Channel Estimation for Interference Mitigation in Millimeter-Wave Multi-Cell Beamspace MIMO Systems

Anzhong Hu

Abstract: **This paper deals with the inter-cell interference (ICI) mitigation in the channel estimation process of millimeter-wave multi-cell beamspace multiple-input multiple-output systems. The known channel estimation approaches cannot distinguish the ICI channels and the in-cell channels. In contrast, the in-cell channels and the ICI channels are estimated together and distinguished with the beam information in this paper. In addition, the beamspace channels are estimated with small error by utilizing the asymptotic property of the beamspace channels. As a result, the ICI can be efficiently suppressed. Performance analysis and numerical results show that the proposed approach outperforms the existing channel estimation approaches.**

Index Terms: **Beamspace, channel estimation, millimeter-wave (mm-wave), multiple-input multiple-output (MIMO).**

I. INTRODUCTION

THE demand on the spectral efficiency and the capacity in future wireless communications will be unprecedentedly large [1], [2]. In the emerging technologies, the millimeter-wave (mm-wave) multiple-input multiple-output (MIMO) technology is a potential candidate for achieving this demand [3]. In mm-wave MIMO systems, there are more available bandwidths for increasing the capacity. In addition, more antennas can be employed owing to the reduced wavelength. Consequently, mm-wave MIMO systems may be massive MIMO systems [3], [4], which can achieve much higher spectral efficiency than conventional MIMO systems [5], [6].

The channel estimation is crucial to the performance of mm-wave MIMO systems. In mm-wave multi-cell MIMO systems, the length of the pilots is constrained by the channel coherence interval, thus the pilots in different cells cannot be orthogonal. Consequently, the channel estimation is affected by the inter-cell interference (ICI), and this causes the data transmissions being affected by the ICI. In fact, this is exactly the pilot contamination phenomenon in multi-cell massive MIMO systems [6].

As the mm-wave channels are sparse [7]–[10], the existing channel estimation approaches mainly adopt the compressed sensing theory [11]–[18] or the matrix decomposition [19]. However, these approaches are proposed for point-to-point mm-wave systems or single cell systems, and are not effective for mitigating the ICI in mm-wave multi-cell MIMO systems. Other ways to exploit the sparsity of mm-wave channels are employing the beamspace principle in [20]–[24] and antenna selection. By selecting a small subset of antennas, the channel dimension is greatly reduced, and the conventional pilot based channel estimation is employed in [25]. In mm-wave beamspace MIMO systems, the basis of the spatial dimension is termed as the beam, and the base station (BS) selects the partial beams that convey most of the energy for transmission/receiving. Thus, the number of radio-frequency chains and the computational complexity of signal processing can be greatly reduced. The traditional channel estimation approach in [26] uses pilots for channel estimation. However, as mentioned previously, in multi-cell scenarios, the pilot based channel estimate is not immune to the ICI, which means the approach proposed in [26] cannot perform as well as expected. With pilot based channel estimate, the beams are selected based on the correlation of channel estimate and the beams in [27]. In order to reduce the number of pilot symbols employed, one beamspace channel estimation approach is proposed in [28]. This approach allocates non-overlap beams for the mobile stations (MSs), thus changes the multiuser transmission into several almost interference-free single user transmissions. However, in multi-cell scenarios, the cell-edge MSs are likely to convey the energy on the overlapped beams, which means the single user transmission in [28] is likely to be affected by the ICI. Hence, the channel estimation approach in [28] is still contaminated by the ICI.

In order to tackle the ICI in the channel estimation process, a beamspace channel estimation approach is proposed for mm-wave multi-cell MIMO systems. The approach firstly utilizes the beamspace idea to select the beams of all the MSs and pick out the beams of the in-cell MSs according to the position information contained in the beams. Then, the spatial frequencies of all the MSs are estimated in the beamspace. These estimates are matched with the beams selected for in-cell MSs to estimate the channels of in-cell MSs. Finally, the path losses are also estimated. The main contributions of this paper are two-fold.

1) The idea of treating the ICI channel the same as the desired channel is proposed. With this idea, the beams of the in-cell MSs and the MSs in other cells are selected together. Then, the beams of the in-cell MSs are selected. The pilots are allocated to keep the MSs with the same pilot away from each other, thus most of the ICI can be mitigated in the beam selection. Moreover, all the channels are estimated

Manuscript received December 17, 2016; approved for publication by Young-Chai Ko, Division II Editor, May 10, 2017.

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LQ16F010007, Project 61601152 and Project U1609216 supported by National Natural Science Foundation of China.

The author is with School of Communication Engineering, Hangzhou Dianzi University, email: huaz@hdu.edu.cn.

The author gratefully acknowledge the many helpful suggestions of the two anonymous reviewers and Editor Dr. Young-Chai Ko.

Digital object identifier: 10.1109/JCN.2017.000061

and matched with the selected beams to estimate the in-cell channels. Thus, there is little ICI in the channel estimates.

2) A beamspace channel estimation approach as precise as that without beamspace processing is proposed. The channel correlation is transformed into the beamspace and decomposed to obtain channel estimates. Since the error caused by the beamspace processing decreases with the increase of the number of the BS antennas, the estimation error is expected to be as small as those obtained without beamspace processing.

This paper is organized as follows. In Section II, the system model and the assumptions are given. In Section III, the proposed channel estimation approach is introduced. In Section IV, the performance analysis of the proposed approach is presented. Finally, the simulation results are given in Section V and the conclusions are drawn in Section VI.

Notations: Lower-case (upper-case) boldface symbols denote vectors (matrices); $\mathbf{I}_K$ represents the $K \times K$ identity matrix; $(\cdot)^*$, $(\cdot)^H$, and $\mathbb{E}\{\cdot\}$ denote the conjugate, the conjugate transpose, and the expectation, respectively; $[\cdot]_j$ is the jth element of a vector or the jth column of a matrix; $[\cdot]_{j,k}$ is the element in the jth row and kth column of a matrix; $\mathrm{tr}(\cdot)$ is the trace of a matrix; $\max\{\cdot\}$ is the maximum of the values in the brackets; $||\cdot||$ is the Euclidean norm of a vector; $\delta(\cdot)$ is the Dirichlet delta function; and i is the imaginary unit.

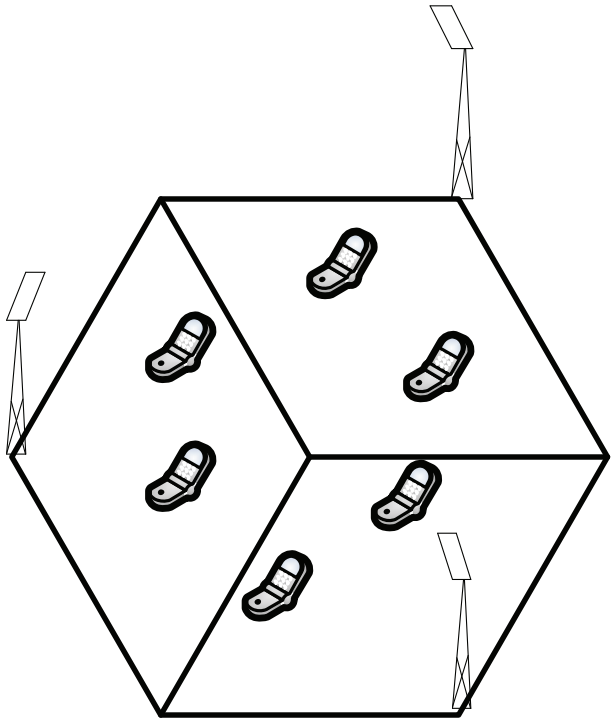

Fig. 1. An example of the considered system model. There are $L = 3$ rhombic cells and $K = 2$ MSs in each cell.

II. SYSTEM MODEL

The mm-wave multi-cell MIMO system model consists of L cells. In each cell, there is one BS simultaneously serving K MSs. The BS is equipped with one uniform rectangular array (URA) and only serves the MSs in front of the array. As can be seen, the L cells should be surrounded by the BSs. Each MS is composed of only one antenna. One typical example of the considered system model is shown in Fig. 1. In this figure, the cells are rhombic and the BSs are at the vertices.

The system operates at time division duplex mode, which means the uplink and the downlink channels are reciprocal. Moreover, the channel is assumed to be constant in one coherence interval. Thus, the transmission process is divided into three subsequent stages, namely the uplink channel estimation, the uplink data transmission, and the downlink data transmission.

According to the system model, the MSs transmit the pilots and the data symbols to the BSs simultaneously. Then, the received pilots and data symbols at the jth BS can be expressed as

$$\begin{aligned}\mathbf{Y}_j^{\mathrm{p}} &= \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{s=0}^{S}\frac{\alpha_{j,l_k,s}}{\sqrt{N}}\mathbf{a}(\theta_{j,l_k,s},\phi_{j,l_k,s})\mathbf{s}_{l,k}^{\mathrm{p}} \\ &+ \mathbf{N}_j^{\mathrm{p}} \in \mathbb{C}^{N\times N_{\mathrm{p}}} \end{aligned} \tag{1}$$

and

$$\begin{aligned}\mathbf{Y}_j^{\mathrm{du}} &= \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{s=0}^{S}\frac{\alpha_{j,l_k,s}}{\sqrt{N}}\mathbf{a}(\theta_{j,l_k,s},\phi_{j,l_k,s})\mathbf{s}_{l,k}^{\mathrm{du}} \\ &+ \mathbf{N}_j^{\mathrm{du}} \in \mathbb{C}^{N\times N_{\mathrm{du}}}, \end{aligned} \tag{2}$$

respectively. In these equations, N is the number of antennas at the BS; $\alpha_{j,l_k,s}$ represents the uplink path loss of the sth path from the kth MS in the lth cell to the BS in the jth cell; $\mathbf{a}(\theta_{j,l_k,s},\phi_{j,l_k,s}) \in \mathbb{C}^{N\times 1}$ is the array manifold, in which $\theta_{j,l_k,s}$ and $\phi_{j,l_k,s}$ are the azimuth direction-of-arrival (DOA) and the elevation DOA of the sth path from the kth MS in the lth cell to the BS in the jth cell; $\mathbf{s}_{l,k}^{\mathrm{p}} \in \mathbb{C}^{1\times N_{\mathrm{p}}}$ is the vector of the transmitted pilot symbols from the kth MS in the lth cell, and satisfies the orthogonality

$$\mathbf{s}_{l',k'}^{\mathrm{p}}(\mathbf{s}_{l,k}^{\mathrm{p}})^H = \sigma_{l,k}\sigma_{l',k'}N_{\mathrm{p}}\delta(k-k'), \tag{3}$$

where $\sigma_{l,k}$ is the power coefficient of the kth MS in the jth cell; $\mathbf{s}_{l,k}^{\mathrm{du}} \in \mathbb{C}^{1\times N_{\mathrm{du}}}$ is composed of the transmitted data symbols from the kth MS in the lth cell, and can be expressed as $\mathbf{s}_{l,k}^{\mathrm{du}} = \sigma_{l,k}\tilde{\mathbf{s}}_{l,k}^{\mathrm{du}}$, where $\tilde{\mathbf{s}}_{l,k}^{\mathrm{du}} \in \mathbb{C}^{1\times N_{\mathrm{du}}}$ is composed of independent and identically distributed (i.i.d.) complex random variables with zero mean and variance one; $\mathbf{N}_j^{\mathrm{p}} \in \mathbb{C}^{N\times N_{\mathrm{p}}}$ and $\mathbf{N}_j^{\mathrm{du}} \in \mathbb{C}^{N\times N_{\mathrm{du}}}$ are received noise matrices, the elements of which are i.i.d. complex Gaussian random variables with zero mean and variance one. It is assumed that the jth BS has the knowledge of the order of magnitude of $\max\{|\alpha_{j,j_k,\tilde{s}_{j,j_k}}|, s = 0,1,\cdots,S\}$, and power control is employed such that $\sigma_{j,k}\max\{|\alpha_{j,j_k,\tilde{s}_{j,j_k}}|, s = 0,1,\cdots,S\}$, $j = 1,2,\cdots,L$, $k = 1,2,\cdots,K$ are on the same order of magnitude. Thus, the differences between the large scale fading coefficients can be compensated with power control.

