IEICE TRANSACTIONS

on Communications

DOI:10.23919/transcom.2023EBN0001

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE COMMUNICATIONS SOCIETY

The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

POSITION PAPER

Overloaded MIMO Spatial Multiplexing Independent of Antenna Setups

Satoshi DENNO[†], ***Senior Member***, **Takumi SUGIMOTO**[††], **Koki MATOBA**[†††], ***Nonmembers***, ***and*** **Yafei HOU**[†*], ***Senior Member***

SUMMARY This paper proposes overloaded MIMO spatial multiplexing that can increase the number of spatially multiplexed signal streams despite of the number of antennas on a terminal and that on a receiver. We propose extension of the channel matrix for the spatial multiplexing to achieve the superb multiplexing performance. Precoding based on the extended channel matrix plays a crucial role in carrying out such spatial multiplexing. We consider three types of QR-decomposition techniques for the proposed spatial multiplexing to improve the transmission performance. The transmission performance of the proposed spatial multiplexing is evaluated by computer simulation. The simulation reveals that the proposed overloaded MIMO spatial multiplexing can implement 6 stream-spatial multiplexing in a 2×2 MIMO system, i.e., the overloading ratio of 3.0. The superior transmission performance is achieved by the proposed overloaded MIMO spatial multiplexing with one of the QR-decomposition techniques.

key words: *Overloaded MIMO, Spatial multiplexing, QR-decomposition, precoding, overloading ratio.*

1. Introduction

Communication speed has been raised to about Gbps by using many cutting-edge techniques even in wireless communication systems. Among them, multiple input multiple output (MIMO) spatial multiplexing has been playing an important role in enhancing communication speed [1]–[3]. For the enhancement, many MIMO techniques have been proposed such as serial interference cancelers based on the minimum mean square error (MMSE), precoders, iterative decoders, and so on [4]–[7]. To multiply the throughput enhancement, lots of antennas are installed on the base station in the fifth generation (5G) cellular system cellular system, which is called "Massive MIMO" [8]–[10]. While those techniques achieve superior performance [11]–[13], those techniques are unable to increase the user throughput. For increasing the user throughput, some techniques have been proposed such as non-orthogonal multiple access [14]–[19], faster-than-Nyquist (FTN) [20], and overloaded MIMO spatial multiplexing [21]. Especially, overloaded MIMO spatial multiplexing can increase the download throughput in massive MIMO systems by making use of the system configuration where lots of antennas are installed on the base stations. The use of massive MIMO in the latest wireless systems such as the 5G cellular system has induced researchers to have an interest in overloaded MIMO spatial multiplexing. Since the number of the spatially multiplexed signal streams is set to be more than the degree of freedom of linear receivers in overloaded MIMO systems, non-linear receivers have been considered for the signal detection at the receivers [22], [23]. However, although non-linear receivers achieve superior transmission performance, because the non-linear receivers execute brute force search, they imposes prohibitive high computational complexity on the receivers. Complexity reduction techniques for them have been proposed [24]–[27]. Some techniques to improve the performance of those complexity reduced non-linear detectors have also been investigated [28], [29]. On the other hand, linear detectors have been considered for overloaded MIMO spatial multiplexing where virtual channels are introduced to assist signal detection in overloaded MIMO channels [30], [31]. Those techniques need complex signal processing, such as oversampling and non-linear optimization, in addition to the linear signal detection and the linear precoding. For further complexity reduction, linear detectors based on serial interference cancellation have been proposed to detect overloaded MIMO spatial multiplexed signal streams [32], [33]. Whereas many different techniques have been utilized in overloaded MIMO systems, they are proposed for the systems where the number of receive antennas is less than that of transmit antennas. Especially, the number of spatially multiplexed signal streams is limited by the number of transmit antennas in all of the conventional overloaded MIMO systems, including the systems with the virtual channels.

This paper proposes overloaded MIMO spatial multiplexing that can increase the number of spatially multiplexed signal streams to more than that of not only transmit antennas but also receive antennas in order to enhance link throughput. Actually, the number of spatially multiplexed signal streams can be raised despite of that of antenna settings on a receiver or a transmitter [†]. For such spatial multiplexing, the channel matrix is extended with some appropriate matrices in the proposed overloaded MIMO spatial multiplexing.

