

# **Sensing RISs: Enabling Dimension-Independent CSI Acquisition for Beamforming**

**(Original paper ID: IT-22-0267)**

## **Response Letter**

Jieao Zhu, *Student Member, IEEE*, Kunzan Liu, *Student Member, IEEE*,  
Zhongzhichao Wan, *Student Member, IEEE*, Linglong Dai, *Fellow, IEEE*,  
Tie Jun Cui, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

## **Response to Editor's Comments**

**Editor:** Manuscript ID IT-22-0267 titled “Sensing RISs: Enabling Dimension-Independent CSI Acquisition for Beamforming” which you submitted to the IEEE Transactions on Information Theory, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter. There may be additional comments as separate files available through your Author Center on the ScholarOne Manuscripts web site, so please check there also to make sure you have received all reviewer comments.

Based on my own detailed reading of the paper and the reviewers' comments, I recommend that you revise and resubmit your manuscript. Specifically, in your revision, the following points must be addressed:

- 1) Two out of the three reviewers have pointed that the paper is missing a strong information-theoretic element. I would recommend to consider adding results that address that concern.
- 2) Another major concern is that methods appear ad-hoc and not particularly novel. Again, it is important to address this concern directly.

One of the reviewers has expressed support for the paper, so I hope you will take a step back and consider revisiting the paper to address all the concerns in a significant manner.

## **Ingredients**

**Theorem 1** Suppose all the entries of  $\mathbf{G} \in \mathbb{C}^{N \times M}$  and  $\mathbf{f} \in \mathbb{C}^{N \times 1}$  obey  $\mathcal{CN}(0, 1)$  distribution, and the channel noise  $\mathbf{N} \sim \mathcal{CN}(0, \sigma^2)$ . Then, the LS channel estimator is given by

$$\hat{\mathbf{H}}_{\text{LS}} = (\mathbf{F}_{N,P}^* \mathbf{F}_{N,P}^\top)^{-1} \mathbf{F}_{N,P}^* \mathbf{Y}_{\text{BS}}^\top, \quad (1)$$

and the MMSE estimator is given by

$$\hat{\mathbf{H}}_{\text{MMSE}} = (\mathbf{F}_{N,P}^* \mathbf{F}_{N,P}^\top + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{F}_{N,P}^* \mathbf{Y}_{\text{BS}}^\top \quad (2)$$

*Proof.* Proof will be provided.

The equivalent model can be expressed by

$$\tilde{\mathbf{y}}_{\text{BS}} = \text{vec}(\mathbf{Y}_{\text{BS}}) = \mathbf{A}\mathbf{h} + \tilde{\mathbf{n}}, \quad (3)$$

where  $\mathbf{h} = \text{vec}(\mathbf{H}^\top) \in \mathbb{C}^{NM}$ ,  $\mathbf{A} = \mathbf{F}_{N,P}^\top \otimes \mathbf{I}_M$ , and  $\tilde{\mathbf{n}} = \text{vec}(\mathbf{N})$ . In this model, if we assume  $\mathbf{C}_h =$

**Theorem 2** With the MMSE channel estimator, the second-order statistics of the estimation error

$$\boldsymbol{\epsilon} = \text{vec}((\hat{\mathbf{H}}_{\text{MMSE}} - \mathbf{H})^\top) \quad (4)$$

is given by

$$\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H | \hat{\mathbf{h}}] = (\mathbf{F}_{N,P}^* \mathbf{F}_{N,P}^\top / \sigma^2 + \mathbf{I}_N)^{-1} \otimes \mathbf{I}_M. \quad (5)$$

Note that the distribution of such an estimation error  $\boldsymbol{\epsilon}$  is not necessarily Gaussian.

**Authors:** We would like to commence by thanking the editor and three professional reviewers for their valuable time in evaluating our submission. Your constructive comments and expert knowledge of the field have helped us to strengthen the manuscript significantly. We endeavored to address all the suggestions and comments, and our reflections are provided below in a point-by-point way. We also indicate how our manuscript has been revised accordingly, and all the revisions have been highlighted in red color in the revised paper. Here, we would like to make a brief summary of the major revisions in the revised paper as below:

1) We have clarified all the unclear statements pointed out by the reviewers, such as “why the lens antenna array can realize the spatial DFT” and “why we design the overall combining matrix  $\bar{\mathbf{W}}$  as a Bernoulli random matrix”.

2) We have justified all impractical assumptions pointed out by the reviewers, such as “the number of path may not be known in advance” and “the user may prefer the conventional antenna array instead of lens antenna array”.

3) We have added the performance analysis of the proposed SSD-based scheme. It shows that this scheme can estimate the key parameters  $\varphi_{l,c}$  for  $l = 1, 2, \dots, L$  with high probability.

**Editor:** As an editor, my major concern is the model (6), where  $\mathbf{W}_q$  is known. Note that  $\mathbf{W}_q$  is an  $N_{\text{RF}} \times N$  matrix combining the adaptive selecting network and RF chains, and  $\tilde{\mathbf{h}}_m$  is majorly the propagation channel. Typically, in wireless communication, RF impairments are captured in the baseband channel, and the propagation channel + RF chain gain/phase are the estimated baseband channel. But, in the assumed model,  $\mathbf{W}_q$  is an  $N_{\text{RF}} \times N$  matrix and has  $N_{\text{RF}} \times N$  elements. We may assume that  $\mathbf{W}_q$  is known using a nominal value, but in this case the unknown impairment or gain/phase offset in  $\mathbf{W}_q$  (from the known nominal  $\mathbf{W}_q$ ) should be able to be captured in  $\tilde{\mathbf{h}}_m$ . But,  $\tilde{\mathbf{h}}_m$  is a column vector, so the unknown impairment or gain/phase offset in  $\mathbf{W}_q$  should be identical for each column. I am not sure about this. If so, please explain this in the revision. If the unknown impairment is elementwise in  $\mathbf{W}_q$ , then the model (6) is not valid.

**Authors:** Thanks for pointing out this important issue. Actually, the RF impairments of  $\mathbf{W}_q$  include two parts. The first one is the impairment incurred by the RF chains. This impairment should be identical for each row of  $\mathbf{W}_q$ , and therefore can be captured in the estimated channel as pointed out by the Editor. The second one is the impairment incurred by the phase shifters in the adaptive selecting network. This impairment is element-wise, and it cannot be fully captured in the estimated channel. We would like to point out that this problem commonly exists in most of the channel estimation schemes for hybrid analog and digital architecture, since the analog module (e.g., phase shifter network or lens) is usually involved in the channel estimation procedure [R1].

Fortunately, when hardware impairment incurred by the analog module is not serious (this is reasonable since only the analog module working stably will be used), it will not result in significant accuracy degradation in channel estimation. This conclusion can be verified by the following Fig. 1, where the parameters are the same as those for Fig. 10 in the paper. In Fig. 1, the dash lines present the case without hardware impairment, i.e.,  $\mathbf{W}_{q,\text{real}} = \mathbf{W}_q$ , while the solid lines present the

case with hardware impairment incurred by the phase shifters in the adaptive selecting network, i.e.,  $\mathbf{W}_{q,\text{real}} = \mathbf{E} \odot \mathbf{W}_q$ , where  $\odot$  denotes the element-wise multiplication,  $\mathbf{E}$  is an  $N_{\text{RF}} \times N$  impairment matrix. Note that each element  $e_{ij}$  of  $\mathbf{E}$  is modeled as  $e_{ij} = |e_{ij}| \exp(j\phi_{ij})$ , where  $|e_{ij}| \sim \mathcal{U}(0.9, 1.1)$  presents the gain with offset,  $\phi_{ij} \sim \mathcal{U}(-10^\circ, 10^\circ)$  presents the phase with offset, and  $\mathcal{U}(a, b)$  denotes the uniform distribution between  $a$  and  $b$  [R1]. Fig. 1 shows that the accuracy degradation caused by the hardware impairment is not significant, and the proposed SSD-based scheme can still achieve the satisfying accuracy close to the ideal oracle LS scheme.

