

Beam-Domain Full-Duplex Massive MIMO Transmission in the Cellular System



Kui Xu, Xiaochen Xia, Yurong Wang, Wei Xie, and Dongmei Zhang

1 Introduction

The ever-growing challenges for significant traffic growth driven by mobile Internet and Internet of Things have made system capacity enhancement one of the most important features in next-generation wireless communication systems. The general consensus is that the aggregate data rate will increase by roughly 1000 by 2020. Massive multiple-input multiple-output (MIMO), which is first advocated in [1], is identified as one of the key enabling technologies to achieve this goal due to its strong potential in boosting the spectral efficiency (SE) of wireless networks [1, 2]. The term massive MIMO indicates that the base station (BS) employs a number of antennas (typically several tens to hundreds) much larger than the number of active data streams per time-frequency resource. The benefits of massive MIMO are twofold. First, massive MIMO produces a large surplus of degrees of freedom, which can be used to create asymptotically orthogonal channels and deliver near interference-free signals for each user equipment (UE). In this way, the network SE is enhanced significantly because more UEs can be served in parallel and each UE suffers from less interference. On the other hand, the tremendous array gain of the large-scale antenna array also helps to save transmit power and thus potentially improves the energy efficiency.

Massive MIMO was originally designed for time-division duplex (TDD) system [1–7], since by exploiting the channel reciprocity in TDD setting, the required channel state information (CSI) for downlink transmission at the BS can be easily obtained via uplink training [1]. The training overhead scales linearly with the number of user equipments (UEs) and is independent with the number of BS antennas, which is acceptable in most of the typical scenarios. As frequency-division

K. Xu (✉) · X. Xia · Y. Wang · W. Xie · D. Zhang
Army Engineering University of PLA, Nanjing, China

duplex (FDD) dominates the current wireless cellular systems, the application of massive MIMO in FDD system is even more desirable. In FDD massive MIMO, the downlink training and corresponding CSI feedback yield an unacceptably high overhead, which poses a significant bottleneck on the achievable SE. One attempt of practical FDD massive MIMO is called joint spatial division and multiplexing (JSDM) [8], where the correlation between channels is exploited to reduce the training and feedback dimensions. Another scheme that enables FDD massive MIMO is called beam division multiple access (BDMA) [9]. The BDMA gets rid of the need of CSI at transmitter and provides strong potential to realize massive MIMO gain in FDD system. Moreover, other innovative approaches, such as the phase-only beamforming [10] and two-stage beamforming [11], are also promising solutions to the FDD massive MIMO.

In TDD and FDD massive MIMO systems (namely, half-duplex (HD) massive MIMO systems), the uplink and downlink UEs must be allocated with orthogonal time slots or frequency bands, which results in insufficient utilization of time-frequency resources. Inspired by the recent development of full-duplex (FD) communication [17], co-time co-frequency uplink and downlink (CCUD) transmission becomes another option in the cellular system. Although attractive in SE, CCUD transmission is considered challenging due to the strong self-interference (SI) caused by the signal leakage between BS transmitter and receiver, especially when the BS is equipped with large-scale antenna arrays. In the small-scale MIMO system, the SI can be mitigated by the active SI cancellation (SIC) scheme, such as digital/circuit domain SIC and spatial suppression [17]. However, the impractical requirement of instantaneous high-dimension SI channel knowledge makes these technologies difficult when applied in the large-scale antenna system. The passive SIC can be applied in the SI channel-unaware environment, but it fails to provide satisfactory SIC level when used alone [17]. On the other hand, to support the CCUD transmission, the BS employs a separate antenna configuration¹ where two separate large-scale antenna arrays are used for transmission and reception, respectively [18]. In this case, the downlink channel reciprocity is commonly considered as unavailable [19]. Without reciprocity, the training overhead to obtain the downlink CSI scales linearly with the number of BS antennas, which poses another big challenge. In [18], to make the system feasible, the authors assumed that each transmit antenna of BS is also connected with a receive radio-frequency chain so that it can receive the pilot signal. In this case, the downlink reciprocity can still be exploited, however, at cost of additional hardware complexity.

Note that the CCUD transmission in the cellular system with massive MIMO BS has been investigated recently in several works (see [12–16] and the references therein). For example, the authors in [12–14] studied the SE performance of

¹We mention that there exists another choice of shared antenna configuration which uses a single antenna array for transmission and reception. However, under the current technologies, the shared configuration is still difficult in the multi-antenna system due to the significant cross talk between different antennas [20]. Therefore, it will not be considered in this chapter.

CCUD transmission in both macro-cell and small-cell environments. The linear beamforming design of the BS for CCUD transmission has been considered in [15] and [16]. However, most of these works are based on the assumption that the SI has been suppressed to a reasonable level and the uplink/downlink channel can be efficiently obtained. As a result, the aforementioned challenges are still not fully addressed.

In this chapter, we investigate the feasibility of CCUD transmission in the cellular system with massive MIMO BS. The contributions are summarized as follows.

1. By exploiting the beam-domain representation of channels based on the basis expansion model [23], we prove that massive MIMO channel matrix (vector) can be represented by a low-dimension effective beam-domain channel matrix (vector). Based on this property, we propose a beam-domain full-duplex (BDFD) massive MIMO scheme (BDFD scheme for short) to enable CCUD transmission in the cellular system. We show that the BDFD scheme achieves significant saving in uplink/downlink training and achieves the uplink and downlink sum capacities simultaneously as the number of BS antennas approaches to infinity.
2. Then, we investigate several important components for the practical implementation of BDFD scheme in the cellular system, including UEs grouping, effective beam-domain channel estimation, beam-domain data transmission, and interference control between uplink and downlink.
3. Finally, we examine the SE of BDFD scheme using the third-generation partnership project long-term evolution (3GPP LTE) simulation model for macro-cell environment. The results demonstrate the superiority of BDFD scheme over the TDD/FDD massive MIMO.

The rest of the paper is organized as follows. The system and channel models are described in Sects. 2, 3, and 4 considering the basic ideal and practical implementation of BDFD scheme, respectively. Section 5 presents the simulation results. Section 6 draws the conclusions.

Notation $\mathbb{E}(\cdot)$ denotes the expectation. $\delta(\cdot)$ denotes the Dirac delta function. $\mathbf{A}^{\{B_1, B_2\}}$ denotes the submatrix of \mathbf{A} by keeping its rows indexed by set B_1 and columns indexed by set B_2 . $\mathbf{A}^{\{B, \cdot\}}$ ($\mathbf{A}^{\{\cdot, B\}}$) denotes the submatrix of \mathbf{A} by keeping its rows (columns) indexed by set B . $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $|\cdot|$, $\|\cdot\|$, and $\text{tr}(\cdot)$ denote transpose, conjugate, conjugate-transpose, determinant, Frobenius norm, and trace of a matrix, respectively. $\mathbf{A} \succeq 0$ means that \mathbf{A} is Hermitian positive semi-definite matrix. The frequently used symbols in this paper are summarized in Table 1.

2 System and Channel Models

Consider a single-cell system with a FD BS, a number of uplink UEs, and a number of downlink UEs as shown in Fig. 1a. We assume that all UEs are HD and have single antenna. To support the CCUD transmission, the BS employs two separate large-scale antenna arrays for transmission and reception, respectively. The uniform

Table 1 Summary of frequently used symbols

Symbols	Descriptions
\mathbf{h}_{k_u} (\mathbf{h}_{k_d})	The channel vector between uplink UE k_u (downlink UE k_d) and BS
\mathbf{H}_{SI}	The SI channel matrix from transmit antenna array to receive antenna array of the BS
\mathbf{H}_{g_u} (\mathbf{H}_{g_d})	The channel matrix between uplink group g_u (downlink group g_d) and BS
N	The number of transmit/receive antennas of BS
M_u, M_d, M_{SI}	The numbers of scattering clusters for uplink, downlink, and SI channels
$[\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]$	The DOA region of uplink signal from UE k_u resulting from the i th scattering cluster
$[\theta_{k_d,i}^{\min}, \theta_{k_d,i}^{\max}]$	The DOD region of downlink signal to UE k_d resulting from the i th scattering cluster
$S_{\omega,i}(\cdot), \omega \in \{k_u, k_d, SI\}$	The product of the large-scale fading and channel power angle spectrum
$g_{u,k}$ ($g_{d,k}$)	The k th UE in the uplink group g_u (downlink group g_d)
B_{k_u} (B_{k_d})	The active beam set of uplink group UE k_u (downlink group UE k_d)
B_{g_u} (B_{g_d})	The active beam set of uplink group g_u (downlink group g_d)
$B_{SI,R}, B_{SI,T}$	The active beam sets of SI channel
G_u (G_d)	The set of all uplink (downlink) group
K_{g_u} (K_{g_d})	The number of UEs in uplink group g_u (downlink group g_d)
b_u (b_d)	The number beams in uplink (downlink) UE group
σ	The variance of AWGN

linear arrays are assumed. In the practical implementation, the transmit and receive antenna arrays of the BS can be deployed on the opposite sides of a building with distance of tens of meters to reduce the SI.

We use $\mathbf{h}_{k_u} \in \mathbb{C}^{N \times 1}$ to denote the channel vector from the uplink UE k_u to the receive antenna array of BS and use $\mathbf{h}_{k_d} \in \mathbb{C}^{N \times 1}$ to denote the channel vector from the transmit antenna array of BS to downlink UE k_d , where N denotes the number of transmit/receive antennas at the BS.² We use $\mathbf{H}_{SI} \in \mathbb{C}^{N \times N}$ to denote the SI channel matrix from transmit antenna array to receive antenna array of the BS.

We consider the general cluster-based channel model [21] where the received signal at the BS from the uplink UE k_u is a sum of the contributions from M_u scattering clusters. The direction of arrival (DOA) of signals resulting from the i th cluster is within the region $[\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]$. Thus, the channel vector between the uplink UE k_u and the BS can be expressed as [21]

$$\mathbf{h}_{k_u} = \sum_{i=1}^{M_u} \int_{\theta_{k_u,i}^{\min}}^{\theta_{k_u,i}^{\max}} \mathbf{a}(\theta) r_{k_u,i}(\theta) d\theta \quad (1)$$

²To simplify the notation, we assume the symmetric antenna deployment at the BS. Extension to the situation with different numbers of transmit and receive antennas is straightforward.

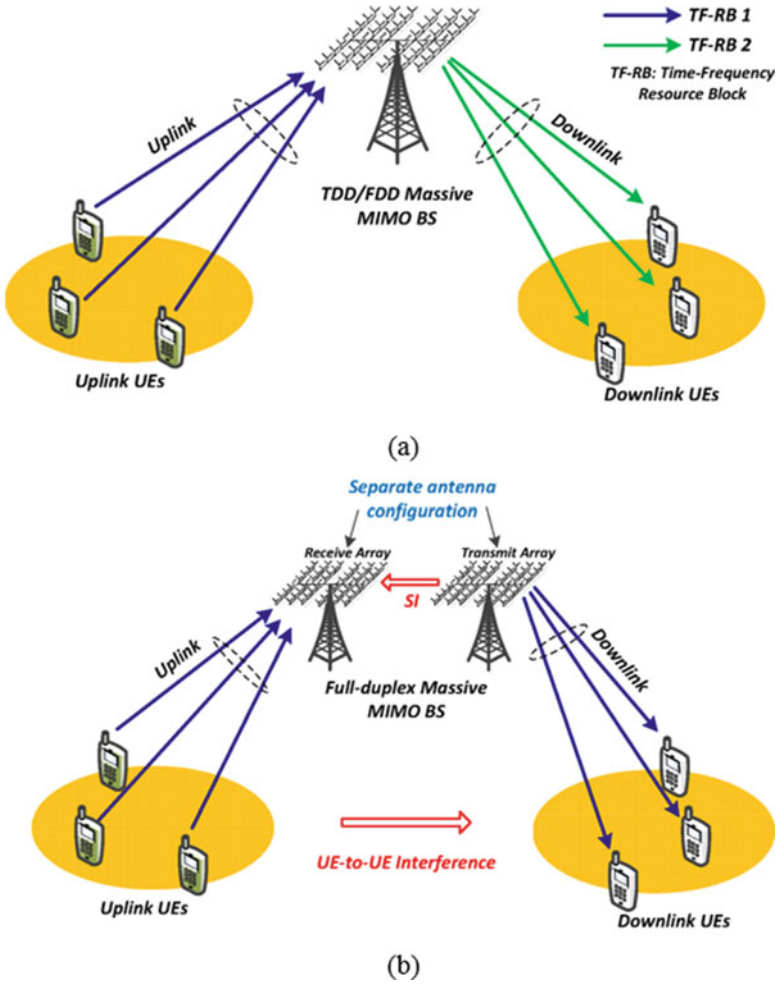


Fig. 1 TDD/FDD massive MIMO and FD massive MIMO systems. (a) TDD/FDD massive MIMO. (b) Full-duplex massive MIMO

where $\mathbf{a}(\theta) = [1, \exp(j2\pi d \sin(\theta)/\lambda), \dots, \exp(j2\pi d(N-1) \sin(\theta)/\lambda)]^T$ is the array response vector with d and λ denoting the antenna spacing and carrier wavelength, respectively. $r_{k_u,i}(\theta)$ denotes the complex-valued response gain. In the above model, the DOA regions of signals from different scattering clusters are disjoint (otherwise, these signals should be considered from the same scattering cluster). Therefore, the number of scattering clusters is finite because $\sum_{i=1}^{M_u} (\theta_{k_u,i}^{\max} - \theta_{k_u,i}^{\min}) \leq 2\pi$.

Similarly, let $[\theta_{k_d,i}^{\min}, \theta_{k_d,i}^{\max}]$ be the direction of departure (DOD) region of signals resulting from the i th scattering clusters and let $r_{k_d,i}(\theta)$ denote the associated complex-valued response gain; the channel vector from the BS to downlink UE k_d can be written as

$$\mathbf{h}_{k_d} = \sum_{i=1}^{M_u} \int_{\theta_{k_d,i}^{\min}}^{\theta_{k_d,i}^{\max}} \mathbf{a}(\theta) r_{k_d,i}(\theta) d\theta \quad (2)$$

The SI signal can be viewed as the contributions of signals from M_{SI} scattering clusters with different DOA and DOD regions. Thus, the SI channel matrix \mathbf{H}_{SI} can be expressed as

$$\mathbf{H}_{SI} = \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} r_{SI,i}(\theta_R, \theta_T) \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) d\theta_R d\theta_T \quad (3)$$

where $r_{SI,i}(\theta_R, \theta_T)$ denotes the complex-valued response gain. In real systems, the BS is commonly elevated at a relatively high altitude, e.g., on the top of a high building or a dedicated tower, so that there are few surrounding scatterers [24]. Moreover, we assume that the passive SIC scheme for infrastructure nodes in [25] has been used, and the direct path between transmit and receive antenna arrays of BS is virtually cancelled. Therefore, in this chapter, we assume that the number of scattering clusters for SI channel is small.

In (1), (2), and (3), the complex-valued response gains with different incidence angles are uncorrelated [22], that is,

$$\mathbb{E}[r_{k_u,i}(\theta) r_{k_u,i}^*(\theta')] = S_{k_u,i}(\theta) \delta(\theta - \theta')$$

$$\mathbb{E}[r_{k_d,i}(\theta) r_{k_d,i}^*(\theta')] = S_{k_d,i}(\theta) \delta(\theta - \theta')$$

$$\mathbb{E}[r_{SI,i}(\theta_R, \theta_T) r_{SI,i}^*(\theta'_R, \theta'_T)] = S_{SI,i}(\theta_R, \theta_T) \delta(\theta_R - \theta'_R) \delta(\theta_T - \theta'_T) \quad (4)$$

where $S_{\omega,i}(\cdot), \omega \in \{k_u, k_d, SI\}$ represents the product of the large-scale fading and channel power angle spectrum. Note that the considered model can be easily transformed into several well-known massive MIMO channel models. For example, by setting $M_u = M_d = 1$, we obtain the ‘‘one-ring’’ model studied in [8]. The ‘‘one-ring’’ model is typically used in the macro-cell environment where the uplink/downlink received signals are resulted from the scattering process in the vicinity of the UEs [21]. Moreover, by setting

$$r_{k_u,i}(\theta) = \sum_j r_{k_u,i,j} \delta(\theta - \theta_j)$$

$$r_{k_u,i}(\theta) = \sum_j r_{k_u,i,j} \delta(\theta - \theta_j)$$

$$r_{SI,i}(\theta_R, \theta_T) = \sum_{j,l} r_{k_{SI},i,j,l} \delta(\theta_R - \theta_{R,j}) \delta(\theta_T - \theta_{T,j}) \quad (5)$$

we arrive at the ray-cluster-based spatial channel model which is usually used for millimeter wave MIMO systems [26]. Therefore, the results in this chapter can be readily applied in these scenarios.