[†]Faculty of Environmental, Life, Natural Science and Technology, Okayama University

[††]Graduate School of Natural Science and Technology, Okayama University

[†††]Graduate School of Environmental, Life, Natural Science and Technology, Okayama University

[†]While the proposed overloaded MIMO spatial multiplexing introduces a virtual transmission signal vector as shown in the following, the proposed multiplexing never makes use of the virtual channels.

Precoding based on the extended channel matrix is applied in order to make the number of the spatially multiplexed signal streams exceed that of the antennas on the terminal or the receiver, which characterizes the proposed overloaded MIMO spatial multiplexing, while the number of the signal streams is limited by the number of antennas in conventional overloaded MIMO systems. Three types of QR-decomposition techniques are considered for the precoding to improve the transmission performance. The performance of the proposed overloaded MIMO spatial multiplexing is evaluated by computer simulation.

Throughout the paper, $\Re[c]$ and $\Im[c]$ represent a real part and an imaginary part of a complex number c, respectively. Superscript T and H indicate transpose and Hermitian transpose of a matrix or a vector, respectively. $\mathrm{tr}[\mathbf{A}]$, $\mathbf{A}_{m,n}$, and $\mathrm{E}[c]$ indicate a trace of a matrix $\mathbf{A}$, an (m,n)-entry of a matrix $\mathbf{A}$, and the ensemble average of c.

2. System Model

We assume an MIMO system where a transmitter with N_T antennas transmits some signal streams for a receiver with N_R antennas. While channel coding is usually used in current wireless communication systems, any channel coding is not assumed in the system. The signal streams from the modulators are provided to a precoder that is explained in the following section. The precoder output signal streams are fed to the N_T antennas for the signal transmission. The signal streams are traveling through fading channels and received at the N_R antennas on the receiver. Let $\mathbf{Y} \in \mathbb{C}^{N_\mathrm{R}}$ represent a received signal vector, the system model is written as follows.

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{N_\mathrm{T}}$, $\mathbf{N} \in \mathbb{C}^{N_\mathrm{R}}$, and $\mathbf{H} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{T}}$ denote a transmission signal vector, an additive white Gaussian noise (AWGN) vector, and a channel matrix defined as,

$$\mathbf{H} = \begin{pmatrix} h(1,1) & h(1,2) & \cdots & h(1,N_\mathrm{T}) \\ h(2,1) & h(2,2) & & \vdots \\ \vdots & & \ddots & \\ h(N_\mathrm{R},1) & \cdots & & h(N_\mathrm{R},N_\mathrm{T}) \end{pmatrix}. \tag{2}$$

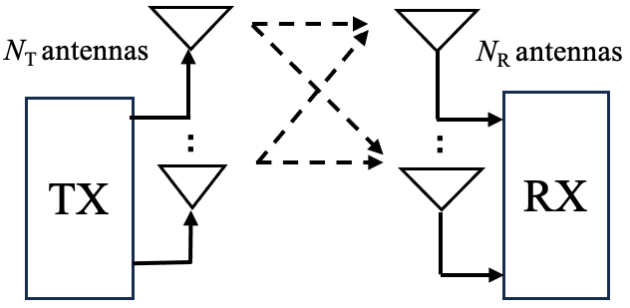


Fig. 1 System Model

In (2), $h(n,m) \in \mathbb{C}$ represents a channel impulse response between the mth antenna on the transmitter and the nth antenna on the receiver, respectively. The system model is illustrated in the figure 1. This system is regarded as one of single-user MIMO systems.

In conventional MIMO systems, the number of spatially multiplexed signal streams N_S is reduced to $\min[N_\mathrm{T}, N_\mathrm{R}]$. As is known, overload MIMO techniques increase the number of signal streams to N_T even when the number of receive antennas N_R is less than that of transmit antennas N_T. The number of the signal streams is restricted by the number of the antennas on a receiver or a transmitter.