Similarly, the received data symbols of the kth MS in the jth

cell can be expressed as

$$\begin{aligned}\mathbf{y}_{j,k}^{\mathrm{dd}} &= \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{s=0}^{S}\frac{\alpha_{l,j_k,s}^{*}}{\sqrt{N}}\mathbf{a}^{H}(\theta_{l,j_k,s},\phi_{l,j_k,s})\mathbf{F}_{l}^{\mathrm{dd}}\mathbf{S}_{l}^{\mathrm{dd}} \\ &+ \mathbf{n}_{j,k}^{\mathrm{dd}} \in \mathbb{C}^{1\times N_{\mathrm{dd}}}, \end{aligned} \tag{4}$$

where $\mathbf{S}_{l}^{\mathrm{dd}} \in \mathbb{C}^{K\times N_{\mathrm{dd}}}$ is composed of the transmitted data symbols from the lth BS to the MSs in the lth cell, and can be expressed as $\mathbf{S}_{l}^{\mathrm{dd}} = \mathbf{Q}_{l}\tilde{\mathbf{S}}_{l}^{\mathrm{dd}}$, where $\mathbf{Q}_{l} \in \mathbb{R}^{K\times K}$ is a diagonal matrix of power coefficients; $\tilde{\mathbf{S}}_{l}^{\mathrm{dd}} \in \mathbb{C}^{K\times N_{\mathrm{dd}}}$ is composed of i.i.d. complex random variables with zero mean and variance one; $\mathbf{F}_{l}^{\mathrm{dd}} \in \mathbb{C}^{N\times K}$ is the precoding matrix for the MSs in the lth cell; $\mathbf{n}_{j,k}^{\mathrm{dd}} \in \mathbb{C}^{1\times N_{\mathrm{dd}}}$ is the received noise vector, the elements of which are i.i.d. complex Gaussian random variables with zero mean and variance one. It is assumed that power allocation is employed such that $[\mathbf{Q}_{j}]_{k,k}\max\{|\alpha_{j,j_k,\tilde{s}_{j,j_k}}|, s=0,1,\cdots,S\}$, $k=1,2,\cdots,K$ are on the same order of magnitude. Thus, the differences between the large scale fading coefficients can be compensated with power allocation.

Before explaining the system model in depth, the expression of the channel should be presented first. According to the expressions of the uplink transmissions in (1) and (2), it is not hard to find that the channel from the kth MS in the lth cell to the BS in the jth cell can be defined as [22]

$$\mathbf{h}_{j,l_k} \triangleq \sum_{s=0}^{S}\frac{\alpha_{j,l_k,s}}{\sqrt{N}}\mathbf{a}(\theta_{j,l_k,s},\phi_{j,l_k,s}) \in \mathbb{C}^{N\times 1}, \tag{5}$$

where $s=0$ corresponds to the line-of-sight (LOS) path and $s=1,2,\cdots,S$ correspond to the multipaths. What's more, the path loss in (5) is defined as [29]

$$\alpha_{j,l_k,s} = \eta_{j,l_k,s}\frac{D(\theta_{j,l_k,s},\phi_{j,l_k,s})\lambda^{2}e^{i\varphi_{j,l_k,s}}}{16\pi^{2}d_{j,l_k,s}^{2}}, \tag{6}$$

where $D(\theta_{j,l_k,s},\phi_{j,l_k,s})$ is the array directivity given in [30], λ is the wavelength, $\varphi_{j,l_k,s}$ is a random phase, $d_{j,l_k,s}$ is the length of the sth path between the kth MS in the lth cell to the BS in the jth cell. Moreover, $\eta_{j,l_k,s}$ is the path loss coefficient. In [31], mm-wave channels are measured, and it is shown that the measurements vary with the scenarios. Here, for the LOS path, $\eta_{j,l_k,0}=0$ and $\eta_{j,l_k,0}=1$ represent a complete blockage and no blockage, respectively. For multipaths, they are usually 15 dB to 20 dB weaker than the LOS path [32], which means that $20\log_{10}\eta_{j,l_k,s}, s=1,2,\cdots,S$ are in the range $[-15,-20]$. In [33], the scattering attenuation is taken as 20 dB. Besides, for URAs, the array manifold is given as

$$\begin{aligned}[\mathbf{a}(\theta_{j,l_k,s},\phi_{j,l_k,s})]_{n} =& e^{i\pi(n_{\mathrm{x}}-0.5(N_{\mathrm{x}}-1))\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s}} \\ &\times e^{i\pi(n_{\mathrm{y}}-0.5(N_{\mathrm{y}}-1))\sin\phi_{j,l_k,s}} \qquad (7) \\ \triangleq& [\mathbf{b}(\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s},\sin\phi_{j,l_k,s})]_{n}, \qquad (8)\end{aligned}$$

where N_{x} and N_{y} are the numbers of antennas in the horizontal and vertical directions of the URA, $n = n_{\mathrm{y}}N_{\mathrm{x}} + n_{\mathrm{x}} + 1$, $n_{\mathrm{x}} = 0,1,\cdots,N_{\mathrm{x}}-1, n_{\mathrm{y}} = 0,1,\cdots,N_{\mathrm{y}}-1$; $\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s}$ and $\sin\phi_{j,l_k,s}$ are the spatial frequencies;

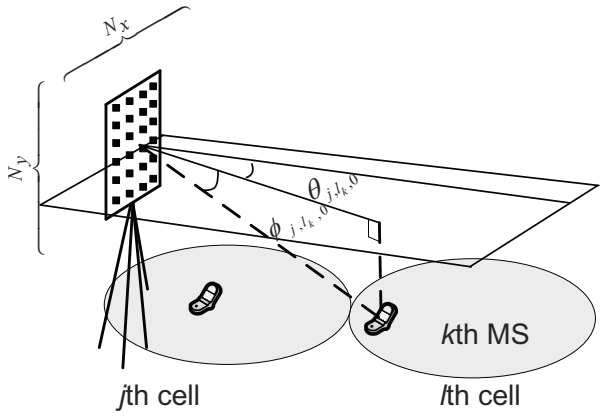

Fig. 2. An illustration of the geometrical relation between the array and the MS.

$\mathbf{b}(\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s},\sin\phi_{j,l_k,s}) \in \mathbb{C}^{N\times 1}$ is another representation of the array manifold, the difference is that the variables in $\mathbf{a}$ are the DOAs and the variables in $\mathbf{b}$ are the spatial frequencies. Obviously, the total number of the antennas in the URA is $N = N_{\mathrm{x}}N_{\mathrm{y}}$. Moreover, the geometrical relation between the array and the MS is demonstrated in Fig. 2. In this figure, the numbers of antennas in the two directions, N_{x} and N_{y}, are depicted. The URA plane and another plane connected with it are orthogonal. Correspondingly, the DOAs of the LOS path are illustrated in the figure. Note that the figure is an example for the case that the URA is vertical to the ground, and the case for tilted URA, i.e., the URA is not vertical to the ground, is more common. In this paper, it is assumed that the coarse spatial frequencies of the strongest path of each MS are known to the BS. This can be realized with other information, such as the channel estimates in the previous channel coherence interval.

After the explanation of the channel model, the sketch of the beamspace principle will be shown in the following. According to (8), it can be seen that

$$\begin{aligned}&\frac{1}{N}\mathbf{b}^{H}(f_m,g_m)\quad \mathbf{b}(f_{m'},g_{m'}) = \delta(m-m'),\forall m\neq m', \\ &f_m = -1+\frac{2m_{\mathrm{x}}}{N_{\mathrm{x}}}, g_m = -1+\frac{2m_{\mathrm{y}}}{N_{\mathrm{y}}}, \\ &m = (m_{\mathrm{y}}-1)N_{\mathrm{x}}+m_{\mathrm{x}}, m_{\mathrm{x}} = 1,2,\cdots,N_{\mathrm{x}}, \\ &m_{\mathrm{y}} = 1,2,\cdots,N_{\mathrm{y}}. \end{aligned} \tag{9}$$

In other words, the vectors listed in (9), i.e., $1/\sqrt{N}\mathbf{b}(f_m,g_m)$, $m=1,2,\cdots,N$, are orthonormal. These vectors are named the beams [22], and are formed into a unitary matrix $\mathbf{U} \in \mathbb{C}^{N\times N}$ as

$$[\mathbf{U}]_{m} = \frac{1}{\sqrt{N}}\mathbf{b}(f_m,g_m), \forall m = 0,1,\cdots,N-1. \tag{10}$$

Obviously, we have the property

$$\mathbf{U}^{H}\mathbf{U} = \mathbf{I}_{N}. \tag{11}$$

In beamspace processing proposed in [22], parts of the N beams are selected for transmission, and the number of beams

selected for each MS is N_{MS}. Correspondingly, the number of beams selected for each cell is $N_{\mathrm{Cell}} = N_{\mathrm{MS}}K$. Then, the selected N_{Cell} beams are formed into the beamforming matrix for each cell, i.e., N_{Cell} columns of $\mathbf{U}$ form the beamforming matrix. As explained in [22], set $N_{\mathrm{MS}} = 1$ is allowable. With $N_{\mathrm{MS}} > 1$, the array gain will be increased and more interference will be mitigated. Thus, N_{Cell} beams rather than N_{MS} are selected for each cell, where $N_{\mathrm{Cell}} > N_{\mathrm{MS}}$.

Here we denote the beamforming matrix for the jth cell as $\tilde{\mathbf{U}}_j \in \mathbb{C}^{N\times N_{\mathrm{Cell}}}$. Based on (11), it can be seen that

$$\tilde{\mathbf{U}}_j^H \tilde{\mathbf{U}}_j = \mathbf{I}_{N_{\mathrm{Cell}}}, \forall j = 1, 2, \cdots, L.$$

After that, beamforming is applied on the received symbol matrix, $\mathbf{Y}_j^{\mathrm{du}}$ in (2), which results into

$$\tilde{\mathbf{Y}}_j^{\mathrm{du}} = \tilde{\mathbf{U}}_j^H \mathbf{Y}_j^{\mathrm{du}} \in \mathbb{C}^{N_{\mathrm{Cell}}\times N_{\mathrm{du}}}. \tag{12}$$

Comparing $\tilde{\mathbf{Y}}_j^{\mathrm{du}}$ and $\mathbf{Y}_j^{\mathrm{du}}$, it can be seen that the dimension of the former is much lower than that of the latter, which is because the the number of selected beams for each cell N_{Cell} is much smaller than the number of beams N. Then, the processed symbol matrix $\tilde{\mathbf{Y}}_j^{\mathrm{du}}$ can be further processed for data detection, such as those shown in [22], [28]. As long as the selected beams convey most of the energy, the beamspace processing will merely degrade system performance and will greatly lower the computational complexity. Substituting the expression of $\mathbf{Y}_j^{\mathrm{du}}$ in (2) to (12) and replacing $\mathbf{a}(\theta_{j,l_k,s}, \phi_{j,l_k,s})$ with $\mathbf{b}(\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s}, \sin\phi_{j,l_k,s})$ yields

$$\begin{aligned}\tilde{\mathbf{Y}}_j^{\mathrm{du}} = &\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{s=0}^{S}\frac{\alpha_{j,l_k,s}}{\sqrt{N}}\tilde{\mathbf{U}}_j^H \\ &\times \mathbf{b}(\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s}, \sin\phi_{j,l_k,s})\mathbf{s}_{l,k}^{\mathrm{d}} + \tilde{\mathbf{U}}_j^H\mathbf{N}_j^{\mathrm{du}}.\end{aligned} \tag{13}$$

Similarly, the received data symbols of the kth MS in the jth cell can be re-written as

$$\begin{aligned}\mathbf{y}_{j,k}^{\mathrm{dd}} = &\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{s=0}^{S}\frac{\alpha_{l,j_k,s}^*}{\sqrt{N}}\mathbf{b}^H(\cos\phi_{l,j_k,s}\sin\theta_{l,j_k,s}, \sin\phi_{l,j_k,s}) \\ &\times \tilde{\mathbf{U}}_l\tilde{\mathbf{F}}_l^{\mathrm{dd}}\mathbf{S}_l^{\mathrm{dd}} + \mathbf{n}_{j,k}^{\mathrm{dd}} \in \mathbb{C}^{1\times N_{\mathrm{dd}}},\end{aligned} \tag{14}$$

where $\tilde{\mathbf{F}}_l^{\mathrm{dd}} = \tilde{\mathbf{U}}_l^H(\tilde{\mathbf{U}}_l\tilde{\mathbf{U}}_l^H)^{-1}\mathbf{F}_l^{\mathrm{dd}} \in \mathbb{C}^{N_{\mathrm{Cell}}\times K}$.