3 Beam-Domain Full-Duplex Transmission Scheme

In this section, we propose a BDFD scheme to realize CCUD transmission in the cellular system. Using the basis expansion model, we first derive the beam-domain channel representation which is the projection of channel vector (matrix) on a common basis. The benefit of the beam-domain representation is that the channel becomes compressible in the beam domain under certain basis. Using this property, channel dimension required to be estimated can be greatly reduced. Moreover, by exploiting the structure of SI channel in the beam domain, it is possible to eliminate the SI without using the instantaneous SI channel knowledge and hence realize efficient CCUD transmission.

3.1 Beam-Domain Channel Representation

Under the basis expansion model [23], the uplink channel vector can be expanded from a set of uniform basis vectors $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N\} \in \mathbb{C}^N \times 1$, that is

$$\mathbf{h}_{k_u} = \sum_{m=1}^N \tilde{h}_{k_u,m} \mathbf{f}_m = \mathbf{F} \tilde{\mathbf{h}}_{k_u} \quad (6)$$

where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N]$. Following [9], the basis vector \mathbf{f}_i is also called a beam, and $\tilde{\mathbf{h}}_{k_u} = [\tilde{h}_{k_u,1}, \tilde{h}_{k_u,2}, \dots, \tilde{h}_{k_u,N}]^T$ is called the beam-domain channel.³ According to (1) and (6), we have

³Note that the idea of beam-domain channel was also studied in [9] for FDD massive MIMO system. However, we investigate the beam-domain properties of a more general channel model and present new capacity achieving scheme using these properties.

$$\tilde{\mathbf{h}}_{k_u} = \mathbf{F}^H \mathbf{h}_{k_u} = \sum_{i=1}^{M_u} \int_{\theta_{k_u,i}^{\min}}^{\theta_{k_u,i}^{\max}} \mathbf{F}^H \mathbf{a}(\theta) r_{k_u,i}(\theta) d\theta \quad (7)$$

To investigate the compressibility of beam-domain channel, we propose the following lemma.

Lemma 1 Consider the basis $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2 \cdots, \mathbf{f}_N]$ with $\mathbf{f}_n = \frac{1}{\sqrt{N}} [1, \exp(j2\pi d\theta_n/\lambda), \cdots, \exp(j2\pi d(N-1)\theta_n/\lambda)]^T$. θ_n is defined as the beam angle of n th beam \mathbf{f}_n , which is selected so that the different beams are orthogonal. As the number of BS antennas N tends to infinity, the average beam-domain channel gain for the uplink UE k_u associated with the n th beam \mathbf{f}_n , i.e., $\mathbb{E} \left[\left| \tilde{h}_{k_u,n} \right|^2 \right]$, has non-negligible value only when $\theta_n \in \cup_{i=1}^{M_u} \left[\sin \theta_{k_u,i}^{\min} - \varepsilon, \sin \theta_{k_u,i}^{\max} + \varepsilon \right]$, where $\varepsilon \geq 0$ and $\lim_{N \rightarrow \infty} \varepsilon = 0$.

Proof Using (7) and [28, Eq. (5)], $\mathbb{E} \left[\left| \tilde{h}_{k_u,n} \right|^2 \right]$ can be written as

$$\begin{aligned} \mathbb{E} \left[\left| \tilde{h}_{k_u,n} \right|^2 \right] &= \sum_{i=1}^{M_u} \int_{\theta_{k_u,i}^{\min}}^{\theta_{k_u,i}^{\max}} \left| \mathbf{f}_n^H \mathbf{a}(\theta) \right|^2 S_{k_u,i}(\theta) d\theta \\ &= \underbrace{\sum_{i=1}^{M_u} N \int_{\theta_{k_u,i}^{\min}}^{\theta_{k_u,i}^{\max}} \text{asinc}_N^2 \left(\frac{d}{\lambda} \theta_n - \frac{d}{\lambda} \sin \theta \right) S_{k_u,i}(\theta) d\theta}_{Y_i} \end{aligned} \quad (8)$$

where $\text{asinc}_N(x)$ is the aliased sinc function, which is defined as $\text{asinc}_N(x) = \sin(N\pi x)/(N \sin(\pi x))$. The envelope of the squared aliased sinc function is shown in Fig. 2. Assuming $\theta_n = \sin \theta_{k_u,i}^{\max} + \varepsilon$ with $\varepsilon > 0$ and $i \in \{1, 2, \cdots, M_u\}$, the i th term in the summation of (8), named as Y_i , can be upper bounded as

$$\begin{aligned} Y_i &\leq N \max_{\theta \in [\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]} S_{k_u,i}(\theta) \int_{\theta_{k_u,i}^{\min}}^{\theta_{k_u,i}^{\max}} \text{asinc}_N^2 \left(\frac{d}{\lambda} \theta_n - \frac{d}{\lambda} \sin \theta \right) d\theta \\ &\leq N \max_{\theta \in [\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]} S_{k_u,i}(\theta) \sum_{l=L_i^{\min}}^{L_i^{\max}} \int_{\vartheta_{l+1}}^{\vartheta_l} \text{asinc}_N^2 \left(\frac{d}{\lambda} \theta_n - \frac{d}{\lambda} \sin \theta \right) d\theta \\ &\leq N (L_i^{\max} - L_i^{\min}) \max_{\theta \in [\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]} S_{k_u,i}(\theta) \int_{\vartheta_{L_i^{\min}+1}}^{\vartheta_{L_i^{\min}}} \text{asinc}_N^2 \left(\frac{d}{\lambda} \theta_n - \frac{d}{\lambda} \sin \theta \right) d\theta \\ &\leq \frac{(L_i^{\max} - L_i^{\min}) (\vartheta_{L_i^{\min}} - \vartheta_{L_i^{\min}+1})}{N(\pi L_i^{\min})^2} \max_{\theta \in [\theta_{k_u,i}^{\min}, \theta_{k_u,i}^{\max}]} S_{k_u,i}(\theta) \end{aligned} \quad (9)$$

where $L_i^{\min} = \left\lceil Nd \left(\theta_n - \sin \theta_{k_u,i}^{\max} \right) / \lambda \right\rceil$ and $L_i^{\max} = \left\lfloor Nd \left(\theta_n - \sin \theta_{k_u,i}^{\min} \right) / \lambda \right\rfloor$. $\vartheta_l = \arcsin(\theta_n - \lambda l / dN)$ is the l th zero point of the function $\text{asinc}_N^2(d(\theta_n - \sin \theta)/\lambda)$.

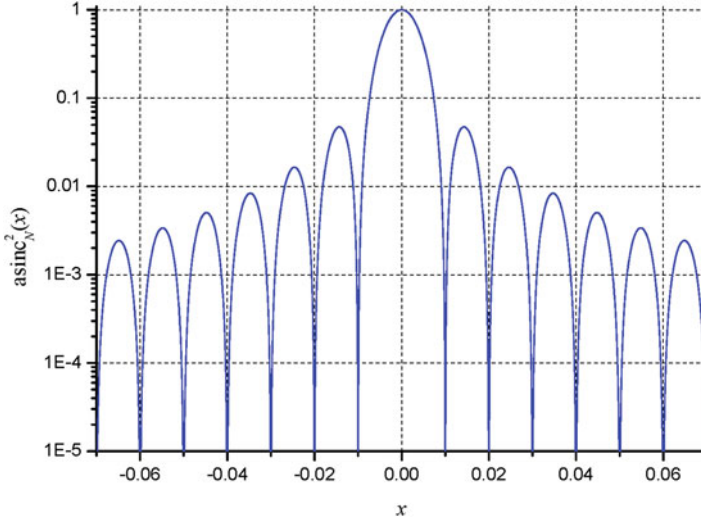


Fig. 2 Envelope of the squared aliased sinc function, where $N = 100$

The l th term in the summation of the second step is the integral over the l th side lobe of the squared aliased sinc function as shown in Fig. 2. The third step is based on the fact that the power of the side lobe of aliased sinc function is a decreasing function of its index. The last step is obtained by using the property that the aliased sinc function converges to the standard sinc function as $N \rightarrow \infty$ [29, Ch. 3] and using the fact $\sin x \leq 1$.

Note that $L_i^{\max} - L_i^{\min}$ scales with $\mathcal{O}(N)$. Moreover, as $N \rightarrow \infty$, $\vartheta_{L_i^{\min}} - \vartheta_{L_i^{\min}+1}$ can be replaced by the differential of the arcsinx at point $x = \theta_n - \lambda L_i^{\min}/dN$, i.e.,

$$\begin{aligned} \vartheta_{L_i^{\min}} - \vartheta_{L_i^{\min}+1} &= \frac{\lambda}{dN} \left. \frac{d(\arcsin x)}{dx} \right|_{x=\theta_n - \frac{\lambda L_i^{\min}}{dN}} \\ &= \frac{\lambda}{dN} \left(1 - \left(\theta_n - \frac{\lambda L_i^{\min}}{dN} \right)^2 \right)^{-1/2} = \mathcal{O}\left(\frac{1}{N}\right) \end{aligned} \quad (10)$$

Recall that we have assumed $\theta_n = \sin \theta_{k_u, i}^{\max} + \varepsilon$. Thus, we have $L_i^{\min} = \lfloor dN\varepsilon/\lambda \rfloor$. Based on the results in the above, the upper bound of Y_i given by the last step of (9) converges to zero when $N \rightarrow \infty$ as long as $\varepsilon \geq \mathcal{O}(N^u)$ with $u > -3/2$. If we can choose $\varepsilon = \mathcal{O}(N^{-1})$, then ε approaches to 0 if $N \rightarrow \infty$. In the same way, we can obtain the similar result if $\theta_n = \sin \theta_{k_u, i}^{\max} - \varepsilon$. Therefore, as $N \rightarrow \infty$, Y_i ($\forall i \in \{1, 2, \dots, M_u\}$) has non-negligible value only if $\theta_n \in \left[\sin \theta_{k_u, i}^{\min} - \varepsilon, \sin \theta_{k_u, i}^{\max} + \varepsilon \right]$, where $\lim_{N \rightarrow \infty} \varepsilon = 0$. This completes the proof.

Note that the upper bound of Y_i in (9) is generally not tight. However, this does not impact the analysis because $Y_i \xrightarrow{N \rightarrow \infty} 0$ as long as its upper bound converges to 0.

If we consider $\sin\theta$ as the virtual DOA of the uplink signal, in Lemma 1 we actually select the beam angle to mimic virtual DOA. That is why only the beam-domain channel elements with beam angles within $\cup_{i=1}^{M_u} \left[\sin \theta_{k_u,i}^{\min} - \varepsilon, \sin \theta_{k_u,i}^{\max} + \varepsilon \right]$ have non-negligible gains. From Lemma 1, the beam-domain channel vector exhibits the desired compressibility in the large N regime when the considered basis is used. Therefore, the basis in Lemma 1 will be employed in the following.

Example 1 As a concrete example of the compressibility, we consider a scenario with $N = 128$, $M_u = 1$, and $\left[\theta_{k_u,1}^{\min}, \theta_{k_u,1}^{\max} \right] = [24.3^\circ, 35.7^\circ]$. This corresponds to “one-ring” model with about 30 m scattering radius and 300 m BS-to-UE distance [21]. The normalized average beam-domain channel gain, which is defined as $\mathbb{E} \left[\left| \tilde{h}_{k_u,n} \right|^2 \right] / \max_{n'=1,2,\dots,N} \mathbb{E} \left[\left| \tilde{h}_{k_u,n'} \right|^2 \right]$, is plotted in Fig. 3. The beam-domain channel elements whose beam angles are within $\left[\sin \theta_{k_u,1}^{\min}, \sin \theta_{k_u,1}^{\max} \right]$ are marked in red. From the figure, we can see that the gains of these channel elements are much higher than the remaining, which matches well with the results in Lemma 1. In fact, from the simulation results, about 96.6% of channel power is captured by these channel elements (less than 10% of the all elements) when $N = 128$. This value becomes 96.9% and 98.1% if we increase N to 256 and 512, respectively.

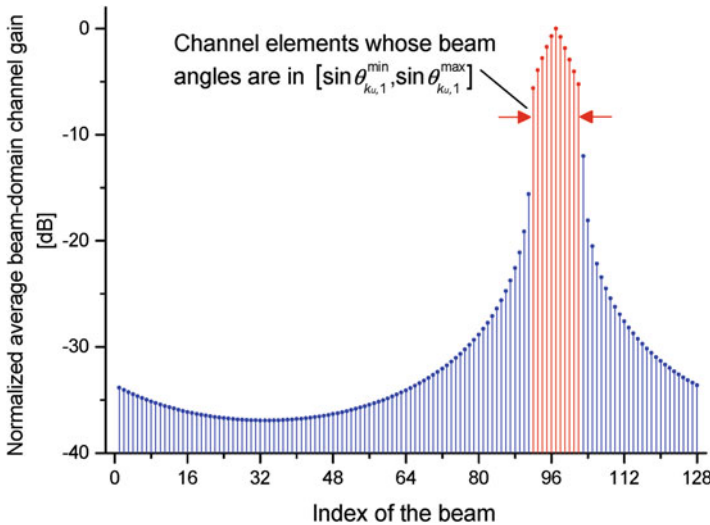


Fig. 3 Normalized average beam-domain channel gain as a function of index of the associated beam, where $\theta_n = \frac{\lambda}{d} \left(\frac{n}{N} - \frac{1}{2} \right)$, and $d = \lambda/2$

Based on Lemma 1, we can approximate the channel vector from uplink UE k_u to BS as

$$\mathbf{h}_{k_u} \approx \sum_{m \in B_{k_u}} \tilde{h}_{k_u, m} \mathbf{f}_m = \mathbf{F}^{\{:, B_{k_u}\}} \tilde{\mathbf{h}}_{k_u}^{\{B_{k_u}, :\}} \quad (11)$$

where B_{k_u} is called the active beam set which contains the indexes of beams with non-negligible beam-domain channel gains. $\mathbf{F}^{\{:, B_{k_u}\}}$ is called the active beam space, whose columns are consisted of the beams in B_{k_u} . The reduced-dimension beam-domain channel vector $\tilde{\mathbf{h}}_{k_u}^{\{B_{k_u}, :\}} \in \mathbf{C}^{|B_{k_u}| \times 1}$ is called the effective beam-domain channel. Note that (11) holds with equality as $N \rightarrow \infty$ according to Lemma 1. Based on (1) and (11), the effective beam-domain channel vector can be expressed as

$$\tilde{\mathbf{h}}_{k_u}^{\{B_{k_u}, :\}} = \left(\mathbf{F}^{\{:, B_{k_u}\}} \right)^H \mathbf{h}_{k_u} = \sum_{i=1}^{M_u} \int_{\theta_{k_u, i}^{\min}}^{\theta_{k_u, i}^{\max}} \left(\mathbf{F}^{\{:, B_{k_u}\}} \right)^H \mathbf{a}(\theta) r_{k_u, i}(\theta) d\theta \quad (12)$$

From (11), the original channel vector can be recovered from the effective beam-domain channel vector if the DOA (and hence the active beam set) information is known. As a result, in order to obtain \mathbf{h}_{k_u} , it is enough to estimate the $|B_{k_u}|$ -dimension effective beam-domain channel during the training phase. This can potentially result in significant saving of the training resource.

