

LETTER

An Adaptive Bit Allocation for Maximum Bit-Rate Tomlinson-Harashima Precoding

Shigenori KINJO^{†a)} and Shuichi OHNO^{††b)}, *Members*

SUMMARY An adaptive bit allocation scheme for zero-forcing (ZF) Tomlinson-Harashima precoding (THP) is proposed. The ZF-THP enables us to achieve feasible bit error rate (BER) performance when appropriate substream permutations are installed at the transmitter. In this study, the number of bits in each substream is adaptively allocated to minimize the average BER in fading environments. Numerical examples are provided to compare the proposed method with eigenbeam space division multiplexing (E-SDM) method.

key words: Tomlinson-Harashima precoding, maximum bit-rate, zero-forcing THP, adaptive bit allocation

1. Introduction

Eigenbeam space division multiplexing (E-SDM) is a well-known linear precoding MIMO scheme, in which precoding and decoding matrices are given by singular-value decomposition (SVD) or eigenvalue decomposition of a MIMO channel matrix [1]. E-SDM is optimum in terms of bit-rate when a water-filling (WF) is applied to determine the transmit power [2]. Computationally reasonable bit and power allocation methods have been studied in [3], [4] for E-SDM. The SVD, however, remains requiring a high computational cost [5], which cannot be ignored from a practical viewpoint.

Tomlinson-Harashima precoding (THP) is a non-linear precoding scheme, which shows feasible bit-rate and bit error rate (BER) performance, whose precoder is given by relatively low complexity algorithms such as QR or Cholesky decompositions [6], [7]. THP has been studied for more than two decades, whose new attractive characteristics have been studied in the last five years [7]–[10].

For MIMO systems, we have proposed a simple precoding scheme based on zero-forcing (ZF) THP [7], [8]. The ZF-THP has two characteristics: permutation of the transmitted signals and unequal bit allocation. More specifically, substream permutations are installed at the transmitter, and its permutation pattern is selected so as to maximize its bit-rate or sum-rate. For convenience, we call the ZF-THP a maximum bit-rate (MBR) THP. In the MBR-THP, a MIMO channel is decomposed into independent paral-

lel substreams, which is similar to the decomposed MIMO channel in E-SDM. Each parallel substream has one gain like a singular value of the MIMO channel matrix. By using this specific characteristic, a different number of bits has been allocated to each substream to improve the BER performance [8]. The scheme has been computationally efficient and achieved relatively good BER performance in comparison to the other precoding schemes if an appropriate bit allocation pattern is selected. The pattern of the bit allocation has been fixed to avoid the additional computational cost for the bit allocation. This fixed bit allocation strategy is, however, generally not appropriate for time-varying fading MIMO channels.

In this letter, we apply an adaptive bit allocation to the MBR-THP to improve its performance. We determine the best bit allocation pattern for a given MIMO channel to keep good performance of the MBR-THP even in slowly time-varying MIMO channels. We apply the bit allocation originally developed for E-SDM [3] to the MBR-THP. However, unlike the scheme in [3], we adopt uniform transmit power allocation since THP is a nonlinear precoder. It has been shown in [11] that the uniform power allocation strategy is appropriate for THP schemes if the transmit power is sufficiently high. The proposed method is different from E-SDM in [3] in two points. The first point is the difference in the decomposition algorithm applied to MIMO channels. We apply modified sorted QR decomposition (MSQRD) [7] while SVD or eigenvalue decomposition is done in [3]. Hence, the computational complexity for the decomposition algorithm is reduced. The second point is the difference in the power allocation strategy. Whereas we apply the uniform power allocation, the scheme in [3] does the WF to determine the optimum power in each substream. The power allocation strategy of the proposed method is therefore simplified. Numerical examples are provided to demonstrate the BER performance of the proposed method.

In this letter, \mathbb{C} and \mathbb{M} denote a set of complex numbers and a set of complex symbols after digital modulation, respectively. Symbols $\mathbb{C}^{N \times 1}$, $\mathbb{M}^{N \times 1}$, and $\mathbb{C}^{N \times M}$ stand for a set of $N \times 1$ complex vectors, a set of $N \times 1$ complex symbol vectors, and a set of $N \times M$ complex matrices. $[\cdot]^T$ and $[\cdot]^H$ indicate the transpose of a matrix and the Hermitian transpose of a complex matrix, whereas $\text{diag}\{\mathbf{x}\}$ represents a diagonal matrix whose diagonal elements equals the corresponding entries of the vector \mathbf{x} . The pseudo-inverse of matrix \mathbf{A} is defined as $\mathbf{A}^+ = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$. $E[\cdot]$ denotes an expectation operator.

