

NL-BMD: Nonlinear Block Multi-Diagonalization Precoding for High SHF Wide-Band Massive MIMO in 5G

Hiroshi NISHIMOTO^{†a)}, Member, Akinori TAIRA[†], Senior Member, Hiroki IURA[†], Shigeru UCHIDA[†], Members, Akihiro OKAZAKI[†], and Atsushi OKAMURA[†], Senior Members

SUMMARY Massive multiple-input multiple-output (MIMO) technology is one of the key enablers in the fifth generation mobile communications (5G), in order to accommodate growing traffic demands and to utilize higher super high frequency (SHF) and extremely high frequency (EHF) bands. In the paper, we propose a novel transmit precoding named “nonlinear block multi-diagonalization (NL-BMD) precoding” for multiuser MIMO (MU-MIMO) downlink toward 5G. Our NL-BMD precoding strategy is composed of two essential techniques: block multi-diagonalization (BMD) and adjacent inter-user interference pre-cancellation (IUI-PC). First, as an extension of the conventional block diagonalization (BD) method, the linear BMD precoder for the desired user is computed to incorporate a predetermined number of interfering users, in order to ensure extra degrees of freedom at the transmit array even after null steering. Additionally, adjacent IUI-PC, as a nonlinear operation, is introduced to manage the residual interference partially allowed in BMD computation, with effectively-reduced numerical complexity. It is revealed through computer simulations that the proposed NL-BMD precoding yields up to 67% performance improvement in average sum-rate spectral efficiency and enables large-capacity transmission regardless of the user distribution, compared with the conventional BD precoding.

key words: 5G, massive MIMO, MU-MIMO, inter-user interference, precoding, nonlinear block multi-diagonalization

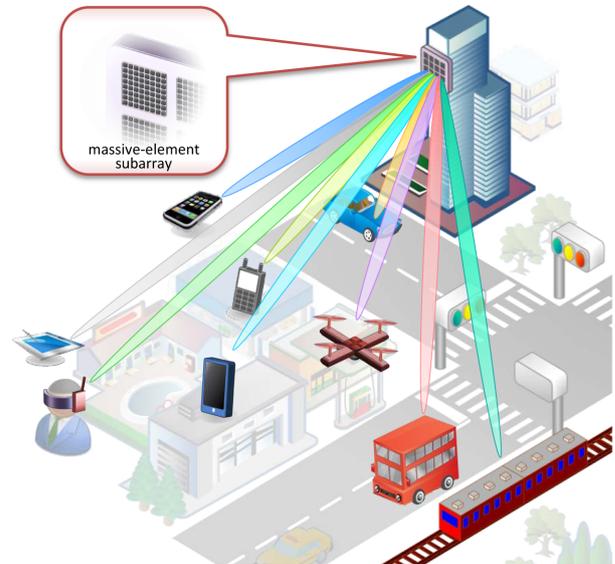


Fig. 1 Envisioned use case of massive MIMO in 5G.

1. Introduction

With the mobile traffic in the early 2020’s being predicted to be 1,000 times greater than that of 2010, the fifth generation mobile communications (5G) are expected to accommodate the forthcoming huge traffic demands [1], [2]. Massive multiple-input multiple-output (MIMO) technology utilizing hundreds of antenna elements has drawn attention as a key antenna configuration for envisioned 5G [3]–[6]. It enables us to drastically earn antenna beam gain and spatial multiplexing capability, and its benefit is of significance especially in higher super high frequency (SHF) and extremely high frequency (EHF) bands. Specifically, multiuser-MIMO (MU-MIMO) must be a spatially-efficient wireless access system exploiting the benefit of massive MIMO, where a base station (BS) equipped with multiple antennas simultaneously makes co-channel communications with multiple users equipped with multiple antennas. Figure 1 illustrates an envisioned use case of massive MIMO in 5G, where we assume that the antenna array at the BS is com-

posed of massive-element subarrays [7], of which each can direct a sharp beam to the corresponding user. As discussed in [7], we believe that the subarray-type analog-digital hybrid beamforming architecture, composed of analog beamforming by massive-element subarrays and digital MIMO signal processing, is a promising solution for realizing massive MIMO in practice.

In MU-MIMO downlink (DL), inter-user interference (IUI) observed at each user is an essential issue due to the broadcasting from BS even if using sharp beams. It is desirable to mitigate IUI at BS in advance because there is a more computational room at BS. Therefore, studies on precoding or preprocessing for IUI reduction have been reported in the past literature.

Our major objective is to have stable and large-capacity links over MU-MIMO DL systems. So we need to establish highly-spatial multiplexing with stability regardless of the given spatial condition. From a viewpoint of physical layer, it is important for the stability to ensure diversity gain in addition to management of IUI.

Approaches of precoding for MU-MIMO DL can be roughly classified into two strategies: i) linear precoding (LP) schemes segmentalizing a channel into per-user beamformed spaces, and ii) nonlinear precoding (NLP) approaches using nonlinear (NL) operation for IUI, namely IUI

Manuscript received August 29, 2016.

Manuscript revised December 13, 2016.

Manuscript publicized February 8, 2017.

[†]The authors are with Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura-shi, 247-8501 Japan.

a) E-mail: hn@m.ieice.org

DOI: 10.1587/transcom.2016FGP0003

pre-cancellation (IUI-PC).

LP can be conventionally regarded as transmit (TX) beamforming at BS. Block diagonalization (BD) is well known as its typical case accomplishing perfect nulling, where we direct prescribed nulls to users except for the target user [8]–[11]. BD enables us to realize IUI-free DL situation in a spatially-uncorrelated scenario and to ease receiver designing. However, since most of the spatial resources at BS are consumed to direct nulls in the BD system, we cannot expect extra TX diversity gain so that the performance of BD may deteriorate considerably in ill-conditioned or spatially-correlated channels. Other LP schemes such as imperfect BD [12], generalized BD [13], regularized BD (RBD) [14], [15], signal-to-leakage-plus-noise power ratio (SLNR) based precoding [16], [17], etc., have been proposed to obtain extra degrees of freedom at BS, by facilitating the constraint on the number of antennas, and also to improve receiving gain. In a scenario with many users, however, LP schemes would need an additional countermeasure for more severe IUI. Although it is known that optimum LP is asymptotically provided by joint iterative optimization of both precoding and postcoding [18], from a practical viewpoint the computation of LP should be uniquely solved with fixed and reasonable numerical complexity.

Alternatively, NLP is a strategy realizing IUI-free DL transmission by canceling IUI observed at users in advance. It achieves near-capacity performance and establishes robust DL transmission irrespective of the given scenario. We note that spatial stability brought by NLP is important to relax constraints in scheduling operation. Dirty paper coding (DPC) is known as an ideal NLP scheme [19], and successive IUI-PC assisted by user-space hierarchization (LQ factorization) is known as its practical scheme [20]. Since signal pre-cancellation causes an increase in amplitude of TX signal chains, we generally need a signal limitation method such as modulo operation performed in Tomlinson-Harashima precoding (THP) [21]–[23], or vector perturbation (VP) applying adaptive signal offset [24]–[26]. Although NLP brings high system throughput in principle, we have to solve the critical issue of high complexity for IUI-PC, of which order is $O(N_{\text{usr}}^2)$, where N_{usr} denotes the number of users.

In [14], the concept of RBD precoding and optimization of system capacity by combining RBD and THP have been discussed. While we agree with an approach to manage IUI through both the LP and NLP strategies, ambiguity in determining RBD precoders and iterative optimization would lead to difficulty in designing device implementation.

Addressing the issues above, in [27] the authors have proposed an alternative LP concept named block multi-diagonalization (BMD) for scenarios with a lot of users. The BMD precoder for the desired user is computed to incorporate a predetermined number of interfering users with fixed complexity comparable to the conventional BD. Consequently, the BMD concept enables us to ensure extra degrees of freedom at the TX array even after null steering. Prior to a study on practical management schemes for residual IUI, in [27] the authors have clarified the significant potential of

the BMD concept through the discussion and numerical evaluation on block bi-diagonalization (BBD) and block lower tri-diagonalization (BLTD), both of which are in the same family of the BMD concept as discussed later.

Based on the background, in order to manage the residual IUI, we here propose a more practical TX precoding named “nonlinear block multi-diagonalization (NL-BMD) precoding,” composed of BMD and adjacent IUI-PC. Technically, by limiting IUI to be pre-canceled to the adjacent users allowed in BMD, the proposed NL-BMD precoding offers advantages especially in scenarios with many users because the total complexity in IUI-PC can be reduced to the order of $O(N_{\text{usr}})$. It is believed that, therefore, the NL-BMD precoding displays its potential force in a dense urban environment, where we have a lot of users simultaneously serviced over spatially-correlated channels. In this paper, through evaluation over orthogonal frequency division multiplexing (OFDM), we reveal the capability of NL-BMD strategy by introducing two reasonable approaches: nonlinear BBD (NL-BBD) and nonlinear BLTD (NL-BLTD).

The rest of the paper is organized as follows. First the handled MU-MIMO DL system model is defined in Sect. 2, and Sect. 3 reviews typical LP and NLP schemes being extensively studied. Section 4 then explains the proposed NL-BMD precoding, and its performance is numerically demonstrated in Sect. 5. Finally, Sect. 6 concludes the paper.

2. MU-MIMO DL System Model

We assume a mobile cellular system where all the equipment is under synchronous control, and the issue we handle is to spatially multiplex signals to the users determined by a scheduler in BS. Also BS is assumed to have DL channel state information of all users, exploiting channel reciprocity in time division duplex (TDD). Let us consider the MU-MIMO DL system where BS has N_{tx} TX baseband ports, user $\#i$ has $N_{\text{rx},i}$ receive (RX) baseband ports, and the number of existing users is N_{usr} . We define the number of system RX ports, namely total RX baseband ports of all users, as $N_{\text{rx}} = \sum_{i=1}^{N_{\text{usr}}} N_{\text{rx},i}$. Also, we assume that $N_{\text{st},i}$ ($1 \leq N_{\text{st},i} \leq N_{\text{rx},i}$) substreams are transmitted to user $\#i$. Note that the number of total substreams transmitted from BS is denoted as $N_{\text{st}} = \sum_{i=1}^{N_{\text{usr}}} N_{\text{st},i}$, and that the number of baseband ports satisfies the condition of $N_{\text{tx}} \geq N_{\text{rx}}$ for fair discussion including conventional LP schemes[†].