Fig. 1. Impact of hardware impairment on the proposed SSD-based scheme.

In the revised paper, we have addressed this issue as:

Finally, it should be noted that the hardware impairment may exist in the adaptive selecting network, leading to the element-wise gain/phase offset in  $\mathbf{W}_q$ , which cannot be fully captured in the estimated channel. This is a common problem existing in most of channel estimation schemes for hybrid analog and digital architecture [4], since the analog module (e.g., phase shifter network) is usually involved in the channel estimation procedure. Fortunately, since the gain/phase offsets are usually not serious in practice, the accuracy degradation in channel estimation caused by this hardware impairment will not be significant.

[R1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

**Authors:** Many thanks again for your valuable time and efforts to review this paper.

Sincerely,

*The Authors*

## Response to Reviewer 1's Comments

**Reviewer 1:** The paper proposes an efficient technique for the wideband beamspace channel estimation problem for mmWave using lens antenna arrays. By exploiting similar beam directions across antennas for a given path, the proposed technique can estimate mmWave channels using lens antennas without assuming a common support set across frequency bands. The proposed technique seems to be novel, and the paper is well written. I have some comments before its publication.

**Authors:** We appreciate the reviewer's concise summary of the key points of this paper and the positive evaluation on our work. We have tried our best to revise the paper according to the reviewer's valuable comments, and our responses are provided in the point-by-point way as below.

**Reviewer 1:** 1) It is not clear at first why, for a given path, the support sets of different antennas are similar as in (9). I think this can be intuitively explained using a figure with antennas and beams, i.e., with far field assumption in microwave, the beam direction is the same for all antennas, with extremely small wavelength of mmWave; however, the beam direction would be different, but similar, across antennas. This conceptual explanation would help understanding the core idea of paper.

**Authors:** Thanks for the suggestion of clearly explaining why the support of each path component has the similar beam indices as in Eq. (9). To answer this question, we first provide the definition of each path component in the beamspace as below:

$$\tilde{\mathbf{c}}_{l,m} = \mathbf{U}_a^H \mathbf{a}(\varphi_{l,m}) = [\Xi(\varphi_{l,m} - \bar{\varphi}_1), \Xi(\varphi_{l,m} - \bar{\varphi}_2), \dots, \Xi(\varphi_{l,m} - \bar{\varphi}_N)]^T,$$

where  $\Xi(x) = \frac{\sin N\pi x}{\sin \pi x}$  is the Dirichlet sinc function as shown in the following Fig. 1. Ideally, when the  $\varphi_{l,m}$  is exactly the same as one of the beam directions  $\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N$  pre-defined by lens antenna array, the power of  $\tilde{\mathbf{c}}_{l,m}$  will be focused on only one beam, as shown by the red dash lines in Fig. 1. However, in practice,  $\varphi_{l,m}$  is arbitrary, which may be different from the pre-defined beam directions. In this case, the path power will be distributed on several pre-defined beams. According to the property of  $\Xi(x)$ , i.e.,  $\Xi(x)$  is larger when  $x$  is closer to 0, we know that the indices of these power-focused beams should be adjacent (similar) as shown by the blue dot dash lines in Fig. 1. More detailed proof can be found in [R1, Lemma 2].

In the revised paper, we have explained Eq. (9) more clearly as:

The reasonability of (9) can be explained as follows. In practice,  $\varphi_{l,m}$  is arbitrary, which is usually different from the pre-defined beam directions  $\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N$ . In this case, the power of  $\tilde{\mathbf{c}}_{l,m}$  will be distributed on several beams. According to the property of  $\Xi(x)$  in  $\tilde{\mathbf{c}}_{l,m}$ , i.e.,  $\Xi(x)$  is larger when  $x$  is closer to 0, we know that the indices of these power-focused beams should be adjacent. More detailed proof can be found in [18, Lemma 2].

Fig. 1. The normalized power distribution of  $\tilde{\mathbf{c}}_{l,m}$ .

[R1] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010-6021, Sep. 2017.

**Reviewer 1:** 2) The paper assumes that the receiver combining matrix  $\bar{\mathbf{W}}$  consists of -1 or +1 with proper normalization. Is this a necessary condition for the proposed technique? If yes, please explicitly mention this in the paper. If not, I think it would be better to generalize the combining matrix.

**Authors:** Thanks for this important comment. The receiver combining matrix  $\bar{\mathbf{W}}$  with normalized elements of -1 or 1 (i.e.,  $\bar{\mathbf{W}}$  is a Bernoulli random matrix) is not a necessary condition for the proposed technique. We use the Bernoulli random matrix due to the following two reasons: i) it can be easily realized by the 1-bit phase shifters with low hardware cost and energy consumption [R1]; ii) it enjoys low mutual-column coherence, and therefore can be expected to achieve high recovery accuracy according to the compressive sensing theory [R2].

However, if  $\bar{\mathbf{W}}$  is not a Bernoulli random matrix, our scheme can still work. One example is shown in the following Fig. 2. Here, we design  $\bar{\mathbf{W}}$  as a matrix, whose elements have the same normalized amplitude  $\frac{1}{\sqrt{QN_{\text{RF}}}}$  but with arbitrary phases (realized by the high-resolution phase shifters). The other parameters for Fig. 2 are the same as those for Fig. 10 in the paper. It shows that our scheme can still achieve higher accuracy than the conventional schemes.

Fig. 2. NMSE comparison against the SNR for channel estimation when the elements of  $\bar{\mathbf{W}}$  have the same normalized amplitude but with arbitrary phases.

To address this comment, in the revised paper, we have modified the description of  $\bar{\mathbf{W}}$  as:

Furthermore, we define  $\bar{\mathbf{W}} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_Q^T]^T$  of size  $QN_{\text{RF}} \times N$  as the overall combining matrix, which is designed according to the hardware realization of the adaptive selecting network. For example, if the adaptive selecting network is realized by the low-cost 1-bit phase shifters as in [18], the elements of  $\bar{\mathbf{W}}$  can be randomly selected from the set  $\frac{1}{\sqrt{QN_{\text{RF}}}} \{-1, +1\}$  with equal probability.

[R2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

**Reviewer 1:** 3) In the first term (inside the integral) in the last line of (20), isn't  $-\Delta\varphi$  correct?

**Authors:** Thanks for the reviewer's careful check. The first term in the last line of Eq. (20) should be  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} - \Delta\varphi \right) d\Delta\varphi$ . However, it can be also equivalently expressed as  $\int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} + \Delta\varphi \right) d\Delta\varphi$  as shown in the previous paper. This can be explained as follows:

$$\begin{aligned} \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} - \Delta\varphi \right) d\Delta\varphi &\stackrel{(a)}{=} \int_{\frac{\alpha_n}{2}}^{-\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} + x \right) d(-x) \\ &= \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} + x \right) dx = \int_{-\frac{\alpha_n}{2}}^{\frac{\alpha_n}{2}} \Xi^2 \left( \frac{\Delta_n}{N} + \Delta\varphi \right) d\Delta\varphi, \end{aligned}$$

where (a) is obtained by letting  $x = -\Delta\varphi$ .