Manuscript received April 5, 2019.

Manuscript revised July 4, 2019.

[†]The author is with the Japan Coast Guard Academy, Kure-shi, 737-8512 Japan.

^{††}The author is with the Graduate School of Engineering, Hiroshima University, Kure-shi, 739-8527 Japan.

a) E-mail: kinjo@jcg.ac.jp

b) E-mail: ohno@hiroshima-u.ac.jp

DOI: 10.1587/transfun.E102.A.1438

2. Maximum Bit-Rate THP

Let us consider a MIMO-OFDM system, where the transmitter sends parallel signals with N_{tx} transmit antennas and the receiver receives them with N_{rx} receive antennas. In the following, we assume that $N_{rx} = N_{tx} = N$ for convenience; but the discussion can be easily extended to general cases where $N_{rx} \geq N_{tx}$. We focus on one subcarrier of the MIMO-OFDM system, which is equivalent to a narrowband MIMO system.

The upper block diagram in Fig. 1 illustrates a ZF-THP with substream permutations [7]. A flat MIMO channel matrix is denoted by $\mathbf{H} \in \mathbb{C}^{N \times N}$, whose (i, j) -th element, $H_{ij} \in \mathbb{C}$, is a channel gain of the flat channel between the transmit antenna j and the receive antenna i . A transmitted symbol vector and a precoded signal vector are represented by $\mathbf{s} \in \mathbb{M}^{N \times 1}$ and $\mathbf{x} \in \mathbb{C}^{N \times 1}$, respectively. Matrices $\mathbf{B} \in \mathbb{C}^{N \times N}$ and $\mathbf{W}_p \in \mathbb{C}^{N \times N}$ are a precoding matrix and a decoding matrix. A permutation matrix \mathbf{E} determines the connection between precoder outputs and transmit antenna elements, whereas δ is a real-valued factor, which attenuates the precoded signals to keep the transmit power within a certain value. A received signal vector \mathbf{y} is expressed as

$$\mathbf{y} = \delta \mathbf{H} \mathbf{E} \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{n} \in \mathbb{C}^{N \times 1}$ is the noise vector whose elements are independent circularly symmetric complex Gaussian random variables with zero mean and variance σ_n^2 [12]. The blocks indicated by "MOD" in Fig. 1 denote the modulo operation, which is installed at the precoder to suppress the amplitude of the precoded signals [13]. The variance of two-dimensional signals with \mathcal{M} -array square constellations, such as \mathcal{M} -QAM, is given by $2(\mathcal{M} - 1)/3$, whereas that of the THP signals is $2\mathcal{M}/3$ [13]. We therefore need to attenuate the transmitted signal \mathbf{x} by

$$\delta = \sqrt{\left(\frac{1}{K} \sum_{j=1}^K \frac{M_j}{M_j - 1} \right)^{-1}} \quad (2)$$

where M_j is the number of constellations of the j -th transmitted symbol, and $K \leq N$ denotes the number of active

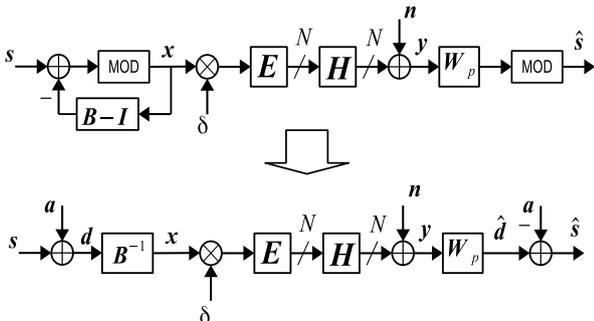


Fig. 1 ZF-THP and its alternative linear representation.

substreams. The active substream is defined as a substream where the number of bits to be allocated is not zero.