Although we deal with MU-MIMO over OFDM transmission in the paper, hereafter let us simplify the discussion by focusing on a subcarrier. Note that the following procedure is common to all the subcarriers. Assuming transmission at a certain subcarrier, for user $\#i$ we here define: an $N_{\text{st},i}$ -dimensional TX signal vector $\mathbf{s}_i(t)$, an $N_{\text{tx}} \times N_{\text{st},i}$ TX precoding (beamforming) matrix \mathbf{B}_i , an $N_{\text{rx},i} \times N_{\text{tx}}$ channel matrix \mathbf{H}_i , an $N_{\text{rx},i}$ -dimensional RX signal vector $\mathbf{r}_i(t)$, and

[†]In this section we presume the condition for the sake of simple discussion whereas in computer simulations discussed in Sect. 5 the number of beam-space RX antennas satisfies the condition instead.

an $N_{\text{rx},i}$ -dimensional additive white Gaussian noise (AWGN) vector $\mathbf{n}_i(t)$. Then the system can be briefly modeled as follows:

$$\bar{\mathbf{r}}(t) = \mathcal{H}\mathcal{B}\bar{\mathbf{s}}(t) + \bar{\mathbf{n}}(t), \quad (1)$$

where,

$$\bar{\mathbf{r}}(t) = \begin{bmatrix} \mathbf{r}_1^T(t) & \mathbf{r}_2^T(t) & \cdots & \mathbf{r}_{N_{\text{usr}}}^T(t) \end{bmatrix}^T, \quad (2)$$

$$\mathcal{H} = \begin{bmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T & \cdots & \mathbf{H}_{N_{\text{usr}}}^T \end{bmatrix}^T, \quad (3)$$

$$\mathcal{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_{N_{\text{usr}}} \end{bmatrix}, \quad (4)$$

$$\bar{\mathbf{s}}(t) = \begin{bmatrix} \mathbf{s}_1^T(t) & \mathbf{s}_2^T(t) & \cdots & \mathbf{s}_{N_{\text{usr}}}^T(t) \end{bmatrix}^T, \text{ and} \quad (5)$$

$$\bar{\mathbf{n}}(t) = \begin{bmatrix} \mathbf{n}_1^T(t) & \mathbf{n}_2^T(t) & \cdots & \mathbf{n}_{N_{\text{usr}}}^T(t) \end{bmatrix}^T, \quad (6)$$

respectively. Note that the matrices \mathcal{H} and \mathcal{B} denote $N_{\text{rx}} \times N_{\text{tx}}$ system channel matrix and $N_{\text{tx}} \times N_{\text{st}}$ system precoding matrix, respectively. Hence the product of \mathcal{H} and \mathcal{B} , namely $\mathcal{H}_e = \mathcal{H}\mathcal{B}$, can be regarded as an effective system channel matrix:

$$\mathcal{H}_e = \begin{bmatrix} \mathbf{H}_1\mathbf{B}_1 & \mathbf{H}_1\mathbf{B}_2 & \cdots & \mathbf{H}_1\mathbf{B}_{N_{\text{usr}}} \\ \mathbf{H}_2\mathbf{B}_1 & \mathbf{H}_2\mathbf{B}_2 & \cdots & \mathbf{H}_2\mathbf{B}_{N_{\text{usr}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_1 & \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_2 & \cdots & \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_{N_{\text{usr}}} \end{bmatrix}. \quad (7)$$

Among submatrices (block entries) in the matrix \mathcal{H}_e , the block diagonal entries $\mathbf{H}_i\mathbf{B}_i$ ($i = 1, \dots, N_{\text{usr}}$) mean the desired effective channel components for users, and the other entries (block off-diagonal entries) $\mathbf{H}_i\mathbf{B}_j$ ($i \neq j$) represent IUI channel components. LP is to spatially reduce IUI channel components by the precoder \mathcal{B} , and NLP is to pre-cancel IUI signal observed at users in addition to LP.

3. Conventional Precoding Schemes

3.1 Block Diagonalization

This subsection explains the BD scheme, which is the typical LP, by showing how to compute the user $\#i$'s precoding matrix $\mathbf{B}_{\text{bd},i}$. We first consider an $(N_{\text{rx}} - N_{\text{rx},i}) \times N_{\text{tx}}$ channel matrix $\mathcal{H}_{\text{bd}}^{(i)}$ composed of user channel matrices except for user $\#i$'s one, and its singular value decomposition (SVD) can be expressed as

$$\mathcal{H}_{\text{bd}}^{(i)} = \begin{bmatrix} \mathbf{H}_1^T & \cdots & \mathbf{H}_{i-1}^T & \mathbf{H}_{i+1}^T & \cdots & \mathbf{H}_{N_{\text{usr}}}^T \end{bmatrix}^T \quad (8)$$

$$= \mathbf{U}_{\text{bd}}^{(i)} \begin{bmatrix} \Sigma_{\text{bd}[s]}^{(i)} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\text{bd}[s]}^{(i)H} \\ \mathbf{V}_{\text{bd}[n]}^{(i)H} \end{bmatrix}, \quad (9)$$

where $\mathbf{U}_{\text{bd}}^{(i)}$ is an $(N_{\text{rx}} - N_{\text{rx},i})$ -dimensional unitary matrix composed of left singular vectors, and an $(N_{\text{rx}} - N_{\text{rx},i}) \times N_{\text{tx}}$ singular value matrix consists of an $(N_{\text{rx}} - N_{\text{rx},i})$ -dimensional diagonal submatrix $\Sigma_{\text{bd}[s]}^{(i)}$, of which diagonal entries are non-negative singular values, and a zero matrix \mathbf{O} . Also, an

N_{tx} -dimensional matrix composed of right singular vectors can be classified into submatrices $\mathbf{V}_{\text{bd}[s]}^{(i)}$ and $\mathbf{V}_{\text{bd}[n]}^{(i)}$, which correspond to the signal space $\Sigma_{\text{bd}[s]}^{(i)}$ and the kernel space \mathbf{O} , respectively. When $\mathbf{V}_{\text{bd}[n]}^{(i)}$ is set to a TX precoder $\mathbf{B}_{\text{bd},i}$, we can direct perfect nulls to all users except for user $\#i$ because the vector space spanned by $\mathbf{V}_{\text{bd}[n]}^{(i)}$ is the mapper to the kernel. Consequently, the precoder $\mathbf{B}_{\text{bd},i}$ ($i = 1, \dots, N_{\text{usr}}$) achieves block diagonalization of the effective system channel matrix as below:

$$\mathcal{H}_{e,\text{bd}} = \mathcal{H}\mathcal{B}_{\text{bd}} \quad (10)$$

$$= \begin{bmatrix} \mathbf{H}_1\mathbf{B}_{\text{bd},1} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{H}_2\mathbf{B}_{\text{bd},2} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_{\text{bd},N_{\text{usr}}} \end{bmatrix}. \quad (11)$$

Over the block-diagonalized channel $\mathcal{H}_{e,\text{bd}}$, we can subsequently apply any kind of single-user (SU) MIMO precoding per user, i.e. intra-user precoding.

Although the BD scheme eliminates IUI, we cannot avoid spending degrees of freedom at the TX array upon the perfect nulling. So, it is not necessarily the case that the obtained TX beams can improve SNR at each user. In particular, we cannot expect extra TX array gain in a dense scenario due to the nulling to a lot of users. Besides, when users are in proximity, conventional LP approaches including the BD scheme generally result in less beam gain.

3.2 Joint Precoding of Block Triangulation and NL Operation

In NLP, user hierarchization is mandatory for practical IUI-PC. When block triangulation (BT) is applied onto the system channel matrix \mathcal{H} , the obtained unitary matrix \mathcal{B}_{bt} enables us to successively pre-cancel IUI [20] (see Appendix A). Using the system precoding matrix \mathcal{B}_{bt} yielding BT, we can obtain the effective system channel matrix as

$$\mathcal{H}_{e,\text{bt}} = \mathcal{H}\mathcal{B}_{\text{bt}} \quad (12)$$

$$= \begin{bmatrix} \mathbf{H}_1\mathbf{B}_{\text{bt},1} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{H}_2\mathbf{B}_{\text{bt},1} & \mathbf{H}_2\mathbf{B}_{\text{bt},2} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_{\text{bt},1} & \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_{\text{bt},2} & \cdots & \mathbf{H}_{N_{\text{usr}}}\mathbf{B}_{\text{bt},N_{\text{usr}}} \end{bmatrix}. \quad (13)$$

Now we briefly explain that \mathcal{B}_{bt} provides ideal MU-MIMO capacity. According to [28], the theoretical sum-rate capacity of MU-MIMO system can be given by

$$C_{\text{sum-rate}} = \log_2 \det \left(\mathbf{I}_{N_{\text{rx}}} + \frac{\gamma}{N_{\text{tx}}} \mathcal{H}_e \mathcal{H}_e^H \right) \quad (14)$$

$$= \log_2 \det \left(\mathbf{I}_{N_{\text{rx}}} + \frac{\gamma}{N_{\text{tx}}} \mathcal{H}\mathcal{B}\mathcal{B}^H\mathcal{H}^H \right), \quad (15)$$

where γ and $\mathbf{I}_{N_{\text{rx}}}$ mean average received SNR in the case of no precoding and N_{rx} -dimensional unit matrix, respectively.

Since \mathcal{B}_{bt} obtained by any approaches of block LQ factorization, e.g. the scheme shown in Appendix A, is a unitary matrix, it satisfies $\mathcal{B}_{bt}\mathcal{B}_{bt}^H = \mathbf{I}_{N_{tx}}$. That is, it can maintain the property of the given channel nature, and it accomplishes lossless linear mapping from TX-port space to user space. We can therefore achieve ideal sum-rate capacity even if applying BT precoding [27].

Joint precoding of BT and IUI-PC (hereafter referred to as NL-BT) can release MU-MIMO's potential. However, since IUI-PC causes a significant increase in TX signal amplitude, introduction of THP[†], where modulo operation limits the amplitude to a required threshold, would be a practical manner. Although use of a modulo operator at TX obliges all users to have the same modulo operator, its impact on hardware implementation can be kept low. When applying BT, the TX signal vector for user $\#i$ preprocessed by THP can be written as

NL-BT :

$$s'_i(t) = \text{modulo}_\tau \left[s_i(t) - (\mathbf{H}_i \mathbf{B}_{bt,i})^+ \sum_{j=1}^{i-1} \mathbf{H}_i \mathbf{B}_{bt,j} s'_j(t) \right], \quad (16)$$

where $(\cdot)^+$ denotes Moore-Penrose (MP) pseudo inverse, and $\text{modulo}_\tau(\cdot)$ means substream-wise modulo operation with modulo boundary τ defined as follows:

$$\text{modulo}_\tau(x) = \left(x_i - \left\lfloor \frac{x_i}{2\tau} + \frac{1}{2} \right\rfloor \cdot 2\tau \right) + j \left(x_q - \left\lfloor \frac{x_q}{2\tau} + \frac{1}{2} \right\rfloor \cdot 2\tau \right), \quad (17)$$

here $x = x_i + jx_q$ is a complex value. Unfortunately, for the desired user $\#i$, IUI-PC utilized together with BT needs removal of IUI signals from all the upper-layer users $\#1, \dots, \#(i-1)$, resulting in higher computational load in proportion to the number of users N_{usr} .