According to the measurement results in [31] and the adopted values in [32], [33], we know that the LOS path loss is usually 15 dB to 20 dB weaker than the multipath loss, and the weakest path loss of the multipath is usually 15 dB weaker than the average path loss of the multipaths. Thus, the path loss of the strongest path is usually much weaker than that of other paths. Since the utilization of all the paths results into much stronger path loss than the utilization of the strongest path, a beamforming strategy that utilizes the strongest path rather than all the paths is beneficial [31]. It is known that beamforming can allocate signal power over different paths. Here is a simple example. By allocating power p_1 over the path with path loss of β_1 and power p_2 over the path with path loss of β_2, the received signal power will be $\beta_1^2 p_1 + \beta_2^2 p_2$. By allocating power $p_1 + p_2$ over the path with path loss of β_1, the received signal power will be $\beta_1^2(p_1 + p_2)$. As long as $\beta_1 > \beta_2$, the received signal power in the second case will be higher.

In this paper, we only utilize this strongest path. Consequently, we only need to estimate the beamspace channel of the strongest path, i.e.,

$$\begin{aligned}&\alpha_{j,j_k,\tilde{s}_{j,j_k}}/\sqrt{N}\tilde{\mathbf{U}}_j^H\mathbf{b}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}}, \\ &\sin\phi_{j,j_k,\tilde{s}_{j,j_k}}), \forall j = 1, 2, \cdots, L, \forall k = 1, 2, \cdots, K,\end{aligned} \tag{15}$$

where $\tilde{s}_{j,l_k} = \arg\max_s\{|\alpha_{j,l_k,s}|, s = 0, 1, \cdots, S\}$ is used to denote the strongest path from the kth MS in the lth cell to the jth BS.

Till now, the system model and the beamspace principle have been introduced. The detection and precoding performance depends on the estimate of the beamspace channel vector in (15). When the beams have been selected, i.e., $\tilde{\mathbf{U}}_j$ is known, the estimation of the beamspace channel vector is equivalent to the estimation of the path loss, $\alpha_{j,j_k,\tilde{s}_{j,j_k}}$, and the spatial frequencies, $\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}, \sin\phi_{j,l_k,\tilde{s}_{j,l_k}}, \forall j = 1, 2, \cdots, L, \forall k = 1, 2, \cdots, K$. Therefore, the aim of beamspace channel estimation is to select the beams and estimate the path loss and the spatial frequencies with the knowledge of the received pilots in (1), the received data symbols in (2), and the vector of pilot symbols $s_{j,k}^{\mathrm{p}}/\sigma_{j,k}$. In the following section, the proposed beamspace channel estimation will be explained in detail.

III. BEAMSPACE CHANNEL ESTIMATION

The existing channel estimation approaches suffer ICI in multi-cell mm-wave beamspace MIMO systems. The approach proposed in [26] relies on orthogonal pilots. However, the channel coherence interval limits the length of the pilots, thus the pilots in different cells are not orthogonal. The approach in [28] transforms the multiuser channel into almost interference-free parallel channels for the MSs, but cell-edge users in adjacent cells may have the highly correlated channels which cause severe ICI to the channel estimation. Thus, it is necessary to develop another beamspace channel estimation approach to mitigate the ICI.

In multi-cell MIMO systems, the mitigation of the ICI in the channel estimation process is usually based on the idea of mitigating the ICI in the received signals. In contrast, here we take the ICI the same as the desired signal, thus estimate the in-cell channels and the ICI channels together. In this way, it is expected that more ICI can be mitigated in the channel estimation process. The proposed approach can be divided into three steps, which will be discussed in the following three subsections.

A. First Step: Beam Selection

In the first step, the beams for the in-cell MSs and the MSs in other cells will be selected. By taking the MSs in other cells the same as the in-cell MSs, the beams for KL MSs in all the cells can be selected. First, the received pilot matrix $\mathbf{Y}_j^{\mathrm{p}}$ in (1)

is processed into

$$\begin{aligned}\mathbf{c}_k &= \frac{1}{N_{\rm p}}\mathbf{U}^H\mathbf{Y}_j^{\rm p}(\mathbf{s}_{j,k}^{\rm p})^H \\ &= \sum_{l=1}^{L}\sum_{s=0}^{S}\frac{\alpha_{j,l_k,s}}{\sqrt{N}}\sigma_{l,k}\sigma_{j,k}\mathbf{U}^H\mathbf{b}(\cos\phi_{j,l_k,s}\sin\theta_{j,l_k,s}, \\ &\quad \sin\phi_{j,l_k,s}) + \frac{1}{N_{\rm p}}\mathbf{U}^H\mathbf{N}_j^{\rm p}(\mathbf{s}_{j,k}^{\rm p})^H \\ &\approx \sum_{l=1}^{L}\frac{\alpha_{j,l_k,\tilde{s}_{j,l_k}}}{\sqrt{N}}\sigma_{l,k}\sigma_{j,k}\mathbf{U}^H\mathbf{b}(\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}, \\ &\quad \sin\phi_{j,l_k,\tilde{s}_{j,l_k}}) + \frac{1}{N_{\rm p}}\mathbf{U}^H\mathbf{N}_j^{\rm p}(\mathbf{s}_{j,k}^{\rm p})^H \in \mathbb{C}^{N\times 1}, \qquad (16)\end{aligned}$$

where (3) and (8) are employed in deriving the second equation, and the approximation is based on the fact that the strongest path is usually much stronger than other paths.

By putting $|[\mathbf{c}_k]_n|, n = 1, 2, \cdots, N$ in the two-dimensional $n_{\rm x} - n_{\rm y}$ space with $n = (n_{\rm y} - 1)N_{\rm x} + n_{\rm x}$, the L largest non-neighboring elements of $|[\mathbf{c}_k]_n|, n = 1, 2, \cdots, N$ are selected, and the corresponding L beams are also selected for the L MSs. When the L MSs in the L cells get close to each other in the horizontal space, the L largest non-neighboring elements of $|[\mathbf{c}_k]_n|, n = 1, 2, \cdots, N$ also get close in the $n_{\rm x} - n_{\rm y}$ space. This will cause ambiguity in the selection the L elements. Therefore, the L MSs in the L cells should be far away from each other. Here, a pilot allocation scheme is also proposed. With the knowledge of the coarse spatial frequencies of the MSs, which is assumed in the previous section, the distances between the kth MS in the jth cell and any $L-1$ MSs in other $L-1$ cells (one MS in one cell) in the $n_{\rm x} - n_{\rm y}$ space are calculated. It is known that the spatial frequencies f_m, g_m and the positions $m_{\rm x}, m_{\rm y}$ are linearly related as the second equation in (9). If the coarse spatial frequencies of two MSs are q_1, r_1 and q_2, r_2, respectively. Then, the positions (coordinates) of the two MSs in the $n_{\rm x} - n_{\rm y}$ space are $(0.5(q_1+1)N_{\rm x}, 0.5(r_1+1)N_{\rm y})$ and $(0.5(q_2+1)N_{\rm x}, 0.5(r_2+1)N_{\rm y})$, respectively. Then, the distance between the two MSs is

$$0.5\sqrt{(q_1-q_2)^2N_{\rm x}^2 + (r_1-r_2)^2N_{\rm y}^2}.$$

As can be seen, there are K^{L-1} cases of the combination of the MSs. First, the cases with distances shorter than a threshold distance, e.g., 6, are discarded. Note that the requirement that the distance should be longer than a threshold distance is to avoid the ambiguity in the beam selection mentioned above. For example, for one MS with spatial frequencies q_1, r_1, the correlation of the array manifold and one beam in (9) is

$$\begin{aligned}&\frac{1}{N}|\mathbf{b}^H(q_1, r_1)\mathbf{b}(f_{m'}, g_{m'})| \\ &= \frac{1}{N}\sum_{n_{\rm x}=1}^{N_{\rm x}} e^{i\pi n_{\rm x}(f_{m'}-q_1)}\sum_{n_{\rm y}=1}^{N_{\rm y}} e^{i\pi n_{\rm y}(g_{m'}-r_1)} \\ &= \frac{1}{N}\left|\frac{\sin(0.5\pi(f_{m'}-q_1)N_{\rm x})}{\sin(0.5\pi(f_{m'}-q_1))}\right|\left|\frac{\sin(0.5\pi(g_{m'}-r_1)N_{\rm y})}{\sin(0.5\pi(g_{m'}-r_1))}\right|.\end{aligned}$$

Taking the correlation as a function of q_1 or r_1, it is easy to find that this function has a curve similar to the sinc function. For $|f_{m'} - q_1| > 12/N_{\rm x}$ and $|f_{m'} - q_1| < 1$, we have

$$\begin{aligned}&\frac{1}{N}|\mathbf{b}^H(q_1, r_1)\mathbf{b}(f_{m'}, g_{m'})| \\ &< \max_{12/N_{\rm x}<|f_{m'}-q_1|<14/N_{\rm x}}\frac{1}{N}\left|\frac{\sin(0.5\pi(f_{m'}-q_1)N_{\rm x})}{\sin(0.5\pi(f_{m'}-q_1))}\right| \\ &\times\left|\frac{\sin(0.5\pi(g_{m'}-r_1)N_{\rm y})}{\sin(0.5\pi(g_{m'}-r_1))}\right| \\ &\to \frac{1}{N}\frac{2N_{\rm x}}{13\pi}\left|\frac{\sin(0.5\pi(g_{m'}-r_1)N_{\rm y})}{\sin(0.5\pi(g_{m'}-r_1))}\right|, \text{ as } N_{\rm x}\to\infty.\end{aligned}$$

For $|f_{m'} - q_1| = 0$, we have

$$\begin{aligned}&\frac{1}{N}|\mathbf{b}^H(q_1, r_1)\mathbf{b}(f_{m'}, g_{m'})| \\ &= \frac{1}{N}N_{\rm x}\left|\frac{\sin(0.5\pi(g_{m'}-r_1)N_{\rm y})}{\sin(0.5\pi(g_{m'}-r_1))}\right|.\end{aligned}$$

It can be seen that the correlation function decreases by $2/(13\pi) \approx 0.04$ percent from $|f_{m'} - q_1| = 0$ to $|f_{m'} - q_1| > 6/N_{\rm x}$ in the region $|f_{m'} - q_1| < 1$. Correspondingly, the distance between q_1, r_1 and $f_{m'}, g_{m'}$ changes from 0 to longer than 6. Therefore, for two MSs with distance longer than 6, the mutual impact on the correlation value is negligible, which means the ambiguity in the beam selection can be avoided. Meanwhile, if the threshold is set too large, too many cases of the the combination of the MSs are discarded, which may make the pilot allocation fail.

