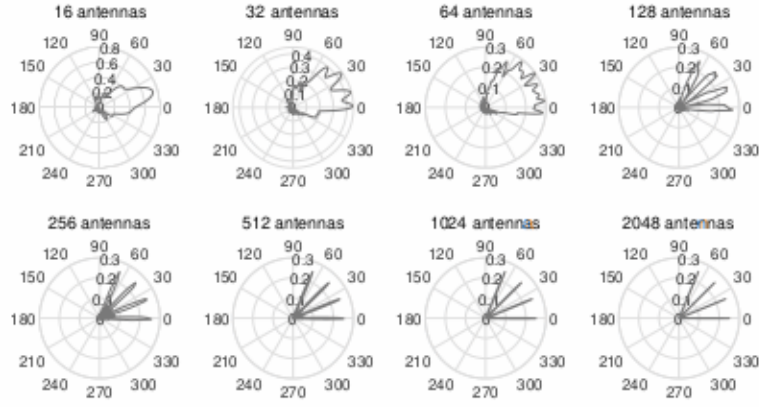
From Fig. ??(c), we can see that the NCPD algorithm is slightly better than the beam combination algorithm when the number of elements is small. However, when the number of elements increases, the NCPD algorithm does not achieve deep depression, and its performance is much better than that of the beam combination algorithm. With the increase in the number of elements, the side lobes of the beam formed by the NCPD algorithm become smaller, the main lobe becomes flatter and the edge of the main lobe becomes sharper, which gradually tends to our design goals. This shows that the NCPD algorithm can solve the problem of deep recesses when the number of elements increases. The NCPD algorithm is applicable to arrays of any size, and there is no need to artificially change the beamforming algorithm for the array size. Therefore, we can conclude that the NCPD algorithm is robust to the number of elements. Therefore, the time complexity of the NCPD algorithm is  $\mathcal{O}(1)$ , and the performance and speed in solving the IRS phase shift matrix in different scenarios are significantly better than those of the beam combination algorithm.

### B. Comparison under Different Beam Width.

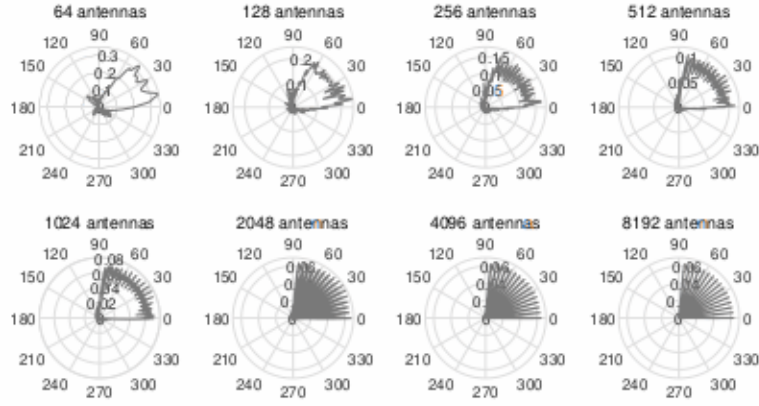
Fig. ?? shows the performance comparison when the beam combination and NCPD algorithms form wide beams with widths of 0.25, 0.5, 1, and 2 on 256 elements, respectively. Figs. ??(a)-??(d) show the performance of the 4-combination, 6-combination, and the NCPD algorithm, respectively. We can see that both the 4-combination and 6-combination algorithms have deep depression depressions they synthesize beams with large widths and wide deep depression of the 4-combination algorithm algorithm is serious. The performance of the NCPD algorithm is slightly better than those of the 4-combination and 6-combination algorithms when they form narrow beams. However, when wide beams are formed, the deep depression effect is better than those of the two beam combination algorithms. From this, it can be concluded that the NCPD algorithm is also robust when forming beams of different widths.