Note that hereinafter we proceed to a discussion assuming that IUI-PC includes modulo operation.

4. Nonlinear Block Multi-Diagonalization

Figure 2 illustrates TX and RX block diagrams in the proposed NL-BMD over OFDM. At the transmitter, a per-substream data sequence is encoded and interleaved before quadrature amplitude modulation (QAM) mapping. Through serial-to-parallel conversion (S/P), at each subcarrier NL operation (IUI-PC) and BMD precoding are applied to all users' QAM-modulated substreams, yielding MU-MIMO multiplexing. After inverse fast Fourier transform (IFFT), parallel-to-serial conversion (P/S), and addition of cyclic prefix (CP), OFDM signal chains are fed to TX ports. At each user (receiver), removal of CP, S/P, and

[†] Although in the past literature NL-BT using modulo operation is generally called as THP, we here distinguish NL-BT from THP because original THP [21], [22] consists of pre-cancellation (or pre-equalization) and modulo operation only.

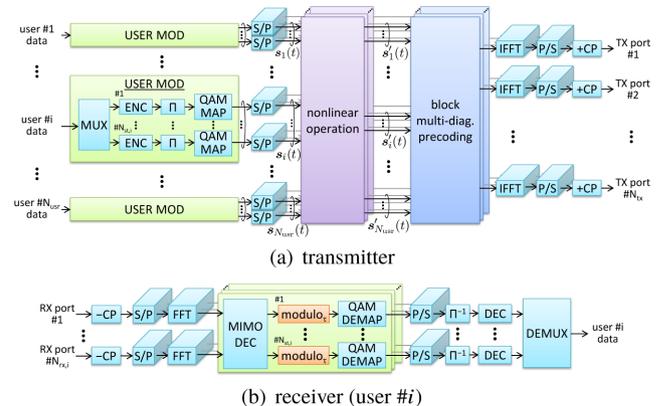


Fig. 2 Block diagrams of NL-BMD over OFDM (equivalent baseband system).

fast Fourier transform (FFT) are applied to received signal chains. In the frequency domain, the receiver performs demultiplexing and detection of the desired substreams, and then computes modulo operation and demapping at each subcarrier. Through P/S, each of estimated sequences is fed to subsequent decoding process.

The proposed NL-BMD is composed of two essential techniques: *i*) *BMD* which can provide extra TX diversity by allowing adjacent IUI, and *ii*) *adjacent IUI-PC* which cancels adjacent users' IUI only with effectively-reduced computational load compared with that in NL-BT. In the following each technique is described.

4.1 Block Multi-Diagonalization (BMD)

While the conventional BD scheme enables per-user spatially-closed signal transmission, it is difficult to ensure extra degrees of freedom at the TX array due to perfect nulling. Also, the BT scheme, with which the whole system space can be losslessly hierarchized per user, is effective for NLP whereas the cascaded IUI-PC process suffers from computational load in proportion to the number of users. Considering massive MIMO with subarray-type analog-digital hybrid beamforming configuration, beams directed to users from BS may be overlapped especially in a dense scenario. In this case we have severe IUI, and conventional LP schemes may sacrifice beam gains to mitigate the IUI.

Extending the BD approach, the BMD concept proposed in [27] is to partially allow IUI in the LP computation process for each desired user. Hereafter we refer to the users prescribed to have the allowed IUI as "IUI users." Partial IUI allowance brings residual degrees of freedom at the TX array even after nulling. Also, for NLP, the concept enables us to reduce the number of IUI signals to be pre-cancelled to the order of $\mathcal{O}(N_{usr})$. Although there are no restrictions on the number of IUI users in principle, in [27] we discussed two valid approaches: BBD, allowing one IUI user, and BLTD, allowing up to two IUI users, from an implementational viewpoint. Actually premising the aforementioned massive MIMO, dominant IUI would arise in a limited area close to

Table 1 Effective system channel matrix, \mathcal{H}_e ($N_{\text{usr}} = 5$).

(a) BD (conventional)					
	#1	#2	#3	#4	#5
#1	$\mathbf{H}_1 \mathbf{B}_1$	\mathbf{O}	\mathbf{O}	\mathbf{O}	\mathbf{O}
#2	\mathbf{O}	$\mathbf{H}_2 \mathbf{B}_2$	\mathbf{O}	\mathbf{O}	\mathbf{O}
#3	\mathbf{O}	\mathbf{O}	$\mathbf{H}_3 \mathbf{B}_3$	\mathbf{O}	\mathbf{O}
#4	\mathbf{O}	\mathbf{O}	\mathbf{O}	$\mathbf{H}_4 \mathbf{B}_4$	\mathbf{O}
#5	\mathbf{O}	\mathbf{O}	\mathbf{O}	\mathbf{O}	$\mathbf{H}_5 \mathbf{B}_5$

(b) BT (conventional)					
	#1	#2	#3	#4	#5
#1	$\mathbf{H}_1 \mathbf{B}_1$	\mathbf{O}	\mathbf{O}	\mathbf{O}	\mathbf{O}
#2	$\mathbf{H}_2 \mathbf{B}_1$	$\mathbf{H}_2 \mathbf{B}_2$	\mathbf{O}	\mathbf{O}	\mathbf{O}
#3	$\mathbf{H}_3 \mathbf{B}_1$	$\mathbf{H}_3 \mathbf{B}_2$	$\mathbf{H}_3 \mathbf{B}_3$	\mathbf{O}	\mathbf{O}
#4	$\mathbf{H}_4 \mathbf{B}_1$	$\mathbf{H}_4 \mathbf{B}_2$	$\mathbf{H}_4 \mathbf{B}_3$	$\mathbf{H}_4 \mathbf{B}_4$	\mathbf{O}
#5	$\mathbf{H}_5 \mathbf{B}_1$	$\mathbf{H}_5 \mathbf{B}_2$	$\mathbf{H}_5 \mathbf{B}_3$	$\mathbf{H}_5 \mathbf{B}_4$	$\mathbf{H}_5 \mathbf{B}_5$

(c) BBD					
	#1	#2	#3	#4	#5
#1	$\mathbf{H}_1 \mathbf{B}_1$	\mathbf{O}	\mathbf{O}	\mathbf{O}	\mathbf{O}
#2	$\mathbf{H}_2 \mathbf{B}_1$	$\mathbf{H}_2 \mathbf{B}_2$	\mathbf{O}	\mathbf{O}	\mathbf{O}
#3	\mathbf{O}	$\mathbf{H}_3 \mathbf{B}_2$	$\mathbf{H}_3 \mathbf{B}_3$	\mathbf{O}	\mathbf{O}
#4	\mathbf{O}	\mathbf{O}	$\mathbf{H}_4 \mathbf{B}_3$	$\mathbf{H}_4 \mathbf{B}_4$	\mathbf{O}
#5	\mathbf{O}	\mathbf{O}	\mathbf{O}	$\mathbf{H}_5 \mathbf{B}_4$	$\mathbf{H}_5 \mathbf{B}_5$

(d) BLTD					
	#1	#2	#3	#4	#5
#1	$\mathbf{H}_1 \mathbf{B}_1$	\mathbf{O}	\mathbf{O}	\mathbf{O}	\mathbf{O}
#2	$\mathbf{H}_2 \mathbf{B}_1$	$\mathbf{H}_2 \mathbf{B}_2$	\mathbf{O}	\mathbf{O}	\mathbf{O}
#3	$\mathbf{H}_3 \mathbf{B}_1$	$\mathbf{H}_3 \mathbf{B}_2$	$\mathbf{H}_3 \mathbf{B}_3$	\mathbf{O}	\mathbf{O}
#4	\mathbf{O}	$\mathbf{H}_4 \mathbf{B}_2$	$\mathbf{H}_4 \mathbf{B}_3$	$\mathbf{H}_4 \mathbf{B}_4$	\mathbf{O}
#5	\mathbf{O}	\mathbf{O}	$\mathbf{H}_5 \mathbf{B}_3$	$\mathbf{H}_5 \mathbf{B}_4$	$\mathbf{H}_5 \mathbf{B}_5$

i -th block row: effective channel observed at user $\#i$ (off-diagonal entries: IUI components)
 j -th block col.: effective channel due to beam to user $\#j$ (no. of significant entries: user diversity order)

the desired user due to overlapped beams. So in practice we should take a few IUI users only into consideration.

In the following we present procedure of BBD computation as an example of the BMD family.

4.1.1 Block Bi-Diagonalization (BBD)

In the BBD computation, user $\#(i+1)$ is set to the IUI user corresponding to the desired user $\#i$. Next, we define $\mathcal{H}_{\text{bbd}}^{(i)}$ by eliminating the channel components of users $\#i$ and $\#(i+1)$ from the system channel matrix \mathcal{H} , and its SVD can be expressed as

$$\mathcal{H}_{\text{bbd}}^{(i)} = \left[\mathbf{H}_1^T \cdots \mathbf{H}_{i-1}^T \mathbf{H}_{i+2}^T \cdots \mathbf{H}_{N_{\text{usr}}}^T \right]^T \quad (18)$$

$$= \mathbf{U}_{\text{bbd}}^{(i)} \left[\boldsymbol{\Sigma}_{\text{bbd}[s]}^{(i)} \quad \mathbf{O} \right] \begin{bmatrix} \mathbf{V}_{\text{bbd}[s]}^{(i)H} \\ \mathbf{V}_{\text{bbd}[n]}^{(i)H} \end{bmatrix}. \quad (19)$$

Here the submatrix $\boldsymbol{\Sigma}_{\text{bbd}[s]}^{(i)}$ in Eq.(19) indicates the user space except for users $\#i$ and $\#(i+1)$, and the zero matrix indicates the kernel space. Therefore, when we use $\mathbf{V}_{\text{bbd}[n]}^{(i)}$ corresponding to the kernel, as a virtual TX weight, we obtain the presumed effective channel matrix as

$$\mathcal{H} \mathbf{V}_{\text{bbd}[n]}^{(i)} = \begin{bmatrix} \vdots \\ \mathbf{O} \\ \mathbf{H}_i \mathbf{V}_{\text{bbd}[n]}^{(i)} \\ \mathbf{H}_{i+1} \mathbf{V}_{\text{bbd}[n]}^{(i)} \\ \mathbf{O} \\ \vdots \end{bmatrix}. \quad (20)$$

That is, $\mathbf{V}_{\text{bbd}[n]}^{(i)}$ achieves perfect null steering to all the users except for users $\#i$ and $\#(i+1)$. The above procedure is basic computation of BBD as inter-user precoding.