Then, the case with the maximal total distance, i.e., the sum of the distances of any two MSs, in the remaining cases is chosen. The MSs in the chosen case are allocated with the same pilot. Additionally, the MSs that have been allocated with pilot will not be chosen in the following allocation process, thus reduces the number of cases for calculating the distances.

According to (8) and (9), each selected beam corresponds to a specific azimuth DOA and a specific elevation DOA, thus corresponds to a specific position in the horizontal space. Thus, the positions corresponding to the L selected beams can be calculated, and the beam with position most close to the jth BS is the beam for the kth MS in the jth cell. Correspondingly, the indices of this selected beam for the kth MS in the jth cell are denoted as $p_{j,k}, q_{j,k}$, i.e., $n_{\rm x} = p_{j,k}$, $n_{\rm y} = q_{j,k}$. Because the strongest path of each MS may be one multipath rather than one LOS path, the position calculated may be the position of the scatterer rather than the position of the MS, which causes error in the beam selection. As the same pilot is allocated to MSs far apart, and the scatter that may establish a link is usually close to the transmitter [31], i.e., the MS, the scatter or MS that is most close to the BS usually corresponds to the link between the MS inside the cell to that BS. Thus, the probability of beam selection error is low.

For each of the L selected beams, the neighboring $N_{\rm MS} - 1$ beams with the largest $|[\mathbf{c}_k]_n|$ are also selected. Thus, there are $N_{\rm MS}$ beams for the kth MS in the jth cell, and $N_{\rm MS}L$ beams for the L MSs with the same pilot in the L cells.

The above process is executed for $k = 1, 2, \cdots, K$ consecutively and the selected beams will not be selected in the subsequent selection process. Then, all the $N_{\rm Cell}L$, i.e., $KLN_{\rm MS}$

Algorithm 1 Proposed beam selection algorithm

for each MS in the jth cell **do**
  select one MS per cell from MSs in other cells
  that have not been assigned pilot;
  denote the combination of the selected MSs
  and the MS in the jth cell as one case;
  for each case **do**
    calculate the distances between the MSs in
    the $n_{\mathrm{x}} - n_{\mathrm{y}}$ space with coarse spatial frequencies;
    discard the case with any distance less than a threshold;
    calculate the total distance of the remaining cases;
  end for
  choose the case with maximal total distance;
  assign the same pilot to the MSs in the chosen case;
end for
for each MS in the jth cell **do**
  process the received pilot matrix as (16);
  put $|[\mathbf{c}_k]_n|, n = 1, 2, \cdots, N$ in the $n_{\mathrm{x}} - n_{\mathrm{y}}$ space;
  select the L largest non-neighboring elements of
  $|[\mathbf{c}_k]_n|, n = 1, 2, \cdots, N$ in the $n_{\mathrm{x}} - n_{\mathrm{y}}$ space;
  select the corresponding L beams for the L MSs;
  select the beam with position most close to the jth BS
  as the beam for the kth MS in the jth cell;
  select $N_{\mathrm{MS}} - 1$ beams that are neighboring to each of
  the L selected beams with the largest $|[\mathbf{c}_k]_n|$;
end for

beams of the KL MSs can be selected, and all the N_{Cell} beams of the K MSs in the jth cell can be selected.

After the beam selection process, the beamforming matrix in the BS of the jth cell for all the KL MSs is constructed by taking the $N_{\mathrm{Cell}}L$ columns of $\mathbf{U}$, i.e., the $N_{\mathrm{Cell}}L$ beams, and is denoted as $\bar{\mathbf{U}}_j \in \mathbb{C}^{N \times N_{\mathrm{Cell}}L}$. Obviously, we have

$$\bar{\mathbf{U}}_j^H \bar{\mathbf{U}}_j = \mathbf{I}_{N_{\mathrm{Cell}}L}, \forall j = 1, 2, \cdots, L. \tag{17}$$

Meanwhile, the beams for the K MSs in the jth cell are used to construct the beamforming matrix in the BS of the jth cell for the K MSs in that cell, i.e., the N_{Cell} columns of $\mathbf{U}$ form the columns of $\tilde{\mathbf{U}}_j$ in (13).

For clarity, the beam selection algorithm is briefly presented in pseudo-code as follows.

B. Second Step: Spatial Frequency Estimation

In the second step, the spatial frequencies of the MSs will be estimated. As shown at the end of Section II, the spatial frequencies are important parameters of the channel vectors. The beams selected in the first step can be used to reduce the dimension of the received signal. Then, the low-dimension signal and the selected beams can be utilized to estimate the spatial frequencies with low computational complexity and high precision.

The received signal in (2) can be re-expressed as

$$\begin{aligned} \mathbf{Y}_j^{\mathrm{du}} &\approx \sum_{l=1}^{L}\sum_{k=1}^{K} \frac{\alpha_{j,l_k,\tilde{s}_{j,l_k}}}{\sqrt{N}} \mathbf{a}(\theta_{j,l_k,\tilde{s}_{j,l_k}}, \phi_{j,l_k,\tilde{s}_{j,l_k}}) \mathbf{s}_{l,k}^{\mathrm{du}} \\ &= \frac{1}{\sqrt{N}} \mathbf{A}_j \mathbf{D}_j \mathbf{S} + \mathbf{N}_j^{\mathrm{du}}, \end{aligned} \tag{18}$$

where the first approximation is based on the fact that the strongest path is usually much stronger than other pahts, $\mathbf{A}_j \in \mathbb{C}^{N \times KL}$ is constructed by taking its columns as

$$\begin{aligned} [\mathbf{A}_j]_{(l-1)\times K + k} = \mathbf{b}(&\cos\phi_{j,l_k,\tilde{s}_{j,l_k}} \sin\theta_{j,l_k,\tilde{s}_{j,l_k}}, \\ &\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}), l = 1, 2, \cdots, L, k = 1, 2, \cdots, K. \end{aligned} \tag{19}$$

$\mathbf{D}_j \in \mathbb{C}^{KL \times KL}$ is a diagonal matrix with diagonal elements being

$$\begin{aligned} [\mathbf{D}_j]_{(l-1)\times K + k, (l-1)\times K + k} = \alpha_{j,l_k,\tilde{s}_{j,l_k}}, \\ l = 1, 2, \cdots, L, k = 1, 2, \cdots, K. \end{aligned}$$

$\mathbf{S} \in \mathbb{C}^{KL \times N_{\mathrm{du}}}$ is constructed by taking its corresponding rows as $\mathbf{s}_{l,k}^{\mathrm{du}}, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$.

By performing beamforming on the received signal $\mathbf{Y}_j^{\mathrm{du}}$ with $\bar{\mathbf{U}}_j$, the received signal is transformed into

$$\begin{aligned} \bar{\mathbf{Y}}_j^{\mathrm{du}} &= \bar{\mathbf{U}}_j^H \mathbf{Y}_j^{\mathrm{du}} \\ &\approx \frac{1}{\sqrt{N}} \bar{\mathbf{U}}_j^H \mathbf{A}_j \mathbf{D}_j \mathbf{S} + \bar{\mathbf{U}}_j^H \mathbf{N}_j^{\mathrm{du}} \in \mathbb{C}^{N_{\mathrm{Cell}}L \times N_{\mathrm{du}}}. \end{aligned}$$

The covariance matrix of each column of $\bar{\mathbf{Y}}_j^{\mathrm{du}}$ is denoted as

$$\begin{aligned} \mathbf{R}_j &= \frac{1}{N_{\mathrm{du}}} \mathbb{E}\{\bar{\mathbf{Y}}_j^{\mathrm{du}} (\bar{\mathbf{Y}}_j^{\mathrm{du}})^H\} \\ &\approx \frac{\sigma_{\mathrm{s}}^2}{N} \bar{\mathbf{U}}_j^H \mathbf{A}_j \mathbf{D}_j \mathbf{\Lambda}_j \mathbf{D}_j^H \mathbf{A}_j^H \bar{\mathbf{U}}_j + \mathbf{I}_{N_{\mathrm{Cell}}L} \\ &\in \mathbb{C}^{N_{\mathrm{Cell}}L \times N_{\mathrm{Cell}}L}, \end{aligned} \tag{20}$$

where $\mathbf{\Lambda}_j \in \mathbb{C}^{K \times K}$ is a diagonal matrix with $[\mathbf{\Lambda}_j]_{k,k} = \sigma_{j,k}^2$.

Then, the partial rows of $\mathbf{Y}_j^{\mathrm{du}}$, $\mathbf{A}_j$, and $\mathbf{N}_j^{\mathrm{p}}$ are taken to form new matrices $\mathbf{X}_j^{(1)}, \mathbf{X}_j^{(2)}, \mathbf{X}_j^{(3)} \in \mathbb{C}^{M \times N_{\mathrm{du}}}$, $\mathbf{B}_j^{(1)}, \mathbf{B}_j^{(2)}, \mathbf{B}_j^{(3)} \in \mathbb{C}^{M \times KL}$, $\mathbf{M}_j^{(1)}, \mathbf{M}_j^{(2)}, \mathbf{M}_j^{(3)} \in \mathbb{C}^{M \times N_{\mathrm{du}}}$, where $M = (N_{\mathrm{x}} - 1)(N_{\mathrm{y}} - 1)$. In particular, the rows of $\mathbf{X}_j^{(1)}$, $\mathbf{B}_j^{(1)}$, and $\mathbf{M}_j^{(1)}$ are the rows of $\mathbf{Y}_j^{\mathrm{du}}$, $\mathbf{A}_j$, and $\mathbf{N}_j^{\mathrm{p}}$ with indices $n_{\mathrm{y}}N_{\mathrm{x}} + n_{\mathrm{x}} + 1$, $n_{\mathrm{x}} = 0, 1, \cdots, N_{\mathrm{x}} - 2, n_{\mathrm{y}} = 0, 1, \cdots, N_{\mathrm{y}} - 2$, respectively. The rows of $\mathbf{X}_j^{(2)}$, $\mathbf{B}_j^{(2)}$, and $\mathbf{M}_j^{(2)}$ are the rows of $\mathbf{Y}_j^{\mathrm{du}}$, $\mathbf{A}_j$, and $\mathbf{N}_j^{\mathrm{p}}$ with indices $n_{\mathrm{y}}N_{\mathrm{x}} + n_{\mathrm{x}} + 1$, $n_{\mathrm{x}} = 1, 2, \cdots, N_{\mathrm{x}} - 1, n_{\mathrm{y}} = 0, 1, \cdots, N_{\mathrm{y}} - 2$, respectively; The rows of $\mathbf{X}_j^{(3)}$, $\mathbf{B}_j^{(3)}$, and $\mathbf{M}_j^{(3)}$ are the rows of $\mathbf{Y}_j^{\mathrm{du}}$, $\mathbf{A}_j$, and $\mathbf{N}_j^{\mathrm{p}}$ with indices $n_{\mathrm{y}}N_{\mathrm{x}} + n_{\mathrm{x}} + 1$, $n_{\mathrm{x}} = 0, 1, \cdots, N_{\mathrm{x}} - 2, n_{\mathrm{y}} = 1, 2, \cdots, N_{\mathrm{y}} - 1$, respectively. According to (18), $\mathbf{X}_j^{(1)}, \mathbf{X}_j^{(2)}, \mathbf{X}_j^{(3)}$ can be expressed as

$$\mathbf{X}_j^{(m)} \approx \frac{1}{\sqrt{N}} \mathbf{B}_j^{(m)} \mathbf{D}_j \mathbf{S} + \mathbf{M}_j^{(m)}, m = 1, 2, 3.$$