The lower figure of Fig. 1 is an alternative representation of the ZF-THP, where $\mathbf{a} \in \mathbb{C}^{N \times 1}$ is a perturbation vector that is produced by the modulo operation [9]. The precoded signal vector is then given by $\mathbf{x} = \mathbf{B}^{-1} \mathbf{d}$, where $\mathbf{d} = \mathbf{s} + \mathbf{a}$. Eq. (1) is rewritten as

$$\mathbf{y} = \delta \mathbf{H} \mathbf{E} \mathbf{B}^{-1} \mathbf{d} + \mathbf{n} = \delta \mathbf{H}_s \mathbf{d} + \mathbf{n}. \quad (3)$$

The decoding matrix (ZF weight matrix) $\mathbf{W}_p \in \mathbb{C}^{N \times N}$ is given by

$$\mathbf{W}_p = \delta^{-1} \mathbf{H}_s^+ = \delta^{-1} \mathbf{B} \mathbf{E}^T \mathbf{H}^+. \quad (4)$$

The estimator of \mathbf{d} is given by $\hat{\mathbf{d}} = \mathbf{W}_p \mathbf{y}$, and the estimated symbol vector is $\hat{\mathbf{s}} = \hat{\mathbf{d}} - \mathbf{a}$.

Suppose that the QR decomposition of $\mathbf{H} \mathbf{E}$ is given by $\mathbf{H} \mathbf{E} = \mathbf{Q} \mathbf{R}$, where $\mathbf{Q} \in \mathbb{C}^{N \times N}$ satisfies $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}$. $\mathbf{R} \in \mathbb{C}^{N \times N}$ is an upper triangular matrix whose (i, j) -th element is r_{ij} . If we define a diagonal matrix, $\mathbf{G} = \text{diag}\{ [r_{11} \ r_{22} \ \cdots \ r_{NN}] \}$, then \mathbf{B} is given by [13]

$$\mathbf{B} = \mathbf{G}^{-1} \mathbf{R}. \quad (5)$$

Equation (3) is re-expressed as

$$\mathbf{y} = \delta \mathbf{Q} \mathbf{R} \mathbf{B}^{-1} \mathbf{d} + \mathbf{n} = \delta \mathbf{Q} \mathbf{G} \mathbf{d} + \mathbf{n}. \quad (6)$$

Multiplying both sides by \mathbf{Q}^H results in

$$\mathbf{y}' = \mathbf{Q}^H \mathbf{y} = \delta \mathbf{G} \mathbf{d} + \mathbf{n}', \quad (7)$$

where $\mathbf{y}' \in \mathbb{C}^{N \times 1}$ and $\mathbf{n}' \in \mathbb{C}^{N \times 1}$.

It follows from Eq. (7) that

$$\hat{\mathbf{d}} = (\delta \mathbf{G})^{-1} \mathbf{y}' = \mathbf{d} + (\delta \mathbf{G})^{-1} \mathbf{n}'. \quad (8)$$

Removing \mathbf{a} from Eq. (8) leads to $\hat{\mathbf{s}} = \mathbf{s} + (\delta \mathbf{G})^{-1} \mathbf{n}'$, since $\mathbf{d} = \mathbf{s} + \mathbf{a}$ and $\hat{\mathbf{d}} = \hat{\mathbf{s}} + \mathbf{a}$. The SNR in the i -th substream is then given by

$$\gamma_i = \frac{\delta^2 r_{ii}^2 P_s}{K \sigma_n^2}, \quad (9)$$

where $P_s = E[\mathbf{s}^H \mathbf{s}]$, since equal transmit power is allocated to each active substream.

The diagonal elements of \mathbf{G} depend on the permutation matrix \mathbf{E} . In the original MBR-THP [7], [8], the modified sorted QR decomposition (MSQRD) is applied to obtain the matrices \mathbf{B} , \mathbf{G} and \mathbf{E} . The diagonal elements of \mathbf{G} then satisfy the relation: $r_{11} \geq r_{22} \geq \cdots \geq r_{NN}$. It is therefore obvious from Eq. (9) that $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N$. A fixed bit pattern has been allocated to each substream of the MBR-THP in [7], [8].

3. Adaptive Bit Allocation for MBR-THP

Although E-SDM is different from the MBR-THP, the relation: $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N$ arises also in E-SDM, which

Table 1 Parameters for BER functions.

m_i	α	β	Modulation
2	1/2	2	QPSK
4	3/8	10	16QAM
6	7/24	42	64QAM

Table 2 Proposed bit allocation scheme for MBR-THP.