Moreover, based on the channel component of the desired user $\#i$, $\mathbf{H}_i \mathbf{V}_{\text{bbd}[n]}^{(i)}$, noted in Eq. (20), we try to direct suitable beam to user $\#i$, namely intra-user precoding. We here presume an eigenvector matrix $\mathbf{V}_{\text{bbd}[e]}^{(i)}$ corresponding to the larger singular values of $\mathbf{H}_i \mathbf{V}_{\text{bbd}[n]}^{(i)}$. Multiplying $\mathbf{V}_{\text{bbd}[n]}^{(i)}$ by $\mathbf{V}_{\text{bbd}[e]}^{(i)}$ from right side gives desirable beamforming to improve SNR for user $\#i$, i.e. eigenmode transmission, over determinate signal space of only users $\#i$ and $\#(i+1)$, exploiting the residual degrees of freedom at the TX array.

Hence the precoding matrix for user $\#i$ in the BBD scheme is computed by $\mathbf{B}_{\text{bbd},i} = \mathbf{V}_{\text{bbd}[n]}^{(i)} \mathbf{V}_{\text{bbd}[e]}^{(i)}$, and the system precoding matrix \mathcal{B}_{bbd} can be obtained by applying the above procedure to all the desired users. We here note that the last user $\#N_{\text{usr}}$ should be handled in a different manner for NLP. User $\#N_{\text{usr}}$ must not have its IUI user when we apply IUI-PC requiring user hierarchization. Therefore, the precoder for user $\#N_{\text{usr}}$ is computed via the same processing of the BD scheme. Consequently, we obtain a purely block bi-diagonalized matrix $\mathcal{H}_{e,\text{bbd}} = \mathcal{H} \mathcal{B}_{\text{bbd}}$, as illustrated in Table 1(c).

Although BLTD is an approach similar to BBD, the difference is to consider up to two IUI users per desired user. That is, by eliminating the channel components of three users $\#i$, $\#(i+1)$ and $\#(i+2)$ from the system channel matrix \mathcal{H} in the first step corresponding to Eq.(18), we can calculate the BLTD precoding matrix $\mathcal{B}_{\text{bltd}}$ in the same manner. Its effective system channel matrix becomes block lower tri-diagonal as shown in Table 1(d), and it enables us to obtain further extra diversity gain for the desired user by allowing IUI up to two users. See [27] for the detailed discussion.

4.2 Adjacent IUI-PC

Although BMD brings higher beam gain due to extra diversity effect, we additionally need countermeasure for the allowed IUI.

In the proposed scheme, we adopt IUI-PC at BS, as well as NL-BT. TX signal vectors after applying IUI-PC prior to BBD and BLTD can be expressed as

NL-BBD :

$$\mathbf{s}'_i(t) = \text{modulo}_{\tau} \left[\mathbf{s}_i(t) - (\mathbf{H}_i \mathbf{B}_{\text{bbd},i})^+ \mathbf{H}_i \mathbf{B}_{\text{bbd},i-1} \mathbf{s}'_{i-1}(t) \right], \quad (21)$$

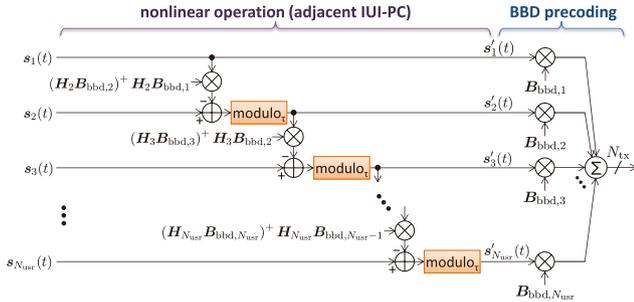


Fig. 3 TX processing circuit diagram in NL-BBD.

Table 2 Complexity at IUI-PC (no. of block off-diagonal entries).

	BD	NL-BT	NL-BBD(proposed)	NL-BLTD(proposed)
complexity	0	$\frac{N_{usr}(N_{usr}-1)}{2}$	$N_{usr}-1$	$2N_{usr}-3$

NL-BLTD :

$$s'_i(t) = \text{modulo}_{\tau} \left[s_i(t) - (\mathbf{H}_i \mathbf{B}_{bld,i})^+ \sum_{j=\max(i-2,1)}^{i-1} \mathbf{H}_i \mathbf{B}_{bld,j} s'_j(t) \right], \quad (22)$$

respectively. Note that $s'_1(t) = s_1(t)$ in the above equations. Figure 3 illustrates a TX processing circuit diagram in NL-BBD. As noticed by Fig. 3, Eqs. (21) and (22), in IUI-PC with BMD the required cancellation is only for adjacent users' substreams corresponding to the allowed IUI in BMD, unlike that with BT shown in Eq. (16). In other words, we can simplify IUI-PC by limiting the target to adjacent IUI only, thanks to BMD property. Therefore, in this paper we refer to the IUI-PC in NL-BMD as adjacent IUI-PC.

We list the number of IUI-PC operation times per symbol required in each scheme in Table 2, where the number is equivalent to the number of block off-diagonal entries $\mathbf{H}_i \mathbf{B}_j \neq \mathbf{O} (i > j)$ in the effective system channel matrix, as seen in Table 1. The complexity order in the proposed NL-BBD and NL-BLTD is $\mathcal{O}(N_{usr})$ whereas NL-BT requires the complexity in proportion to the square of the number of users, namely $\mathcal{O}(N_{usr}^2)$. We can say that the proposed NL-BMD has more implementability compared with the conventional NL-BT.

5. Computer Simulations

5.1 Scenario Setup

In this section we evaluate the performance of the proposed NL-BMD by link-level simulations.

Figure 4 illustrates scenario setup in this paper. We assumed MU-MIMO DL transmission at radio frequency of 28 GHz, of which wavelength is $\lambda = 10.7$ mm. Here we define ϕ and θ as azimuth and zenith angles, respectively. BS (TX) was installed at 21 m height, and users with 1 m height were located within a horizontal distance of 20 m–75 m and an azimuth angle range of $-60^\circ \leq \phi \leq +60^\circ$ from BS. The numbers of TX&RX baseband ports and users were

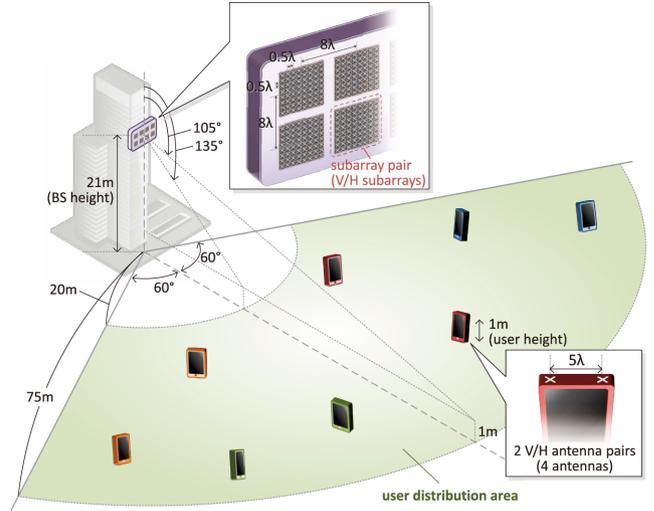


Fig. 4 Scenario setup.

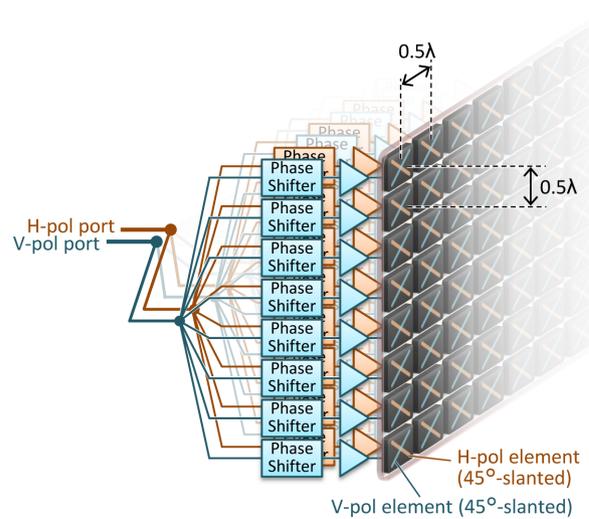


Fig. 5 Configuration of each V/H subarray pair.

set to $N_{Tx} = 16$, $N_{Rx} = 32$, and $N_{usr} = 8$, respectively, where the number of RX ports per user was fixed to $N_{rx,i} = 4$ irrespective of users. Targeting the analog-digital hybrid configuration with multiple subarrays as massive MIMO [7], [29], an antenna corresponding to a TX port at BS was assumed to be a planar active phased array antenna (APAA), so the TX array was composed of 16 APAA subarrays arranged in a plane, where we exactly had $2 \times 4 = 8$ pairs of ideally-isolated cross-polarized (45° -slanted V/H) subarrays as shown in Figs. 4 and 5. Each subarray had $\lambda/2$ -spaced $8 \times 8 = 64$ single-polarized antenna elements, of which radiation pattern was based on the prescription in Table 7.3-1 in [30], namely maximum directional gain of 8 dBi and angular full width at half maximum (FWHM) of 65° for both azimuth and elevation planes. Therefore a subarray pair had $64 \times 2 = 128$ antenna elements, and the TX array had $64 \times 16 = 1,024$ antenna elements in total. A signal from a TX port was fed to a subarray and divided into 64 antenna elements via