### C. Performance of NCPD Algorithm on Giant IRS

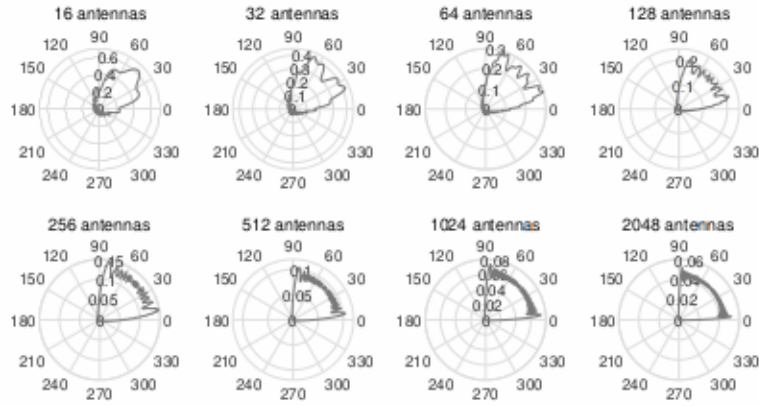
Fig. ?? shows the results of the NCPD algorithm forming beams of different widths for the number of elements of 10000, where we can see that the beam has almost no side lobes, its edge is sharply cut off, and the power in it is flatly distributed. This shows the superiority of the



(a) 4-combination algorithm.



(b) 16-combination algorithm.



(c) NCPD algorithm.

Fig. 5. Comparison of beam combination algorithm and NCPD algorithm in different sizes of IRS.



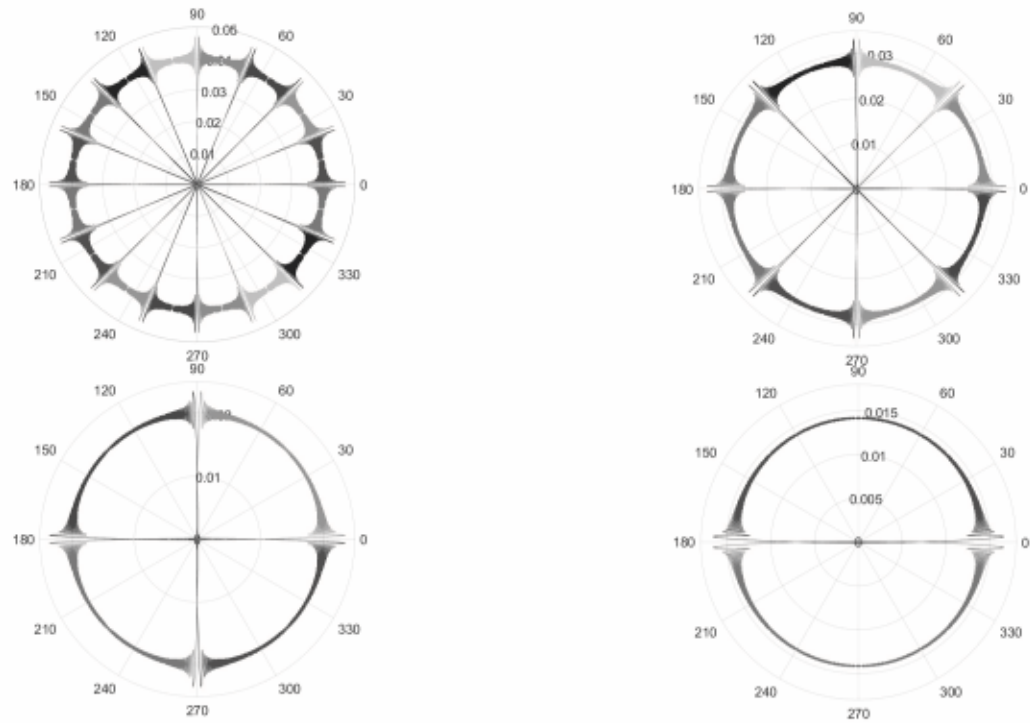


Fig. 7. The beams formed by NCPD algorithm on 10000 element IRS.