	H , M and possible bit allocation patterns are given.
1	Produce G , B and E by the MSQRD.
2	Calculate \mathcal{P}_b over the entire bit allocation patterns.
3	Find a bit allocation pattern which minimizes \mathcal{P}_b .
4	Transmit a signal based on the determined bit allocation.

has been utilized to develop an adaptive bit allocation for E-SDM [3]. We develop here a similar adaptive bit allocation to obtain a further performance improvement for the MBR-THP.

Let us define a BER averaged over all the substreams as

$$\mathcal{P}_b = \frac{1}{M} \sum_{i=1}^N m_i P_b(m_i, \gamma_i), \quad M = \sum_{i=1}^N m_i, \quad (10)$$

where $P_b(\cdot)$ is the BER function depending on modulation schemes [3] and m_i is the number of the bits in the i -th substream, which should satisfy the relation: $m_1 \geq m_2 \geq \dots \geq m_N$ since $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_N$. Our objective is to find a bit allocation pattern that minimizes \mathcal{P}_b .

The approximated BER P_b for QAM is given by [3]

$$P_b \approx \alpha \operatorname{erfc}\left(\sqrt{\frac{\gamma}{\beta}}\right), \quad (11)$$

where γ denotes an SNR, α and β depend on the number of bit m_i as shown in Table 1, and $\operatorname{erfc}(\cdot)$ is the complementary error function. It is known that the upper-bound of P_b is given by

$$P_b \leq 2\alpha e^{\gamma/\beta}. \quad (12)$$

This upper-bound for Eq. (11) has been utilized in [3] to simplify the calculation. We also use the upper-bound to calculate \mathcal{P}_b in Eq. (10).

Table 2 summarizes the proposed adaptive bit allocation, which requires the information about the MIMO channel matrix, H , and the total number of bits, M . We list the bit allocation patterns utilized for our scheme in Table 3, where a combination of N and M is denoted by (N, M) , and $(m_1 m_2 \dots m_N)$ represents the number of bits in each substream. In Table 3, all the possible bit allocation patterns subject to $m_1 \geq m_2 \geq \dots \geq m_N$ are listed for a given (N, M) .

In the first step of the adaptive bit allocation, we decompose H into the matrices G , B and E by using the MSQRD. We then calculate \mathcal{P}_b by using Eq. (10) over the entire bit allocation patterns in Table 3 for a given (N, M) in the second step. In Eq. (12), γ in the i -th substream is given by Eq. (9). In the third step, we determine a bit allocation pattern that minimizes \mathcal{P}_b , and we transmit a signal vector

Table 3 Bit allocation patterns.

No.	$(N, M) = (4, 16)$	$(N, M) = (6, 24)$	$(N, M) = (8, 32)$
1	(6640)	(666600)	(66666200)
2	(6442)	(666420)	(66664220)
3	(6622)	(666222)	(66662222)
4	(4444)	(664440)	(66664400)
5		(664422)	(66644420)
6		(644442)	(66644422)
7		(444444)	(66444440)
8			(66444422)
9			(64444442)
10			(44444444)

Table 4 Computational complexity in the second step.

Calculations of Eq. (12)	Summations in Eq. (10)
$3N$	LN
24 ($N = 8$)	80 ($N = 8, L = 10$)

based on the determined bit allocation pattern in the final step.

Now let us evaluate the computational complexity of the scheme in Table 2. In the first step, the MSQRD requires about $2N^3$ flops, which is much less than $22N^3$ flops for SVD [5] in E-SDM. In the second step, we calculate the average BER in Eq. (10) over all the bit allocation patterns, L , where L denotes the maximum number of the bit allocation patterns. To execute this process, we need first to calculate the upper-bound of the BER in Eq. (12) for three modulation schemes defined in Table 1, and for given SNR in the i -th substream, γ_i , where $i = 1, 2, \dots, N$. Equation (12) consists of one real division, the calculation of the exponential function and one real multiplication. The total number of Eq. (12) to be calculated is given by $3N$. The number is independent of the maximum number of bit allocation patterns, L . Next, we have to calculate the average BER in Eq. (10) for all the bit allocation patterns. The total number of the summations required for calculating all the average BER is LN . Multiplying the integer value m_i in Eq. (10) can be incorporated into α_i in Eq. (12), then we can ignore its cost. The result of the complexity analysis is shown in Table 4. We also show the values when $N = 8$ and $L = 10$. Since the MSQRD and SVD require respectively about 1024 and 11264 flops for its execution when $N = 8$, we can say that the computational complexity of the second step is much lower than that of the decomposition algorithms. Regarding the third and fourth steps, we except them from the analysis since their complexity is obviously negligible.