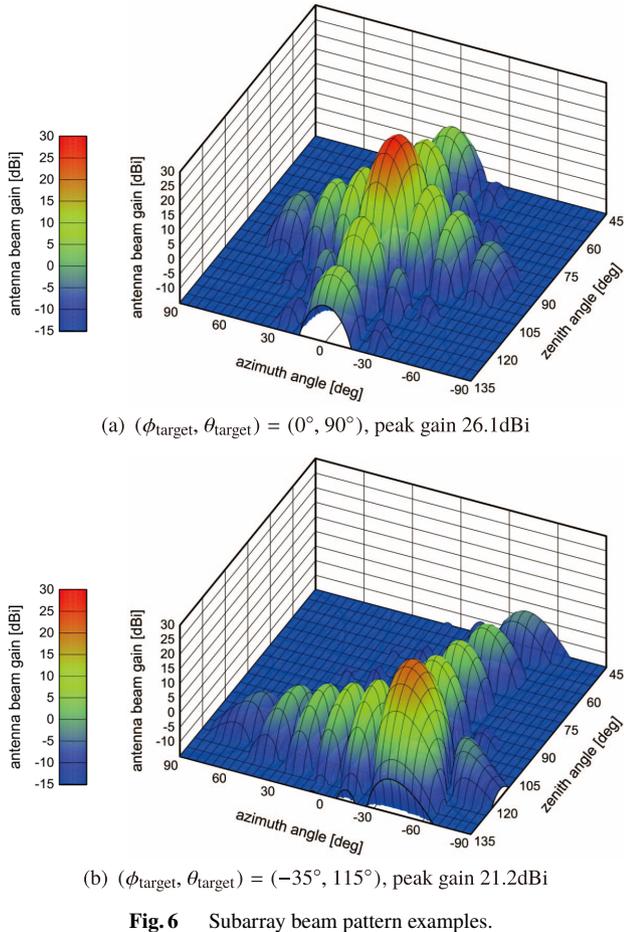


Fig. 6 Subarray beam pattern examples.

phase shifters, controlled so as to direct an analog beam to the target direction. Note that the power of an amplifier at each antenna element was common to the whole antenna elements and among subarrays. TX subarray spacing between adjacent pairs was set to 8λ , and each pair of V/H subarrays was associated with a target user to ideally direct V/H analog beams to the corresponding direction. We did not give tilt angle to the TX array in installation, i.e. its frontface direction was fixed to $\phi = 0^\circ$ & $\theta = 90^\circ$. We here show two examples of subarray beam pattern in Fig. 6, where target azimuth and zenith angles to be analog-beamformed were set to (a) $\phi_{\text{target}} = 0^\circ$ & $\theta_{\text{target}} = 90^\circ$ and (b) $\phi_{\text{target}} = -35^\circ$ & $\theta_{\text{target}} = 115^\circ$.

Also, we assumed that each user had two ideally-isolated cross-polarized antenna pairs with inter-pair spacing of 5λ , where each element pattern was assumed to be ideally isotropic with gain of 0 dBi, for the sake of simplicity. Array orientation of each user was randomly set in horizontal plane, and we assumed no traveling for all users to model quasi-static block fading.

Figure 7 shows user scenarios evaluated in this paper. Scenario A is a configuration where users were randomly distributed within the area, equivalently assuming a non-correlated case. Since each of beams from BS can be directed to different directions, this scenario is favorable for

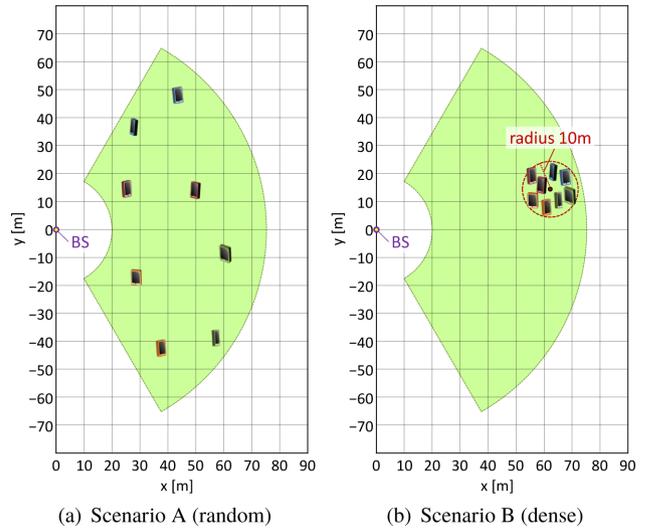


Fig. 7 User scenarios (top view).

the system to spatially multiplex users. Scenario B, on the other hand, is a configuration where all users were located within a circular spot of which radius was 10 m, assuming a densely-populated urban environment. Although for stable system operation a scheduler would generally tend to group relatively-distant users for simultaneous access, this scenario is to evaluate an extreme case that it cannot avoid unfavorable scheduling. Note that the circular spot for geometrically constraining all users was randomly dropped so as not to exceed the area.

The employed channel model for link-level simulations was clustered delay line model type-E (CDL-E) specified in [30], which models line-of-sight (LOS) channels in urban micro cells for above 6 GHz frequencies. We note that the LOS angular values defined in the first cluster (see Table 7.7.1-5 in [30]) were replaced by instantaneous LOS angular values between BS subarrays and user antennas depending on the given geometric condition, in order to simulate the aforementioned MU-MIMO DL scenarios. Root mean square (RMS) delay spread was set at 32 ns as a 28 GHz short-delay urban micro cell prescribed in Table 7.7.3-2 in [30].

5.2 Simulation Conditions

Table 3 lists simulation parameters. We conducted OFDM simulations to evaluate wide-band MU-MIMO DL transmission, where the number of subcarriers and subcarrier spacing were set at $N_{\text{sc}} = 1,200$ and 60 kHz, respectively. In order to solve computation of linear precoders at BS in the system antenna configuration of $N_{\text{tx}} = 16$ and $N_{\text{rx}} = 32$, we first degenerated the per-user RX antenna domain from four physical antenna ports to two beam-space antennas, by using two major eigenvectors maximizing effective channel gain as postcoders at all the subcarriers [31]. Then BS computed an LP matrix \mathcal{B} from a 16×16 beam-space system channel matrix at each subcarrier, under the ideal condition of per-

Table 3 MU-MIMO OFDM simulation parameters.

radio frequency	28 GHz
no. of users	$N_{\text{usr}} = 8$
no. of TX ports	$N_{\text{tx}} = 16$ (V:8, H:8)
no. of RX ports	$N_{\text{rx}} = 32$ ($N_{\text{rx},i} = 4$ (V:2, H:2))
TX antenna type	APAA ($8 \times 8 = 64$ elements with $\lambda/2$ spacing)
RX antenna type	isotropical
no. of substreams	$N_{\text{st}} = 16$ ($N_{\text{st},i} = 2$), 15 ($N_{\text{st},8} = 1$ for user #8)
no. of subcarriers	$N_{\text{sc}} = 1,200$
subcarrier spacing	60 kHz
channel bandwidth	$\text{BW}_{\text{ch}} = 72$ MHz
FFT/IFFT points	2,048
cyclic prefix	144
OFDM symbol length	$T_s = 17.84 \mu\text{s}$
modulation schemes	QPSK, 16QAM, 64QAM, 256QAM
channel coding	turbo code ($K = 4$, $R = 1/3 - 5/6$) coding per substream max-log MAP decoding with 8 iterations
no. of MCSs	19 (see Table 4)
max. sum-rate	$6.229 \times 16 = 99.66$ bps/Hz
precoding	inter-user: BD, NL-BT, NL-BBD, NL-BLTD intra-user: eigenmode (SVD)
MIMO detection	MMSE [32]
modulo boundary	$\tau = 1.225$ ($\sqrt{3}/2$), ∞ (w/o modulo operation)
channel estimation / synchronization	ideal
channel model	CDL-E [30] (quasi-static)

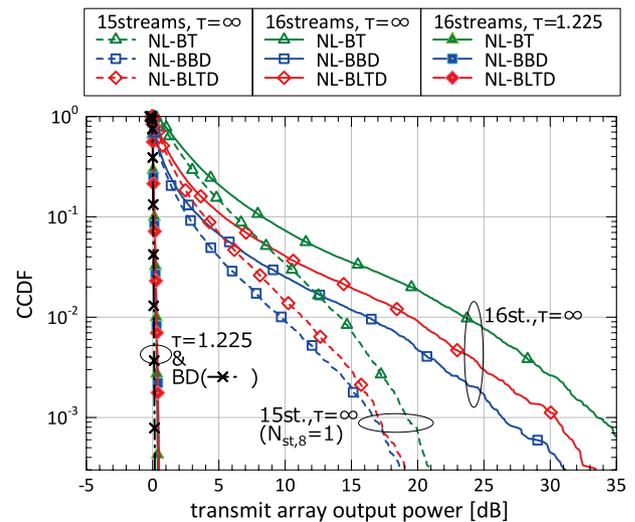
fect knowledge on channel state information at TX (CSIT) exploiting time reciprocity. Basically $N_{\text{st},i} = 2$ substreams per user were transmitted from BS. Therefore 16-substream transmission over 8-user multiplexing was performed. For comparison, we also simulated 15-substream transmission where only the last user #8 was limited to use single substream. Turbo coding was applied to each substream independently, where each codeword was closed within each OFDM symbol. As listed in Table 4, we prepared 19 patterns of modulation and coding schemes (MCSs) in total from combinations of modulation schemes and coding rates in order to evaluate achievable spectral efficiency for every substream. Here, per-substream spectral efficiency for MCS index # i is derived from

$$\eta_i = \frac{(\log_2 M_i) \times R_i \times N_{\text{sc}}}{T_s \times \text{BW}_{\text{ch}}} \quad [\text{bps/Hz}], \quad (23)$$

where M_i is modulation order of index # i ($M_i = 4, 16, 64, 256$ for QPSK, 16QAM, 64QAM, and 256QAM, respectively), R_i is coding rate of index # i , T_s is OFDM symbol time length, and BW_{ch} is channel bandwidth, as specified in Table 3. The maximum sum-rate spectral efficiency in 16-substream transmission is, therefore, $16\eta_{19} = 99.66$ bps/Hz. When we tried MCS index # i upon a substream and found no bit errors after turbo decoding, we could count η_i as an instantaneous spectral efficiency of the substream. Note that, for the sake of simplicity, equal power allocation over users and substreams was adopted, and adaptive user ordering discussed in [27] was not applied. Modulo boundary was set to $\tau = \sqrt{3}/2 = 1.225$, so that modulo operation did not cause an increase in TX signal amplitude (see Appendix B). In this case, TX power in NLP is almost the

Table 4 Per-substream MCS table.