**Reviewer 1:** 4) In Fig. 5, it would be better to have some intuitive explanations why  $\Delta_n$  varies with  $n$ .

**Authors:** Thanks for this suggestion. In the revised paper, we have explained why  $\Delta_n$  varies with  $n$  as:

Moreover, Fig. 5 also shows that the derived  $\Delta_n$  varies when  $\varphi_{l,c}$  equals different  $\bar{\varphi}_n$  (leading to different  $C_n$  defined in **Lemma 2**). The reason is that different  $\varphi_{l,c}$  incur different degrees of power leakage due to beam squit. Therefore, we need to change  $\Delta_n$  to make sure that we can always capture most of the power of  $C_n$ .

**Reviewer 1:** 5) In the simulation section, please define NMSE mathematically.

**Authors:** Thanks for this suggestion. In the revised paper, we have added the mathematical definition of NMSE in Section V (Simulation Results) as:

Finally, we use the normalized mean square error (NMSE) to measure the accuracy of channel estimation for each user, which is mathematically defined as

$$\mathbb{E} \left( \frac{1}{M} \sum_{m=1}^M \frac{\|\tilde{\mathbf{h}}_m - \tilde{\mathbf{h}}_m^e\|_2^2}{\|\tilde{\mathbf{h}}_m\|_2^2} \right), \quad (77)$$

where  $\tilde{\mathbf{h}}_m^e$  is the estimated beamspace channel at sub-carrier  $m$ .

**Reviewer 1:** 6) The authors explained why there exists an error floor at high SNR in Fig. 9. It is not clear with this explanation why “Oracle LS scheme” also experiences such a floor. Please elaborate this.

**Authors:** Thanks for this insightful comment. Actually, in the oracle LS scheme, we also regard the elements of beamspace channel with very low power as zeros, as we have done in all the considered schemes. The only difference is that the positions of the elements with large power are assumed to be perfectly known in the oracle LS scheme, and that is why the word “oracle” is used. Therefore, even for the oracle LS scheme, it still experiences a NMSE floor at high SNR. In the revised paper, we have explained this issue more clearly as:

Note that for all the schemes mentioned above, we regard the elements of wideband beamspace channel with indices outside the support as zeros.

**Reviewer 1:** 7) In Fig. 11, why is the performance order of SOMP and OMP different with channel estimation SNRs?

**Authors:** Thanks for pointing out this important issue. When the SNR for data transmission is low, the sum-rate performance of wideband beam selection is dominated by noise. In this case, all the schemes



enjoy the same performance order even if the SNR for channel estimation is changed. However, when the SNR for data transmission is high, the sum-rate performance is dominated by channel estimation error. Therefore, different SNRs for channel estimation will lead to different performance order. In the revised paper, we have addressed this issues as:

Finally, we observe that the performance orders of all schemes change with the SNR for channel estimation. This is due to the fact that when the SNR for data transmission is high, the sum-rate performance of wideband beam selection is dominated by channel estimation error.

**Reviewer 1:** 8). In Fig. 13, the SNR for channel estimation should be quite high, the value that might be impractical, even for “Oracle LS scheme”. With this result, it seems a bit pessimistic to have accurate CSI with multiple antennas at UEs. Is there any way to improve this?

**Authors:** Thanks for this insightful comment. It is true that the required SNR for channel estimation is quite high in the case with multiple-antenna users. This is because that the dimension of beamspace channel is significantly increased, and the number of nonzero elements to be estimated is much larger than the number in the case with single-antenna users. This problem can be partly solved by increasing the pilot overhead. Moreover, as we can see from Fig. 11 in the previous paper (Fig. 12 in the revised paper), even if the SNR for channel estimation is not high enough (e.g., 0 dB or 15 dB), wideband beam selection can still achieve satisfying sum-rate. This is due to the fact that for data transmission, only the reduced-dimensional beamspace channel with much smaller size is effective. Therefore, when the SNR for channel estimation is not high enough, although the total NMSE performance is not satisfying, the reduced-dimensional beamspace channel may already be accurate. This means that we do not need to estimate the beamspace channel so accurately at the cost of high SNR in practice.

Fig. 11. (Fig. 12 in the revised paper) Sum-rate comparison against the SNR for data transmission.

In the revised paper, we have addressed this issue as:

Moreover, we would like to point out that in the case with multiple-antenna users, the required SNR for channel estimation to achieve satisfying NMSE is usually larger than the case with single-antenna users. However, as we can see from Fig. 12, even if the SNR for channel estimation is not high enough (e.g., 0 dB or 15 dB), wideband beam selection can still achieve satisfying sum-rate. The reason is that for data transmission, only the reduced-dimensional beamspace channel with much smaller size is effective. Although the NMSE performance may not be good enough with low SNR for channel estimation, the reduced-dimensional beamspace channel is already accurate enough for data transmission. Therefore, in practice we do not need to estimate the beamspace channel so accurately at the cost of high SNR.

**Authors:** Many thanks again for your valuable time and efforts to review this paper.

Sincerely,

*The Authors*

## Response to Reviewer 2's Comments

**Reviewer 2:** In this paper, the authors study a millimeter wave lens antenna array system. In particular, the authors focus on the beamspace channel estimation in wideband communication. It firstly shows that the antenna support of the beamspace channel depends on the frequency, and then proposes a successive support detection based scheme for channel estimation. The paper is well written in general. The studied topic is interesting and timely. The contribution is solid. I have the following comments.

**Authors:** We appreciate the reviewer's concise summary of the key contributions of our paper and the positive evaluation on our work as well. To address your professional comments, we have made several major revisions in the revised paper, which are described one by one as below.

**Reviewer 2:** 1) It is unclear what is the fundamental principle of lens antenna array so as to achieve the spatial discrete fourier transform (DFT). In particular, for wideband channels, is the response of the lens antenna array also frequency-dependent, such as the expression given in (1)?

**Authors:** Thanks for this insightful comment. The reason why lens antenna array realizes the function of spatial DFT has been analyzed in [R1, Lemma 1]. Briefly, the electromagnetic lens is similar to the optical lens, which is able to change the propagation directions of the electromagnetic rays to achieve the energy focusing or beam collimation capability. As shown in the following Fig. 1, this means that all the plane electromagnetic rays with a incidence  $\phi$  should arrive at one feed antenna  $B$  located on the lens focal surface with identical phase for constructive superposition. To achieve this target, we need to design the phase shift profile  $\Phi(y, z)$ , which is realized by the spatial phase shifter at any point  $(0, y, z)$  on the lens's aperture. With some mathematical derivations, it can be finally shown that based on the carefully designed phase shift profile, the electromagnetic rays received at the lens focal surface can be well-approximated as the spatial DFT of the ones received at the lens's aperture.

Next, we will explain whether the response of the lens antenna array is frequency-dependent. As we have mentioned above, realizing the power-focusing capability of lens requires us to carefully design the phase shift profile  $\Phi(y, z)$  of the spatial phase shifters on the lens's aperture. In general, the spatial phase shifters can be realized by different analog modules, such as a class of bandpass surfaces or

even the conventional phase shifters [R2]. However, most of this analog modules can not be adjusted according to different frequencies [R3]. As a result, when the electromagnetic rays with different frequencies arrive at the lens aperture, they will be focused on different feed antennas, which means that the response of lens antenna array can not be frequency-dependent as the expression given in Eq. (1) in the paper.