In the practical scenario with arbitrary finite number of BS antennas, the active beam set B_{k_u} can be obtained by solving a cardinality minimization problem, with constraint that most of the channel power is captured by the effective beam-domain channel vector, i.e.,

$$\min_{B_{k_u}} |B_{k_u}| \quad \text{s.t.} \quad \begin{cases} \frac{\mathbb{E} \left[\left\| \tilde{\mathbf{h}}_{k_u}^{\{B_{k_u}, :\}} \right\|^2 \right]}{\mathbb{E} \left[\left\| \tilde{\mathbf{h}}_{k_u} \right\|^2 \right]} = \frac{\sum_{i=1}^{M_u} \int_{\theta_{k_u, i}^{\min}}^{\theta_{k_u, i}^{\max}} \left\| \left(\mathbf{F}^{\{:, B_{k_u}\}} \right)^H \mathbf{a}(\theta) \right\|^2 S_{k_u, i}(\theta) d\theta}{\sum_{i=1}^{M_u} \int_{\theta_{k_u, i}^{\min}}^{\theta_{k_u, i}^{\max}} \left\| \mathbf{F}^H \mathbf{a}(\theta) \right\|^2 S_{k_u, i}(\theta) d\theta} \geq \eta \\ \eta < 1 \end{cases} \quad (13)$$

where $\eta < 1$ denotes the threshold and should be chosen closed to 1 in real implementation. The problem is combinatorial and difficult to solve in closed form. However, since (13) is related only with DOA information which is slow time-varying,⁴ we can build off-line table of B_{k_u} for different DOA regions to reduce the computation load. If N is large enough, to further reduce the complexity, we can

⁴Since the DOA/DOD information is slow time-varying, we assume that these parameters can be obtained perfectly at the BS through the long-term estimation [27].

simply select the beams whose angles are in $\theta_m \in \cup_{i=1}^{M_{SI}} \left[\sin \theta_{R,i}^{\min} - \varepsilon, \sin \theta_{R,i}^{\max} + \varepsilon \right]$ to constitute B_{k_d} . According to Lemma 1 (also demonstrated by in Example 1), the beam-domain channel elements associated with these beams contain almost all the channel power in the large N regime.

Similarly, with the basis expansion model, the channel vector from the BS to downlink UE k_d and the associated effective beam-domain channel vector can be expressed as

$$\begin{aligned} \tilde{\mathbf{H}}_{SI}^{\{B_{SI,R}, B_{SI,T}\}} &= \left(\mathbf{F}^{\{:, B_{SI,R}\}} \right)^H \mathbf{H}_{SI} \mathbf{F}^{\{:, B_{SI,T}\}} \\ &= \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} r_{SI,i}(\theta_R, \theta_T) \left(\mathbf{F}^{\{:, B_{SI,R}\}} \right)^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \mathbf{F}^{\{:, B_{SI,T}\}} d\theta_R d\theta_T \end{aligned} \quad (14)$$

where the active beam set B_{k_d} can be design by solving a similar problem as that in (13).

To exploit the compressibility of SI channel, the basis expansion is performed for column and row spaces of \mathbf{H}_{SI} simultaneously, which results in $\mathbf{H}_{SI} = \mathbf{F} \tilde{\mathbf{H}}_{SI} \mathbf{F}^H$. As a generalization of Lemma 1, we have the following Lemma.

Lemma 2 The average beam-domain SI channel gain $\mathbb{E} \left[\left| \left[\tilde{\mathbf{H}}_{SI} \right]_{m,n} \right|^2 \right] = \mathbb{E} \left[\left| \mathbf{f}_m^H \mathbf{H}_{SI} \mathbf{f}_n \right|^2 \right]$ has non-negligible value only when the beam angle of \mathbf{f}_m lies in $\cup_{i=1}^{M_{SI}} \left[\sin \theta_{R,i}^{\min} - \varepsilon, \sin \theta_{R,i}^{\max} + \varepsilon \right]$, and meanwhile the beam angle of \mathbf{f}_n lies in $\cup_{i=1}^{M_{SI}} \left[\sin \theta_{T,i}^{\min} - \varepsilon, \sin \theta_{T,i}^{\max} + \varepsilon \right]$, where ε approaches to zero in the large N regime.

Lemma 2 can be simply proved by using a similar procedure at that in the proof of Lemma 1. Therefore, the detailed proof is omitted due to space limitation. From Lemma 2, the SI channel matrix can be approximated as $\mathbf{H}_{SI} \approx \mathbf{F}^{\{:, B_{SI,R}\}} \tilde{\mathbf{H}}_{SI}^{\{B_{SI,R}, B_{SI,T}\}} \left(\mathbf{F}^{\{:, B_{SI,T}\}} \right)^H$, where the effective beam-domain SI channel matrix $\tilde{\mathbf{H}}_{SI}^{\{B_{SI,R}, B_{SI,T}\}} \in \mathbf{C}^{|B_{SI,R}| \times |B_{SI,T}|}$ can be expressed as

$$\begin{aligned} \tilde{\mathbf{H}}_{SI}^{\{B_{SI,R}, B_{SI,T}\}} &= \left(\mathbf{F}^{\{:, B_{SI,R}\}} \right)^H \mathbf{H}_{SI} \mathbf{F}^{\{:, B_{SI,T}\}} \\ &= \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} r_{SI,i}(\theta_R, \theta_T) \left(\mathbf{F}^{\{:, B_{SI,R}\}} \right)^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \mathbf{F}^{\{:, B_{SI,T}\}} d\theta_R d\theta_T \end{aligned} \quad (15)$$

The active beam sets $B_{SI,R}$ and $B_{SI,T}$ can be determined by solving the problem (16) on the top of the next page, where $\tilde{\mathbf{H}}_{SI}$ is defined as $\tilde{\mathbf{H}}_{SI} = \mathbf{F}^H \mathbf{H}_{SI} \mathbf{F}$ and $\eta < 1$ denotes the threshold in (16), shown at the bottom of the page.

$$\begin{aligned}
& \min_{B_{SI,R}, B_{SI,T}} \max \{ |B_{SI,R}|, |B_{SI,T}| \} \text{ s.t. } \frac{\mathbb{E} \left[\left\| \tilde{\mathbf{H}}_{SI}^{\{B_{SI,R}, B_{SI,T}\}} \right\|^2 \right]}{\mathbb{E} \left[\left\| \tilde{\mathbf{H}}_{SI} \right\|^2 \right]} \\
& = \frac{\sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} S_{SI,i}(\theta_R, \theta_T) \left\| \left(\mathbf{F}^{\{B_{SI,R}\}} \right)^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \mathbf{F}^{\{B_{SI,T}\}} \right\|^2 d\theta_R d\theta_T}{\sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} S_{SI,i}(\theta_R, \theta_T) \left\| \mathbf{F}^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \mathbf{F} \right\|^2 d\theta_R d\theta_T} \geq \eta
\end{aligned} \tag{16}$$

3.2 Beam-Domain Full-Duplex Transmission

The key idea of the BDFD scheme lies in partitioning UEs according to their active beam sets to realize efficient CCUD transmission. In particular, we divide the UEs into groups according the following two criteria (UE grouping criteria):

1. Criterion 1: The uplink/downlink UEs with the same active beam set are collected in the same group. The active beam sets of different uplink/downlink groups are non-overlapping. Mathematically, letting B_{g_u} and $B_{g'_u}$ be the active beam sets of two arbitrary uplink groups g_u and g'_u , and letting B_{g_d} and $B_{g'_d}$ be the active beam sets of two arbitrary downlink groups g_d and g'_d , we have $B_{g_u} \cap B_{g'_u} = \emptyset$ and $B_{g_d} \cap B_{g'_d} = \emptyset$.
2. Criterion 2: Let G_u and G_d be the sets of uplink UE groups and downlink UE groups, respectively. The active beam sets B_{g_u} and B_{g_d} satisfy $(\cup_{g_u \in G_u} B_{g_u}) \cap B_{SI,R} = \emptyset$ or $(\cup_{g_d \in G_d} B_{g_d}) \cap B_{SI,T} = \emptyset$.

For the sake of illustration, in the following we assume that each active beam set contains the same number of beams, i.e., $|B_{g_u}| = b_u$ and $|B_{g_d}| = b_d$. Define the index $g_{u,k} = k + \sum_{g'_u=1}^{g_u-1} K_{g'_u}$ to denote the k th uplink UE of the uplink group g_u , where K_{g_u} is the number of UEs in the group g_u . Similarly, letting K_{g_d} be the number of UEs in the downlink group g_d , we can define the $g_{d,k} = k + \sum_{g'_d=1}^{g_d-1} K_{g'_d}$ to denote the k th downlink UE of the group g_d . Define $\mathbf{H}_{g_u} = [\mathbf{h}_{g_{u,1}}, \dots, \mathbf{h}_{g_{u,K_{g_u}}}]$ and $\mathbf{H}_{g_d} = [\mathbf{h}_{g_{d,1}}, \dots, \mathbf{h}_{g_{d,K_{g_d}}}]$ as the channel matrix from the uplink group g_u to the BS and that from the BS to the downlink group g_d , respectively, and define $\tilde{\mathbf{H}}_{g_u}$ and $\tilde{\mathbf{H}}_{g_d}$ as the corresponding beam-domain channel matrices.

During the data transmission phase, the uplink UEs transmit data to the BS, and meanwhile, the BS transmits data to the downlink UEs. Assuming the above UE grouping criteria, the received signals at the BS and downlink group g_d can be expressed as

$$\mathbf{y}_u = \sum_{g'_u \in G_u} \mathbf{H}_{g'_u} \mathbf{s}_{g'_u} + \mathbf{H}_{SI} \sum_{g'_d \in G_d} \mathbf{x}_{g'_d} + \mathbf{n}_u \tag{17}$$

$$\mathbf{y}_{g_d} = \mathbf{H}_{g_d}^H \sum_{g'_d \in G_d} \mathbf{x}_{g'_d} + \sum_{g'_u \in G_u} \mathbf{H}_{g'_u \rightarrow g_d} \mathbf{s}_{g'_u} + \mathbf{n}_{g_d} \quad (18)$$

where \mathbf{n}_u and \mathbf{n}_{g_d} denote the additive white Gaussian noises (AWGNs) with variance σ . $\mathbf{H}_{g'_u \rightarrow g_d}$ denotes the interference channel from uplink group g_u to downlink group g_d . Since the UEs have single antenna and are geographically distributed, the elements of $\mathbf{H}_{g'_u \rightarrow g_d}$ are assumed to be independent Gaussian random variables with zero mean. $\mathbf{s}_{g_u} \in \mathbb{C}^{K_{g_u} \times 1}$ denotes the transmit signal of uplink group g_u . $\mathbf{x}_{g_d} \in \mathbb{C}^{N \times 1}$ denotes the precoded transmit signal of the BS. In the BDFD scheme, the UEs of the downlink group g_d detect signal only on their active beam space. Therefore, we let $\mathbf{x}_{g_d} = \mathbf{F}^{\{:\cdot, B_{g_d}\}} \tilde{\mathbf{x}}_{g_d}$, where $\tilde{\mathbf{x}}_{g_d} \in \mathbb{C}^{b_d \times 1}$ is referred to as the beam-domain precoded transmit signal.

By multiplying both sides of (17) with $\left(\mathbf{F}^{\{:\cdot, B_{g_u}\}}\right)^H$ and using the definition $\mathbf{x}_{g_d} = \mathbf{F}^{\{:\cdot, B_{g_d}\}} \tilde{\mathbf{x}}_{g_d}$ on (18), we arrive at the beam-domain received signal at the BS from uplink UE group g_u and the beam-domain received signal at downlink group g_d

$$\tilde{\mathbf{y}}_{g_u} = \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}, \cdot\}} \mathbf{s}_{g_u} + \sum_{g'_u \in G_u / \{g_u\}} \tilde{\mathbf{H}}_{g'_u}^{\{B_{g_u}, \cdot\}} \mathbf{s}_{g'_u} + \sum_{g'_d \in G_d} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u}, B_{g'_d}\}} \tilde{\mathbf{x}}_{g'_d} + \tilde{\mathbf{n}}_{g_u} \quad (19)$$

$$\tilde{\mathbf{y}}_{g_d} = \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, \cdot\}}\right)^H \tilde{\mathbf{x}}_{g_d} + \sum_{g'_d \in G_d / \{g_d\}} \left(\tilde{\mathbf{H}}_{g'_d}^{\{B_{g_d}, \cdot\}}\right)^H \tilde{\mathbf{x}}_{g'_d} + \sum_{g'_u \in G_u} \mathbf{H}_{g'_u \rightarrow g_d} \mathbf{s}_{g'_u} + \mathbf{n}_{g_d} \quad (20)$$

where $\tilde{\mathbf{n}}_{g_u} = \left(\mathbf{F}^{\{:\cdot, B_{g_u}\}}\right)^H \mathbf{n}_u$. Note that (20) is in fact the same with (18). To emphasize that (20) is the beam-domain received signal and be consistent with (19), we introduce the new notation $\tilde{\mathbf{y}}_{g_d}$. The second terms on the right-hand sides (RHSs) of (19) and (20) indicate the inter-group interferences (IGIs). The third term of RHS of (19) denotes the received SI. From (19), b_u should satisfy $b_u \geq K_{g_u}$ in order to support K_{g_u} independent data streams. Similarly, we require $b_d \geq K_{g_d}$ in downlink according to (20). As will be shown in Sect. 4.4, with simple UE scheduling, the interference from uplink UEs to downlink UEs can be made negligible compared to AWGN. Thus, we temporarily neglect this interference term in the analysis below. In this case, we build the optimality of the BDFD scheme using the following theorem.

Theorem 1 Assuming the UE grouping criteria is satisfied and the effective beam-domain channel matrices for all uplink and downlink groups are perfectly known,

the BDFD scheme achieves the uplink and downlink sum capacities⁵ simultaneously as the number of BS antennas approaches to infinity.

Proof Recalling the UE grouping Criterion 1 and Lemma 1, we can deduce that the IGI approaches to zero in the large N regime. Since we require $(\cap_{g_u \in G_u} B_{g_u}) \cap B_{SI,R} = \emptyset$ or $(\cap_{g_d \in G_d} B_{g_d}) \cap B_{SI,T} = \emptyset$ in UE grouping Criterion 2, the elements of $\tilde{\mathbf{H}}_{SI}^{\{B_{g_u}, B_{g_d}\}}$ SI converge to zero⁶ for very large N according to Lemma 2. Therefore, as $N \rightarrow \infty$ the beam-domain received signal $\tilde{\mathbf{y}}_{g_u}$ reduces to

$$\tilde{\mathbf{y}}_{g_u} = \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}, \cdot\}} \mathbf{s}_{g_u} + \tilde{\mathbf{n}}_{g_u} \quad (21)$$

Let $R_u^{\text{sum}}(P_u)$ be the uplink achievable sum rate with total power constraint P_u , and let $\Lambda_{g_u} = \mathbb{E}[\mathbf{s}_{g_u} \mathbf{s}_{g_u}^H]$ be the diagonal input covariance matrix of the uplink group g_u . Assuming the minimum mean square error with successive interference cancellation (MMSE-SIC) is employed to detect \mathbf{s}_{g_u} from (23), the uplink achievable sum rate can be expressed as (22) [30] on the top of the next page, where the second step is based on the property $\mathbf{I} + \mathbf{A}\mathbf{B}^{-1} = \mathbf{I} + \mathbf{B}\mathbf{A}^{-1}$. The last step follows from Lemma 1, i.e., $\mathbf{H}_{g_u} = \mathbf{F}^{\{\cdot, B_{g_u}\}} \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}, \cdot\}}$ is satisfied as $N \rightarrow \infty$. Note that the last line of (22) is exactly the uplink sum capacity [30].

$$\begin{aligned} R_u^{\text{sum}}(P_u) &= \max_{\Lambda_{g_u} \succcurlyeq 0, \sum_{g_u \in G_u} \text{tr}(\Lambda_{g_u}) \leq P_u} \sum_{g_u \in G_u} \log_2 \left| \mathbf{I}_{b_u} + \frac{1}{\sigma^2} \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}\}} \Lambda_{g_u} \left(\tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}\}} \right)^H \right| \\ &= \max_{\Lambda_{g_u} \succcurlyeq 0, \sum_{g_u \in G_u} \text{tr}(\Lambda_{g_u}) \leq P_u} \sum_{g_u \in G_u} \log_2 \left| \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{F}^{\{B_{g_u}\}} \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}\}} \Lambda_{g_u} \left(\tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}\}} \right)^H \left(\mathbf{F}^{\{B_{g_u}\}} \right)^H \right| \\ &= \max_{\Lambda_{g_u} \succcurlyeq 0, \sum_{g_u \in G_u} \text{tr}(\Lambda_{g_u}) \leq P_u} \sum_{g_u \in G_u} \log_2 \left| \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{H}_{g_u} \Lambda_{g_u} \mathbf{H}_{g_u}^H \right| \end{aligned} \quad (22)$$

Similarly, as $N \rightarrow \infty$, according to Criterion 1 and Lemma 1, the downlink beam-domain received signal $\tilde{\mathbf{y}}_{g_d}$ reduces to

$$\tilde{\mathbf{y}}_{g_d} = \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, \cdot\}} \right)^H \tilde{\mathbf{x}}_{g_d} + \mathbf{n}_{g_d} \quad (23)$$

⁵Herein, the uplink and downlink sum capacities indicate the maximum achievable rates of standard MIMO multiple access channel (MAC) and MIMO broadcasting (BC) channels, respectively [30].