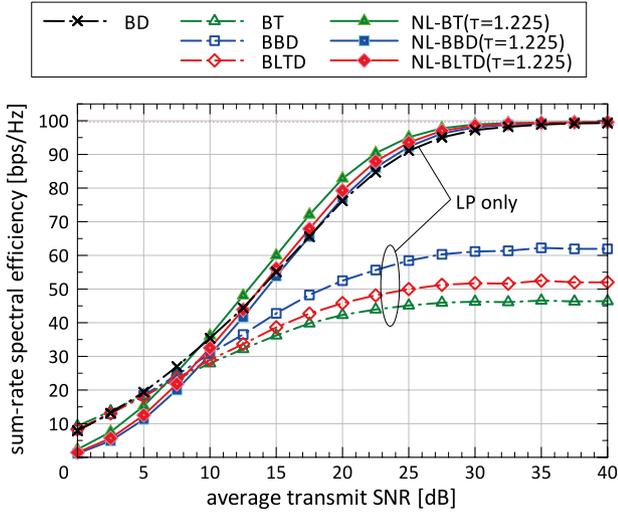
index	modulation	coding rate R	spectral efficiency η [bps/Hz]
#1	QPSK	1/3	0.623
#2	QPSK	1/2	0.934
#3	QPSK	2/3	1.246
#4	QPSK	3/4	1.401
#5	QPSK	4/5	1.495
#6	QPSK	5/6	1.557
#7	16QAM	1/2	1.869
#8	16QAM	2/3	2.491
#9	16QAM	3/4	2.803
#10	16QAM	4/5	2.990
#11	16QAM	5/6	3.114
#12	64QAM	2/3	3.737
#13	64QAM	3/4	4.204
#14	64QAM	4/5	4.485
#15	64QAM	5/6	4.672
#16	256QAM	2/3	4.983
#17	256QAM	3/4	5.606
#18	256QAM	4/5	5.980
#19	256QAM	5/6	6.229

**Fig. 8** CCDF of instantaneous TX array output power (Scenario A).

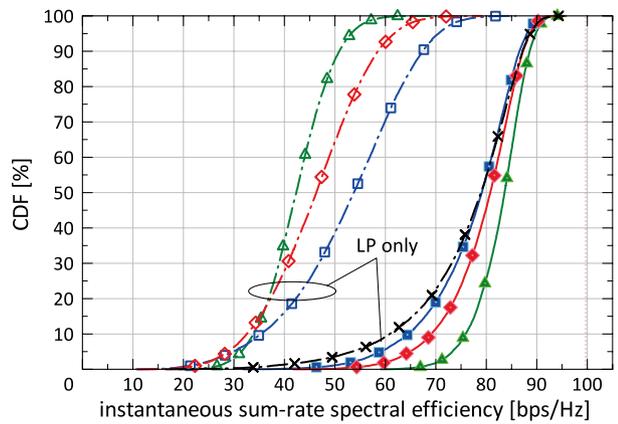
same as that in BD, of which precoders are normalized vectors. At users, MIMO demultiplexing was performed by minimum mean square error (MMSE) spatial filtering based on precoded channel estimates for the desired user, where we employed amplitude correction of spatial filter outputs so that the subsequent modulo operation works well (see Appendix C in [32]). When using modulo operation, in demapping at each user bit log-likelihood ratios (LLRs) were computed by expanding the replica signal constellation on the basis of modulo boundary [23]. Also, we obtained performance without modulo operation, namely $\tau = \infty$, for comparison.

5.3 Array Output Power Performance

Figure 8 shows CCDF of TX array output power $\|\mathcal{B}\bar{s}(t)\|^2$. Instantaneous TX array output power in abscissa is normalized by average power in BD, $E[\|\mathcal{B}_{\text{bd}}\bar{s}(t)\|^2]$. In 16-

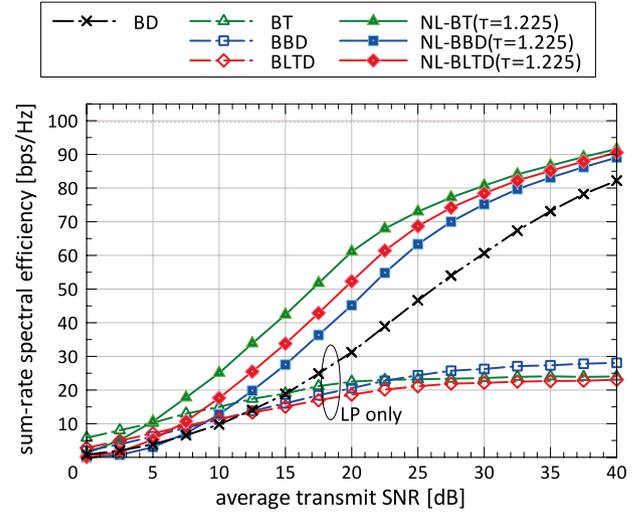


(a) average sum-rate spectral efficiency vs. average TX SNR

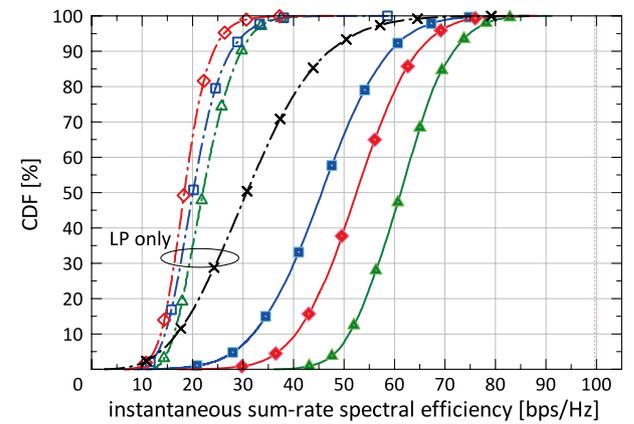


(b) CDF of instantaneous spectral efficiency at average TX SNR of 20 dB

Fig. 9 Sum-rate spectral efficiency performance in Scenario A.



(a) average sum-rate spectral efficiency vs. average TX SNR



(b) CDF of instantaneous spectral efficiency at average TX SNR of 20 dB

Fig. 10 Sum-rate spectral efficiency performance in Scenario B.

substream transmission without modulo operation ($\tau = \infty$), channel variation may lead to extremely-high TX array output powers: more than 15 dB at CCDF of 10^{-2} , and more than 25 dB at CCDF of 10^{-3} . As formulated in Eqs. (16), (21) and (22), the IUI-PC term includes the inverse of the effective channel for the desired user. In particular, the second substream, corresponding to the lower eigenmode, of the last user #8 does not have diversity gain in any of BT, BBD, and BLTD precoding. Therefore, an increase in TX power due to IUI-PC without modulo operation is dependent on the diversity order and is dominated by the behavior of the last substream. On the other hand, in 15-substream transmission the TX power can be suppressed thanks to disuse of the no-diversity substream. For instance, NL-BBD in 15-substream transmission restrains the increase in TX power to 10 dB at CCDF of 10^{-2} . In this way, rank adaptation (resource allocation) based on diversity order should be effective for TX power suppression in NLP. However, the increase in TX power still remains an unacceptable issue in NLP without modulo operation. In contrast, it is noticeable that applying modulo operation with $\tau = 1.225$ significantly reduces

TX power to the level equivalent to BD even in full-stream transmission. Meanwhile, modulo operation may lead to detection errors in demapping. We verify the influence on throughput in the next subsection.

5.4 Spectral Efficiency Performance

Figures 9 and 10 demonstrate sum-rate spectral efficiency performance in Scenarios A and B, respectively. Each figure shows (a) average sum-rate spectral efficiency vs. average TX SNR, and (b) CDF of instantaneous sum-rate spectral efficiency at average TX SNR of 20 dB, where “LP only” means the performance without NL operation. Also, all the graphs evaluated here are in 16-substream transmission. Here, average TX SNR is defined on the basis of power sum of all users’ QAM-mapped signals, $E[\|\bar{s}(t)\|^2]$. It is also equivalent to the SNR observed in open-loop (non-precoded) transmission. We notice that the proposed NL-BBD and NL-BLTD yield higher spectral efficiencies even when applying modulo operation. However, their improvement gain in Scenario A is marginal. In comparison with BD at average

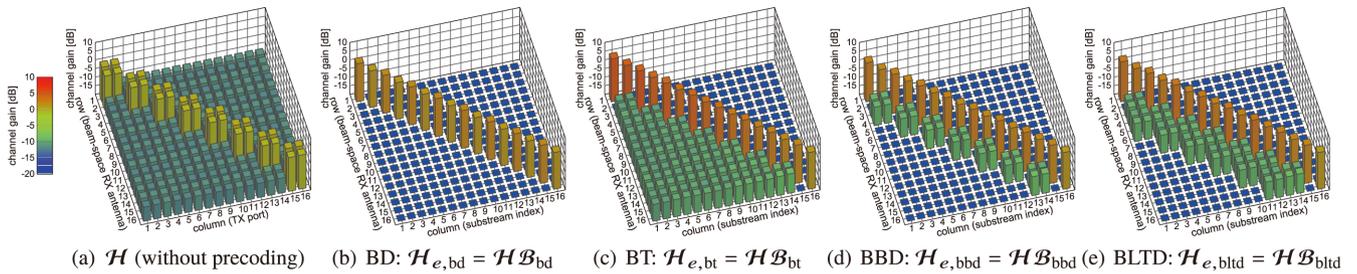


Fig. 11 Average element-wise channel gain with and without precoding in Scenario A.

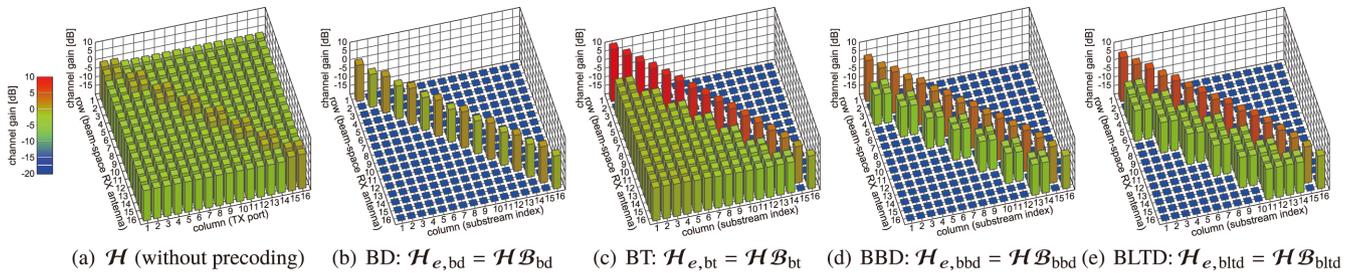


Fig. 12 Average element-wise channel gain with and without precoding in Scenario B.