Fig. 1. A planar lens placed in the y-z plane with the feed antennas located on its focal surface.

Finally, we would like to admit that it is possible but challenging that a frequency-dependent lens antenna array compensating for the effect of beam squint can be manufactured. In this case, our proposed SSD-based scheme can be further simplified to the narrowband version, since we do not need to generate several beamspace windows to estimate the spatial direction of each path at the carrier frequency any more.

In the revised paper, we have addressed this issues as:

The reason why lens antenna array realizes the spatial DFT can be found in [10, **Lemma 1**]. It shows that the power-focusing capability of lens relies on the spatial phase shifters on lens's aperture, which usually cannot be adjusted according to different frequencies. As a result, the response of lens antenna array cannot be frequency-dependent as in (1). However, we would like to admit that it is possible but challenging that a frequency-dependent lens antenna array compensating for the effect of beam squint can be manufactured. In this case, the proposed SSD-based scheme can be further simplified to the narrowband version as we have proposed in [18], since the OFDM is adopted.

[R1] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557-1571, Apr. 2016.

[R2] M. A. Al-Joumayly and N. Behdad, "Wideband planar microwave lenses using sub-wavelength spatial phase shifters," *IEEE Trans. Antennas Propag.*, vol. 59, no. 12, pp. 4542-4552, Dec. 2011.

[R3] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal

processing techniques for millimeter wave MIMO systems,” *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

Fig. 3. (in the revised paper) Beam power variation over frequency.

**Reviewer 2:** 2) How is Fig. 3 related to the definition of the beamspace  $\tilde{\mathbf{c}}_{l,m}$  given in (5)? It seems that (5) is given only in the discrete frequencies at each sub-carrier  $m$ , whereas Fig. 3 is in continuous frequency. Besides, what does the index  $n$  represent?

**Authors:** Thanks for pointing out this important issue. In this revised paper, we have modified Fig. 3 to the discrete frequencies case (also provided as below) to make it consistent with Eq. (5). Moreover, we have also explained what is the index  $n$  as:

Fig. 3 shows the effect of beam squint from another perspective, where the parameters are the same as Fig. 2, and the curve indexed by  $n$  represents the power variation of the  $n$ th element (beam) of  $\tilde{\mathbf{c}}_{l,m}$  over frequency.

**Reviewer 2:** 3) In Fig. 1, an adaptive selecting network is shown after the lens antenna array, whereas below equation (7), a combiner matrix with 1-bit phase shifters is used. This seems to be inconsistent as we would simply expect antenna selection based on Fig. 1.

**Authors:** Thanks for this comment. In this paper, we utilize 1-bit phase shifters instead of the conventional switches to realize the adaptive selecting work as in [R4] due to the following two facts: i) 1-bit phase shifters also have low cost as the conventional switches. Therefore, the additional hardware cost and energy consumption incurred by 1-bit phase shifters are not significant [R4]; ii) the correspondingly designed  $\bar{\mathbf{W}}$  has been proved to enjoy low mutual-column coherence. Therefore, it can be expected to achieve high recovery accuracy according to the compressive sensing theory [R5].

Fig. 2. Typical architecture of phase shifter [R6].

However, we would like to point out that even if the adaptive selecting network is realized by 1-bit phase shifters, it can still be configured to realize the traditional function of beam selection during the data transmission. Specifically, we can turn off some phase shifters to realize “unselect” and set some

phase shifters to shift the phase 0 degree to realize “select” in beam selection. Next, we will explain how to turn off the phase shifters in practice. The typical architecture of phase shifter is shown in Fig. 2 as below, where the phase shifts are quantified and controlled by digital signals [R6]. Fig. 2 shows that the phase shifter can be considered as the combination between switches and transmission lines with different lengths. Therefore, when the phase shifter should be turned off, we can just turn off the corresponding switches.

In the revised paper, we have addressed this issue as:

It is worth pointing out that during the data transmission, such adaptive selecting network with 1-bit phase shifters can also be configured to realize the conventional beam selection [12]. To do this, we can turn off some phase shifters to realize “unselect” [32] and set some phase shifters to shift the phase 0 degree to realize “select”.

[R4] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010-6021, Sep. 2017.

[R5] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[R6] N. Tucker, “A low cost electro-mechanical phase shifter design, including a brief summary of solid state methods,” available at [www.activefrance.com](http://www.activefrance.com).

**Reviewer 2:** 4) Furthermore, the authors are advised to explain why a randomly selected combining matrix  $\bar{\mathbf{W}}$  is a good choice.

**Authors:** Thanks for this suggestion. In the revised paper, we have explained why we design  $\bar{\mathbf{W}}$  as a randomly selected matrix by:

The reason we adopt the randomly selected matrix is that it has been proved to enjoy low mutual-column coherence, and therefore can be expected to achieve high recovery accuracy according to

the compressive sensing theory [34].

**Reviewer 2:** 5) For the proposed SSD-based channel estimation scheme, is there any requirement on a prior knowledge of the channel, such as the total number of the resolvable paths  $L$ ?

**Authors:** Yes, for the proposed SSD-based scheme, we assume that the number of resolvable paths  $L$  is known. This is due to the fact that a suggested  $L$  can be usually obtained by channel measurement in advance. For example, a measurement report in New York City has shown that the average number of resolvable paths in 28 GHz propagation environment is 6.8 with a standard deviation of 2.2 [R7]. Therefore, for the specific environment,  $L$  can be directly set as the largest possible value (e.g., in New York City, we set  $L = 6.8 + 2.2 = 9$ ). In practice, the actual number  $L_a$  of resolvable paths should be a little smaller than  $L$ , but this will not significantly affect the performance of the proposed SSD-based scheme. One example is shown in the following Fig. 3, where the parameters are the same as those for Fig. 10 in the paper except that  $L \geq L_a = 3$ . We observe that even if  $L$  is a little larger than  $L_a$ , our scheme still enjoys satisfying accuracy.

Fig. 3. NMSE comparison against the assumed number of resolvable paths  $L$ .

Moreover, we would like to point out that the prior knowledge of  $L$  is not the necessary condition in the proposed SSD-based scheme. When  $L$  cannot be obtained in advance, we can borrow the idea of the classical OMP and SOMP algorithms, and run the proposed SSD-based scheme for several times [R8]. Specifically, during the  $t$ -th running ( $t = 1, 2, \dots$ ), we set  $L = t$ , and define the channel estimation update as  $\frac{1}{MN} \left\| \tilde{\mathbf{H}}^{(t)} - \tilde{\mathbf{H}}^{(t-1)} \right\|_F^2$ , where  $\tilde{\mathbf{H}}^{(t-1)}$  and  $\tilde{\mathbf{H}}^{(t)}$  are the previous estimated channel at the  $(t-1)$ -th running and the current estimated channel at the  $t$ -th running, respectively. If the update is smaller than a threshold  $\zeta$ , then we will terminate. The following Fig. 4 shows the NMSE comparison between the SSD-based scheme with known  $L$  and the one with unknown  $L$ , where we empirically set  $\zeta = 0.1$ . We observe that by utilizing this method, we can obtain the solution quite close to the one when  $L$  is known in advance.

Fig. 4. (Fig. 10 in the revised paper) NMSE comparison between the SSD-based scheme with known and unknown  $L$ .