⁶In (III-B) and (III-B), the terms of IGIs and SI are not exactly equal to zero for the general N . So we keep these terms in the equations.

Let $R_d^{\text{sum}}(P_d)$ be the downlink achievable sum rate with total power constraint P_d . Assuming the beam-domain transmit signal $\tilde{\mathbf{x}}_{g_d}$ is generated according to the rule of dirty paper code and using the MAC-BC duality [30], we have

$$R_d^{\text{sum}}(P_d) = \max_{\Lambda_{g_d} \succ 0, \sum_{g_d \in G_d} \text{tr}(\Lambda_{g_d}) \leq P_d} \left| \mathbf{I}_N + \frac{1}{\sigma} \sum_{g_d \in G_d} \mathbf{F}\{:, B_{g_d}\} \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \Lambda_{g_d} \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g_d}\} \right)^H \right| \quad (24)$$

where Λ_{g_d} denotes the diagonal input covariance matrix of the dual MAC channel. To show the optimality of the BDFD scheme, we examine the following equality. Letting $G'_d \subset G_d$ and $G'_d \neq \emptyset$, for arbitrary $g_d \in G'_d$, we have

$$\begin{aligned} & \left| \mathbf{I}_N + \frac{1}{\sigma} \sum_{g_d \in G_d} \mathbf{F}\{:, B_{g_d}\} \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \Lambda_{g_d} \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g_d}\} \right)^H \right| \\ &= |\mathbf{K}_{g_d}| \left| \mathbf{I}_N + \frac{1}{\sigma} \mathbf{K}_{g_d}^{-1} \sum_{g'_d \in G_d / \{g_d\}} \mathbf{F}\{:, B_{g'_d}\} \tilde{\mathbf{H}}_{g'_d}^{\{B_{g'_d}, :\}} \Lambda_{g'_d} \left(\tilde{\mathbf{H}}_{g'_d}^{\{B_{g'_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g'_d}\} \right)^H \right| \\ &= |\mathbf{K}_{g_d}| \left| \mathbf{I}_N + \frac{1}{\sigma} \left(\mathbf{I}_N - \mathbf{F}\{:, B_{g_d}\} \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \left(\Lambda_{g_d}^{-1} + \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^{-1} \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \right. \right. \\ & \quad \left. \left. \times \left(\mathbf{F}\{:, B_{g_d}\} \right)^H \right) \sum_{g'_d \in G_d / \{g_d\}} \mathbf{F}\{:, B_{g'_d}\} \tilde{\mathbf{H}}_{g'_d}^{\{B_{g'_d}, :\}} \Lambda_{g'_d} \left(\tilde{\mathbf{H}}_{g'_d}^{\{B_{g'_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g'_d}\} \right)^H \right| \\ &= |\mathbf{K}_{g_d}| \left| \mathbf{I}_N + \frac{1}{\sigma} \sum_{g'_d \in G_d / \{g_d\}} \mathbf{F}\{:, B_{g'_d}\} \tilde{\mathbf{H}}_{d, g'}^{\{B_{g'_d}, :\}} \Lambda_{d, g'} \left(\tilde{\mathbf{H}}_{d, g'}^{\{B_{g'_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g'_d}\} \right)^H \right| \end{aligned} \quad (25)$$

where $\mathbf{K}_{g_d} = \mathbf{I}_N + \frac{1}{\sigma} \mathbf{F}\{:, B_{g_d}\} \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \Lambda_{g_d} \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g_d}\} \right)^H$. The second step is obtained by applying the matrix inversion lemma on $\mathbf{K}_{g_d}^{-1}$ and the third step is based on Criterion 1. Using (25) repeatedly, we can rewrite the achievable downlink sum rate (24) as

$$\begin{aligned} R_d^{\text{sum}}(P_d) &= \max_{\Lambda_{g_d} \succ 0, \sum_{g_d \in G_d} \text{tr}(\Lambda_{g_d}) \leq P_d} \sum_{g_d \in G_d} \log_2 \left| \mathbf{I}_N + \frac{1}{\sigma} \mathbf{F}\{:, B_{g_d}\} \tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \Lambda_{g_d} \left(\tilde{\mathbf{H}}_{g_d}^{\{B_{g_d}, :\}} \right)^H \left(\mathbf{F}\{:, B_{g_d}\} \right)^H \right| \quad (26) \\ &= \max_{\Lambda_{g_d} \succ 0, \sum_{g_d \in G_d} \text{tr}(\Lambda_{g_d}) \leq P_d} \sum_{g_d \in G_d} \log_2 \left| \mathbf{I}_N + \frac{1}{\sigma} \mathbf{H}_{g_d} \Lambda_{g_d} \mathbf{H}_{g_d}^H \right| \end{aligned}$$

which is exactly the sum capacity of the dual MAC channel with total power constraint P_d .

Theorem 1 reveals that only the reduced-dimension effective beam-domain CSI is enough for the BDFD scheme to achieve the uplink and downlink capacities simultaneously in the large N regime. Therefore, the BDFD scheme reduces the difficulty of channel acquisition. Note that conventional TDD/FDD massive MIMO can only achieve the uplink or downlink capacity on each time-frequency unit, even full CSI is available. Meanwhile, the BDFD scheme avoids the deployment of active SIC (which is hardware and energy costly) and does not need the instantaneous knowledge of the SI channel. However, in the practical application, the passive SIC may still be needed to suppress the residual SI under the SI channel-unaware environment.

Remark 1 Note that Theorem 1 is valid for single-cell system. In multicell system with FD BS, each transmission experiences more interferences compared to the single-cell situation, which include SI, UE-to-UE interference from both within the cell and neighboring cells, and BS-to-BS interference. To realize the gain of FD massive MIMO, efficient multicell interference mitigation technologies from different aspects, which may include UE scheduling, power control, and multiuser precoding with limited interference channel knowledge, should be studied.

4 Practical Implementation of BDFD Scheme

In this section, we consider several key components of BDFD scheme in the practical implementation, which include UE grouping, effective beam-domain channel acquisition, beam-domain data transmission, and interference control between uplink and downlink.

4.1 *K-Means-Based UE Grouping*

In real cellular system, UEs will not naturally partition in groups with exactly the same active beam set. In order to implement the BDFD scheme efficiently, the UEs with different active beam sets must be partitioned so that the UE grouping criteria are satisfied as close as possible. In this subsection, we propose a UE grouping scheme to achieve this task. Our scheme consists of the following three steps.

Step 1: Compute the active beam sets of all UEs and the SI channel based on the DOA/DOD information using the method in Sect. 3.1.

Step 2: The aim of the second step is to gather the UEs with similar active beam spaces into a group based on the K-means principle. Without loss of generality, we consider the uplink UEs and the operation for downlink UEs is similar. To apply the K-means algorithm, we need first to define the “distance” between UEs. In the

proposed scheme, we employ the chordal distance between the active beam spaces of UEs. In particular, the distance between UEs k_u and k'_u can be expressed as

$$D_{\text{chordal}} \left(\mathbf{F}^{(:,B_{k_u})}, \mathbf{F}^{(:,B_{k'_u})} \right) = \left\| \mathbf{F}^{(:,B_{k_u})} \left(\mathbf{F}^{(:,B_{k_u})} \right)^H - \mathbf{F}^{(:,B_{k'_u})} \left(\mathbf{F}^{(:,B_{k'_u})} \right)^H \right\|^2 \\ = \left| B_{k_u} \right| + \left| B_{k'_u} \right| - 2 \left| B_{k_u} \cap B_{k'_u} \right| \quad (27)$$

where the second equality is based on the orthogonality between columns of $\mathbf{F}^{(:,B_{k_u})}$ and $\mathbf{F}^{(:,B_{k'_u})}$. Moreover, for a group of UEs \mathcal{U} , the ‘‘centroid’’ of their active beam spaces is defined as [31]

$$\bar{\mathbf{F}} = \text{eig}_{b_u} \left\{ \sum_{k_u \in \mathcal{U}} \mathbf{F}^{(:,B_{k_u})} \left(\mathbf{F}^{(:,B_{k_u})} \right)^H \right\} \quad (28)$$

where $\text{eig}_{b_u} \{\mathbf{A}\}$ indicates the unitary matrix whose columns are composed of b_u dominant eigenvectors of matrix \mathbf{A} . Since the columns of \mathbf{F} are all eigenvectors of $\sum_{k_u \in \mathcal{U}} \mathbf{F}^{(:,B_{k_u})} \left(\mathbf{F}^{(:,B_{k_u})} \right)^H$, we can obtain $\bar{\mathbf{F}} = \mathbf{F}^{(:,bar{B})}$, where \bar{B} denotes the active beam set of ‘‘centroid.’’ By examining (28), it is easy to see that \bar{B} can be expressed as

$$\bar{B} = \{f_1, f_2, \dots, f_{b_d}\} \quad (29)$$

where f_i is the index of the i th most frequent appeared beam in the sets $\{B_{k_u}\}_{k_u \in \mathcal{U}}$. With (27) and the notion of active beam set, we can conduct a very simple UE grouping algorithm based on the K-means principle, as shown in Algorithm 1. Note that as the output of Algorithm 1, the active beam set for group ‘‘centroid’’ is treated as the active beam set of that group.

Step 3: After step 2, we get a set of uplink UE groups, a set of downlink UE groups, and their active beam sets. To meet the SI cancellation condition in Criterion 2, for each uplink group, if its active beam set (denoted by B_u) is (partially) overlapped with $B_{SI,R}$, we update B_u as $B_u = B_u / (B_u \cap B_{SI,R})$. The active beam sets for downlink groups keep unchanged. Alternatively, we can also update the active beam set of downlink group (denoted by B_d) as $B_d = B_d / (B_d \cap B_{SI,T})$, if $B_d \cap B_{SI,T} \neq \emptyset$, while keeping the active beam sets for uplink groups unchanged. On the other hand, the active beam sets of different uplink/downlink UE groups after step 2 may also partially overlapped, which is not allowed according to Criterion 1. To deal with this problem, we further classify the uplink/downlink UE groups into several clusters so that the active beam sets of uplink/downlink UE groups in the same cluster are non-overlapping with certain guard interval. On certain time-frequency resource, only one uplink cluster and one downlink cluster are served

using the BDFD scheme. Moreover, the UEs groups from different clusters are served using orthogonal time-frequency resources to eliminate the interference.

Remark 2 Here, we mention that orthogonal time-frequency resource allocation for different clusters does not mean more time-frequency resource consumption. The reason is that, in the cellular system, the number of UEs within each cell is commonly large. Thus it is impossible to serve all UEs using the same time-frequency resource. To access to the network, UEs which cannot be served simultaneously should be allocated to other time-frequency resources using technologies such as orthogonal frequency-division multiple access. Thus, the UE clustering operation can actually be viewed as an additional constraint on time-frequency resource allocation and will not degrade the system SE significantly.

Remark 3 When the active beam sets of uplink or downlink groups are (partially) overlapped with that of the SI channel, these groups get less beams due to Criterion 2. This may cause some problems in fairness between uplink and downlink UEs. Note that Criterion 2 can be satisfied by reducing the active beams of uplink UE groups or downlink groups in step 3. Therefore, if we reduce the active beams of uplink group to meet Criterion 2 in odd time slots, and reduce the active beams of downlink group to meet Criterion 2 in even time slots. The fairness can be improved to some extent. More intelligently, the resource allocation algorithm in time-frequency dimension can be investigated to achieve some kind of fairness (e.g., max-min fairness) among all UEs. This is interesting for future research.

4.2 Full-Duplex Effective Beam-Domain Channel Estimation

In this subsection, we propose a full-duplex channel estimation scheme to estimate the effective beam-domain channels. During the training phase, all the uplink UEs transmit pilot signals to the BS, and meanwhile, the BS transmits the pilot signals to the downlink UEs. Let $\Phi_u \in \mathbf{C}^{\tau_u \times \max_{g_u \in G_u} K_{g_u}}$ be the orthogonal pilot sequence set for uplink training, where τ_u denotes the length of pilot sequence which satisfies $\tau_u \geq \max_{g_u \in G_u} K_{g_u}$. The pilot sequences allocated for group g_u can be given by $\Phi_{g_u} = \Phi_u^{\{:, 1:K_{g_u}\}}$. Meanwhile, let $\Phi_d \in \mathbf{C}^{\tau_d \times b_d}$ be the orthogonal downlink pilot sequences, where $\tau_d \geq b_d$ denotes the length of pilot sequence. The downlink pilot sequence for group g_d is precoded by multiplying the matrix $\mathbf{F}^{\{:, B_{g_d}\}}$. This operation is essential to suppress the IGI during training phase as will be seen below. The received pilot signals at the BS and the downlink group g_d can be expressed as

$$\mathbf{Y}_u = \sum_{g'_u \in G_u} \mathbf{H}_{g'_u} \Phi_{g'_u}^T + \mathbf{H}_{SI} \sum_{g'_d \in G_d} \mathbf{F}^{\{:, B_{g'_d}\}} \dot{\Phi}_d^T + \mathbf{N}_u \quad (30)$$

$$\mathbf{Y}_{g_d} = \mathbf{H}_{g_d}^H \sum_{g'_d \in G_d} \mathbf{F}^{\{::, B_{g'_d}\}} \Phi_d^T + \mathbf{N}_{g_d} \quad (31)$$

where $\dot{\Phi}_d = \Phi_d^{\{1:\tau_u\}}$ if $\tau_d \geq \tau_u$ and $\dot{\Phi}_d = \left[\Phi_d^T, \mathbf{0}_{(\tau_u - \tau_d) \times b_d}^T \right]^T$ if $\tau_d < \tau_u$. \mathbf{N}_u and \mathbf{N}_d denote the AWGNs with variance σ .