TX SNR of 20 dB, in Scenario A, NL-BBD and NL-BLTD with modulo boundary of $\tau = 1.225$ give 77.4 bps/Hz and 79.5 bps/Hz (1.2% and 4.0% improvements over BD), respectively, whereas BD achieves 76.4 bps/Hz. This implies that, when users are randomly distributed and illuminated by custom-tailored beams from massive-element antennas, in general the given system channel tends to naturally become semi-diagonal across users so that the conventional BD is sufficient to mitigate IUI. On the other hand, improvement by the proposal is visibly significant in Scenario B, where users are densely located: NL-BBD and NL-BLTD with $\tau = 1.225$ give 45.2 bps/Hz and 52.3 bps/Hz (44.7% and 67.5% improvements over BD), respectively, whereas BD shows 31.2 bps/Hz. This is because BMD ensures beam gain by partially allowing IUI even if users are in proximity, and then adjacent IUI-PC compensates the residual IUI, resulting in the equivalent of flexible beamforming.

To investigate the phenomena aforementioned in detail, we check analog-beamformed channel gain with and without LP. Figures 11 and 12 visually show element-wise channel gain in channel matrices averaged over subcarriers and trials. In each figure, (a) shows average element-wise gain of the 16×16 analog-beamformed system channel matrix \mathbf{H} , and (b), (c), (d), and (e) show that of the 16×16 effective system channel matrix \mathbf{H}_e when applying BD, BT, BBD, and BLTD, respectively, as LP. All the values of channel gain are normalized by the average gain of analog-beamformed channels associated with the desired users. We note that, when applying LP schemes, per-user subchannels in block-diagonal entries are diagonalized due to SVD used as intra-user precoding. In Scenario A, thanks to analog beams directed to distant users, the channel \mathbf{H} can be already roughly block-diagonalized even without precoding, where undesired power (IUI power) is reduced to

less than -10 dB compared to desired power (channel gain of block-diagonal entries). So, with the conventional BD we can obtain relatively good effective channels without losing gain for the desired users, whereas BBD and BLTD methods provide marginal gain. On the other hand, in Scenario B, we have high gain at block off-diagonal entries, i.e. IUI components, in the channel \mathbf{H} due to overlapping of analog beams directed to similar directions. As shown in Fig. 12(b), thus, BD sacrifices channel gain for the desired users in order to reduce the IUI components. Meanwhile, we see that the other schemes including BBD and BLTD yield better channel gain for the desired users by the benefit of the extra degrees of freedom after partial nulling. Since IUI caused by the residual block off-diagonal entries can be canceled by IUI-PC, the proposed NL-BBD and NL-BLTD consequently provide better performance than the conventional BD, especially in Scenario B.

We see that NL-BT provides excellent performance. As discussed in Sect. 3.2, BT achieves ideal capacity of MU-MIMO, and in the study here we assume quasi-static fading and ideal IUI-PC. Therefore, NL-BT performance in Figs. 9 and 10 can be regarded as nearly optimal upper bound of sum-rate spectral efficiency under our simulation conditions. In the meantime, when considering numerical complexity in IUI-PC, while NL-BT requires removal of 56 substreams per symbol per subcarrier, NL-BBD takes removal of just 14 substreams (1/4 complexity of NL-BT). The fact indicates that our proposal has higher implementability. It also implies that, thanks to the fact that less IUI needs to be managed, the proposed NL-BMD may be more robust toward degrading factors such as calculation/estimation errors and channel variation. This behavior will be studied in future work.

In the following, we deal with NL-BBD with $\tau = 1.225$ as our proposal for the sake of simple discussion. Setting the

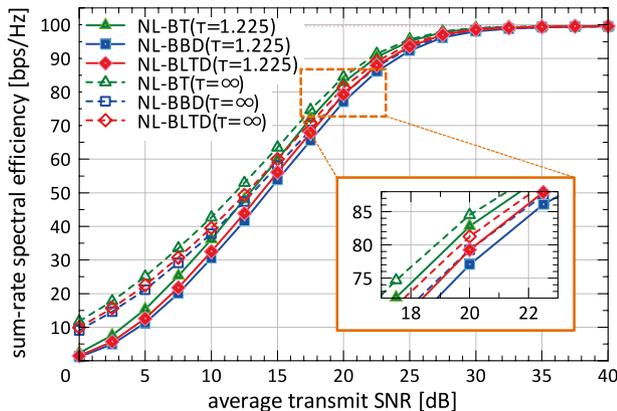


Fig. 13 Average sum-rate spectral efficiency performance in Scenario A, to verify an impact of modulo operation.

target sum-rate spectral efficiency to 60 bps/Hz, when comparing average spectral efficiency performance with BD, NL-BBD gives about 5.7 dB gain in Scenario B while keeping the same SNR as BD in Scenario A. Also, CDF performance shows that NL-BBD provides remarkable gain at lower probabilities, thanks to successful diversity effect, compared with BD. Especially in Scenario B, at 10th percentile, NL-BBD gives 32.0 bps/Hz (190.2% improvement over BD) whereas BD shows 16.8 bps/Hz. In addition to improvement in user experience, such stability in the performance is expected to relax complexity in scheduling computation in an actual operational system.

Finally, we evaluate an influence of modulo operation on the performance. Figure 13 shows results of average sum-rate spectral efficiencies with and without modulo operation. Note that, due to the definition of average TX SNR in the abscissa, an increase in TX signal power due to IUI-PC without modulo operation, discussed in the previous subsection, is not considered in the graph. In the region of average TX SNR of over 20 dB, performance degradation due to modulo operation is less than 1 dB. Considering the result in Fig. 8 additionally, we can say that modulo operation is an excellent approach to reduce TX power dramatically while maintaining the transmission performance.

6. Conclusions

In this paper, we proposed NL-BMD, a novel TX precoding scheme for MU-MIMO DL consists of BMD and adjacent IUI-PC. Through numerical evaluation over OFDM, it has been clarified that, NL-BMD yields up to 67% performance improvement in sum-rate spectral efficiency on average, compared with the conventional BD. Moreover, we have verified that, at lower probabilities in CDF of sum-rate spectral efficiency, NL-BMD provides excellent performance, nearly twice better than BD, thanks to extra diversity gain.

Next step of our study is a more practical evaluation that considers time reciprocity, dynamic fading, and 5G waveform candidates. Also, we will study to refine our precoding schemes so as to be appropriate for various user scenarios

and conditions. In addition, an optimum solution of user ordering in NL-BMD briefly discussed in [27] should be investigated as our future work.

Acknowledgements

The work includes a part of the results of “The research and development project for realization of the fifth-generation mobile communications system” commissioned by Japan’s Ministry of Internal Affairs and Communications.

References

- [1] ITU-R, “IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond,” Recommendation ITU-R M.2083-0, Sept. 2015 (<http://www.itu.int/rec/R-REC-M.2083>).
- [2] 5GMF, “White paper (ver.1.0.1),” 5G Mobile Communications Systems for 2020 and beyond, July 2016 (http://5gmf.jp/wp/wp-content/uploads/2016/07/5GMF_WP101_All.pdf).
- [3] T.L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol.9, no.11, pp.3590–3600, Nov. 2010.
- [4] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol.30, no.1, pp.40–60, Jan. 2013.
- [5] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?,” *IEEE J. Sel. Areas Commun.*, vol.31, no.2, pp.160–171, Feb. 2013.
- [6] T. Obara, S. Suyama, J. Shen, and Y. Okumura, “Joint processing of analog fixed beamforming and CSI-based precoding for super high bit rate massive MIMO transmission using higher frequency bands,” *IEICE Trans. Commun.*, vol.E98-B, no.8, pp.1474–1481, Aug. 2015.
- [7] A. Taira, S. Yamaguchi, K. Tsutsumi, S. Shinjo, A. Okazaki, and A. Okamura, “A hybrid beamforming architecture for high SHF wide-band massive MIMO in 5G,” *Proc. IEEE APWCS 2016*, S3-3, pp.295–299, Aug. 2016.
- [8] M. Rim, “Multi-user downlink beamforming with multiple transmit and receive antennas,” *Electron. Lett.*, vol.38, no.25, pp.1725–1726, Dec. 2002.
- [9] K.-K. Wong, R.D. Murch, and K. Letaief, “Performance enhancement of multiuser MIMO wireless communication systems,” *IEEE Trans. Commun.*, vol.50, no.12, pp.1960–1970, Dec. 2002.
- [10] L.U. Choi and R.D. Murch, “A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach,” *IEEE Trans. Wireless Commun.*, vol.3, no.1, pp.20–24, Jan. 2004.
- [11] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Process.*, vol.52, no.2, pp.461–471, Feb. 2004.
- [12] H. Nishimoto, S. Kato, Y. Ogawa, T. Ohgane, and T. Nishimura, “Imperfect block diagonalization for multiuser MIMO downlink,” *Proc. IEEE PIMRC 2008*, Sept. 2008.
- [13] T. Sada, J. Webber, T. Nishimura, T. Ohgane, and Y. Ogawa, “A generalized approach to block diagonalization for multiuser MIMO downlink,” *Proc. IEEE PIMRC 2010*, pp.504–509, Sept. 2010.
- [14] V. Stankovic and M. Haardt, “Generalized design of multi-user MIMO precoding matrices,” *IEEE Trans. Wireless Commun.*, vol.7, no.3, pp.953–961, March 2008.
- [15] K. Zu, R.C. de Lamare, and M. Haardt, “Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser MIMO systems,” *IEEE Trans. Commun.*, vol.61, no.10, pp.4232–4242, Oct. 2013.
- [16] M. Sadek, A. Tarighat, and A.H. Sayed, “A leakage-based precod-