After that, the partial rows of $\bar{\mathbf{U}}_j$ are taken to form a new matrix $\mathbf{V}_j \in \mathbb{C}^{M \times N_{\mathrm{Cell}}L}$, whose rows are the rows of $\bar{\mathbf{U}}_j$ with indices $n_{\mathrm{y}}N_{\mathrm{x}} + n_{\mathrm{x}} + 1$, $n_{\mathrm{x}} = 0, 1, \cdots, N_{\mathrm{x}} - 2, n_{\mathrm{y}} = 0, 1, \cdots, N_{\mathrm{y}} - 2$. Then, $\mathbf{X}_j^{(m)}$ is processed into

$$\begin{aligned} \bar{\mathbf{X}}_j^{(m)} &= \mathbf{V}_j^H \mathbf{X}_j^{(m)} \\ &\approx \frac{1}{\sqrt{N}} \mathbf{V}_j^H \mathbf{B}_j^{(m)} \mathbf{D}_j \mathbf{S} + \mathbf{V}_j^H \mathbf{M}_j^{(m)} \in \mathbb{C}^{N_{\mathrm{Cell}}L \times N_{\mathrm{du}}}, \\ &\quad m = 1, 2, 3. \end{aligned} \tag{21}$$

Accordingly, the covariance matrix of any column of $\bar{\mathbf{X}}_j^{(m)}$ with any column of $\bar{\mathbf{X}}_j^{(1)}$ is denoted as

$$\begin{aligned}\mathbf{R}_j^{(m)} &= \frac{1}{N_{\mathrm{du}}}\mathbb{E}\{\bar{\mathbf{X}}_j^{(m)}(\bar{\mathbf{X}}_j^{(1)})^H\} \\ &\approx \frac{1}{N}\mathbf{V}_j^H\mathbf{B}_j^{(m)}\mathbf{D}_j\mathbf{\Lambda}_j\mathbf{D}_j^H(\mathbf{B}_j^{(1)})^H\mathbf{V}_j + \mathbf{V}_j^H\mathbf{V}_j \\ &\quad \in \mathbb{C}^{N_{\mathrm{Cell}}L\times N_{\mathrm{Cell}}L}, \forall m = 1,2,3. \end{aligned} \tag{22}$$

Note that the covariance matrix can be estimated with a sample covariance matrix

$$\frac{1}{N_{\mathrm{du}}}\sum_{n_{\mathrm{d}}=1}^{N_{\mathrm{du}}}[\bar{\mathbf{X}}_j^{(m)}]_{n_{\mathrm{d}}}([\bar{\mathbf{X}}_j^{(1)}]_{n_{\mathrm{d}}})^H.$$

Before continuing the derivation, two properties of $\mathbf{B}_j^{(m)}$, $\forall m = 1,2,3$ and $\mathbf{V}_j$ should be given first.

Property 1: $\mathbf{B}_j^{(2)} = \mathbf{B}_j^{(1)}\mathbf{\Phi}_2$, $\mathbf{B}_j^{(3)} = \mathbf{B}_j^{(1)}\mathbf{\Phi}_3$, where $\mathbf{\Phi}_2 \in \mathbb{C}^{KL\times KL}$ and $\mathbf{\Phi}_3 \in \mathbb{C}^{KL\times KL}$ are diagonal matrices with diagonal elements being $e^{i\pi\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$, and $e^{i\pi\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$, respectively.

Proof: According to the definition of $\mathbf{A}_j$ in (19) and the definition of $\mathbf{B}_j^{(m)}$ below (20), it can be seen that $[\mathbf{B}_j^{(2)}]_{(l-1)*K+k} = [\mathbf{B}_j^{(1)}]_{(l-1)*K+k}e^{i\pi\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}}$ and $[\mathbf{B}_j^{(3)}]_{(l-1)*K+k} = [\mathbf{B}_j^{(1)}]_{(l-1)*K+k}e^{i\pi\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}}$. □

Property 2: In the limit of $N_{\mathrm{x}}\to\infty$ and $N_{\mathrm{y}}\to\infty$, we have $\mathbf{V}_j^H\mathbf{V}_j = \mathbf{I}_{N_{\mathrm{Cell}}L}$.

Proof: According to the definition of $\bar{\mathbf{U}}_j$ in (17) and the definition of $\mathbf{V}_j$ above (21), it can be seen that

$$\begin{aligned}&[\mathbf{V}_j]_{n_{\mathrm{y}}(N_{\mathrm{x}}-1)+n_{\mathrm{x}}+1,q} = [\mathbf{U}]_{n_{\mathrm{y}}N_{\mathrm{x}}+n_{\mathrm{x}}+1,a(q)}, \\ &n_{\mathrm{x}} = 0,1,\cdots,N_{\mathrm{x}}-2, n_{\mathrm{y}} = 0,1,\cdots,N_{\mathrm{y}}-2, \\ &q = 1,2,\cdots,N_{\mathrm{Cell}}L,\end{aligned}$$

where $a(q)$ is the index of the selected column of $\mathbf{U}$. According to the definition of $\mathbf{U}$ in (10), we have

$$[\mathbf{V}_j]_{n_{\mathrm{y}}(N_{\mathrm{x}}-1)+n_{\mathrm{x}}+1,q} = \frac{1}{\sqrt{N}}[\mathbf{b}(f_{a(q)},g_{a(q)})]_{n_{\mathrm{y}}N_{\mathrm{x}}+n_{\mathrm{x}}+1}. \tag{23}$$

According to the definition in (8), it can be seen that

$$\begin{aligned}[\mathbf{V}_j]_{n_{\mathrm{y}}(N_{\mathrm{x}}-1)+n_{\mathrm{x}}+1,q} &= \frac{1}{\sqrt{N}}e^{i\pi(n_{\mathrm{x}}-0.5(N_{\mathrm{x}}-1))f_{a(q)}} \\ &\times e^{i\pi(n_{\mathrm{y}}-0.5(N_{\mathrm{y}}-1))g_{a(q)}}. \end{aligned} \tag{24}$$

Then, we have

$$\begin{aligned}\lim_{N_{\mathrm{x}},N_{\mathrm{y}}\to\infty}[\mathbf{V}_j]_q^H[\mathbf{V}_j]_q &= \lim_{N_{\mathrm{x}},N_{\mathrm{y}}\to\infty}\frac{(N_{\mathrm{x}}-1)(N_{\mathrm{y}}-1)}{N} \\ &= 1. \\ \lim_{N_{\mathrm{x}},N_{\mathrm{y}}\to\infty}[\mathbf{V}_j]_q^H[\mathbf{V}_j]_{q'} &= \lim_{N_{\mathrm{x}},N_{\mathrm{y}}\to\infty}\frac{1}{N}(\mathbf{b}^H(f_{a(q)},g_{a(q)}) \\ &\times \mathbf{b}(f_{a(q')},g_{a(q')})) = 0, \forall q\neq q'.\end{aligned}$$

□

By employing Property 1, we have $\mathbf{V}_j^H\mathbf{B}_j^{(2)} = \mathbf{V}_j^H\mathbf{B}_j^{(1)}\mathbf{\Phi}_2$ and $\mathbf{V}_j^H\mathbf{B}_j^{(3)} = \mathbf{V}_j^H\mathbf{B}_j^{(1)}\mathbf{\Phi}_3$. According to the fact that the number of antennas at the BS in mm-wave MIMO systems is extremely large, Property 2 can be employed to approximate $\mathbf{R}_j^{(m)}$ in (22) as

$$\begin{aligned}\mathbf{R}_j^{(m)} &\approx \frac{1}{N}\mathbf{V}_j^H\mathbf{B}_j^{(m)}\mathbf{D}_j\mathbf{\Lambda}_j\mathbf{D}_j^H(\mathbf{B}_j^{(1)})^H\mathbf{V}_j + \mathbf{I}_{N_{\mathrm{Cell}}L}, \\ \forall m &= 1,2,3.\end{aligned}$$

Then, the estimation of signal parameters via rotational invariance techniques (ESPRIT) in [35] can be employed to estimate the diagonal elements of $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$. The details can be found in [35] and will not be shown here. As a result, the spatial frequencies, $\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$ and $\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$, can be estimated. Their estimates are denoted as $\tilde{\beta}_{j,l_k,\tilde{s}_{j,l_k}}$ and $\tilde{\gamma}_{j,l_k,\tilde{s}_{j,l_k}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$.

There are three facts that need to be emphasized here. First, this estimation process treats the spatial frequencies of in-cell MSs the same as that of the MSs in other cells, i.e., the in-cell channels and the ICI channels are estimated together. Thus, the ICI will not influence the channel estimation. Second, the estimation utilizes the beamspace principle to decrease the dimension of the covariance matrix $\mathbf{R}_j^{(m)}$. Hence, the beamspace principle is employed to decrease the computational complexity. Third, the estimated spatial frequencies are obtained in random order. Thus, we cannot find the correspondence between the spatial frequencies and the MSs. In the following, we will continue to solve this problem.

In the first step, we have got the the indices of the K beams with the largest correlation value for the kth MS in the jth cell, which are $p_{j,k}, q_{j,k}$. Here, these coarse estimates will be matched with the estimated spatial frequencies obtained in the second step, which are $\tilde{\beta}_{j,l_k,\tilde{s}_{j,l_k}}$ and $\tilde{\gamma}_{j,l_k,\tilde{s}_{j,l_k}}, l = 1,2,\cdots,L, k = 1,2,\cdots,K$. According to (16), the $z_{j,k}$th element of $\mathbf{c}_k$, where $z_{j,k} = (q_{j,k}-1)N_{\mathrm{x}} + p_{j,k}$, can be approximated as

$$\begin{aligned}[\mathbf{c}_k]_{z_{j,k}} &\approx \sum_{l=1}^{L}\frac{\alpha_{j,l_k,\tilde{s}_{j,l_k}}}{N}\sigma_{l,k}\sigma_{j,k}\mathbf{b}^H(f_{z_{j,k}},g_{z_{j,k}}) \\ &\times \mathbf{b}(\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}},\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}) \\ &+ \frac{1}{N_{\mathrm{p}}\sqrt{N}}\mathbf{b}^H(f_{z_{j,k}},g_{z_{j,k}})\mathbf{N}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H.\end{aligned}$$

Since the selected beams $\mathbf{b}(f_{z_{j,k}},g_{z_{j,k}})$ are mostly correlated with $\mathbf{b}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}},\sin\phi_{j,j_k,\tilde{s}_{j,j_k}})$, it is expected that the value of $[\mathbf{c}_k]_{z_{j,k}}$ is almost determined by the term with $\mathbf{b}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}},\sin\phi_{j,j_k,\tilde{s}_{j,j_k}})$. Then, we have

$$\left|[\mathbf{c}_k]_{z_{j,k}}\right| \approx \eta_{z_{j,k}} + n_{z_{j,k}}, \tag{25}$$

where

$$\eta_{z_{j,k}} = \frac{|\alpha_{j,j_k,\tilde{s}_{j,j_k}}|}{N}\sigma_{j,k}^2|\mathbf{b}^H(f_{z_{j,k}}, g_{z_{j,k}}) \times \mathbf{b}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}}, \sin\phi_{j,j_k,\tilde{s}_{j,j_k}})| = \frac{|\alpha_{j,j_k,\tilde{s}_{j,j_k}}|}{N}\sigma_{j,k}^2 \left|\frac{\sin(0.5\pi N_{\mathrm{y}}(\sin\phi_{j,j_k,\tilde{s}_{j,j_k}} - g_{z_{j,k}}))}{\sin(0.5\pi(\sin\phi_{j,j_k,\tilde{s}_{j,j_k}} - g_{z_{j,k}}))}\right| \times \left|\frac{\sin(0.5\pi N_{\mathrm{x}}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}} - f_{z_{j,k}}))}{\sin(0.5\pi(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}} - f_{z_{j,k}}))}\right| \quad (26)$$

is the high correlation value, and

$$n_{z_{j,k}} = \left|\frac{1}{N_{\mathrm{p}}\sqrt{N}}\mathbf{b}^H(f_{z_{j,k}}, g_{z_{j,k}})\mathbf{N}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H\right|$$

is the noise. In the above derivation, (8) is used in deriving (26). According to the above expression, we can take $\left|[\mathbf{c}_k]_{z_{j,k}}\right|$ as an estimate of $\eta_{z_{j,k}}$.