Uplink Effective Beam-Domain Channel Estimation

By multiplying both sides of (30) with $\left(\mathbf{F}^{\{::, B_{g_u}\}} \right)^H$, we arrive at the beam-domain receive pilot signal from uplink group g_u

$$\begin{aligned} \mathbf{Y}_{g_u} &= \left(\mathbf{F}^{\{::, B_{g_u}\}} \right)^H \mathbf{Y}_u \\ &= \tilde{\mathbf{H}}_{g_u}^{\{B_{g_u}, \cdot\}} \Phi_{g_u}^T + \sum_{g'_u \in G_u / \{g_u\}} \tilde{\mathbf{H}}_{g'_u}^{\{B_{g_u}, \cdot\}} \Phi_{g'_u}^T + \sum_{g'_d \in G_d} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u}, B_{g'_d}\}} \dot{\Phi}_d^T + \left(\mathbf{F}^{\{::, B_{g_u}\}} \right)^H \mathbf{N}_u \end{aligned} \quad (32)$$

With (32), the least squares (LS) estimator of the effective beam-domain channel vector for the uplink UE $g_{u,k}$ can be obtained as

$$\begin{aligned} \tilde{\mathbf{h}}_{g_{u,k}, \text{LS}}^{\{B_{g_u}, \cdot\}} &= \frac{1}{\tau_u p_u} \mathbf{Y}_{g_u} \Phi_u^* \mathbf{e}_k \\ &= \tilde{\mathbf{h}}_{g_{u,k}}^{\{B_{g_u}, \cdot\}} + \sum_{g'_u \in G_u / \{g_u\}} \tilde{\mathbf{h}}_{g'_{u,k}}^{\{B_{g_u}, \cdot\}} + \frac{1}{\tau_u p_u} \sum_{g'_d \in G_d} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u}, B_{g'_d}\}} \dot{\Phi}_d^T \Phi_u^* \mathbf{e}_k \\ &\quad + \frac{1}{\tau_u p_u} \left(\mathbf{F}^{\{::, B_{u,g}\}} \right)^H \mathbf{N}_u \Phi_u^* \mathbf{e}_k \end{aligned} \quad (33)$$

where p_u denotes the power of each uplink pilot symbol, i.e., $\left| [\Phi_{g_u}]_{i,j} \right|^2 = p_u$. The second term of RHS of (33) indicates the pilot contamination due to the use of same pilot sequences over all the uplink groups. The third term is the SI due to the simultaneous uplink and downlink training. Recalling Criterion 1 and Criterion 2 in the last section, and using Lemma 1 and Lemma 2, we can deduce that the pilot contamination and SI approach to zero in the large N regime.

In the practical scenario with finite number of BS antennas, the LS estimate in (33) can be further refined by a linear minimum mean square error (LMMSE) procedure to mitigate the residual pilot contamination and SI. Based on the general expression of LMMSE estimator [32, Ch. 12], the refined estimates can be expressed as

$$\begin{aligned}
\tilde{\mathbf{h}}_{g_u,k,\text{LM}}^{\{B_{g_u},\cdot\}} &= \sum_{i=1}^{M_u} \int_{\theta_{g_u,k,i}^{\min}}^{\theta_{g_u,k,i}^{\max}} \left(\mathbf{F}^{\{:\cdot, B_{g_u}\}} \right)^H \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{F}^{\{:\cdot, B_{g_u}\}} S_{g_u,k,i}(\theta) d\theta \\
&\times \left(\frac{\sigma}{\tau_u p_u} \mathbf{I}_{b_u} + \sum_{g'_u \in G_u} \sum_{i=1}^{M_u} \int_{\theta_{g_u,k,i}^{\min}}^{\theta_{g'_u,k,i}^{\max}} \left(\mathbf{F}^{\{:\cdot, B_{g_u}\}} \right)^H \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{F}^{\{:\cdot, B_{g_u}\}} S_{g'_u,k,i}(\theta) d\theta \right. \\
&+ \left. \left(\frac{1}{\tau_u p_u} \right)^2 \sum_{g'_d, g''_d \in G_d} \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} \mathbf{G}_{\theta_R, \theta_T}^{\{B_{g_u}, B_{g'_d}\}} \Psi_k \left(\mathbf{G}_{\theta_R, \theta_T}^{\{B_{g_u}, B_{g''_d}\}} \right)^H \right. \\
&\times S_{SI,i}(\theta_R, \theta_T) d\theta_R d\theta_T \Big)^{-1} \tilde{\mathbf{h}}_{g_u,k,\text{LS}}^{\{B_{g_u},\cdot\}}
\end{aligned} \tag{34}$$

where $\mathbf{G}_{\theta_R, \theta_T}^{\{B_{g_u}, B_{g'_d}\}} \triangleq \left(\mathbf{F}^{\{:\cdot, B_{g_u}\}} \right)^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \mathbf{F}^{\{:\cdot, B_{g'_d}\}}$ and $\Psi_k \triangleq \left(\frac{1}{\tau_u p_u} \right)^2 \Phi_d^T \Phi_u^* \mathbf{e}_k \mathbf{e}_k^H \Phi_u^T \Phi_d^*$.

Downlink Effective Beam-Domain Channel Estimation

By exploiting the beam-domain presentation and using (31), the LS estimator for the effective beam-domain channel vector of the downlink UE $g_{d,k}$ can be obtained as

$$\begin{aligned}
\tilde{\mathbf{h}}_{g_{d,k},\text{LS}}^{\{B_{g_d},\cdot\}} &= \frac{1}{\tau_d p_d} (\mathbf{Y}_{g_d} \Phi_d^*)^H \mathbf{e}_k \\
&= \tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g_d},\cdot\}} + \sum_{g'_d \in G_d} \tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g'_d},\cdot\}} + \frac{1}{\tau_d p_d} (\mathbf{N}_{g_d} \Phi_d^*)^H \mathbf{e}_k
\end{aligned} \tag{35}$$

where p_d denotes the power of each downlink pilot symbol, i.e., $|\Phi_d|_{i,j}|^2 = p_d$. The second term indicates the pilot contamination due to the use of same pilot sequences over all the downlink UE groups. Using Criterion 1 and Lemma 1, the pilot contamination converges to zero as $N \rightarrow \infty$. Similarly, we can refine the estimates with the LMMSE procedure, resulting in

$$\begin{aligned}
\tilde{\mathbf{h}}_{g_{d,k},\text{LM}}^{\{B_{g_d},\cdot\}} &= \sum_{g'_d \in G_d} \sum_{i=1}^{M_d} \int_{\theta_{g_d,k,i}^{\min}}^{\theta_{g'_d,k,i}^{\max}} \left(\mathbf{F}^{\{:\cdot, B_{g_d}\}} \right)^H \mathbf{a}(\theta) \\
&\times \mathbf{a}^H(\theta) \mathbf{F}^{\{:\cdot, B_{g'_d}\}} S_{g_d,k,i}(\theta) d\theta \left(\sum_{g'_d, g''_d \in G_d} \sum_{i=1}^{M_d} \int_{\theta_{g_d,k,i}^{\min}}^{\theta_{g'_d,k,i}^{\max}} \left(\mathbf{F}^{\{:\cdot, B_{g'_d}\}} \right)^H \mathbf{a}(\theta) \right. \\
&\times \mathbf{a}^H(\theta) \mathbf{F}^{\{:\cdot, B_{g''_d}\}} S_{g_d,k,i}(\theta) d\theta + \frac{\sigma}{\tau_d p_d} \Big)^{-1} \tilde{\mathbf{h}}_{g_{d,k},\text{LS}}^{\{B_{g_d},\cdot\}}
\end{aligned} \tag{36}$$

Table 2 Minimum required lengths of pilot sequences in the proposed estimation scheme and conventional schemes

	BDFD	TDD massive MEMO (linear transceiver [3])	FDD massive MIMO (JSDM [8])	FD massive MIMO (linear transceiver [18])
Minimum length of pilot sequence	$\max \left\{ \max_{g_u \in G_u} K_{g_u}, b_d \right\}$	$\sum_{g_u \in G_u} K_{g_u} + \sum_{g_d \in G_d} K_{g_d}$	$\max_{g_u \in G_u} K_{g_u} + b_d$ (Approximate)	$\sum_{g_u \in G_u} K_{g_u} + N$

To estimate the channels of all $\sum_{g_u \in G_u} K_{g_u}$ uplink UEs and $\sum_{g_d \in G_d} K_{g_d}$ downlink UEs, the minimum required lengths of pilot sequences in the proposed full-duplex estimation scheme and conventional schemes used in the TDD/FDD/FD massive MIMO systems are summarized in Table 2. It is seen that the proposed scheme improves the training efficiency significantly.

Example 2 Considering the channel model in Example 1, we have $b_d \approx 12$ if the BS is equipped with $N = 128$ antennas. If three uplink groups and three downlink groups are scheduled and each group contains five UEs, the minimum required length of pilot sequences in the proposed scheme is 12 (symbol times).

However, this number becomes 30, 17, and 143 (symbol times), respectively, in the reference schemes listed in Table 1. After downlink channel estimation, the estimated CSI should be feedback to BS in order to perform downlink transmission. This can affect the system from two aspects. First, the feedback error due to quantization error, noise, and feedback delay decreases the accuracy of downlink CSI. Moreover, CSI feedback increases the load of feedback channel and, hence, can degrade the overall system SE. However, the results in [33] showed that the CSI error (in term of mean-square error) due to imperfect feedback can be made much smaller than that caused by estimation error in downlink training phase, especially in the high signal-to-noise ratio (SNR) region. Moreover, since the effective downlink channel dimension is greatly reduced in proposed BDFD scheme, we assume that the additional load caused by CSI feedback is negligible when compared with the other feedback information. Therefore, for simplicity, we consider the optimistic situation of error-free CSI feedback and neglect the SE penalty due to feedback. A similar approach is also adopted in [8].

4.3 Beam-Domain Data Transmission and Achievable Rate with Noisy CSI

To keep the complexity low, we assume that the BS employs linear processing in the beam domain. In uplink, to detect the signals from group g_u , the BS combines the

beam-domain received signal (Sect. 3.2) by multiplying the receive beamforming matrix $\mathbf{W}_{g_u} = [\mathbf{w}_{g_u,1}, \mathbf{w}_{g_u,2}, \dots, \mathbf{w}_{g_u,K_g}] \in \mathbf{C}^{b_u \times K_g}$, i.e., $\dot{\mathbf{y}}_{g_u} = \mathbf{W}_{g_u}^H \tilde{\mathbf{y}}_{g_u}$. The k th entry of $\dot{\mathbf{y}}_{g_u}$

$$\begin{aligned} \dot{y}_{g_u,k} &= \mathbf{w}_{g_u,k}^H \tilde{\mathbf{h}}_{g_u,k}^{\{B_{g_u,\cdot}\}} s_{g_u,k} + \mathbf{w}_{g_u,k}^H \sum_{k'=1, k' \neq k}^{K_{g_u}} \tilde{\mathbf{h}}_{g_u,k'}^{\{B_{g_u,\cdot}\}} s_{g_u,k'} \\ &+ \mathbf{w}_{g_u,k}^H \sum_{g'_u \in G_u / \{g_u\}} \tilde{\mathbf{H}}_{g'_u}^{\{B_{g_u,\cdot}\}} \mathbf{s}_{g'_u} + \mathbf{w}_{g_u,k}^H \sum_{g'_d \in G_d} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u}, B_{g'_d}\}} \tilde{\mathbf{x}}_{g'_d} + \mathbf{w}_{g_u,k}^H \tilde{\mathbf{n}}_{g_u} \end{aligned} \quad (37)$$

is used to decode the symbol of UE $g_{u,k}$. In downlink, the intended signal of group g_d , i.e., $\mathbf{s}_{g_d} \in \mathbf{C}^{K_{g_d} \times 1}$, are precoded by the beamforming matrix $\mathbf{W}_{g_d} = [\mathbf{w}_{g_d,1}, \mathbf{w}_{g_d,2}, \dots, \mathbf{w}_{g_d,K_{g_d}}] \in \mathbf{C}^{b_d \times K_{g_d}}$ in the beam domain. Thus, the beam-domain transmit signal vector for group g_d can be expressed as $\tilde{\mathbf{x}}_{g_d} = \mathbf{W}_{g_d} \mathbf{s}_{g_d}$. Using these on (III-B), the beam-domain received signal at UE $g_{d,k}$ can be expressed as

$$\begin{aligned} \tilde{y}_{g_{d,k}} &= \left(\tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g_d,\cdot}\}} \right)^H \mathbf{w}_{g_{d,k}} s_{g_{d,k}} + \left(\tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g_d,\cdot}\}} \right)^H \sum_{k'=1, k' \neq k}^K \mathbf{w}_{g_{d,k'}} s_{g_{d,k'}} \\ &+ \sum_{g'_d \in G_d / \{g_d\}} \left(\tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g'_d,\cdot}\}} \right)^H \mathbf{W}_{g'_d} \mathbf{s}_{g'_d} + n_{g_{d,k}} \end{aligned} \quad (38)$$

The optimal beamforming scheme to maximize the sum rate has been proved NP-hard [34]. Thus, we consider the suboptimal scheme to provide a bound on the system performance. In general, the (suboptimal) beamforming matrices can be designed with different criteria, e.g., maximizing the desired signal power which corresponds to the eigen beamforming or minimizing the inter-UE interference which corresponds to the zero-forcing (ZF) beamforming. In this work, we adopt the latter one since the ZF beamforming is known to approach the asymptotic limit of achievable rate faster as the number of BS antennas increases [2]. Assuming the channel estimators in (34) and (36), the transmit and receive beamforming matrices of the BS can be expressed as

$$\begin{aligned} \mathbf{W}_{g_u} &= \tilde{\mathbf{H}}_{g_u, \text{LM}}^{\{B_{g_u,\cdot}\}} \left(\left(\tilde{\mathbf{H}}_{g_u, \text{LM}}^{\{B_{g_u,\cdot}\}} \right)^H \tilde{\mathbf{H}}_{g_u, \text{LM}}^{\{B_{g_u,\cdot}\}} \right)^{-1} \\ \mathbf{W}_{g_d} &= \tilde{\mathbf{H}}_{g_d, \text{LM}}^{\{B_{g_d,\cdot}\}} \left(\left(\tilde{\mathbf{H}}_{g_d, \text{LM}}^{\{B_{g_d,\cdot}\}} \right)^H \tilde{\mathbf{H}}_{g_d, \text{LM}}^{\{B_{g_d,\cdot}\}} \right)^{-1} \Upsilon_{g_d}^{-1/2} \end{aligned} \quad (39)$$

where Υ_{g_d} is a diagonal normalized matrix with $[\Upsilon_{g_d}]_{l,l} = \mathbf{e}_l^H \left(\left(\tilde{\mathbf{H}}_{g_d, \text{LM}}^{B_{g_d, \cdot}} \right)^H \tilde{\mathbf{H}}_{g_d, \text{LM}}^{B_{g_d, \cdot}} \right)^{-1} \mathbf{e}_l$. Due to the requirement of matrix inversion, the complexity of ZF beamforming becomes high when the number of UEs in each group is large. Some low complexity linear beamforming schemes, such as eigen beamforming, can be employed to deal with this problem at the cost of a few performance loss. To do so, we need just replace (39) with the beamforming matrices of these schemes. No other change is required.

According to (37) and (38) and using the bounding technique in [35], the average achievable rates at the uplink UE $g_{u,k}$ and downlink UE $g_{d,k}$ can be expressed as

$$R_{g_{u,k}} = \frac{T - \max\{\tau_u, \tau_d\}}{T} \log_2 \left(1 + \frac{p_{g_{u,k}}}{\mathbb{E}[\text{CE}_{g_{u,k}}] + \mathbb{E}[\text{IUI}_{g_{u,k}}] + \mathbb{E}[\text{IGI}_{g_{u,k}}] + \mathbb{E}[\text{SI}_{g_{u,k}}] + \mathbb{E}[\|\mathbf{w}_{u,g_k}\|^2]} \right)$$

$$R_{g_{d,k}} = \frac{T - \max\{\tau_u, \tau_d\}}{T} \log_2 \left(1 + \frac{p_{g_{d,k}} \mathbb{E}[\Upsilon_{g_d}^{-1}]_{k,k}}{\mathbb{E}[\text{CE}_{g_{d,k}}] + \mathbb{E}[\text{IUI}_{g_{d,k}}] + \mathbb{E}[\text{IGI}_{g_{d,k}}] + 1} \right) \quad (40)$$

where T denotes the channel coherent time. $p_{g_{u,k}} = \mathbb{E}[|s_{g_{u,k}}|^2]$ and $p_{g_{d,k}} = \mathbb{E}[|s_{g_{d,k}}|^2]$ denote the transmit powers. CE_i , IUI_i , IGI_i , and SI_i ($i \in \{g_{u,k}, g_{d,k}\}$) denote the powers of channel estimation error, inter-UE interference (IUI) within the group, IGI, and SI, respectively, whose expressions are summarized in Table 3.