- ing scheme for downlink multi-user MIMO channels,” IEEE Trans. Wireless Commun., vol.6, no.5, pp.1711–1721, May 2007.
- [17] M.C.H. Lim, M. Ghogho, and D.C. McLernon, “Spatial multiplexing in the multi-user MIMO downlink based on signal-to-leakage ratios,” Proc. IEEE GLOBECOM 2007, pp.3634–3638, Dec. 2007.
- [18] A. Tölli, M. Codreanu, and M. Juntti, “Linear multiuser MIMO transceiver design with quality of service and per-antenna power constraints,” IEEE Trans. Signal Process., vol.56, no.7, pp.3049–3055, June 2008.
- [19] M. Costa, “Writing on dirty paper,” IEEE Trans. Inf. Theory, vol.29, no.3, pp.439–441, May 1983.
- [20] G. Caire and S. Shamai (Shitz), “On the achievable throughput of a multi-antenna Gaussian broadcast channel,” IEEE Trans. Inf. Theory, vol.49, no.7, pp.1691–1706, July 2003.
- [21] M. Tomlinson, “New automatic equaliser employing modulo arithmetic,” Electron. Lett., vol.7, no.5, pp.138–139, March 1971.
- [22] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference,” IEEE Trans. Commun., vol.20, no.4, pp.774–780, Aug. 1972.
- [23] E.C.Y. Peh and Y.-C. Liang, “Power and modulo loss tradeoff with expanded soft demapper for LDPC coded GMD-THP MIMO systems,” IEEE Trans. Wireless Commun., vol.8, no.2, pp.714–724, Feb. 2009.
- [24] C.B. Peel, B.M. Hochwald, and A.L. Swindlehurst, “A vector-perturbation technique for near-capacity multi-antenna multiuser communication—Part I: Channel inversion and regularization,” IEEE Trans. Commun., vol.53, no.1, pp.195–202, Jan. 2005.
- [25] B.M. Hochwald, C.B. Peel, and A.L. Swindlehurst, “A vector-perturbation technique for near-capacity multi-antenna multiuser communication—Part II: Perturbation,” IEEE Trans. Commun., vol.53, no.3, pp.537–544, March 2005.
- [26] S. Sonobe, S. Tsukamoto, T. Maeda, K. Yano, H. Ban, M. Uno, and K. Kobayashi, “Field experiments of LTE-Advanced-based 8×8 multiuser MIMO system with vector perturbation,” Proc. IEEE WiOpt 2013, pp.83–88, May 2013.
- [27] H. Nishimoto, H. Iura, A. Taira, A. Okazaki, and A. Okamura, “Block lower multi-diagonalization for multiuser MIMO downlink,” Proc. IEEE WCNC 2016 Workshop, pp.342–347, April 2016.
- [28] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” IEEE Trans. Inf. Theory, vol.49, no.10, pp.2658–2668, Oct. 2003.
- [29] A. Taira, H. Iura, K. Nakagawa, S. Uchida, K. Ishioka, A. Okazaki, S. Suyama, T. Obara, Y. Okumura, and A. Okamura, “Performance evaluation of 44 GHz band massive MIMO based on channel measurement,” Proc. IEEE GLOBECOM 2015 Workshop, Dec. 2015.
- [30] 3GPP, TR 38.900 (V14.1.0), “Channel model for frequency spectrum above 6 GHz,” Sept. 2016 (<http://www.3gpp.org/DynaReport/38900.htm>).
- [31] H. Iura, K. Nakagawa, S. Uchida, A. Taira, A. Okazaki, S. Suyama, T. Obara, Y. Okumura, and A. Okamura, “Transmit performance of analog beamforming MIMO system in practical outdoor environment using high frequency band,” IEICE Technical Report, A-P2015-116, RCS2015-201, Nov. 2015 (in Japanese).
- [32] H. Nishimoto, Studies on MIMO spatial multiplexing for high-speed wireless communications, Ph.D. thesis, Hokkaido University, Japan, 2007 (<http://hdl.handle.net/2115/42630>).

Appendix A: BT Algorithm Example

Table A·1 summarizes a block triangulation algorithm as an example. Here we employ an approach of successive mapping to kernel space of prior users. User $\#(i + 1)$ and its posteriori users can form the signal space which never interferes their priori users since the kernel space of user $\#i$ is

Table A·1 Block triangulation algorithm example.

$$\begin{aligned}
 & \text{set } \mathcal{B}_{\text{bt}} \leftarrow \mathbf{I}_{N_{\text{tx}}} \\
 & \text{set } \mathbf{V}_{\text{bt}[n]}^{(0)} \leftarrow \mathbf{I}_{N_{\text{tx}}} \\
 & \text{for } i = 1, \dots, N_{\text{usr}} \\
 & \quad \text{set } \mathbf{H}_{\text{bt}}^{(i)} \leftarrow \mathbf{H}_i \mathbf{V}_{\text{bt}[n]}^{(0)} \dots \mathbf{V}_{\text{bt}[n]}^{(i-1)} = \mathbf{H}_i \prod_{j=0}^{i-1} \mathbf{V}_{\text{bt}[n]}^{(j)} \\
 & \quad \text{SVD: } \mathbf{H}_{\text{bt}}^{(i)} = \mathbf{U}_{\text{bt}}^{(i)} \Sigma_{\text{bt}}^{(i)} \mathbf{V}_{\text{bt}}^{(i)H} = \mathbf{U}_{\text{bt}}^{(i)} \begin{bmatrix} \mathbf{L}^{(i)} & \mathbf{O} \\ \mathbf{V}_{\text{bt}[s]}^{(i)H} & \mathbf{V}_{\text{bt}[n]}^{(i)H} \end{bmatrix} \\
 & \quad \text{set } \mathbf{V}_{N_{\text{tx}} \times N_{\text{tx}}}^{(i)} \leftarrow \begin{bmatrix} \mathbf{I}_{\sum_{j=1}^{i-1} N_{\text{rx},j}} & \mathbf{O} \\ \mathbf{O} & \sqrt{\frac{\mathbf{V}_{\text{bt}[s]}^{(i)} \mathbf{V}_{\text{bt}[n]}^{(i)}}{\mathbf{V}_{\text{bt}[s]}^{(i)} \mathbf{V}_{\text{bt}[n]}^{(i)}}}} \end{bmatrix} \\
 & \quad \text{set } \mathcal{B}_{\text{bt}} \leftarrow \mathcal{B}_{\text{bt}} \mathbf{V}^{(i)} = \prod_{j=1}^i \mathbf{V}^{(j)} \\
 & \text{next } i
 \end{aligned}$$

spanned over the kernel space of all the priori users. Consequently we can obtain \mathcal{B}_{bt} providing a triangulated effective system channel matrix, as formulated in Eq. (13).

Appendix B: On Signal Power After Modulo Operation

Assuming that pre-canceled signal after modulo operation, $\text{modulo}_{\tau}\{s'(t)\}$, is uniformly distributed within $[-\tau, +\tau]$ over I/Q plane, its expected signal power can be approximately estimated as follows:

$$\begin{aligned}
 P_{\tau} &= E \left[|\text{modulo}_{\tau}\{s'(t)\}|^2 \right] \\
 &\approx \int_{-\tau}^{\tau} \int_{-\tau}^{\tau} p(x, y) |s'(t)|^2 \partial x \partial y \\
 &= 4 \int_0^{\tau} \int_0^{\tau} \frac{1}{4\tau^2} (x^2 + y^2) \partial x \partial y \\
 &= \frac{2}{3} \tau^2. \tag{A·1}
 \end{aligned}$$

That is, we have $P_{\tau} = 1$ when setting modulo boundary to $\tau = \sqrt{3/2}$.



Hiroshi Nishimoto received his B.E., M.E., and Ph.D. degrees from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively. As well, he was also a Research Fellow of the Japan Society for the Promotion of Science (JSPS) from 2005 to 2008. In 2008, he joined Mitsubishi Electric Corporation, Japan. His general research interests are in digital wireless communication systems, and recent primary topics are in 5G-related wireless technologies including massive MIMO and linear cell designing

for high-speed vehicles. He received the IEICE Young Researchers' Award in FY2006, the 2007 IEEE VTS Japan Chapter Student Paper Award, the 2008 IEEE AP-S Japan Chapter Young Engineer Award, and the 2008 IEEE Sapporo Section Young Scientist Award. He is a member of the IEEE.



Akinori Taira received the B.E., M.E., and Ph.D. degrees in electronics and communication engineering from Waseda University, Japan, in 1994, 1996, and 2008, respectively. In 1996, he joined Mitsubishi Electric Corporation, where he has been engaged in research and development on broadband wireless communication systems, especially OFDM and MIMO technologies. From 2013 to 2015, he was the Associate Professor of the Research Institute of Electrical Communication, Tohoku University, Japan. He

is currently the Chief Researcher of the Communication Technology Department, Information Technology R&D Center, Mitsubishi Electric Corporation. He received the IEICE Young Researchers' Award in FY2000. He is a member of the IEEE.



Atsushi Okamura received his B.E. and M.E. degrees from the University of Tsukuba, Japan, in 1987 and 1989, respectively, and Ph.D. degree from the Keio University, Japan, in 2000. Since he joined Mitsubishi Electric Corporation, Japan, in 1989, he has been engaged in research and development on radar systems for avionics, especially through digital signal processing, adaptive signal processing, and array antenna techniques. Currently, he is the Senior Manager of the Communication Technology Department,

Information Technology R&D Center. He received the IEICE Young Researchers' Award in 1996 and served as a secretary of IEICE Technical Committee on Space, Aeronautical and Navigational Electronics (SANE) from 2006 to 2008. He is a senior member of the IEEE.



Hiroki Iura received his B.E. and M.E. degrees from Niigata University, Japan, in 2004 and 2006, respectively. In 2006, he joined Mitsubishi Electric Corporation, Japan. His active research areas are digital wireless communication systems, millimeter-wave antenna propagation measurements and its modelings for 5G. He received the IEICE Young Researchers' Award in FY2009.



Shigeru Uchida received the B.E. degree in the Department of Electrical and Electronic Engineering and the M.E. degree in the Department of Advanced Applied Electronics from Tokyo Institute of Technology, Japan, in 1996 and 1998, respectively. And he joined Mitsubishi Electric Corporation, Japan, in 1998. Since then, he has been engaged in research and development related to mobile broadband communications systems and home network. He is currently a head researcher of the Communication Tech-

nology Department, Information Technology R&D Center.



Akihiro Okazaki received the B.E. and M.E. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 1996 and 1998, respectively. He joined Mitsubishi Electric Corporation in 1998. From May 2006 to February 2009, he was with Mitsubishi Electric R&D Centre Europe (MERCE). Since 1998, he has been engaged in research and development on digital mobile radio communication systems. He is currently the Manager of the Wireless Multiple Access Group

in the Communication Technology Department, Information Technology R&D Center, Mitsubishi Electric Corporation. He received the Ministry of Education, Culture, Sports, Science and Technology Award in 2010. He also received the IEICE Young Researchers' Award in 2004.