We propose a technique to overcome the restriction in this paper. In a word, the proposed technique enables the number of spatially multiplexed signal streams N_S to be greater than the maximum of the number of the antennas on receiver and that on transmitter, i.e., $N_\mathrm{S} > \max[N_\mathrm{T}, N_\mathrm{R}]$.

3. Overloaded MIMO Spatial Multiplexing

While the system model is written in (1), the system model can be rewritten as follows.

$$\overline{\mathbf{Y}} = \begin{pmatrix} \mathbf{H} & \mathbf{0}_{N_\mathrm{R}\times N_\mathrm{B}} \\ \mathbf{0}_{N_\mathrm{A}\times N_\mathrm{T}} & \mathbf{0}_{N_\mathrm{A}\times N_\mathrm{B}} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \tilde{\mathbf{X}} \end{pmatrix} + \begin{pmatrix} \mathbf{N} \\ \mathbf{0}_{N_\mathrm{A}} \end{pmatrix} \tag{3}$$

In (3), $\mathbf{0}_{N\times M}$, $\tilde{\mathbf{X}} \in \mathbb{C}^{N_\mathrm{B}\times 1}$, and $\overline{\mathbf{Y}} \in \mathbb{C}^{(N_\mathrm{R}+N_\mathrm{A})\times 1}$ denote an $N \times M$-dimensional null matrix, an N_B-dimensional virtual transmission signal vector, and an extended received signal vector defined as $\overline{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y}^\mathrm{T} & \mathbf{0}_{N_\mathrm{A}}^\mathrm{T} \end{pmatrix}^\mathrm{T}$ where $\mathbf{0}_N$ represents the N-dimensional null vector. We introduce an extended channel matrix $\overline{\mathbf{H}} \in \mathbb{C}^{(N_\mathrm{R}+N_\mathrm{A})\times(N_\mathrm{T}+N_\mathrm{B})}$ in this paper, which is defined as,

$$\overline{\mathbf{H}} = \begin{pmatrix} \mathbf{H} & \mathbf{J}_1 \\ \mathbf{J}_2 & \mathbf{J}_3 \end{pmatrix}. \tag{4}$$

In (4), $\mathbf{J}_1 \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{B}}$, $\mathbf{J}_2 \in \mathbb{C}^{N_\mathrm{A}\times N_\mathrm{T}}$, and $\mathbf{J}_3 \in \mathbb{C}^{N_\mathrm{A}\times N_\mathrm{B}}$ indicate matrices, which are shown as examples in the following section. Let $\overline{\mathbf{X}} \in \mathbb{C}^{(N_\mathrm{T}+N_\mathrm{B})\times 1}$ denote an extended transmission signal vector defined as $\overline{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^\mathrm{T} & \tilde{\mathbf{X}}^\mathrm{T} \end{pmatrix}^\mathrm{T}$, the system model can be further rewritten in the following.

$$\overline{\mathbf{Y}} = \overline{\mathbf{H}}\,\overline{\mathbf{X}} + \overline{\mathbf{N}}, \tag{5}$$

where $\overline{\mathbf{N}} \in \mathbb{C}^{(N_\mathrm{R}+N_\mathrm{A})\times 1}$ represents an extended noise vector defined as follows.

$$\overline{\mathbf{N}} = \begin{pmatrix} \left(\mathbf{N} - \mathbf{J}_1\tilde{\mathbf{X}}\right)^\mathrm{T} & \left(-\mathbf{J}_2\mathbf{X} - \mathbf{J}_3\tilde{\mathbf{X}}\right)^\mathrm{T} \end{pmatrix}^\mathrm{T} \tag{6}$$

While the system defined in (1) comprises the transmitter with N_T antennas and the receiver with N_R antennas, the system in (5) looks extended to consist of the transmitter with $N_\mathrm{T} + N_\mathrm{B}$ antennas and the receiver with $N_\mathrm{R} + N_\mathrm{A}$ antennas. In other words, not only the number of the transmit antennas but also that of the receive antennas are increased in the extended system model. This extended system model enables overloaded MIMO spatial multiplexing independent of antennas settings, in principle. Moreover, we propose a

linear overloaded MIMO system, which can be implemented with small computational complexity. The detail is described in the following sections.