Table 3 Expressions of powers of channel estimation error, IUI within the group, IGI, and SI

	Uplink	Downlink
Useful signal power	$p_{g_{u,k}}$	$p_{g_{d,k}} [\Upsilon_{g_d}^{-1}]_{k,k}$
Channel estimation error	$\text{CE}_{g_{u,k}} = p_{g_{u,k}} \left \mathbf{w}_{g_{u,k}}^H \Delta \tilde{\mathbf{h}}_{g_{u,k}}^{B_{g_u, \cdot}} \right ^2$	$\text{CE}_{g_{d,k}} = p_{g_{d,k}} \left \left(\Delta \tilde{\mathbf{h}}_{g_{d,k}, \text{LM}}^{B_{g_d, \cdot}} \right)^H \mathbf{w}_{g_{d,k}} \right ^2$
Inter-UE interference	$\text{IUI}_{g_{u,k}} = \sum_{k'=1, k' \neq k}^{K_{g_u}} p_{g_{u,k'}} \left \mathbf{w}_{g_{u,k}}^H \Delta \tilde{\mathbf{h}}_{g_{u,k'}}^{B_{g_u, \cdot}} \right ^2$	$\text{IUI}_{g_{d,k}} = \sum_{k'=1, k' \neq k}^{K_{g_d}} p_{g_{d,k'}} \left \left(\Delta \tilde{\mathbf{h}}_{g_{d,k}, \text{LM}}^{B_{g_d, \cdot}} \right)^H \mathbf{w}_{g_{d,k'}} \right ^2$
Inter-group interference	$\text{IGI}_{g_{u,k}} = \sum_{g'_u \in G_u / \{g_u\}} \left\ \mathbf{w}_{g_{u,k}}^H \tilde{\mathbf{H}}_{g'_u}^{B_{g_u, \cdot}} \Lambda_{g'_u}^{1/2} \right\ ^2$	$\text{IGI}_{g_{d,k}} = \sum_{g'_d \in G_d / \{g_d\}} \left\ \left(\tilde{\mathbf{h}}_{g'_d}^{B_{g_d, \cdot}} \right)^H \mathbf{w}_{g'_d} \Lambda_{g'_d}^{1/2} \right\ ^2$
Self-interference	$\text{SI}_{g_{u,k}} = \sum_{g'_d \in G_d} \left\ \mathbf{w}_{g_{u,k}}^H \tilde{\mathbf{H}}_{g'_d}^{B_{g_u, \cdot}, B_{g'_d, \cdot}} \mathbf{w}_{g'_d} \Lambda_{g'_d}^{1/2} \right\ ^2$	

The exact expressions of the average achievable rates are difficult to obtain under the considered channel model. Instead, in the following we focus on the question how do the above negative factors (i.e., channel estimation error, IUI, IGI, and SI) affect the achievable rate performance in the BDFD scheme. To answer this question, we present the scaling behaviors for powers of channel estimation error, IUI, IGI, and SI in the following theorem.

Theorem 2 Assume the cardinalities of the active beam sets scale linearly with N , i.e., $\lim_{N \rightarrow \infty} \frac{b_u}{N} > 0$ and $\lim_{N \rightarrow \infty} \frac{b_d}{N} > 0$. The scaling behaviors for average powers of channel estimation error, IUI, IGI, and SI in the large N regime are given by Table 4.

Proof See the Appendix.

Theorem 2 reveals that, in the BDFD scheme, the powers of IGI and SI decrease faster than other terms when the number of BS antennas increases. As a result, the effect of IGI and SI diminishes in the large N regime. In this sense, the BDFD scheme in fact decomposes the original system into several lower dimension uplink or downlink massive MIMO systems operating on the (asymptotically) orthogonal beam spaces. Another important observation from Theorem 2 is that the SI power decreases faster than $\mathcal{O}(N^{-1})$ in the BDFD scheme. This is quite different from the FD massive MIMO with linear transceiver [18], where the SI power changes exactly with $\mathcal{O}(N^{-1})$ in the large N regime. The reason is that, with the UE grouping criteria (Criterion 2), the signals of uplink or downlink groups occupy asymptotically orthogonal beam spaces with the SI. Thus, better SI suppression can be achieved in the BDFD scheme.

Remark 4 In theorem 2, we have assumed that b_u and b_d scale linearly with N . This is a standard assumption in the field of massive MIMO [8] in order to use the analytic tools developed for large-scale antenna systems. The assumption indicates that b_u and b_d , and hence the required length of pilot sequences, tend to infinity as $N \rightarrow \infty$, which is contrary to the purpose of this paper. However, in the practical implementation, the BS cannot be equipped with too many antennas due to the realistic constraints on hardware complexity and power consumption. With reasonable N , the training overhead is still low (See the Example 2 in Sect. 4.2).

Table 4 Scaling behaviors of channel estimation error, IUI within the group, IGI, and SI in the large N regime

	Useful signal power	Channel estimation error	Inter-UE interference	Inter-group interference	Self-interference
Uplink	$\mathcal{O}(1)$	$\mathcal{O}(N^{-1})$	$\mathcal{O}(N^{-1})$	$\begin{cases} \mathcal{O}(N^{\sigma_{\text{IGI}}^u}) \\ \sigma_{\text{IGI}}^u < -1 \end{cases}$	
Downlink	$\mathcal{O}(N)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\begin{cases} \mathcal{O}(N^{\sigma_{\text{IGI}}^d}) \\ \sigma_{\text{IGI}}^d < 0 \end{cases}$	$\begin{cases} \mathcal{O}(N^{\sigma_{\text{SI}}}) \\ \sigma_{\text{SI}} < -1 \end{cases}$

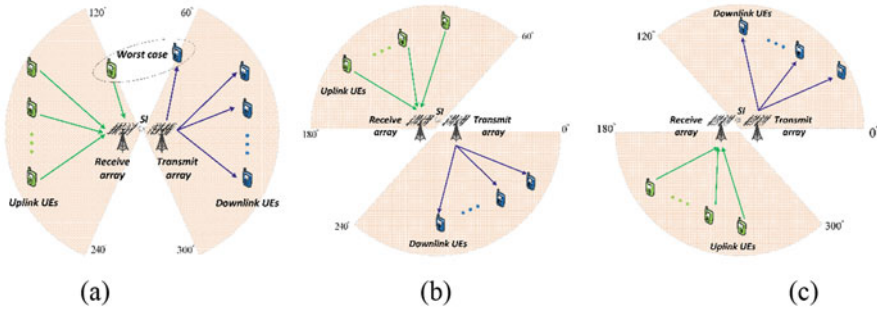


Fig. 4 Cell sectorization

4.4 Interference Control Between Uplink and Downlink

When multiple uplink UEs and downlink UEs are active simultaneously at the same frequency band, the resultant network suffers from increased interferences from uplink UEs to downlink UEs. One simple approach to alleviate the interference is cell sectorization. As shown in Fig. 4a, on the particular time-frequency resource, we only schedule the uplink UEs and downlink UEs in two opposite 120° sectors. This ensures that the uplink UE and downlink UE with small distance will not be scheduled on the same time-frequency resource. In the worst case where the uplink and downlink UEs are both located on the boundaries of the sectors as shown in Fig. 4a, the interference channel between uplink and downlink is still much weaker than the useful channel. For example, when the distance between BS and UEs is 300 m, the interference channel between two boundary UEs is 46 dB weaker than the useful channel according to the 3GPP LTE BS-to-UE and UE-to-UE path loss models [36, Table 6.4–1] (note that the UE-to-UE channel suffers from more path loss than the BS-to-UE channel even though the transmission distances are the same [36]). On the other hand, to cover the whole cell evenly, we can schedule the UEs in the rotated sectors, as shown in Fig. 4b, c, using different time-frequency resources.

5 Simulation Results

In this section, the performance of BDFD scheme is evaluated using the 3GPP LTE simulation model for macro-cell environment [36]. The simulation parameters are summarized in Table 5. It is assumed that the passive SIC scheme for infrastructure nodes proposed in [25] has been employed at the BS. In such scheme, the suppression is from two parts, namely, (i) the path loss introduced by the 20 m separation between transmit and receive antenna arrays and (ii) an additional cancellation of 45 dB provided by techniques, such as radio-frequency absorber material and cross-polarization. No other active SIC scheme is used.

Table 5 Simulation parameters

Parameter	Value
Bandwidth	20 MHz
Central frequency	2.4GHz
Thermal noise density	-174 dBm/Hz
Channel coherent time	200(Symbol times)
Path loss (BS-to-UE)	$2.7+42.8\log_{10}(R)[dB]$ R: distance in meter
Path loss(UE-to-UE)	$55.78+40\log_{10}(R)[dB]$ R: distance in meter
Number of scattering clusters	$M_u = M_d = 1$

We first consider a scenario where the uplink/downlink UEs gather perfectly in three groups and the DOA/DOD regions of UEs in each group are identical. We assume that each group contains five UEs. The DOA regions of three uplink groups are $[-33^\circ, -23^\circ]$, $[7^\circ, 17^\circ]$, and $[34^\circ, 44^\circ]$, respectively. Since we assume $M_u = 1$, the DOA region of uplink group is $[a, b]$ means $[\theta_{g_u,k,1}^{\min}, \theta_{g_u,k,1}^{\max}] = [a, b]$. Similarly, the DOD regions of three downlink groups are $[-39^\circ, -29^\circ]$, $[10^\circ, 20^\circ]$, and $[19^\circ, 30^\circ]$, respectively.

The DOA and DOD regions of SI channel are set to $[-15^\circ, -5^\circ]$, $[54^\circ, 66^\circ]$, and $[-25^\circ, -35^\circ]$, $[19^\circ, 30^\circ]$, respectively. The resulting active beam sets for all the groups satisfy the UE grouping criteria.

Figure 5 compares the SEs⁷ of BDFD scheme, TDD massive MIMO with linear transceiver [3], FDD massive MIMO with JSDM [8], FD massive MIMO with linear transceiver [18], and FD massive MIMO with spatial SI suppression [37]. For the scheme with spatial SI suppression [37], the instantaneous CSI of SI channel is required at the BS in order to perform SI cancellation in spatial domain. With perfect effective beam-domain CSI, it is seen that the SE of BDFD scheme approaches the sum of uplink and downlink capacities as the number of BS antennas increases. With estimated effective beam-domain channels, the performance gap increases as N becomes larger. The reason is that, although the BDFD scheme can reduce the required length of pilot sequence significantly, the training overhead still increases linearly with N . Due to the same reason, when downlink reciprocity is available (at the cost of higher hardware complexity), the reference schemes in [18] and [37] achieve better SE over the BDFD scheme in the large N region. On the other hand, significant SE gain can be achieved by the BDFD scheme over the TDD and FDD massive MIMO systems. Interestingly, the performance gain can even be greater than $2 \times$ (e.g., $2.08 \times$ gain is observed over the TDD massive MIMO when $N = 200$), which is impossible in the conventional FD system. This is because the TDD massive MIMO spends more resource for pilot signaling as discussed in Sect.

⁷The SEs of BDFD scheme and FD massive MIMO with linear transceiver are defined as the sum of achievable rates of all uplink and downlink UEs. The SEs of TDD/FDD massive MIMO are defined in the same way but penalized by a factor of 1/2 due to the orthogonal uplink/downlink resource allocation.

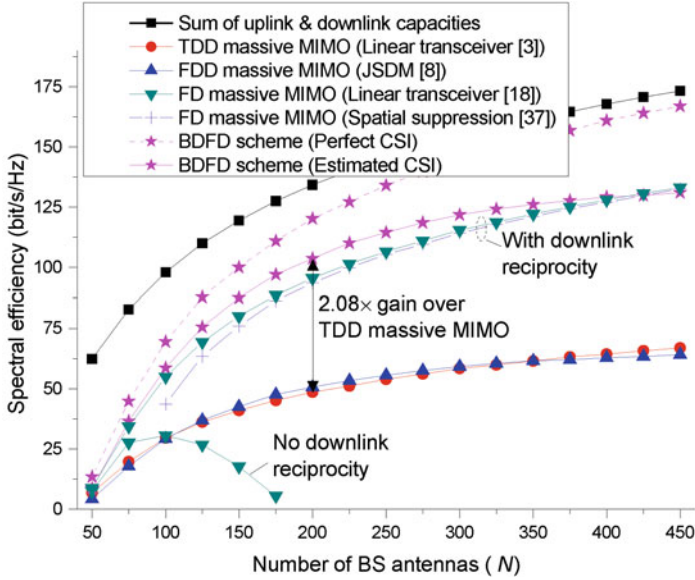
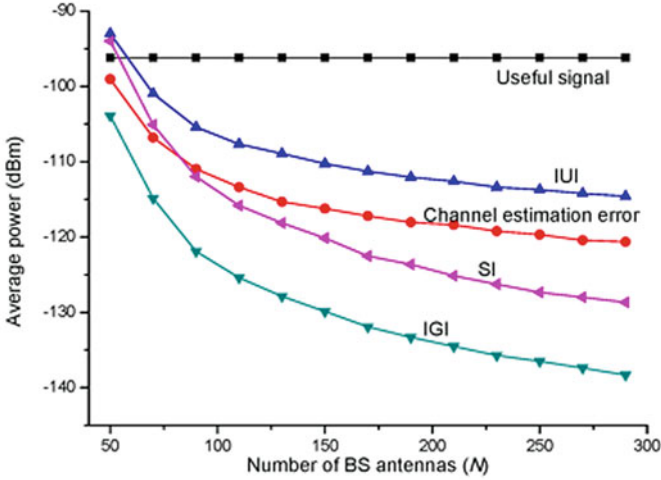


Fig. 5 Spectral efficiency with perfect UE grouping. The transmit powers of BS and uplink UEs are 20.2 dBm. The distance between UEs and BS is set to 500 m. This setup ensures the average uplink/downlink receive SNR is 3 dB

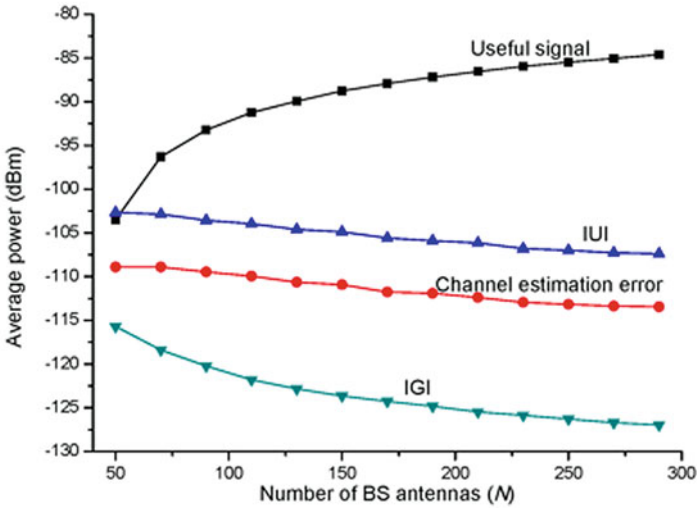
4.2. At last, without downlink reciprocity, it is observed that the FD massive MIMO with linear transceiver becomes infeasible if the number of BS antennas exceeds 175, since almost all the time resource is allocated for downlink training.