From (9), it is known that $f_{z_{j,k}+1} = f_{z_{j,k}} + 2/N_{\mathrm{x}}$, $g_{z_{j,k}+1} = g_{z_{j,k}}$, $f_{z_{j,k}+N_{\mathrm{x}}} = f_{z_{j,k}}$, $g_{z_{j,k}+N_{\mathrm{x}}} = g_{z_{j,k}} + 2/N_{\mathrm{y}}$. Hence, $\eta_{z_{j,k}}$ in (26) satisfies the two equations listed at the top of the next page.

In the first step, we have selected the beam with the largest correlation value for each MS, and the indices are $p_{j,k}$ and $p_{j,k}$, which are used to calculate $z_{j,k} = (q_{j,k} - 1)N_{\mathrm{x}} + p_{j,k}$. Thus, the values of $f_{z_{j,k}}$, $f_{z_{j,k}+1}$, $g_{z_{j,k}}$, and $g_{z_{j,k}+N_{\mathrm{x}}}$ are known. In the second step, we have estimated the spatial frequencies of all the MSs, $\cos\phi_{j,l_k,\tilde{s}_{j,l_k}}\sin\theta_{j,l_k,\tilde{s}_{j,l_k}}, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$ and $\sin\phi_{j,l_k,\tilde{s}_{j,l_k}}, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$, the estimates of which are denoted as $\tilde{\beta}_{j,l_k,\tilde{s}_{j,l_k}}$ and $\tilde{\gamma}_{j,l_k,\tilde{s}_{j,l_k}}, l = 1, 2, \cdots, L, k = 1, 2, \cdots, K$. Thus, we can pick out parts of these estimates as the estimates of the spatial frequencies of the MSs in the jth cell by matching $\eta_{z_{j,k}}/\eta_{z_{j,k}+1}$ with $u(\tilde{\beta}_{j,l_k,\tilde{s}_{j,l_k}}, f_{z_{j,k}})$, and matching $\eta_{z_{j,k}}/\eta_{z_{j,k}+N_{\mathrm{x}}}$ with $v(\tilde{\gamma}_{j,l_k,\tilde{s}_{j,l_k}}, g_{z_{j,k}})$. Since we only have the estimates of $\eta_{z_{j,k}}$, which is $\left|[\mathbf{c}_k]_{z_{j,k}}\right|$, we replace $\eta_{z_{j,k}}$ with $\left|[\mathbf{c}_k]_{z_{j,k}}\right|$ in the matching process.

Denote the estimates of the spatial frequencies of the in-cell MSs as $\hat{\beta}_{j,j_k,\tilde{s}_{j,j_k}}$ and $\hat{\gamma}_{j,j_k,\tilde{s}_{j,j_k}}, k = 1, 2, \cdots, K$, then the match process can be expressed in equations as

$$\hat{\beta}_{j,j_k,\tilde{s}_{j,j_k}} = \underset{\tilde{\beta}_{j,l_{k'},\tilde{s}_{j,l_{k'}}},\forall l,k'}{\operatorname{argmin}} \left|\frac{\left|[\mathbf{c}_k]_{z_{j,k}}\right|}{\left|[\mathbf{c}_k]_{z_{j,k}+1}\right|} - u(\tilde{\beta}_{j,l_{k'},\tilde{s}_{j,l_{k'}}}, f_{z_{j,k}})\right|^2,$$

$$\hat{\gamma}_{j,j_k,\tilde{s}_{j,j_k}} = \underset{\tilde{\gamma}_{j,l_{k'},\tilde{s}_{j,l_{k'}}},\forall l,k'}{\operatorname{argmin}} \left|\frac{\left|[\mathbf{c}_k]_{z_{j,k}}\right|}{\left|[\mathbf{c}_k]_{z_{j,k}+N_{\mathrm{x}}}\right|} - v(\tilde{\gamma}_{j,l_{k'},\tilde{s}_{j,l_{k'}}}, g_{z_{j,k}})\right|^2.$$

C. Third Step: Path Loss Estimation

In the third step, the path loss of each MS is estimated. As shown at the end of Section II, despite the spatial frequencies estimated in the second step, the path loss coefficients are the remaining parameters that determine the channel vectors. With the estimated spatial frequencies and the received pilot signal, the path loss can be estimated according to its relation with these known variables.

Similar to (25), we have

$$[\mathbf{c}_k]_{z_{j,k}} \approx \tilde{\eta}_{z_{j,k}} + \tilde{n}_{z_{j,k}},$$

where

$$\begin{aligned}\tilde{\eta}_{z_{j,k}} &= \frac{\alpha_{j,j_k,\tilde{s}_{j,j_k}}}{N}\sigma_{j,k}^2\mathbf{b}^H(f_{z_{j,k}}, g_{z_{j,k}}) \\ &\times \mathbf{b}(\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}}, \sin\phi_{j,j_k,\tilde{s}_{j,j_k}}), \\ \tilde{n}_{z_{j,k}} &= \frac{1}{N_{\mathrm{p}}\sqrt{N}}\mathbf{b}^H(f_{z_{j,k}}, g_{z_{j,k}})\mathbf{N}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H.\end{aligned}$$

In the expression of $\tilde{\eta}_{z_{j,k}}$, $\tilde{\eta}_{z_{j,k}}$ has been estimated with $[\mathbf{c}_k]_{z_{j,k}}$, $\cos\phi_{j,j_k,\tilde{s}_{j,j_k}}\sin\theta_{j,j_k,\tilde{s}_{j,j_k}}$, and $\sin\phi_{j,j_k,\tilde{s}_{j,j_k}}$ have been estimated with $\hat{\beta}_{j,j_k,\tilde{s}_{j,j_k}}$ and $\hat{\gamma}_{j,j_k,\tilde{s}_{j,j_k}}$. The other parameters but $\alpha_{j,j_k,\tilde{s}_{j,j_k}}$ are known. As can be seen, the unknown path loss, $\alpha_{j,j_k,\tilde{s}_{j,j_k}}$, can be estimated as

$$\hat{\alpha}_{j,j_k,\tilde{s}_{j,j_k}} = \frac{[\mathbf{c}_k]_{z_{j,k}}N}{\sigma_{j,k}^2\mathbf{b}^H(f_{z_{j,k}}, g_{z_{j,k}})\mathbf{b}(\hat{\beta}_{j,j_k,\tilde{s}_{j,j_k}}, \hat{\gamma}_{j,j_k,\tilde{s}_{j,j_k}})}.$$

Till now, the complete beamspace channel estimation process has been presented, in which the beamspace principle has been utilized to estimate the channel parameters and to reduce computational complexity. In the following section, the performance of the proposed approach will be discussed.

Table 1. Performance comparison of the channel estimation approaches.

Approaches	Main ideas/Application scenarios
Beamspace based [28]	Select beams with large correlation, estimate channels with correlation match/ Single cell, short channel coherence interval
Pilot based [26]	Estimated channel with pilot orthogonality/ Single cell
Proposed	Select inside beams with large correlation, estimate channels with data symbols, pick out channels with correlation match/ A large number of BS antennas and data symbols

IV. PERFORMANCE ANALYSIS

In this section, the performance and the computational complexity of the proposed approach will be compared with that of other existing approaches.

A. Comparison of Performance

The existing approach for mm-wave beamspace MIMO systems is the approach in [28]. Additionally, the approach proposed in [26] can be modified for mm-wave beamspace MIMO systems. In Table 1, the two approaches are compared with the approach proposed in this paper, where the approach in [28], the modification of the approach proposed in [26], and the approach proposed in this paper are denoted as "Beamspace based [28]", "Pilot based [26]", and "Proposed" in this table.

The approach in [28] is also based on the beamspace principle. This approach selects the beams with the larger correlation values of the obtained signal vector and the beam vectors. Then, the spatial frequencies of the MSs are estimated by searching for the spatial frequencies that minimize the difference between the observed correlation value and the theoretical correlation value. As shown in the analysis of [28], the channel estimation error in the single-cell LOS scenario tends to zero as the number of the BS antennas tends to infinity. In multi-cell scattered scenarios, the cell-edge MSs in adjacent cells may be close to cell-edge MSs inside the cell and the LOS path may be blocked, the path loss of the strongest path of the ICI channel may be lower than that of the in-cell channel, i.e.,

$$\max_s |\alpha_{j,j_k,s}| < \max_s |\alpha_{j,l_k,s}|.$$

Thus, the received signal power from the in-cell MS may be lower than that from MSs in other cells, i.e.,

$$\max_s |\alpha_{j,j_k,s}|\sigma_{j,k}^2 < \max_s |\alpha_{j,l_k,s}|\sigma_{j,k}\sigma_{l,k}.$$

In this case, the strongest path of the ICI channel, $\mathbf{h}_{j,l_{k'}}$, rather than the strongest path of the channel of the in-cell MS, $\mathbf{h}_{j,j_k}$, is estimated. In the limit of infinite BS antennas, the K paths with the highest received signal power, including the in-cell channels and the ICI channels, are estimated with no error. If the strongest path of the ICI channel $\mathbf{h}_{j,l_{k'}}$ is taken as the channel estimate of the channel $\mathbf{h}_{j,j_k}$, the combination (k,k') is included in the set $\mathbb{C}_{j,l}$. Then, for the jth cell, the channel estimation error is approximately

$$\sum_{l\neq j}\sum_{(k,k',l)\in\mathbb{C}_{j,l}} ||\mathbf{h}_{j,j_k} - \mathbf{h}_{j,l_{k'}}||^2, \tag{27}$$

where the estimation error caused by multipaths is ignored as they are usually much smaller than the error above.

In case that there is only one cell, the estimation error is only caused by the multipaths and is relatively small. Additionally, the advantage of this approach is that only one pilot symbol is needed. Therefore, this approach only suits for single-cell scenarios with short channel coherence interval.

When the approach in [26] is modified for the considered system, the orthogonal pilots are employed in each cell and are reused among cells. The number of pilot symbols required for this approach is K. Then, the orthogonality of the pilots inside the cell is used to obtain the channel estimate. The received pilot matrix $\mathbf{Y}_j^{\mathrm{p}}$ in (1) is processed into

$$\frac{1}{N_{\mathrm{p}}}\mathbf{Y}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H = \mathbf{h}_{j,j_k} + \sum_{l\neq j}\mathbf{h}_{j,l_k} + \frac{1}{N_{\mathrm{p}}}\mathbf{N}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H,$$

which is taken as the channel estimate for the kth MS in the jth cell. Thus, the ICI remains in the channel estimate. For the jth cell, the channel estimation error is

$$\sum_{k=1}^{K}\left\|\sum_{l\neq j}\mathbf{h}_{j,l_k} + \frac{1}{N_{\mathrm{p}}}\mathbf{N}_j^{\mathrm{p}}(\mathbf{s}_{j,k}^{\mathrm{p}})^H\right\|^2. \tag{28}$$

As can be seen, the ICI are contained in the channel estimate. Finally, the correlations of the beam vectors and the channel estimate are used to select the beams for the in-cell MSs. Since the large scale fading coefficients and the DOAs of the cell-edge MSs are close, these correlations are also close. Thus, the beams for MSs in other cells may be selected. In case that there is only one cell, only the noise results into estimation error. Thus, the approach in [26] suits for single-cell scenarios.