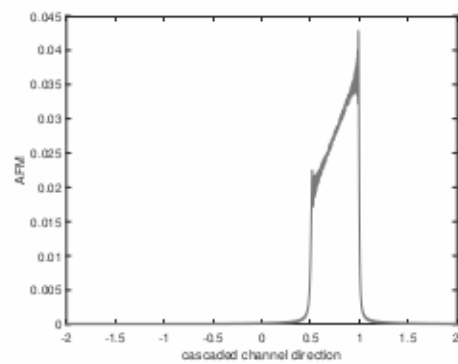


Fig. 8. The trapezoidal beams formed by NCPD algorithm on 10000 element IRS.

#### D. Comparison of Misalignment Rates between NCPD and Beam Combination Algorithms

Next, we compare the misalignment rates of wide beams formed by the NCPD, beam combination algorithms and the beam in the ideal case. The beam combination algorithm is the BMW-SS algorithm proposed in [?]. In the BMW-SS algorithm, a binary tree is also used for implementing a hierarchical codebook, in which the bottom layer also uses a narrow beam. In the case of wide beams, the BMW-SS algorithm uses  $2^{\lfloor \frac{l+1}{2} \rfloor}$  sub-beams for synthesis, where  $l$  represents the number of layers away from the bottom layer. For example, when forming a wide beam of the penultimate layer,  $l = 1$ , that is, the array is divided into two sub-arrays, and two sub-beams are respectively realized, and then the two sub-beams are synthesized into the required wide beam. Since the wider the beam, the smaller is the average power in each direction, the codeword of the top layer has the highest probability of misalignment so under the same signal-to-noise ratio (SNR). Therefore, the misalignment rate of the hierarchical codebook mainly depends on that of the first layer of codewords. Hence, we show the comparison of the misalignment only of the first layer of codewords. The ideal case means the ideal beams in the coverage of  $(-2, 0)$  and  $(0, 2)$  with the power of  $2N_{ver}$  or  $2N_{hor}$ , which are the target wide beams of the first layer of the hierarchical codebook. It is worth mentioning that there're no codewords corresponding to ideal case. We just suppose the formed beams are in the ideal shape and test the misalignment rate with the use of beams in this shape.

We conducted two experiments to show the comparison between the NCPD and the BMW-SS algorithms under different IRS received SNRs and different IRS array sizes. The IRS received SNR refers to the ratio of the total received power of IRS to the noise power of the cascaded channel. Since we choose the total received SNR of IRS rather than the received SNR on a single element, we can exclude the gain caused by the increase in the IRS array size, and show only the beamforming gain of IRS.

Fig. ?? shows the misalignment curves of the first-layer codewords of the NCPD and the BMW-SS algorithms and the ideal case under different IRS received SNRs on the uniform linear array (ULA) of 256 elements. We can see that when SNR is less than 0dB, the misalignment rates of the two algorithms are high. This is because the noise power at this stage is too large, and IRS cannot use the beamforming algorithm to offset the noise interference. But when SNR is greater than 0dB, the misalignment rate of the NCPD algorithm is significantly lower than that of the BMW-SS algorithm as the power distribution in the beam formed by the NCPD

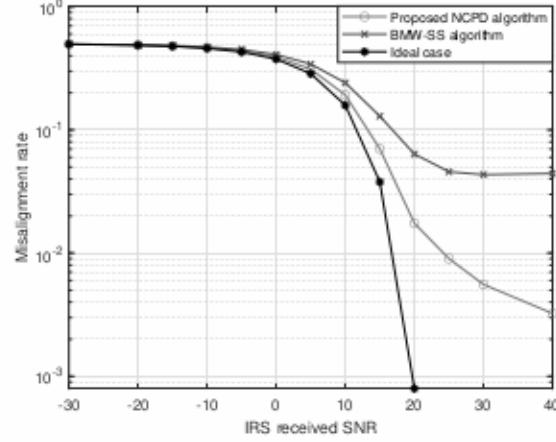


Fig. 9. Comparison of beam misalignment rates under different IRS received SNR.