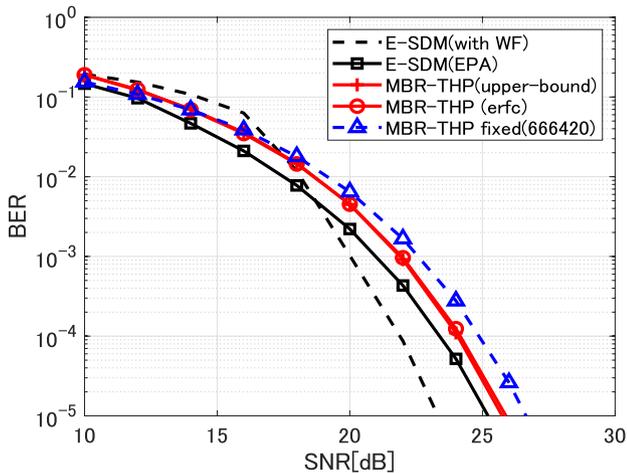
4. Numerical Examples

We demonstrate some numerical examples to evaluate the proposed adaptive bit allocation. The simulation parameters are listed in Table 5. The MIMO channel model is based on the Kronecker model [14].

Figure 2 shows BER curves of E-SDM and the MBR-THP as a function of SNR. We assume uncorrelated MIMO channel, where the correlation factors ρ_t and ρ_r for the trans-

Table 5 Simulation parameters.

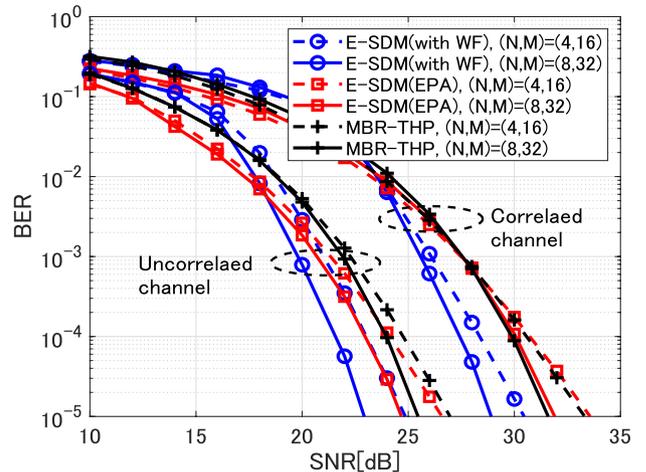
Channel	Kronecker model
Antennas	$N = 4, 6, 8$
Modulation	QPSK, 16-QAM, 64-QAM
M	16, 24, 32
FEC	No coding

**Fig. 2** BER curves when $(N, M) = (6, 24)$.

mitter and the receiver sides are zero. Two cases are demonstrated for E-SDM. In one case, we apply the power allocation using the WF, and in the other case, we do the equal power allocation (labeled as EPA in the figure). The upper-bound in Eq. (12) is applied to calculate average BERs for the adaptive bit allocation. For the MBR-THP, the cases of using Eq. (11) and Eq. (12) for the average BER calculation are demonstrated. A BER curve with the fixed bit allocation is also shown for the reference, where we take (666420) as the bit allocation pattern since it has shown the best BER performance when $(N, M) = (6, 24)$ [8].

We can see first that applying the adaptive bit allocation to the MBR-THP leads to the performance improvement. As for the MBR-THP with the adaptive bit allocation, the difference in the BER calculation gives no impact. E-SDM with the power allocation using the WF shows the best BER performance. E-SDM with EPA is degraded by 1.5 dB from the best BER curve at the BER of 10^{-4} , where the MBR-THP with adaptive bit allocation is degraded by about 0.7 dB from the BER of E-SDM with EPA.