3.1 QR-decomposition For Extended Channel Matrix

Triangulation of a channel matrix is introduced in not only wireless communication systems but also wired communication systems in order to improve the transmission performance or to reduce computational complexity. The triangulation can be performed not only at a receiver but also at a transmitter, because linear signal processing such as the triangulation can be transferred between a receiver and a transmitter in linear communication systems. The triangulation is usually implemented with QR-decomposition based on some signal processing techniques such as Gram-Schmidt orthonormalization. When the QR-decomposition is applied to the extended channel matrix, the channel matrix is decomposed into a semi-orthogonal matrix $\mathbf{Q}_{\mathrm{R}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$, a diagonal matrix $\mathbf{\Gamma}_{\mathrm{R}} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{B}})\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$, and an upper triangular matrix $\mathbf{R}_{\mathrm{R}} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{B}})\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$, which is expressed in the following.

$$\overline{\mathbf{H}}\mathbf{S}_{\mathrm{R}} = \mathbf{Q}_{\mathrm{R}}\mathbf{\Gamma}_{\mathrm{R}}\mathbf{R}_{\mathrm{R}} \tag{7}$$

However, the diagonal elements in the upper triangular matrix $\mathbf{R}_{\mathrm{R}}$ are all ones. $\mathbf{S}_{\mathrm{R}} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{B}})\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$ in (7) denotes a transform matrix. For successful triangulation, the extended channel matrix has to be slim, which results in the following requirement.

$$N_{\mathrm{R}} + N_{\mathrm{A}} \geq N_{\mathrm{T}} + N_{\mathrm{B}} \tag{8}$$

When the requirement in (8) is not satisfied, we consider to apply triangulation to the Hermite transpose of the extended channel matrix $\overline{\mathbf{H}}$ as follows.

$$\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{S}_{\mathrm{T}} = \mathbf{Q}_{\mathrm{T}}\mathbf{\Gamma}_{\mathrm{T}}\mathbf{R}_{\mathrm{T}} \tag{9}$$

Similar as the triangulation in (7), $\mathbf{S}_{\mathrm{T}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$, $\mathbf{Q}_{\mathrm{T}} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{B}})\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$, $\mathbf{\Gamma}_{\mathrm{T}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$, and $\mathbf{R}_{\mathrm{T}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$ denote a transform matrix, a semi-orthogonal matrix, a diagonal matrix, and an upper triangular matrix with ones in the diagonal positions. Same to (7), the Hermite transform of the extended channel matrix has to be slim, which can be expressed in the following equation.

$$N_{\mathrm{R}} + N_{\mathrm{A}} \leq N_{\mathrm{T}} + N_{\mathrm{B}} \tag{10}$$

As is shown above, the transform matrices are required for the QR-decomposition, which is truly necessary for the linear precoding and the THP as described in the following sections.

3.2 QR-decomposition Techniques

Whereas the Gram-Schmidt orthonormalization is one of QR-decomposition techniques, other QR-decomposition techniques have been applied to MIMO systems. Some representatives of them are listed below. The notation of $\mathbf{S}_{\Omega}$ Ω = R or T is used for describing $\mathbf{S}_{\mathrm{R}}$ or $\mathbf{S}_{\mathrm{T}}$. The same notation is applied for $\mathbf{\Gamma}_{\mathrm{R}}$ or $\mathbf{\Gamma}_{\mathrm{T}}$ and $\mathbf{R}_{\mathrm{R}}$ or $\mathbf{R}_{\mathrm{T}}$ in the following.

(a) Sorted QR-decomposition [34]
This technique tries to arrange the diagonal elements of the diagonal matrix $\mathbf{\Gamma}_{\Omega}$ in ascending order, although it is not guaranteed to carry out the arrangement. In other words, let $\mathbf{\Gamma}_{\Omega}$ be defined as $\mathbf{\Gamma}_{\Omega} = \mathrm{diag}\left[\gamma_{\Omega}(1)\cdots\gamma_{\Omega}(N_{\Omega})\right]$ where $\gamma_{\Omega}(m) \in \mathbb{R}$ denotes the mth diagonal elements of the matrix $\mathbf{\Gamma}_{\Omega}$, the arrangement is mathematically described as,

$$|\gamma_{\Omega}(1)| \lesssim |\gamma_{\Omega}(2)| \lesssim \cdots \lesssim |\gamma_{\Omega}(N_{\Omega})|. \tag{11}$$

In the sorted QR-decomposition (SQRD), $\mathbf{S}_{\Omega}$ is reduced to a permutation matrix.