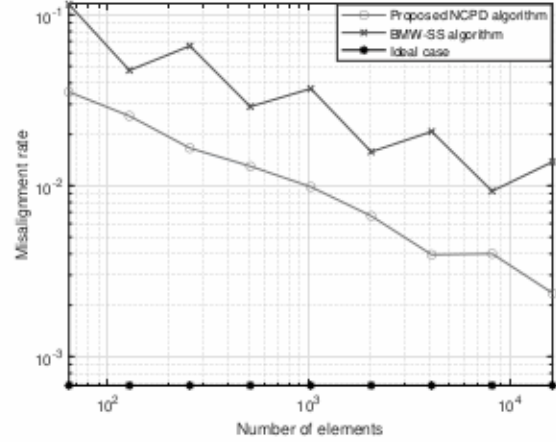


Fig. 10. The misalignment rates on different IRS size when IRS received SNR is 20 dB.

algorithm is more uniform. At the same time, we noticed that when SNR is greater than 30dB, the misalignment rate of the NCPD algorithm is still declining, while that in the BMW-SS algorithm is almost unchanged. This is because when SNR is high, the deep depression problem becomes the main reason for limiting the performance of the BMW-SS algorithm, while the performance of the NCPD algorithm is gradually improved with increasing SNR as it has no deep depression problem.

Fig. ?? and Fig. ?? show the misalignment rates of the NCPD and BMW-SS algorithms on ULAs of different sizes when the IRS received SNR is 10dB and 20dB, respectively. The reason why the curve of the BMW-SS algorithm is a broken line is that the relationship between the

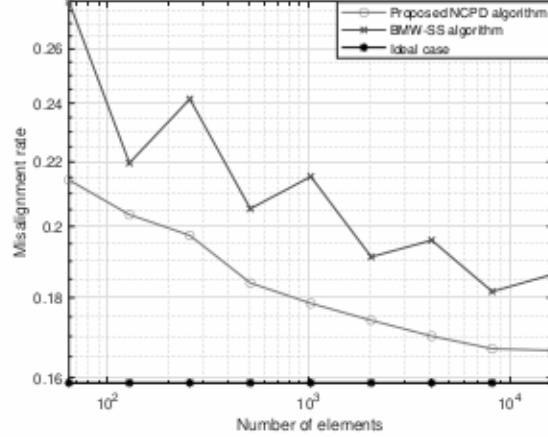


Fig. 11. The misalignment rates on different IRS size when IRS received SNR is 10 dB.

number of sub-arrays and number of array elements is not a continuous function. For example, when the number of elements is 128 or 256, the BMW-SS algorithm divides IRS into 16 sub-arrays. But when the number of elements is 512, it divides IRS into 32 sub-arrays. When the IRS array size is in the range from 128 to 256, the number of sub-arrays does not change. But due to the increase in the total number of elements, the number of elements in each sub-array increases, which leads to a more serious deep depression problem in the combined beam and an increase in the misalignment rate of the beam. This also shows that the deep depression problem is the performance bottleneck when dealing with large-scale array problems using the conventional beam combination algorithm. The performance of the NCPD algorithm is gradually improved with the increase in the array size. In all the experimental array sizes, the misalignment of the NCPD algorithm is lower than that of the BMW-SS algorithm, which proves that the NCPD algorithm is superior in solving the beamforming problem of large-scale IRS arrays.

## VIII. CONCLUSION

Aiming at the characteristics of IRS-aided communication scenarios, in this paper we proposed a new array factor model. In this model, we proposed the variables of the cascaded channel direction to replace the conventional elevation and azimuth angles for describing the AoA and AoD of IRS which can reduce the number of parameters. In addition, we used the integral form to replace the summation form in the conventional array factor. In order to control the shape of the array factor function of IRS, we proposed the NCPD algorithm and an analytical solution of