To examine the scaling results in Theorem 2, Fig. 6 simulates the average powers of useful signal, channel estimation error, IUI, IGI, and SI in the BDFD scheme. From the figure, it is seen that the powers of IGI and SI decrease faster than other negative factors in the large N regime, which coincides with Theorem 2. Then, in Figs. 7 and 8, we consider a more realistic scenario where the UEs are not naturally partitioned in groups with exactly the same active beam set. We assume that 50 uplink UEs and 50 downlink UEs are located in two opposite 120° sectors, as shown in Fig. 4a. The BS is equipped with $N = 128$ transmit/receive antennas. The signal of each uplink/downlink UE is within a 10° DOA/DOD region which is randomly distributed in the sectors. The distance between uplink/downlink UE and BS is randomly distributed in the interval [200, 1000] m. After UE grouping, three UE clusters are formed using the method in Sect. 4.1, and the UEs groups in different clusters are served with orthogonal time-frequency resources. Without loss of generality, the active beam sets of uplink groups which are (partially) overlapped that of the SI channel are updated using the method in Sect. 4.1 (step 3).

Figure 7 depicts the SE of BDFD scheme as a function of average receive SNR. Since the UE groups are divided into three clusters, the SE is defined as the average of SEs for three clusters. The number of scattering cluster for SI channel is set to



(a)



(b)

Fig. 6 Average powers of useful signal, channel estimation error, and interferences in the BDFD scheme. The simulation setup is the same with Fig. 5. (a) Uplink. (b) Downlink

$M_{SI} = 2$. The DOA and DOD regions of SI channel are $[-15^\circ, -5^\circ]$, $[54^\circ, 66^\circ]$, and $[-25^\circ, -35^\circ]$, $[19^\circ, 30^\circ]$, respectively. Again, it is seen that the BDFD scheme achieves the best performance. In particular, the BDFD scheme achieves $1.80\times$ and $1.87\times$ SE gain over the TDD massive MIMO when the SNRs are 3 dB and 12 dB,

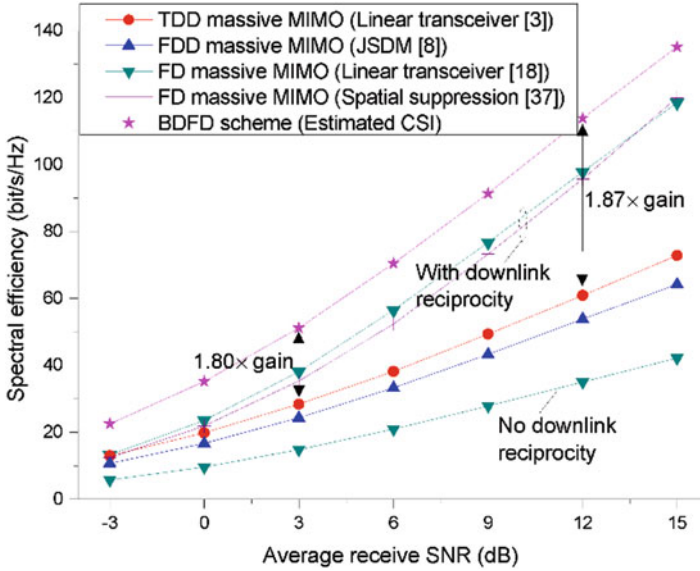


Fig. 7 Spectral efficiency of BDFD scheme with imperfect UE grouping for different average receive SNRs

respectively. The performance gain is generally smaller compared with that in Fig. 5. The reason is that, different from Fig. 5, the UEs in each group may not have exactly the same active beam set. Thus, not all the beams in the active beam set of that group can be fully used by all the UEs. This will result in some performance degradation. Moreover, there is no performance floor in the large SNR region for the BDFD scheme, even the interference from uplink UE to downlink exist. At last, we mention that, with the UE scheduling scheme in Sect. 4.4, the powers of interferences from uplink UEs to downlink UEs are much smaller than the background noise. That is why no obvious performance floor is observed for BDFD scheme in the large SNR region.

In the previous simulations, the number of scattering clusters for SI channel is fixed to $M_{SI} = 2$. In Fig. 8, we consider the SE of BDFD scheme with larger M_{SI} . In particular, we let M_{SI} increase from 2 to 10. The SI signal from each scattering cluster is within a 10° DOA/DOD region which is randomly distributed $[-90^\circ, 90^\circ]$. It is seen the SE of BDFD scheme approaches to that of the TDD/FDD massive MIMO as M_{SI} increases. This is because the numbers of uplink groups and the active beams for each uplink group decrease according to algorithm in Sect. 4.1. In fact, for large M_{SI} , the performance gain of BDFD scheme over TDD/FDD massive MIMO is mainly due to the saving in the training resources. Moreover, since the scheme in [37] cancels the SI completely using the instantaneous CSI, it achieves better SE when $M_{SI} \geq 6$. However, the gain can be realized only when the instantaneous SI channel can be efficiently estimated.

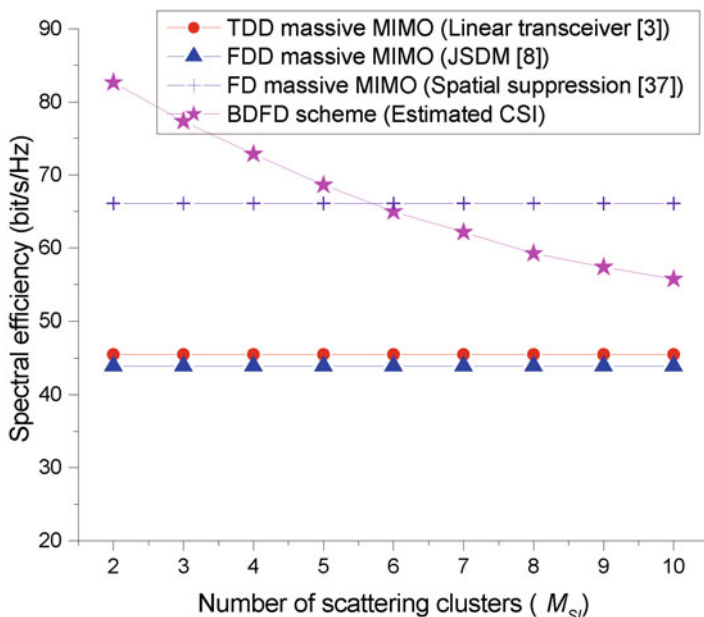


Fig. 8 Spectral efficiency of BDFD scheme for different numbers of scattering clusters for SI channel

6 Conclusion

This paper proposes a BDFD massive MIMO scheme to realize CCUD transmission in the cellular system. By exploiting the compressibility of beam-domain channel, the BDFD scheme can eliminate SI due to CCUD transmission efficiently. The simulation results show that the BDFD massive MIMO scheme outperforms the TDD/FDD massive MIMO and FD massive MIMO with linear transceiver significantly in the macro-cell environment. Due to the above advantages, we suggest BDFD massive MIMO as a potential enabling technology for evolution toward future wireless cellular system.

Acknowledgment This work was supported in part by National Natural Science Foundation of China under Grant 61671472 and in part by Jiangsu Province Natural Science Foundation under Grant BK20160079.

Appendix

As in [6], we assume that the number of channel paths is very large within the DOA/DOD regions. As a result, we can assume that the uplink/downlink channel

is Gaussian distributed from the law of large numbers, i.e., $\mathbf{h}_{g_{u,k}} \sim \mathcal{CN}(0, \mathbf{C}_{g_{u,k}})$ and $\mathbf{h}_{g_{d,k}} \sim \mathcal{CN}(0, \mathbf{C}_{g_{d,k}})$. According to (1) and (2), the correlation matrices can be expressed as

$$\begin{aligned}\mathbf{C}_{g_{u,k}} &= \mathbb{E} \left[\mathbf{h}_{g_{u,k}} \mathbf{h}_{g_{u,k}}^H \right] = \sum_{i=1}^{M_u} \int_{\theta_{g_{u,k},i}^{\min}}^{\theta_{g_{u,k},i}^{\max}} \mathbf{a}(\theta) \mathbf{a}^H(\theta) S_{g_{u,k},i}(\theta) d\theta \\ \mathbf{C}_{g_{d,k}} &= \mathbb{E} \left[\mathbf{h}_{g_{d,k}} \mathbf{h}_{g_{d,k}}^H \right] = \sum_{i=1}^{M_d} \int_{\theta_{g_{d,k},i}^{\min}}^{\theta_{g_{d,k},i}^{\max}} \mathbf{a}(\theta) \mathbf{a}^H(\theta) S_{g_{d,k},i}(\theta) d\theta\end{aligned}\quad (41)$$

Based on the above assumption, we have $\tilde{\mathbf{h}}_{g_{u,k}}^{\{B_{g'_u, \cdot}\}} \sim \mathcal{CN}(0, \tilde{\mathbf{C}}_{g_{u,k}}^{g'_u})$ and $\tilde{\mathbf{h}}_{g_{d,k}}^{\{B_{g'_d, \cdot}\}} \sim \mathcal{CN}(0, \tilde{\mathbf{C}}_{g_{d,k}}^{g'_d})$, where $\tilde{\mathbf{C}}_{g_{u,k}}^{g'_u} = \left(\mathbf{F}^{\{ \cdot, B_{g'_u} \}} \right)^H \mathbf{C}_{g_{u,k}} \mathbf{F}^{\{ \cdot, B_{g'_u} \}}$ and $\tilde{\mathbf{C}}_{g_{d,k}}^{g'_d} = \left(\mathbf{F}^{\{ \cdot, B_{g'_d} \}} \right)^H \mathbf{C}_{g_{d,k}} \mathbf{F}^{\{ \cdot, B_{g'_d} \}}$. With this and the standard result for LMMSE estimator [32, Ch. 12], the LMMSE estimates for uplink and downlink channels have distributions $\tilde{\mathbf{h}}_{g_{u,k}, \text{LM}}^{\{B_{g'_u, \cdot}\}} \sim \mathcal{CN}(0, \tilde{\mathbf{C}}_{g_{u,k}, \text{LM}}^{g'_u})$ and $\tilde{\mathbf{h}}_{g_{d,k}, \text{LM}}^{\{B_{g'_d, \cdot}\}} \sim \mathcal{CN}(0, \tilde{\mathbf{C}}_{g_{d,k}, \text{LM}}^{g'_d})$, where the correlation matrices can be expressed as

$$\begin{aligned}\tilde{\mathbf{C}}_{g_{u,k}, \text{LM}}^{g'_u} &= \tilde{\mathbf{C}}_{g_{u,k}}^{g'_u} \left(\frac{\sigma}{\tau_u p_u} \mathbf{I}_{b_u} + \sum_{g'_u \in G_u} \tilde{\mathbf{C}}_{g_{u,k}}^{g'_u} + \sum_{g'_d, g''_d \in G_d} \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} \mathbf{G}_{\theta_R, \theta_T}^{\{B_{u,g}, B_{d,g'}\}} \right. \\ &\quad \left. \times \Psi_k \left(\mathbf{G}_{\theta_R, \theta_T}^{\{B_{u,g}, B_{d,g''}\}} \right)^H S_{SI,i}(\theta_R, \theta_T) d\theta_R d\theta_T \right)^{-1} \tilde{\mathbf{C}}_{g_{u,k}}^{g'_u} \\ \tilde{\mathbf{C}}_{g_{d,k}, \text{LM}}^{g'_d} &= \sum_{g'_d \in G_d} \sum_{i=1}^{M_d} \int_{\theta_{g_{d,k},i}^{\min}}^{\theta_{g_{d,k},i}^{\max}} \left(\mathbf{F}^{\{ \cdot, B_{g'_d} \}} \right)^H \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{F}^{\{ \cdot, B_{g'_d} \}} S_{g_{d,k},i}(\theta) d\theta \\ &\quad \times \left(\sum_{g', g'' \in G_d} \sum_{i=1}^{M_d} \int_{\theta_{g_{d,k},i}^{\min}}^{\theta_{g_{d,k},i}^{\max}} \left(\mathbf{F}^{\{ \cdot, B_{g'_d} \}} \right)^H \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{F}^{\{ \cdot, B_{g''_d} \}} S_{g_{d,k},i}(\theta) d\theta + \frac{\sigma}{\tau_d p_d} \mathbf{I}_{b_d} \right)^{-1} \\ &\quad \times \sum_{g'_d \in G_d} \sum_{i=1}^{M_d} \int_{\theta_{g_{d,k},i}^{\min}}^{\theta_{g_{d,k},i}^{\max}} \left(\mathbf{F}^{\{ \cdot, B_{g'_d} \}} \right)^H \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{F}^{\{ \cdot, B_{g'_d} \}} S_{g_{d,k},i}(\theta) d\theta\end{aligned}\quad (42)$$

Moreover, as $N \rightarrow \infty$, we have [3]

$$\begin{aligned} \frac{1}{N} \left(\tilde{\mathbf{h}}_{g_u,k,\text{LM}}^{\{B_{g_u},:\}} \right)^H \tilde{\mathbf{h}}_{g_u,k,\text{LM}}^{\{B_{g_u},:\}} &= \frac{1}{N} \text{tr} \left(\tilde{\mathbf{C}}_{g_u,k,\text{LM}}^{g_u} \right) \\ \frac{1}{N} \left(\tilde{\mathbf{h}}_{g_d,k,\text{LM}}^{\{B_{g_d},:\}} \right)^H \tilde{\mathbf{h}}_{g_d,k,\text{LM}}^{\{B_{g_d},:\}} &= \frac{1}{N} \text{tr} \left(\tilde{\mathbf{C}}_{g_d,k,\text{LM}}^{g_d} \right) \end{aligned} \quad (43)$$

With (41), (42), and (43), the scaling behaviors for average powers of useful signal, channel estimation, IUI, and IGI can be readily obtained by using the technique in [38, Proof of Theorem 4] as shown in (44) at the bottom of the page. Then we focus on the scaling behavior of SI power. According to Table 3, the asymptotic result in (43) and the property $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, we can rewrite the average SI power as

$$\begin{aligned} \mathbb{E} [\text{SI}_{g_u,k}] &= \left(\text{tr} \left(\tilde{\mathbf{C}}_{g_u,k,\text{LM}}^{g_u} \right) \right)^{-2} \sum_{g'_d \in G_d} \sum_{k'=1}^{K_{g'_d}} p_{g'_d,k'} \left(\text{tr} \left(\tilde{\mathbf{C}}_{g'_d,k',\text{LM}}^{g'_d} \right) \right)^{-1} \\ &\times \text{tr} \left(\underbrace{\mathbb{E} \left[\tilde{\mathbf{h}}_{g_u,k,\text{LM}}^{\{B_{g_u},:\}} \left(\tilde{\mathbf{h}}_{g_u,k,\text{LM}}^{\{B_{g_u},:\}} \right)^H \tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \tilde{\mathbf{h}}_{g'_d,k',\text{LM}}^{\{B_{g'_d},:\}} \left(\tilde{\mathbf{h}}_{g'_d,k',\text{LM}}^{\{B_{g'_d},:\}} \right)^H \left(\tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \right)^H \right]}_{X_{g_u,k,g'_d,k}} \right) \end{aligned} \quad (44)$$