(b) Lattice Reduction [35]
While several techniques to implement the lattice reduction have been proposed and evaluated, the Lenstra–Lenstra–Lovász (LLL) algorithm [36] has been widely used in the field of communications [†], because it takes only a polynomial time length to find a near optimum vector from the view point of the lattice reduction [††]. Let the upper triangular matrix $\mathbf{R}_{\Omega}$ be defined as,

$$\mathbf{R}_{\Omega} = \begin{pmatrix} 1 & r_{\Omega}(1,2) & \cdots & r_{\Omega}(1,N_{\Omega}) \\ 0 & 1 & r_{\Omega}(2,3) & \vdots \\ \vdots & 0 & \ddots & r_{\Omega}(N_{\Omega}-1,N_{\Omega}) \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{12}$$

The LLL algorithm achieves the following equations.

$$\Re\left[r_{\Omega}(n,m)\right] \leq \frac{1}{2}, \quad \Im\left[r_{\Omega}(n,m)\right] \leq \frac{1}{2} \tag{13}$$

$$\delta\left|\gamma_{\Omega}(m-1)\right|^2 \leq \left|\gamma_{\Omega}(m)\right|^2 + \left|\gamma_{\Omega}(m-1)\, r_{\Omega}(m-1,m)\right|^2 \tag{14}$$

$\delta \in \mathbb{R}$ in (14) denotes a parameter, which is set as $\frac{1}{2} < \delta < 1$ [37].

(c) Equal Gain Transform [38]
The equal gain transform [38] equalizes the diagonal elements with the transform matrix as,

$$\gamma_{\Omega}(1) = \gamma_{\Omega}(2) = \cdots = \gamma_{\Omega}(N_{\Omega}), \tag{15}$$

The matrix $\mathbf{S}_{\Omega}$ is served as a unitary matrix, i.e., $\mathbf{S}_{\Omega}^{\mathrm{H}}\mathbf{S}_{\Omega} = \mathbf{I}_{N_{\Omega}}$.

The derivation of those algorithms is described in the papers [34], [35], [38].

We propose overloaded MIMO spatial multiplexing

[†]The LLL algorithm has been applied to overloaded MIMO systems, and its superior performance has been shown [32], [33]. This is the reason why the LLL algorithm is applied to our proposed overloaded MIMO spatial multiplexing.

[††]The LLL algorithm can be applied to any matrix as far as the matrix can be successfully QR-decomposed, i.e., the matrix is slim.

with QR-decomposition where the above three techniques are applied in the following section [†].

3.2.1 Linear Precoding

If the above QR-decomposition techniques are applied to the extended channel matrix defined in (4), we can realize that the transform matrix $\mathbf{S}_{\mathrm{R}}$ can be used as a linear precoding matrix.

$$\overline{\mathbf{X}} = g_{\mathrm{R}}\mathbf{S}_{\mathrm{R}}\mathbf{D}, \tag{16}$$

where $\mathbf{D} \in \mathbb{C}^{(N_{\mathrm{T}}+N_{\mathrm{B}})\times 1}$ and $g_{\mathrm{R}} \in \mathbb{R}$ represent a modulation signal vector and a normalization factor that keeps the transmission power constant. Let the transform matrix $\mathbf{S}_{\mathrm{R}}$ be decomposed as $\mathbf{S}_{\mathrm{R}} = \begin{pmatrix}\mathbf{S}_{\mathrm{R},1}^{\mathrm{T}} & \mathbf{S}_{\mathrm{R},2}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$ where $\mathbf{S}_{\mathrm{R},1} \in \mathbb{C}^{N_{\mathrm{T}}\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$ and $\mathbf{S}_{\mathrm{R},2} \in \mathbb{C}^{N_{\mathrm{B}}\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$ denote an upper and a lower part of the transform matrix $\mathbf{S}_{\mathrm{R}}$, the transmission signal vector $\mathbf{X}$ and the virtual transmission signal vector $\tilde{\mathbf{X}}$ can be written as,

$$\mathbf{X} = g_{\mathrm{R}}\mathbf{S}_{\mathrm{R},1}\mathbf{D} \quad \text{and} \quad \tilde{\mathbf{X}} = g_{\mathrm{R}}\mathbf{S}_{\mathrm{R},2}\mathbf{D}. \tag{17}$$