In contrast, the proposed approach treats the MSs in other cells the same as the in-cell MSs and estimate the spatial frequencies of all the MSs. When the number of data symbols and the number of BS antennas tend to infinity, the effect of finite sample and noise diminishes, and only other paths from the same MS affects the estimate of the spatial frequencies and the path loss of the strongest path. Since the effect of other paths is related to the path losses and the spatial frequencies of other paths, the channel estimate of the kth MS in the lth cell can be expressed as

$$\frac{\alpha_{j,l_k,\tilde{s}_{j,l_k}} + \delta_{j,l_k}^{\alpha}}{\sqrt{N}}\mathbf{a}(\theta_{j,l_k,\tilde{s}_{j,l_k}} + \delta_{j,l_k}^{\theta}, \phi_{j,l_k,\tilde{s}_{j,l_k}} + \delta_{j,l_k}^{\phi}),$$

where δ_{j,l_k}^{α}, δ_{j,l_k}^{θ}, and δ_{j,l_k}^{ϕ} are the estimation error of the path loss and the spatial frequencies, and are functions of $\alpha_{j,l_k,s}$, $\theta_{j,l_k,s}$, and $\phi_{j,l_k,s}, s = 0, 1, \cdots, S$. Then, the partial estimates obtained in the second step are selected to form the final channel estimates according to the beams selected in the first step. Note that the beams are selected by picking out beams that correspond to positions inside the cell. As long as the in-cell MS and MSs in other cells that are allocated with the same pilot are far away from each other, the selected beam will correspond to position near the in-cell MS. As a result, the match of the beams selected in the first step and the channel estimates obtained in the second step is of little error, which means most of the ICI can be canceled. In the ideal case that there is no error in the match process, the channel estimate of the kth MS in the jth cell can be expressed as

$$\frac{\alpha_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\alpha}}{\sqrt{N}}\mathbf{a}(\theta_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\theta}, \phi_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\phi}).$$

Thus, the channel estimation error is

$$\sum_{k=1}^{K}\left\|\sum_{s=0}^{S}\frac{\alpha_{j,j_k,s}}{\sqrt{N}}\mathbf{a}(\theta_{j,j_k,s},\phi_{j,j_k,s}) - \frac{\alpha_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\alpha}}{\sqrt{N}} \times\mathbf{a}(\theta_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\theta}, \phi_{j,j_k,\tilde{s}_{j,j_k}} + \delta_{j,j_k}^{\phi})\right\|^2.$$

In case that the strongest path is much stronger than the other paths, i.e.,

$$|\alpha_{j,j_k,\tilde{s}_{j,j_k}}| \gg |\alpha_{j,j_k,s}|, \forall s \neq \tilde{s}_{j,j_k}, \tag{29}$$

we have $|\delta_{j,j_k}^{\alpha}| \approx 0$, $\delta_{j,j_k}^{\theta} \approx 0$, $\delta_{j,j_k}^{\phi} \approx 0$, and

$$\sum_{s=0}^{S}\frac{\alpha_{j,j_k,s}}{\sqrt{N}}\mathbf{a}(\theta_{j,j_k,s},\phi_{j,j_k,s}) \approx \frac{\alpha_{j,j_k,\tilde{s}_{j,j_k}}}{\sqrt{N}}\mathbf{a}(\theta_{j,j_k,\tilde{s}_{j,j_k}},\phi_{j,j_k,\tilde{s}_{j,j_k}}).$$

Table 2. Computational complexity comparison of the channel estimation approaches.

Approaches	Computational complexity
[28]	$O(N^2)$ multiplications, $O(K)$ additions and divisions
Modified [26]	$O(N^2K)$ multiplications and additions
Proposed	$O(\max(K^L, N^2K, N^2N_{\mathrm{du}}))$ additions, $O(\max(K^L, N^2K, N_{\mathrm{Cell}}^3L^3))$ multiplications, and $O(K)$ divisions

As can be seen, the channel estimation error is approximately zero. Since the assumption of channel in (29) is of high probability, which is known from the measurements in [31], the channel estimation error of the proposed approach is approximately zero. Therefore, the proposed channel estimation approach suits for scenarios with a large number of BS antennas and data symbols.

B. Comparison of Computational Complexity

In Table 2, the computational complexities of the three approaches are compared. The notation $O(n)$ in [36] is used to denote that the number of floating-point operations for each cell is linear in $n \in \mathbb{R}^+$.

For the approach in [28], the main calculations include correlation of the received pilot vector and the beams, matching of these correlation values and the theoretical values, and estimation of path gains. The first calculation costs $O(N^2)$ multiplications, the second calculation costs $O(K)$ additions, multiplications, and divisions, the third calculation costs $O(K)$ multiplications and divisions. As the number of antennas, N, is much larger than the number of MSs, K, the approach in [28] costs $O(N^2)$ multiplications, $O(K)$ additions and divisions.

When the approach in [26] is modified for the considered system, the main calculation is the channel estimation and beam selection. The channel estimation costs $O(NN_{\mathrm{p}}K)$ multiplications and additions, the beam selection costs $O(N^2K)$ multiplication and additions. As $N \gg N_{\mathrm{p}}$, the modified [26] costs $O(N^2K)$ multiplications and additions.

For the proposed approach, the main calculations include the pilot allocation, the beam correlation, the spatial frequency estimation, and the path loss estimation. There are $(K-k+1)^{L-1}$ cases for pilot allocation of the kth MS in the jth cell. Thus, there are $O(K^L)$ additions and multiplications in pilot allocation. The beam correlation is the same as that in [26], and costs $O(N^2K)$ multiplications and additions. The spatial frequency estimation include the calculation of sample covariance matrices, the singular value decomposition of the covariance matrices, and the match process, which cost $O(N_{\mathrm{Cell}}^3L^3)$ multiplications, $O(\max(K^2, N^2N_{\mathrm{du}}))$ additions, and $O(K)$ divisions, respectively. Therefore, the proposed approach costs $O(\max(K^L, N^2K, N_{\mathrm{Cell}}^3L^3))$ multiplications, $O(\max(K^L, N^2K, N^2N_{\mathrm{du}}))$ additions, and $O(K)$ divisions, respectively.

In the typical case of $K = 50$, $L = 3$, $N = 4,832$, $N_{\mathrm{du}} = 300$, the approach in [28] costs $O(10^7)$ multiplications, $O(10^2)$ additions and divisions, respectively. The modified [26] costs $O(10^9)$ multiplications and additions. The proposed approach costs $O(10^9)$ multiplications, $O(10^9)$ additions, and $O(10^2)$ divisions, respectively. As can be seen, the proposed approach has close computational complexity with the other two approaches and is applicable in mm-wave beamspace MIMO systems.

V. SIMULATION RESULTS

In this section, simulations are conducted and the results will show the performance of the proposed approach and other existing approaches, including the modified [26], the approach in [28], and the perfect channel estimation. The perfect channel estimation means that the channels are completely known at the BS and no pilot symbol is needed. This case is denoted as "Perfect" and corresponds to the upper bound of channel estimation performance. For the two beamspace approaches, the estimated path loss is normalized to the order of the magnitude of the path loss if the estimated path loss is of higher order of magnitude. This normalization can avoid large error in the estimate of the path loss. Since the time division multiple access (TDMA) can be employed to cancel the pilot contamination, it is combined with the modified [26] and the approach in [28] to form another two approaches. Specifically, in the considered TDMA system, the MSs in each cell transmit in $1/L$ of the total transmission period, i.e., $(T_c - N_{\mathrm{p}})/L$, and the transmission period of each cell is non-overlap. The cell topology is the same as the one in Fig. 1. The simulation parameters are listed in Table 3, which are similar to that in [22]. According to the measurements in [31], [37], the channel coefficients are set as follows. The number of multipaths of each channel is 3. The path loss coefficients of multipaths, $20\log_{10}\eta_{j,l_k,s}, s = 1, 2, \cdots, S$, cf. (6), are i.i.d. uniformly distributed in the range $[-20, -15]$, which is adopted in [32]. Since the attenuation of the scattering is severe is both cases, which means the multipaths are of much weaker strength. the simulation results are expected to be similar. Moreover, the azimuth angle differences between the multipaths and the LOS path are i.i.d. uniformly distributed in the range $[-10°, 10°]$, the elevation angle differences between the multipaths and the LOS path are i.i.d. uniformly distributed in the range $[-3°, 3°]$. The path loss coefficient of LOS, $\eta_{j,l_k,0}$, is a random variable, the probability that $\eta_{j,l_k,0} = 0$ (complete blockage) is 0.1, and the probability that $\eta_{j,l_k,0} = 1$ (no blockage) is 0.9.

In the simulations, both the uplink and the downlink transmissions are considered and they occupy the same length of transmission interval. Thus, the number of uplink/downlink data symbols is $N_{\mathrm{du}} = N_{\mathrm{dd}} = 0.5(T_c - N_{\mathrm{p}})$, and the spectral efficiency is averaged over the channel coherence interval. Additionally, the minimum mean square error (MMSE) criterion is employed in the two-way transmission. For the downlink, the MMSE precoding matrix, i.e., $\mathbf{F}_l^{\mathrm{dd}}$ in (4), is

$$\mathbf{F}_l^{\mathrm{dd}} = \tilde{\mathbf{U}}_l\tilde{\mathbf{U}}_l^H\hat{\mathbf{H}}_l\left(\hat{\mathbf{H}}_l^H\tilde{\mathbf{U}}_l\tilde{\mathbf{U}}_l^H\hat{\mathbf{H}}_l + \frac{K}{\rho_{\mathrm{BS}}}\mathbf{I}_K\right)^{-1},$$

where $\hat{\mathbf{H}}_l \in \mathbb{C}^{N\times K}$ is the estimate of the channels from the MSs in the lth cell to the lth BS, and $[\hat{\mathbf{H}}_l]_k$ is the estimate of $\mathbf{h}_{l,l_k}$; ρ_{BS} is the transmission power of each BS, i.e., $\rho_{\mathrm{BS}} = \mathrm{tr}(\mathbf{F}_l^{\mathrm{dd}}\mathbf{Q}_l^2(\mathbf{F}_l^{\mathrm{dd}})^H)$. Correspondingly, the downlink signal-to-noise ratio (SNR) of each MS is defined as ρ_{BS}/K. For the

Table 3. Simulation parameters.

Number of cells L	3
Number of MSs in each cell K	50
Number of pilot symbols N_{p}	50
Number of beams for each MS N_{MS}	2
Number of antennas (horizontal) N_{x}	32
Number of antennas (vertical) N_{y}	151
BS height	10 m
Minimal MS to BS foot distance	10 m
Maximal MS to BS foot distance	100 m
SNR	0 dB
Transmission frequency	80 GHz
Channel coherence interval T_c	500 slots

uplink, the received signal in (2) is processed as

$$\check{\mathbf{Y}}_j^{\mathrm{du}} = \mathbf{R}_j^H(\hat{\mathbf{H}}_j^H\tilde{\mathbf{U}}_j\tilde{\mathbf{U}}_j^H\hat{\mathbf{H}}_j\mathbf{R}_j + \mathbf{I}_K)^{-1}\hat{\mathbf{H}}_j^H\tilde{\mathbf{U}}_j\tilde{\mathbf{U}}_j^H\mathbf{Y}_j^{\mathrm{du}} \in \mathbb{C}^{K\times N_{\mathrm{du}}},$$

where $\mathbf{R}_j \in \mathbb{R}^{K\times K}$ is a diagonal matrix of the variances of the data symbols, i.e., $[\mathbf{R}_j]_{k,k} = \sigma_{j,k}^2$. Correspondingly, the uplink SNR of the kth MS in the jth cell is defined as $|\alpha_{j,j_k,\tilde{\mathrm{s}}_{j,j_k}}|^2\sigma_{j,k}^2$. In the simulations, the uplink SNRs of the MSs are the same.