In the revised paper, we have addressed this issue as:

Actually, a suggested  $L$  can be usually obtained in advance by channel measurement [24]. For example, a measurement report in New York City has shown that the average number of resolvable paths in 28 GHz propagation environment is 6.8 with a standard deviation of 2.2, and therefore we can set  $L$  as the largest possible value as  $L = 9$  in this case. In practice, the actual number of resolvable paths should be a little smaller than  $L$ , but this will not significantly affect the performance of the proposed SSD-based scheme. Moreover, it is worth pointing out that the prior knowledge of  $L$  is not a necessary condition in our scheme. When  $L$  cannot be obtained in advance, we can borrow the idea of the classical OMP and SOMP algorithms, and run the proposed SSD-based scheme for several times [35]. Specifically, during the  $t$ -th running ( $t = 1, 2, \dots$ ), we set  $L = t$ , and define the channel estimation update as  $\frac{1}{MN} \left\| \tilde{\mathbf{H}}^{(t)} - \tilde{\mathbf{H}}^{(t-1)} \right\|_F^2$ , where  $\tilde{\mathbf{H}}^{(t-1)}$  and  $\tilde{\mathbf{H}}^{(t)}$  are the previous estimated channel at the  $(t-1)$ -th running and the current estimated channel at the  $t$ -th running, respectively. If the update is smaller than a threshold  $\zeta$  (e.g.,  $\zeta = 0.1$ ), then we will terminate. In Section V, we will verify that by utilizing this method, the proposed SSD-based scheme can still achieve satisfying accuracy without the knowledge of  $L$ .

[R7] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335-349, May 2013.

[R8] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053-4085, Sep. 2011.

**Reviewer 2:** 6) In Section IV, the users are assumed to employ also lens antenna array. However, lens antenna array is usually bulky compared to conventional array. Is it feasible to be employed at the user side? Another possibility is for the user to employ a small conventional antenna array. Is the proposed scheme applicable to such setup as well?

**Authors:** Thanks for pointing out this important issue. The reviewer is right that in practice the users are more likely to employ the conventional antenna array, since currently the lens antenna array is



usually bulky. However, we would like to point out that the proposed SSD-based scheme can also work when the users employ the conventional antenna arrays. In this case, the power of wideband beamspace channel at one sub-carrier is focused on a small number of rows instead of a small number of low-dimensional sub-matrices. One example is shown in the following Fig. 5, where we set the number of BS antennas  $N = 64$  and the number of user antennas  $U = 16$ . This property enables us to further simplify our scheme described in Section IV (Extension to Multiple-Antenna Users). Specifically, when the conventional antenna array is employed at the user side, we do not need to estimate the column support  $\mathcal{S}_{l,m}$  of each path component any more. After we have estimated the row support  $\mathcal{T}_{l,m}$  by Eq. (72) and Eq. (74) in the paper, the support of the  $l$ -th path component can be directly obtained as  $\mathcal{D}_{l,m} = \{n + (u - 1)N \mid n \in \mathcal{T}_{l,m}, u \in \{1, 2, \dots, U\}\}$  instead of  $\mathcal{D}_{l,m} = \{n + (u - 1)N \mid n \in \mathcal{T}_{l,m}, u \in \mathcal{S}_{l,m}\}$ .

In the revised paper, we have addressed this issue as:

In the end, we would like to point out that in practice, the users are also likely to employ the conventional antenna array, since currently the lens antenna array is usually bulky. In this case, the power of wideband beamspace channel at one sub-carrier will be focused on a small number of rows instead of a small number of low-dimensional sub-matrices. This property enables us to further simplify our scheme. Specifically, when the conventional antenna array is employed at the user side, we do not need to estimate the column support  $\mathcal{S}_{l,m}$  any more. After we have estimated the row support  $\mathcal{T}_{l,m}$  by (72) and (74), the support  $\mathcal{D}_{l,m}$  of the  $l$ -th path component can be directly obtained as  $\mathcal{D}_{l,m} = \{n + (u - 1)N \mid n \in \mathcal{T}_{l,m}, u \in \{1, 2, \dots, U\}\}$ .

Fig. 5. The power distribution of wideband beamspace channel at one sub-carrier when (a) the lens antenna array is used at the user side; (b) the conventional antenna array is used at the user side.

**Authors:** Many thanks again for your valuable time and efforts to review this paper.

Sincerely,

*The Authors*

## Response to Reviewer 3's Comments

**Reviewer 3:** This paper considers a wideband mmWave MU-MIMO uplink system and proposes a channel estimation algorithm for the system. The beam squint effect is newly considered in this paper, compared to the conventional wideband mmWave systems. The scope of the paper is timely relevant. The comments of the reviewer are as follows.

**Authors:** We appreciate the reviewer's positive evaluation on the novelty of our work. To address your professional comments, we have made several major revisions in the revised paper, which are described one by one as below.

**Reviewer 3:** 1) Any analytic reason for the problem formulation (18) to design  $\Delta_n$ , a key parameter for the proposed algorithm, is not provided in this paper. Why do we need to consider (18) and choose the solution? The reviewer thinks the paragraph above (18) cannot be justified (and seems to be not technically sounded) and the authors need to address another justification or provide any technical relationship between the problem formulation and the channel estimation performance.

**Authors:** Thanks for this valuable suggestion. We agree with the reviewer that the justification of Eq. (18) in the previous paper is not technically sounded enough. In the revised paper, we have rewritten the paragraph above Eq. (18) to justify its rationality more clearly. The revised version is also listed here for you to check.

The next problem is how to design  $\Delta_n$  in the Bwin  $\Upsilon_n$ . Note that our target is to estimate  $\varphi_{l,c}$  by using different BWins to capture the power of the  $l$ -th path component, and we assume that  $\varphi_{l,c}$  belongs to the set  $\{\bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N\}$  pre-defined by the lens antenna array (the corresponding quantization error is negligible when  $N$  is large, e.g.,  $N = 256$  [8]). For the specific case where  $\varphi_{l,c} = \bar{\varphi}_n$  (i.e.,  $\mathbf{C}_n = [\tilde{\mathbf{c}}_{l,1}, \tilde{\mathbf{c}}_{l,2}, \dots, \tilde{\mathbf{c}}_{l,M}]$ ),  $\varphi_{l,c}$  can be correctly estimated only when the following condition is satisfied

$$\|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 > \max_{n' \neq n} (\|\mathbf{C}_n(\Upsilon_{n'}, :)\|_F^2). \quad (18)$$

In practice,  $\mathbf{C}_n$  may be corrupted by interference or noise. To overcome this problem,  $\Upsilon_n$  should

be designed to capture the power  $\mathbf{C}_n$  as much as possible, i.e.,

$$\Upsilon_n^* = \arg \max_{\Upsilon_n} \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2, \quad (19)$$

while  $\Upsilon_{n'}$  for  $n' \neq n$  should be designed to capture the power of  $\mathbf{C}_n$  as less as possible, i.e.,

$$\Upsilon_{n'}^* = \arg \min_{\Upsilon_{n'}} \|\mathbf{C}_n(\Upsilon_{n'}, :)\|_F^2. \quad (20)$$

Consider all the cases  $\varphi_{l,c} = \bar{\varphi}_1, \bar{\varphi}_2, \dots, \bar{\varphi}_N$  and collect the optimization problems related to  $\Upsilon_n$ , we know that  $\Upsilon_n$  should be designed to optimize the following problem

$$\max_{\Delta_n} \left( \|\mathbf{C}_n(\Upsilon_n, :)\|_F^2 - \max_{n' \neq n} \|\mathbf{C}_{n'}(\Upsilon_n, :)\|_F^2 \right), \quad (21)$$

where  $\mathbf{C}_{n'}$  is constructed with  $\varphi_{l,c} = \bar{\varphi}_{n'}$ , and we replace the optimization variable  $\Upsilon_n$  by  $\Delta_n$ , since designing  $\Upsilon_n$  is equivalent to designing  $\Delta_n$ .