The expression of $X_{g_u,k,g'_d,k}$ can be rewritten as

$$\begin{aligned} X_{g_u,k,g'_d,k} &= \text{tr} \left(\mathbb{E} \left[\tilde{\mathbf{C}}_{g_u,k,\text{LM}}^{g_u} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \tilde{\mathbf{C}}_{g'_d,k',\text{LM}}^{g'_d} \left(\tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \right)^H \right] \right) \\ &\leq \text{tr} \left(\mathbb{E} \left[\tilde{\mathbf{C}}_{g_u,k}^{g_u} \tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \tilde{\mathbf{C}}_{g'_d,k'}^{g'_d} \left(\tilde{\mathbf{H}}_{SI}^{\{B_{g_u},B_{g'_d}\}} \right)^H \right] \right) \\ &= \text{tr} \left(\tilde{\mathbf{C}}_{g_u,k}^{g_u} \sum_{i=1}^{M_{SI}} \int_{\theta_{R,i}^{\min}}^{\theta_{R,i}^{\max}} \int_{\theta_{T,i}^{\min}}^{\theta_{T,i}^{\max}} S_{SI,i}(\theta_R, \theta_T) \left(\mathbf{F}^{\{:,B_{g_u}\}} \right)^H \mathbf{a}(\theta_R) \mathbf{a}^H(\theta_T) \right. \\ &\quad \left. \times \mathbf{F}^{\{:,B_{g_d}\}} \tilde{\mathbf{C}}_{g'_d,k'}^{g'_d} \left(\mathbf{F}^{\{:,B_{g'_d}\}} \right)^H \mathbf{a}(\theta_T) \mathbf{a}^H(\theta_R) \mathbf{F}^{\{:,B_{g_u}\}} d\theta_R d\theta_T \right) \\ &\leq I_R \times I_T \max_{\theta_R \in \bigcup_{i=1}^{M_{SI}} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}], \theta_T \in \bigcup_{i=1}^{M_{SI}} [\theta_{T,i}^{\min}, \theta_{T,i}^{\max}]} S_{SI,i}(\theta_R, \theta_T) \end{aligned} \quad (45)$$

where

$$I_R = \int_{\theta_R \in \bigcup_{i=1}^{MSI} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}]} \mathbf{a}^H(\theta_R) \mathbf{F}^{(:,B_{gu})} \tilde{\mathbf{C}}_{g_{u,k}}^{g_u} \left(\mathbf{F}^{(:,B_{gu})} \right)^H \mathbf{a}(\theta_R) d\theta_R$$

$$I_T = \int_{\theta_T \in \bigcup_{i=1}^{MSI} [\theta_{T,i}^{\min}, \theta_{T,i}^{\max}]} \mathbf{a}^H(\theta_T) \mathbf{F}^{(:,B_{g'_d})} \tilde{\mathbf{C}}_{g_{d',k'}}^{g'_d} \left(\mathbf{F}^{(:,B_{g'_d})} \right)^H \mathbf{a}(\theta_T) d\theta_T \quad (46)$$

In (45), the first step is based on the independence between $\tilde{\mathbf{h}}_{g_{u,k},\text{LM}}^{\{B_{gu},:\}}$, $\tilde{\mathbf{h}}_{g'_{d,k},\text{LM}}^{\{B_{g'_d},:\}}$ and $\tilde{\mathbf{H}}_{SI}^{\{B_{gu},B_{g'_d}\}}$. The second step is based on the relation $\tilde{\mathbf{C}}_{g_{m,k},\text{LM}}^{g_m} = \tilde{\mathbf{C}}_{g_{m,k}}^{g_m} - \left(\tilde{\mathbf{C}}_{g_{m,k}}^{g_m} - \tilde{\mathbf{C}}_{g_{m,k},\text{LM}}^{g_m} \right)$ ($m \in \{u, d\}$) and the positive definiteness of $\tilde{\mathbf{C}}_{g_{m,k}}^{g_m}$ and $\tilde{\mathbf{C}}_{g_{m,k},\text{LM}}^{g_m} - \tilde{\mathbf{C}}_{g_{m,k},\text{LM}}^{g_m}$. The third step is obtained by using the equation $\tilde{\mathbf{H}}_{SI}^{\{B_{gu},B_{g'_d}\}} = \left(\mathbf{F}^{(:,B_{gu})} \right)^H \mathbf{H}_{SI} \mathbf{F}^{(:,B_{g'_d})}$ and the SI channel model (3).

By substituting (41) into (46), the integral I_R can be rewritten as

$$\begin{aligned} \frac{1}{N} I_R &= \frac{1}{N} \sum_{i=1}^{M_u} \int_{\theta_{g_{u,k},i}^{\min}}^{\theta_{g_{u,k},i}^{\max}} S_{g_{u,k},i}(\theta) \int_{\theta_R \in \bigcup_{i=1}^{MSI} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}]} \mathbf{a}^H(\theta_R) \mathbf{a}(\theta) \mathbf{a}^H(\theta) \mathbf{a}(\theta_R) d\theta_R d\theta \\ &= N \sum_{i=1}^{M_u} \int_{\theta_{g_{u,k},i}^{\min}}^{\theta_{g_{u,k},i}^{\max}} S_{g_{u,k},i}(\theta) \int_{\theta_R \in \bigcup_{i=1}^{MSI} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}]} \text{sinc}_N^2 \left(\frac{d}{\lambda} \sin \theta_R - \frac{d}{\lambda} \sin \theta \right) d\theta_R d\theta \end{aligned} \quad (47)$$

Note that according to Lemma 1 and Lemma 2, we have $\bigcap_{i=1}^{M_u} [\theta_{g_{u,k},i}^{\min}, \theta_{g_{u,k},i}^{\max}] \cap \bigcap_{i=1}^{MSI} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}] = \emptyset$ or $\bigcap_{i=1}^{M_d} [\theta_{g_{d,k},i}^{\min}, \theta_{g_{d,k},i}^{\max}] \cap \bigcap_{i=1}^{MSI} [\theta_{T,i}^{\min}, \theta_{T,i}^{\max}] = \emptyset$, otherwise, Criterion 2 will be violated. If $\bigcap_{i=1}^{M_u} [\theta_{g_{u,k},i}^{\min}, \theta_{g_{u,k},i}^{\max}] \cap \bigcap_{i=1}^{MSI} [\theta_{R,i}^{\min}, \theta_{R,i}^{\max}] = \emptyset$, with a same procedure as that in the proof of Lemma 1, we can obtain $\frac{1}{N} I_R < \mathcal{O}(1)$ or $I_R < \mathcal{O}(N)$ as $N \rightarrow \infty$; otherwise, $I_R = \mathcal{O}(N)$. In the same way, we can prove that $I_T < \mathcal{O}(N)$ if $\bigcap_{i=1}^{M_d} [\theta_{g_{d,k},i}^{\min}, \theta_{g_{d,k},i}^{\max}] \cap \bigcap_{i=1}^{MSI} [\theta_{T,i}^{\min}, \theta_{T,i}^{\max}] = \emptyset$, and $I_T = \mathcal{O}(N)$ otherwise. Combining the results in the above, we have $I_R \times I_T < \mathcal{O}(N^2)$.

Moreover, it has been shown in [38] that $\text{tr}(\tilde{\mathbf{C}}_{g_{u,k},\text{LM}}^{g_u})$ and $\text{tr}(\tilde{\mathbf{C}}_{g_{d',k'},\text{LM}}^{g_{d'}})$ scale with $\mathcal{O}(N)$ as $N \rightarrow \infty$. Using these results on (44), we have $\mathbb{E}[\text{SI}_{g_{u,k}}] < \mathcal{O}(N^{-1})$, which is exactly the result in Table 3.

References

1. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of BS antennas. *IEEE Trans. Wirel. Commun.* **9**(11), 3590–3600 (2010)
2. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. *IEEE Trans. Commun.* **61**(4), 1436–1449 (2013)
3. J. Hoydis, S. Brink, M. Debbah, Massive MIMO in UL/DL of cellular networks: how many antennas do we need? *IEEE J. Sel. Areas Commun.* **31**(2), 160–171 (2013)
4. J. Zhang, C.K. Wen, S. Jin, X. Gao, K. Wong, On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas. *IEEE J. Sel. Areas Commun.* **31**(2), 133–148 (2013)
5. H. Cui, L. Song, B. Jiao, Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas. *IEEE Trans. Wirel. Commun.* **13**(5), 2636–2645 (2014)
6. L. You, X. Gao, X.G. Xia, N. Ma, Y. Peng, Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels. *IEEE Trans. Wirel. Commun.* **14**(6), 3352–3366 (2015)
7. S. Jin, X. Liang, K.K. Wong, X. Gao, Q. Zhu, Ergodic rate analysis for multipair massive MIMO two-way relay networks. *IEEE Trans. Wirel. Commun.* **14**(3), 1480–1491 (2015)
8. A. Adhikary, J. Nam, J.-Y. Ahn, G. Caire, Joint spatial division and multiplexing: the large-scale array regime. *IEEE Trans. Inf. Theory* **59**(10), 6441–6463 (2013)
9. C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, C. Xiao, Beam division multiple access transmission for massive MIMO communications. *IEEE Trans. Commun.* **63**(6), 2170–2184 (2015)
10. A. Liu, V. Lau, Phase only RF precoding for massive MIMO systems with limited RF chains. *IEEE Trans. Signal Process.* **62**(17), 4505–4515 (2014)
11. D. Kim, G. Lee, Y. Sung, Two-stage beamformer design for massive MIMO downlink by trace quotient formulation. *IEEE Trans. Commun.* **63**(6), 2200–2211 (2015)
12. Y. Jang, K. Min, S. Park, S. Choi, Spatial resource utilization to maximize uplink spectral efficiency in full-duplex massive MIMO. In *Proc. IEEE Int. Conf. Commun.*, London, UK, 2015, pp. 1583–1588
13. Y. Li, P. Fan, L. Anatolii, L. Liu, On the spectral and energy efficiency of full-duplex small cell wireless systems with massive MIMO. *IEEE Trans. Veh. Technol.* **66**(3), 2339–2353 (2017)
14. H. Tabassum, A.H. Sakr, E. Hossain, Massive MIMO-enabled wireless backhauls for full-duplex small cells. In *Proc. IEEE Global Commun. Conf.*, San Diego, 2015, pp. 1–6
15. R. Mai, D.H.N. Nguyen, T. Le-Ngoc, Joint MSE-based hybrid precoder and equalizer design for full-duplex massive MIMO systems. In *Proc. IEEE Int. Conf. Commun.*, Kuala Lumpur, Malaysia, 2016, pp. 1–6
16. B. Li, D. Zhu, P. Liang, Small cell in-band wireless backhaul in massive MIMO systems: a cooperation of next-generation techniques. *IEEE Trans. Wirel. Commun.* **14**(12), 7057–7069 (2015)
17. A. Sabharwal, P. Schniter, D. Guo, D.W. Bliss, S. Rangarajan, R. Wichman, In-band full-duplex wireless: challenges and opportunities. *IEEE J. Sel. Areas Commun.* **32**(9), 1637–1652 (2014)
18. H.Q. Ngo, H.A. Suraweera, M. Matthaiou, E.G. Larsson, Multipair full-duplex relaying with massive arrays and linear processing. *IEEE J. Sel. Areas Commun.* **32**(9), 1721–1737 (2014)

19. G. Zheng, Joint beamforming optimization and power control for full duplex MIMO two-way relay channel. *IEEE Trans. Signal Process.* **63**(3), 555–566 (2015)
20. D. Bharadia, E. McMillin, S. Katti, Full duplex radios. In *Proc ACM Sigcomm*, Hong Kong, China, Aug. 2013, pp. 375–386
21. 3GPP, Universal mobile telecommunications system (UMTS); spatial channel model for multiple input multiple output (MIMO) simulations. In *v.12.0.0, Tech. Rep.* TR 25.996, Jun. 2012. [Online]. Available: www.3gpp.org
22. K.I. Pedersen, P.E. Mogensen, B.H. Fleury, A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments. *IEEE Trans. Veh. Technol.* **49**(2), 437–447 (2000)
23. G.B. Giannakis, C. Tepedelenlioglu, Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proc. IEEE* **86**(10), 1969–1986 (1998)
24. S. Goyal, P. Liu, S.S. Panwar, R.A. Difazio, R. Yang, E. Bala, Full duplex cellular systems: will doubling interference prevent doubling capacity? *IEEE Commun. Mag.* **53**(5), 121–127 (2015)
25. E. Everett, A. Sahai, aA. Sabharwal, Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Trans. Wirel. Commun.* **13**(2), 680–694 (2014)
26. J. Singh, S. Ramakrishna, On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays. *IEEE Trans. Wirel. Commun.* **14**(5), 2670–2683 (2015)
27. Z.M. Liu, Y.Y. Zhou, A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections. *IEEE Trans. Signal Process.* **61**(15), 3786–3798 (2013)
28. J.C. Shen, J. Zhang, E. Alsusa, K.B. Letaief, Compressed CSI acquisition in FDD massive MIMO with partial support information. In *Proc. IEEE Int. Conf. Commun.*, London, UK, Jun. 2015, pp. 1459–1464
29. J.O. Smith, *Spectral Audio Signal Processing*, W3, 2011. [Online]. Available: <http://books.w3k.org/>
30. S. Vishwanath, N. Jindal, A. Goldsmith, Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans. Inf. Theory* **49**(10), 2658–2668 (2003)
31. A. Barg, D.Y. Nogin, Bounds on packings of spheres in the Grassmann manifold. *IEEE Trans. Inf. Theory* **48**(9), 2450–2454 (2002)
32. S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Englewood Cliffs, 1993)
33. G. Caire, N. Jindal, M. Kobayashi, N. Ravindran, Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Trans. Inf. Theory* **56**(6), 2845–2866 (2010)
34. D. Nguyen, L.-N. Tran, P. Pirinen, M. Latva-aho, Precoding for full duplex multiuser MIMO systems: Spectral and energy efficiency maximization. *IEEE Trans. Signal Process.* **61**(16), 4038–4050 (2013)
35. B. Hassibi, B.M. Hochwald, How much training is needed in multiple antenna wireless links? *IEEE Trans. Inf. Theory* **49**(4), 951–963 (2003)
36. 3GPP, Further enhancements to LTE time division duplex (TDD) for downlink-uplink (DL-UL) interference management and traffic adaptation. In *v.11.0.0, Tech. Rep.* 36.828, Jun. 2012. [Online]. Available: www.3gpp.org
37. H.A. Suraweera, I. Krikidis, G. Zheng, C. Yuen, P.J. Smith, Low complexity end-to-end performance optimization in MIMO full-duplex relay systems. *IEEE Trans. Wirel. Commun.* **13**(2), 913–927 (2014)
38. X. Xia, D. Zhang, K. Xu, W. Ma, Y. Xu, Hardware impairments aware transceiver for full-duplex massive MIMO relaying. *IEEE Trans. Signal Process.* **63**(24), 6565–6580 (2015)



Kui Xu was born in 1982. He received his B.S. degree in wireless communications and Ph.D. degree in software-defined radio from the PLA University of Science and Technology, Nanjing, China, in 2004 and 2009, respectively. He is currently an associate professor in the College of Communications Engineering, Army Engineering University of PLA. Since 2013, he has been a postdoctoral fellow with the PLA University of Science and Technology. His research interests include broadband wireless communications, signal processing for communications, network coding, and wireless communication networks. He has authored about 50 papers in refereed journals and conference proceedings and holds five patents in China. He is currently serving on the Technical Program Committee of the IEEE WCSP 2014. He received the URSI Young Scientists Award in 2014 and the 2010 Ten Excellent Doctor Degree Dissertation Award of PLAUST. He also serves as the reviewer of the IEEE Transactions on Wireless Communication, the IEEE Transaction Vehicle Technology, the IEEE Communications Letter, and the IEEE Signal Processing Letters.



Xiaochen Xia was born in China in 1987. He received his B.S. degree in electronic science and technology from Tianjin University (TJU) in 2010, and M.S. degree in communication and information system from Army Engineering University of PLA, in 2013. He is currently working toward the Ph.D. degree in Institution of Communications Engineering, Army Engineering University of PLA. His research interests include relaying network, full-duplex communication, network coding, and MIMO techniques. He received the 2013 excellent master degree dissertation award of Jiangsu Province, China.



Yurong Wang was born in China in 1990. She received her B.S. degree in computer science and technology from Beijing Institute of Technology (BIT) in 2013. She is now working toward the Ph.D. degree in Institution of Communications Engineering, Army Engineering University of PLA. Her research interests include full-duplex communication, massive MIMO system, and broadband wireless communications.



Wei Xie received his M.S. degree in communication and information system from PLA University of Science and Technology (PLAUST) in 2002. He is currently a lecturer in the PLAUST. His research interests include wireless communication systems and networks, cooperative communications, and cognitive radio.



Dongmei Zhang was born in 1972. She received her B.S. and M.S. degrees in communication engineering in 1993 and 2005, respectively, and Ph.D. degree in communications and information system from the PLA University of Science and Technology, Nanjing, China, in 2013. She is currently a professor in the College of Communications Engineering, Army Engineering University of PLA, Nanjing, China. Her research interests include new generation wireless mobile communication system, radio resource management, and network coding in wireless communication.