In Fig. 3, we compare E-SDM and the MBR-THP when uncorrelated and correlated channels are applied. The correlation factors are $\rho_t = \rho_r = 0.7$ for the correlated channel. The upper-bound in Eq. (12) is used for the adaptive bit allocation in all the schemes. The BER performance degrades in correlated MIMO channels regardless of the precoding scheme since the channel condition tends to be poor. When we compare the performance of the precoding schemes, E-SDM with optimum power allocation by using the WF outperforms the other cases. E-SDM with EPA outperforms the MBR-THP by about 1 dB at the BER of 10^{-4} when $(N, M) = (8, 32)$ in the uncorrelated channel case, but the

**Fig. 3** BER curves with uncorrelated and correlated channels.

difference decreases when $(N, M) = (4, 16)$. On the other hand, in the correlated channel case with EPA, E-SDM and the MBR-THP show almost the same BER performance. In the correlated channel condition, MIMO channel matrices tend to be ill-conditioned with high probability. This means that, in E-SDM, the SNR in the first substream tends to be extremely higher than that of the other substreams. The SNRs of the substreams except for the first substream are then degraded significantly so that they are almost equivalent to that of the MBR-THP. As shown in Fig. 3, the difference of the BER performance between two schemes can be small since the BER performance is dominated by these degraded substreams.

These results show that the MBR-THP is comparable with E-SDM in the BER performance if the equal power allocation is adopted.

5. Conclusions

The adaptive bit allocation for the MBR-THP has been proposed. The number of bits in each substream was determined adaptively so as to minimize the average BER. The numerical examples demonstrated that the BER performance was improved by using the proposed bit allocation in comparison to the fixed bit allocation. In addition, the proposed method achieved comparable BER performance with E-SDM if an equal power is allocated in the correlated channel.

References

- [1] T. Ohgane, T. Nishimura, and Y. Ogawa, "Applications of space division multiplexing and those performance in a MIMO channel," *IEICE Trans. Commun.*, vol.E88-B, no.5, pp.1843–1851, May 2005.
- [2] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H.V. Poor, *MIMO Wireless Communications*, Cambridge University Press, Cambridge, 2007.
- [3] K. Miyashita, T. Nishimura, T. Ohgane, Y. Ogawa, Y. Takatori, and K. Cho, "High data-rate transmission with eigenbeam-space division multiplexing (E-SDM) in a MIMO channel," *Proc. IEEE 56th Vehicular Tech. Conf.*, vol.3, pp.1302–1306, Vancouver, Canada, Sept.

- 2002.
- [4] K. Ban, M. Katayama, T. Yamazato, and A. Ogawa, "Joint optimization of transmitter/receiver with multiple transmit/receive antennas in band-limited channels," *IEICE Trans. Commun.*, vol.E83-B, no.8, pp.1697–1704, Aug. 2000.
- [5] G.H. Golub and C.F.V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1996.
- [6] C. Windpassinger, R.F.H. Fischer, T. Vencel, and J.B. Huber, "Precoding in multi-antenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol.3, no.4, pp.1305–1316, July 2004.
- [7] S. Kinjo and S. Ohno, "Tomlinson-Harashima precoding with sub-stream permutations based on the bit rate maximization for single-user MIMO systems," *IEICE Trans. Fundamentals*, vol.E98-A, no.5, pp.1095–1104, May 2015.
- [8] S. Kinjo and S. Ohno, "A new multiuser MIMO system with sum-rate maximization," *IEICE Commun. Express*, vol.7, no.8, pp.297–302, Aug. 2018.
- [9] L. Cao and S. Denno, "Nonlinear precoding for XOR physical layer network coding in bi-directional MIMO relay systems," *IEICE Trans. Commun.*, vol.E100-B, no.3, pp.440–448, March 2017.
- [10] Y. Shinbo, N. Hiruma, and F. Maehara, "Performance evaluation of MU-MIMO THP with user scheduling in terms of system capacity and fairness," *IEEE International Symposium on Antenna and Propagation*, pp.1–2, Phuket, Thailand, Oct. 2017.
- [11] M. Payaro, A. Perez-Neira, and M.A. Lagunas, "Achievable rates for generalized spatial Tomlinson-Harashima precoding in MIMO systems," *Proc. IEEE 60th Vehicular Tech. Conf.*, pp.2462–2466, Los Angeles, U.S.A., Sept. 2004.
- [12] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol.24, no.3, pp.528–541, March 2006.
- [13] R.F.H. Fischer, *Precoding and Signal Shaping for Digital Transmission*, Wiley-Interscience, New York, 2002.
- [14] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on the NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol.53, no.3, pp.655–665, May 2004.
-