If the spectral efficiencies of one approach in the uplink data transmission period and the downlink data transmission period are R_{u} and R_{d}, respectively, then the spectral efficiency of this approach depends on the employed approach. For the proposed approach, the spectral efficiency is $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})(T_c-N_{\mathrm{p}})/T_c$, since N_{p} symbol slots are used for pilot transmission. For the modified [26], the spectral efficiency is $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})(T_c-K)/T_c$, since K symbol slots are used for pilot transmission. For the combination of the modified [26] and TDMA, the spectral efficiency is $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})(T_c-K)/T_c/L$. For the approach in [28], the spectral efficiency is $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})(T_c-1)/T_c$, as only one pilot is used. For the combination of the approach in [28] and TDMA, the spectral efficiency is $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})(T_c-1)/T_c/L$. For the upper bound obtained with perfect channel state information, the spectral efficiency is also $0.5(R_{\mathrm{u}}+R_{\mathrm{d}})$, as the pilot is not needed.

Despite the spectral efficiency, the normalized mean square error (NMSE) is another metric for measuring the channel estimation performance. For the channel from the kth MS in the jth cell to the jth BS, the NMSE of the channel estimation is

$$\frac{||\hat{\mathbf{h}}_{j,j_k}-\mathbf{h}_{j,j_k}||^2}{||\mathbf{h}_{j,j_k}||^2},$$

where $\hat{\mathbf{h}}_{j,j_k} \in \mathbb{C}^{N\times 1}$ is the estimate of the channel $\mathbf{h}_{j,j_k}$. In this paper, the NMSE of one channel estimation approach is defined as

$$\sum_{k=1}^{K}\frac{||\hat{\mathbf{h}}_{j,j_k}-\mathbf{h}_{j,j_k}||^2}{||\mathbf{h}_{j,j_k}||^2}.$$

In the first simulation, the SNR changes from -20 dB to 20 dB. Fig. 3 demonstrates the spectral efficiencies and NMSEs of the four approaches versus the SNR. It can be seen that the spectral efficiency of proposed approach surpasses that of the modified [26], the approach in [28], and their variations to TDMA. In addition, the NMSE of the proposed approach is smaller than that of the other two approaches with the high SNR. This means that the proposed approach has mitigated more ICI in the channel estimates than the other two approaches.

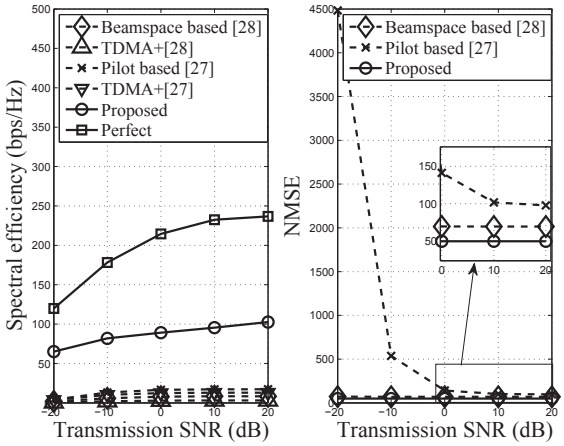

Fig. 3. Spectral efficiency versus the transmission SNR.

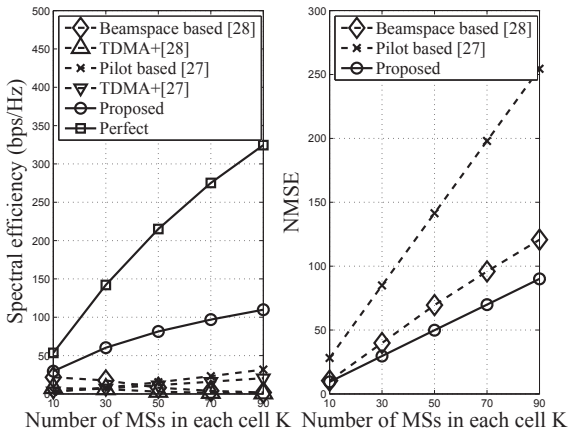

Fig. 4. Spectral efficiency versus the number of MSs in each cell.

In the second simulation, the number of MSs in each cell changes from 10 to 90. Fig. 4 shows the spectral efficiencies and NMSEs versus the number of MSs in each cell. As can be seen, the proposed approach is of higher spectral efficiency than the approach in [28], the modified [26], and their variations to TDMA. In addition, the NMSE of the proposed approach is smaller than that of the other two approaches. Meanwhile, the NMSEs of the three approaches increase with the increase of the number of MSs. For the two beamspace approaches, this is because the distances between MSs decreases as the number of MSs increases, which leads to ambiguity in the beam selection for the MSs in the first step of the proposed approach. For the pilot based approach, the reason is that the NMSE is linearly related to the number of the MSs.

In the third simulation, the number of BS antennas changes from 252 to $4,832$. More specifically, the numbers of antennas in the horizontal direction of the URA are $N_{\mathrm{x}} = 7, 11, 16, 24, 32$, the numbers of antennas in the vertical direc-

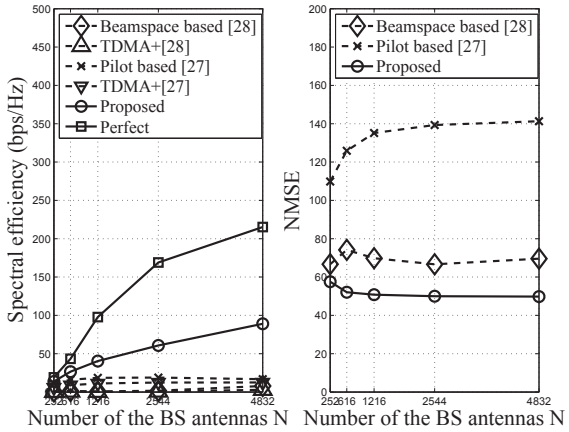

Fig. 5. Spectral efficiency versus the number of BS antennas.

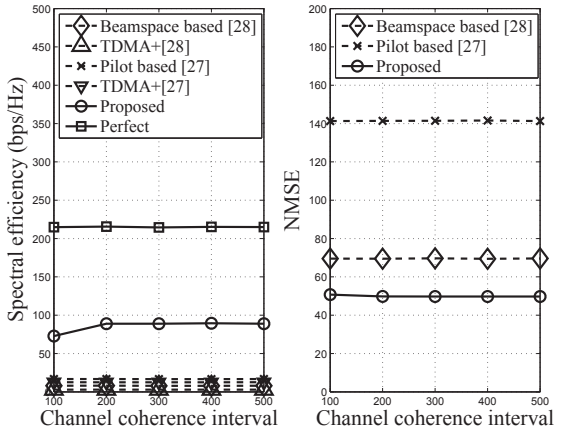

Fig. 6. Spectral efficiency versus the channel coherence interval.

tion of the URA are $N_{\mathrm{y}} = 36, 56, 76, 106, 151$. The spectral efficiencies and NMSEs versus the number of BS antennas are revealed in Fig. 5. As can be seen, the spectral efficiency of the proposed approach surpasses that of the approach in [28], the modified [26], and their variations. Moreover, the gap increases with the increase of the number of BS antennas. The reason is that the distances between the MSs increases in the beamspace, which benefits the beam selection in the first step of the proposed approach, and the channel estimation error in the second step of the proposed approach also decreases. This result verifies the analysis given in deriving the second step of the proposed approach.

In the last simulation, the channel coherence interval changes from 100 to 500. Fig. 6 depicted the relation between the spectral efficiencies, the NMSEs, and the channel coherence interval. When the channel coherence interval is short, the proposed approach is of low spectral efficiency. Meanwhile, the NMSE is also relatively larger with a shorter channel coherence interval. This is because the channel coherence interval is almost occupied by the transmission of the pilot symbols, leaving little symbol slots for data transmission, and the estimation accuracy depends on the accuracy of the sample covariance matrix below (22). As the channel coherence interval increases, the proposed approach benefits from the increased interval for data transmission. Additionally, the proposed approach performs better than the approach in [28] and the modified [26].

VI. CONCLUSION

In this paper, a channel estimation approach is proposed for mm-wave multi-cell beamspace MIMO systems. The estimation process is divided into three steps. In the first step, the beams of all the MSs are selected, and the pilots are allocated to distinguish the beams. As the beams indicate the positions of the MSs or scatters, the beams of in-cell MSs are picked out. In the second step, the beamspace channel estimation can achieve close performance as that without beamspace processing. The estimated spatial frequencies are matched with the selected beams to find the channel estimate of each MS. Finally, the path losses are estimated. The ICI processing idea and the beamspace idea are combined to mitigate the ICI efficiently.

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, June 2011.

[2] T. S. Rappaport *et al*., "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] A. L. Swindlehurst *et al*., "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sept. 2014.

[4] R. W. Heath Jr. *et al*., "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] F. Rusek *et al*., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] O. E. Ayach *et al*., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[8] A. Adhikary *et al*., "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[9] J. Wang, Z. Lan, and C. W. Pyo, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 3–4, Oct. 2009.

[10] A. I. Sulyman *et al*., "Radio propagation path loss models for 5G cellular networks in the 28 GHZ and 38 GHZ millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 78–86, Sept. 2014.

[11] W. U. Bajwa *et al*., "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[12] A. Alkhateeb *et al*., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, May 2014.

[13] D. C. Araújo *et al*., "Channel estimation for millimeter-wave very-large MIMO systems," in *Proc. 22nd European Signal Process. Conf.*, 2014.

[14] J. Mo *et al*., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilomar Conf. Signals, Syst. and Computers*, 2014.

[15] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Compressed-sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015.

[16] Y. Peng, Y. Li, and P. Wang, "An enhanced channel estimation method for millimeter wave systems with massive antenna arrays," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1592–1595, Sept. 2015.

[17] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[18] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.
[19] Z. Zhou *et al*., "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.
[20] Y. Han *et al*., "A joint SDMA and interference suppression multiuser transmission scheme for millimeter-wave massive MIMO systems," in *Proc. Sixth Int. Conf. Wireless Commun. and Signal Process.*, 2014.
[21] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.
[22] J. Brady and A. Sayeed, "Beamspace MU-MIMO for high-density gigabit small cell access at millimeter-wave frequencies," in *Proc. IEEE 15th Int. Workshop Signal Process. Advances Wireless Commun.*, 2014.
[23] G. H. Song, J. Brady, and A. Sayeed, "Beamspace MIMO transceivers for low-complexity and near-optimal communication at mm-wave frequencies," in *Proc. 2013 IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2013.
[24] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, June 2015.
[25] L. Yang, Y. Zeng, and R. Zhang, "Efficient channel estimation for millimeter wave MIMO with limited RF chains," in *Proc. 2016 IEEE Int. Conf. commun.*, 2016.
[26] H. Q. Ngo, E. G. Larsson, and T. L. Marzettat, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. of 49th Annual Allerton Conf. Commun., Control, and Computing*, 2011.
[27] J. Hogan and A. Sayeed, "Beam selection for performance-complexity optimization in high-dimensional MIMO systems," in *Proc. 2016 Annu. Conf. Inform. Sci. Syst.*, 2016.
[28] A. Hu, "Estimation of sparse channels in millimeter-wave MU-MIMO systems," *KSII Trans. Internet and Inform. Syst.*, vol. 10, no. 5, pp. 2102–2123, May 2016.
[29] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.
[30] C. A. Balanis, *Antenna Theory: Analysis and Design.* Wiley-Interscience, 2005.
[31] T. S. Rappaport *et al*., "38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications," in *Proc. 2012 IEEE Int. Conf. Commun.*, 2012.
[32] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies for 5G millimeter-wave MIMO systems with large antenna arrays," submitted to IEEE Commun. Mag., available online at arXiv:1607.04559.
[33] X. Gao *et al*., "Near-optimal beam selection for beamspace mmwave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.
[34] J. Chen *et al*., "Enhanced beam selection for multi-user mm-wave massive MIMO systems," *Electron. Lett.*, vol. 52, no. 14, pp. 1268–1270, July 2016.
[35] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, July 1989.
[36] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.
[37] M. R. Akdeniz *et al*., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

Anzhong Hu received the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China. His current research interests include channel estimation and array processing.