Finally, it is worth pointing out that the method we proposed to estimate  $\varphi_{l,c}$  is heuristic, which may not be the optimal one. It has been verified in the paper that our method is simple and efficient. However, designing the optimal method to estimate  $\varphi_{l,c}$  is also interesting, and we will leave it for our future work. In the revised paper, we have addressed this issue as:

Note that the method described above is heuristic. In Section V, we will verify that this method is simple and efficient. Designing the optimal method to estimate  $\varphi_{l,c}$  is interesting, which will be left for our future work.

**Reviewer 3:** 2) In the proposed algorithm, the number of paths  $L$  is a priori known to the BS. The third footnote of this paper indicates that it is usually obtained in advance, but the reviewer thinks this assumption needs to be further justified since it is an important factor to run the proposed algorithm. How to obtain the number of paths in the channel measurement? Why do we separately estimate channels and identify  $L$ ? A solution to combine them would be beneficial. Or it would be better to modify the proposed algorithm to run without the parameter  $L$ .

**Authors:** Thanks for this insightful comment. For the proposed SSD-based scheme, we assume that the number of resolvable paths  $L$  is known. This is due to the fact that a suggested  $L$  can be usually obtained by channel measurement in advance. For example, a measurement report in New York City has shown that the average number of resolvable paths in 28 GHz propagation environment is 6.8 with a standard deviation of 2.2 [R1]. Therefore, for the specific environment,  $L$  can be directly set as the largest possible value (e.g., in New York City, we set  $L = 6.8 + 2.2 = 9$ ). In practice, the actual number  $L_a$  of resolvable paths should be a little smaller than  $L$ , but this will not significantly affect the performance of the proposed SSD-based scheme. One example is shown in the following Fig. 1, where the parameters are the same as those for Fig. 10 in the paper except that  $L \geq L_a = 3$ . We observe that even if  $L$  is a little larger than  $L_a$ , our scheme still enjoys satisfying accuracy.

Fig. 1. NMSE comparison against the assumed number of resolvable paths  $L$ .

Moreover, as suggested by the reviewer, we have also modified our scheme to make it work without the knowledge of  $L$ . Specifically, borrowing the idea of the classical OMP and SOMP algorithms, we can run the proposed SSD-based scheme for several times [R2]. During the  $t$ -th running ( $t = 1, 2, \dots$ ), we set  $L = t$ , and define the channel estimation update as  $\frac{1}{MN} \left\| \tilde{\mathbf{H}}^{(t)} - \tilde{\mathbf{H}}^{(t-1)} \right\|_F^2$ , where  $\tilde{\mathbf{H}}^{(t-1)}$  and  $\tilde{\mathbf{H}}^{(t)}$  are the previous estimated channel at the  $(t-1)$ -th running and the current estimated channel at the  $t$ -th running, respectively. If the update is smaller than a threshold  $\zeta$ , then we will terminate. The following Fig. 2 shows the NMSE comparison between the SSD-based scheme with known  $L$  and the one with unknown  $L$ , where we empirically set  $\zeta = 0.1$ . We observe that by utilizing this method, we can obtain the solution quite close to the one when  $L$  is known in advance.

Fig. 2. (Fig. 10 in the revised paper) NMSE comparison between the SSD-based scheme with known and unknown  $L$ .

In the revised paper, we have addressed this issue as:

Actually, a suggested number of resolvable paths  $L$  can be usually obtained in advance by channel measurement [24]. For example, a measurement report in New York City has shown that the average number of resolvable paths in 28 GHz propagation environment is 6.8 with a stan-

dard deviation of 2.2, and therefore we can set  $L$  as the largest possible value as  $L = 9$  in this case. In practice, the actual number of resolvable paths should be a little smaller than  $L$ , but this will not significantly affect the performance of the proposed SSD-based scheme. Moreover, it is worth pointing out that the prior knowledge of  $L$  is not a necessary condition in our scheme. When  $L$  cannot be obtained in advance, we can borrow the idea of the classical OMP and SOMP algorithms, and run the proposed SSD-based scheme for several times [35]. Specifically, during the  $t$ -th running ( $t = 1, 2, \dots$ ), we set  $L = t$ , and define the channel estimation update as  $\frac{1}{MN} \left\| \tilde{\mathbf{H}}^{(t)} - \tilde{\mathbf{H}}^{(t-1)} \right\|_F^2$ , where  $\tilde{\mathbf{H}}^{(t-1)}$  and  $\tilde{\mathbf{H}}^{(t)}$  are the previous estimated channel at the  $(t-1)$ -th running and the current estimated channel at the  $t$ -th running, respectively. If the update is smaller than a threshold  $\zeta$  (e.g.,  $\zeta = 0.1$ ), then we will terminate. In Section V, we will verify that by utilizing this method, the proposed SSD-based scheme can still achieve satisfying accuracy without the knowledge of  $L$ .

[R1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335-349, May 2013.

[R2] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053-4085, Sep. 2011.

**Reviewer 3:** 3) In the simulation section, the performance of the proposed method is superior to that of the other conventional methods. It is a straightforward result since the system model further considers the beam squint effect. New methods suitable for new system models outperform the conventional ones. The reviewer hopes that any performance analysis of the proposed algorithm will be provided to validate its superiority or to show any demerit. The paper does not provide any analytical result for the proposed algorithm except the complexity analysis.

**Authors:** Thanks for this valuable comment. We agree with the reviewer that the analytical results are necessary for the high-quality journals like IEEE TSP. To this end, we have added a subsection (i.e., Section III-D. Performance analysis) in the revised paper to mathematically analyze the performance of the proposed SSD-based scheme. For you to check, we also provide this subsection as follows with

some slight modifications, e.g., the numbers of equations, lemmata, and figures are different from the those in the revised paper.

Fig. 3. Transformation from  $\tilde{\mathbf{H}}$  to  $\mathbf{Z}$ .

We will prove that the proposed SSD-based scheme can correctly estimate the key parameters  $\varphi_{l,c}$  for  $l = 1, 2, \dots, L$  with a certain probability.

To do this, we first rewrite  $\tilde{\mathbf{H}}$  as  $\tilde{\mathbf{H}} = \mathbf{T}\mathbf{Z}$ . Here, we define  $\mathbf{Z}$  of size  $\sum_{n=1}^N |\Upsilon_n| \times M$  as an enlarged version of  $\tilde{\mathbf{H}}$ , which can be presented as  $\mathbf{Z} = [\mathbf{Z}_1^H, \mathbf{Z}_2^H, \dots, \mathbf{Z}_N^H]^H$  with

$$\mathbf{Z}_n = \begin{cases} \tilde{\mathbf{H}}(\Upsilon_n, :), & \text{if } n \in \{n_1, n_2, \dots, n_L\}, \\ \mathbf{0}_{|\Upsilon_n| \times M}, & \text{if } n \notin \{n_1, n_2, \dots, n_L\}, \end{cases} \quad (6)$$

and  $\tilde{\mathbf{H}}(\Upsilon_n, :) = \beta_l \mathbf{C}_{n_l}(\Upsilon_n, :)$ . Here, for the  $l$ -th path component, we assume that  $\varphi_{l,c} = \bar{\varphi}_{n_l}$  and all its power can be captured by  $\mathbf{C}_{n_l}(\Upsilon_n, :)$  (this assumption leads negligible performance loss as we have proved in the paper). Correspondingly,  $\mathbf{T}$  of size  $N \times \sum_{n=1}^N |\Upsilon_n|$  is the transformation matrix. More specifically,  $\mathbf{T}$  can be presented as  $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_N]$ , where  $\mathbf{T}_n$  is of size  $N \times |\Upsilon_n|$  and its  $i$ -th column ( $i = 1, 2, \dots, |\Upsilon_n|$ ) only has one nonzero element 1 at the location  $\Upsilon_n(i)$  with  $\Upsilon_n(i)$  presenting the  $i$ -th element selected from the set  $\Upsilon_n$ . By utilizing this transformation, we can transfer each path component in  $\tilde{\mathbf{H}}$  to a specific block in  $\mathbf{Z}$  as illustrated in Fig. 3 above, where different blocks are non-overlapped.

According to the definitions of  $\mathbf{Z}$  and  $\mathbf{T}$ , we have

$$\bar{\mathbf{Y}} = \bar{\mathbf{W}}\mathbf{T}\mathbf{Z} + \mathbf{N} = [\bar{\mathbf{W}}(:, \Upsilon_1), \bar{\mathbf{W}}(:, \Upsilon_2), \dots, \bar{\mathbf{W}}(:, \Upsilon_N)] \mathbf{Z} + \mathbf{N}. \quad (7)$$

Then, the key matrix  $\mathbf{A}_l(\Upsilon_n, :)$  used for estimating  $\varphi_{l,c}$  (i.e., step 2 of **Algorithm 1** in the paper) can be presented as

$$\mathbf{A}_l(\Upsilon_n, :) = \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{Y}} = \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N} + \sum_{i=1}^N \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}^H(:, \Upsilon_i) \mathbf{Z}_i. \quad (8)$$

Next, we define two auxiliary parameters  $\mu$  and  $\mu_B$  as

$$\mu \triangleq \max_{i \neq j} |\bar{\mathbf{W}}^H(:, i) \bar{\mathbf{W}}(:, j)|, \quad (9)$$

and

$$\mu_B \triangleq \max_{i \neq j} \frac{1}{\sqrt{|\Upsilon_i| |\Upsilon_j|}} \left\| \bar{\mathbf{W}}^H(:, \Upsilon_i) \bar{\mathbf{W}}(:, \Upsilon_j) \right\|_2, \quad (10)$$

respectively. Note that  $\mu$  is exactly the same as the coherence of the dictionary  $\bar{\mathbf{W}}$  in compress sensing theory [R3], while  $\mu_B$  can be regarded as a generalized version of the block coherence introduced in [R4].

Then, based on the discussion above, we have the following **Lemma 1**.

**Lemma 1.** *For the  $l$ -th path component, assume that  $\varphi_{l,c} = \bar{\varphi}_{n_l}$  and*

$$\frac{(1 - (|\Upsilon_{n_l}| - 1) \mu + \mu_B |\Upsilon_{n_l}|) |\beta_l|}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{C}_{n_l}(\Upsilon_{n_l}, :)\|_F \geq 2M\sigma\sqrt{\alpha} + 2\mu_B \sum_{n_i \in \mathcal{L}} |\Upsilon_{n_i}| |\beta_i| \|\mathbf{C}_{n_i}(\Upsilon_{n_i}, :)\|_F \quad (11)$$

for some constant  $\alpha$ , where  $\mathcal{L} = \{n_1, n_2, \dots, n_L\}$ . Then, with the probability exceeding

$$\prod_{n=1}^N (1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}), \quad (12)$$

the proposed SSD-based scheme can correctly estimate  $\varphi_{l,c}$ .

*Proof:* To prove **Lemma 1**, we first list two useful lemmata, which have been proved in [R5].

**Lemma 2.** *Let  $\mathbf{u}$  be a Gaussian random vector of size  $d \times 1$ . Assume that  $\mathbf{u}$  has mean  $\mathbf{0}$  and covariance  $\mathbf{I}_d$ . Then, we have*

$$\Pr \left\{ \|\mathbf{u}\|_2^2 \geq t^2 \right\} \leq 0.8dt^{d-2}e^{-t^2/2}. \quad (13)$$

*Proof:* Please refer to [R5, **Lemma 4**].

**Lemma 3.** *Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$  be  $M$  jointly Gaussian random vectors. Assume that  $\mathbb{E}(\mathbf{v}_m) = \mathbf{0}$  for  $m = 1, 2, \dots, M$ , but that the covariances of the vectors are unspecified and that the vectors are not necessarily independent. Then, we have*

$$\Pr \left\{ \|\mathbf{v}_1\|_2 \leq c_1, \|\mathbf{v}_2\|_2 \leq c_2, \dots, \|\mathbf{v}_M\|_2 \leq c_M \right\} \geq \prod_{m=1}^M \Pr \left\{ \|\mathbf{v}_m\|_2 \leq c_m \right\}. \quad (14)$$

*Proof:* Please refer to [R5, **Lemma 3**].

Based on **Lemma 2** and **Lemma 3**, we have the following **Lemma 4**.

**Lemma 4.** Assume that each column  $\mathbf{n}_m^{\text{eff}}$  of  $\mathbf{N}$  in (7) is a Gaussian random vector of size  $QN_{\text{RF}} \times 1$  with mean  $\mathbf{0}$  and covariance  $\sigma^2 \mathbf{I}_{QN_{\text{RF}}}$ . Then, we have

$$\Pr \left( \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right) \geq 1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}, \quad (15)$$

where we define

$$\tau_n^2 = \sigma^2 (1 + (|\Upsilon_n| - 1) \mu) \alpha, \quad (16)$$

and  $\alpha$  is a constant value to guarantee that  $0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2} \leq 1$  for  $n = 1, 2, \dots, N$ .

*Proof:* Note that  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}}$  is also a Gaussian random vector with mean  $\mathbf{0}$  and covariance  $\sigma^2 \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$ . Define a auxiliary vector  $\mathbf{u}$  as

$$\mathbf{u} \triangleq \frac{1}{\sigma} (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n))^{-\frac{1}{2}} \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}}. \quad (17)$$

We know that  $\mathbf{u}$  should be a Gaussian random vector of size  $|\Upsilon_n| \times 1$  with mean  $\mathbf{0}$  and covariance  $\mathbf{I}_{|\Upsilon_n|}$ . As a result, we have

$$\begin{aligned} \Pr \left\{ \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right\} &= \Pr \left\{ \sigma^2 \left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n))^{\frac{1}{2}} \mathbf{u} \right\|_2^2 \leq \tau_n^2 \right\} \\ &\stackrel{(a)}{\geq} \Pr \left\{ \sigma^2 \left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)) \right\|_2 \left\| \mathbf{u} \right\|_2^2 \leq \tau_n^2 \right\}. \end{aligned} \quad (18)$$

where (a) is due to the spectral norm property of a matrix [R6]. Note that all the diagonal elements of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  are equal to 1, while all the off-diagonal elements of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  have the amplitudes smaller than  $\mu$  according to the definition in (9). Therefore, by the Gershgorin circle theorem [R6], we know that the largest singular value of  $\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)$  should be upper-bounded by  $1 + (|\Upsilon_n| - 1) \mu$ , which means that

$$\left\| (\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_n)) \right\|_2 \leq 1 + (|\Upsilon_n| - 1) \mu. \quad (19)$$

Correspondingly, we have

$$\Pr \left\{ \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right\} \geq \Pr \left\{ \left\| \mathbf{u} \right\|_2^2 \leq \frac{\tau_n^2}{\sigma^2 (1 + (|\Upsilon_n| - 1) \mu)} \right\}. \quad (20)$$



According to the definition of  $\tau_n^2$  in (16), the right side of (20) can be rewritten as

$$\Pr \left\{ \|\mathbf{u}\|_2^2 \leq \frac{\tau_n^2}{\sigma^2 (1 + (|\Upsilon_n| - 1) \mu)} \right\} = \Pr \{ \|\mathbf{u}\|_2^2 \leq \alpha \}. \quad (21)$$

Based on **Lemma 2**, we can conclude that

$$\Pr \left( \left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \right\|_2^2 \leq \tau_n^2 \right) \geq 1 - \Pr \{ \|\mathbf{u}\|_2^2 \geq \alpha \} = 1 - 0.8 |\Upsilon_n| \alpha^{|\Upsilon_n|/2-1} e^{-\alpha/2}. \quad (22)$$

which completes the proof.

Next, we continue the proof of **Lemma 1**. Specifically, for the  $l$ -th path component,  $\varphi_{l,c}$  can be correctly estimated only if

$$\frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}}, \quad (23)$$

where  $\mathbf{A}_l(\Upsilon_n, :)$  is given by (8).

Consider the event  $B = \{ \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{n}_m^{\text{eff}} \leq \tau_n^2, n = 1, 2, \dots, N \}$ . Based on **Lemma 3** and **Lemma 4**, we know that event B will occur with the probability exceeding (12). When it occurs, the right side of (23) can be upper-bounded by

$$\begin{aligned} \max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}} &= \max_{n \notin \mathcal{L}} \frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N} + \sum_{n_i \in \mathcal{L}} \bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F}{\sqrt{|\Upsilon_n|}} \\ &\leq \max_{n \notin \mathcal{L}} \frac{\|\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N}\|_F}{\sqrt{|\Upsilon_n|}} + \max_{n \notin \mathcal{L}} \frac{\sum_{n_i \in \mathcal{L}} \|\bar{\mathbf{W}}^H(:, \Upsilon_n) \bar{\mathbf{W}}(:, \Upsilon_{n_i})\|_2 \|\mathbf{Z}_{n_i}\|_F}{\sqrt{|\Upsilon_n|}}. \end{aligned} \quad (24)$$

Since  $\mathbf{N}$  has  $M$  Gaussian random columns and event B occurs, we know that

$$\|\bar{\mathbf{W}}^H(:, \Upsilon_n) \mathbf{N}\|_F \leq M \tau_n \stackrel{(a)}{\leq} M \sigma \sqrt{|\Upsilon_n| \alpha}, \quad (25)$$

where (a) is true due to the fact that  $\mu \leq 1$ . Combining this result and the definition of  $\mu_B$  in (10), (25) can be further upper-bounded by

$$\max_{n \notin \mathcal{L}} \frac{\|\mathbf{A}_l(\Upsilon_n, :)\|_F}{\sqrt{|\Upsilon_n|}} \leq M \sigma \sqrt{\alpha} + \mu_B \sum_{n_i \in \mathcal{L}} |\Upsilon_{n_i}| \|\mathbf{Z}_{n_i}\|_F, \quad (26)$$

On the other hand, the left side of (23) can be lower-bounded by

$$\begin{aligned} \frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} &= \frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \mathbf{N} + \sum_{i \in \mathcal{L}} \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F}{\sqrt{|\Upsilon_{n_l}|}} \\ &\geq \frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_l}) \mathbf{Z}_{n_l} \right\|_F - \left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \mathbf{N} \right\|_F - \left\| \sum_{n_i \in \mathcal{L} \setminus n_l} \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_i}) \mathbf{Z}_{n_i} \right\|_F}{\sqrt{|\Upsilon_{n_l}|}}. \end{aligned} \quad (27)$$

Also according to the Gershgorin circle theorem as we have used in **Lemma 4**, the first term of (27) can be lower-bounded by

$$\frac{\left\| \bar{\mathbf{W}}^H(:, \Upsilon_{n_l}) \bar{\mathbf{W}}(:, \Upsilon_{n_l}) \mathbf{Z}_{n_l} \right\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \frac{(1 - (|\Upsilon_{n_l}| - 1)\mu)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F. \quad (28)$$

Since event B occurs, the second term of (27) is lower-bounded by  $-M\sigma\sqrt{\alpha}$ . Finally, similar to the manipulation in (26), the third term of (27) can be lower-bounded by

$$-\mu_B \left( \sum_{n_i \in \mathcal{L}} |\Upsilon_{n_i}| \|\mathbf{Z}_{n_i}\|_F - |\Upsilon_{n_l}| \|\mathbf{Z}_{n_l}\|_F \right). \quad (29)$$

Combining these results, we have

$$\frac{\|\mathbf{A}_l(\Upsilon_{n_l}, :)\|_F}{\sqrt{|\Upsilon_{n_l}|}} \geq \frac{(1 - (|\Upsilon_{n_l}| - 1)\mu)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F - M\sigma\sqrt{\alpha} - \mu_B \left( \sum_{n_i \in \mathcal{L}} |\Upsilon_{n_i}| \|\mathbf{Z}_{n_i}\|_F - |\Upsilon_{n_l}| \|\mathbf{Z}_{n_l}\|_F \right). \quad (30)$$

Considering (26) and (30) together, we can conclude that when

$$\frac{(1 - (|\Upsilon_{n_l}| - 1)\mu + \mu_B |\Upsilon_{n_l}|)}{\sqrt{|\Upsilon_{n_l}|}} \|\mathbf{Z}_{n_l}\|_F \geq 2M\sigma\sqrt{\alpha} + 2\mu_B \sum_{n_i \in \mathcal{L}} |\Upsilon_{n_i}| \|\mathbf{Z}_{n_i}\|_F, \quad (31)$$

$\varphi_{l,c}$  can be correctly estimated with probability exceeding (12). Substituting  $\|\mathbf{Z}_{n_l}\|_F = |\beta_l| \|\mathbf{C}_{n_l}(\Upsilon_{n_l}, :)\|_F$  into (31), we can finally complete this proof.

□

From **Lemma 1**, we observe that when the noise power  $\sigma^2$  is large, to guarantee the probability of correctly estimating  $\varphi_{l,c}$ , the power of the  $l$ -th path component  $|\beta_l|^2 \|\mathbf{C}_{n_l}(\Upsilon_{n_l}, :)\|_F^2$  should be large. On the other hand, when the number of instants  $Q$  for pilot transmission is large,  $\bar{\mathbf{W}}$  of size  $QN_{\text{RF}} \times N$  can be expected to have lower  $\mu$  and  $\mu_B$ . As a result, the power of the  $l$ -th path component is not necessary to be large to correctly estimate  $\varphi_{l,c}$ . These conclusions are consistent with the expectations and will be further verified by simulation results.

- [R3] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [R4] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042-3054, Jun. 2010.
- [R5] Z. Ben-Haim and Y. C. Eldar, “Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements,” *IEEE J. Sel. Topics Signal Process.*, vol. 59, no. 5, pp. 1032-1047, Sep. 2011.
- [R6] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012.

**Authors:** Many thanks again for your valuable time and efforts to review this paper.

Sincerely,

*The Authors*