The normalization factor g_{R} can be defined with only the transmission signal vector $\mathbf{X}$ because the vector is actually transmitted from the antennas.

$$g_{\mathrm{R}} = \sqrt{\frac{P_0}{\sigma_{\mathrm{d}}^2 \mathrm{tr}\left[\mathbf{S}_{\mathrm{R},1}^{\mathrm{H}}\mathbf{S}_{\mathrm{R},1}\right]}} = \sqrt{\frac{P_0}{N_{\mathrm{T}}\sigma_{\mathrm{d}}^2}} \tag{18}$$

$P_0 \in \mathbb{R}$ and $\sigma_{\mathrm{d}}^2 \in \mathbb{R}$ in (18) represent power of the transmission signal and that of the modulation signal defined as $\mathrm{E}\left[\mathbf{D}\mathbf{D}^{\mathrm{H}}\right] = \sigma_{\mathrm{d}}^2\mathbf{I}_{N_{\mathrm{T}}+N_{\mathrm{B}}}$.

In order to apply the above linear precoding, the system has to satisfy the requirement written in (8). When the system does not meet the requirement, another precoding that meets the requirement written in (10) is necessary, which is proposed in the next section.

3.2.2 THP Based on MMSE

Since precoding based on the MMSE criterion achieves superior transmission performance [39], we introduce precoding based on the MMSE for the system defined in (5). Let $\mathbf{A} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ represent a precoder input signal vector, a precoder based on the MMSE provides an extended transmission signal vector $\overline{\mathbf{X}}$ defined as,

$$\begin{aligned}\overline{\mathbf{X}} &= g_{\mathrm{T}}\overline{\mathbf{H}}^{\mathrm{H}}\left(\overline{\mathbf{H}}\,\overline{\mathbf{H}}^{\mathrm{H}}\right)^{-1}\mathbf{A} \\ &= g_{\mathrm{T}}\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{S}_{\mathrm{T}}\left(\{\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{S}_{\mathrm{T}}\}^{\mathrm{H}}\{\overline{\mathbf{H}}^{\mathrm{H}}\mathbf{S}_{\mathrm{T}}\}\right)^{-1}\mathbf{S}_{\mathrm{T}}^{\mathrm{H}}\mathbf{A}.\end{aligned} \tag{19}$$

In (19), $g_{\mathrm{T}} \in \mathbb{R}$ represents a normalization factor. In the above equation, the extended channel matrix is transformed with the transform matrix $\mathbf{S}_{\mathrm{T}}$. Since we assume that the extended channel matrix satisfies the requirement in (10) as is inferred at the end of the previous section, the extended channel matrix can be transformed in a manner defined in (9). Besides, the precoder input signal $\mathbf{A}$ is defined with a vector $\dot{\mathbf{D}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ as $\mathbf{A} = \mathbf{S}_{\mathrm{T}}\dot{\mathbf{D}}$ where the vector $\dot{\mathbf{D}}$ is defined below. For the definition of the transmission signal vector $\mathbf{X}$, the semi-orthogonal matrix $\mathbf{Q}_{\mathrm{T}}$ is decomposed as $\mathbf{Q}_{\mathrm{T}} = \begin{pmatrix}\mathbf{Q}_{\mathrm{T},1}^{\mathrm{T}} & \mathbf{Q}_{\mathrm{T},2}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$ where $\mathbf{Q}_{\mathrm{T},1} \in \mathbb{C}^{N_{\mathrm{T}}\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$ and $\mathbf{Q}_{\mathrm{T},2} \in \mathbb{C}^{N_{\mathrm{B}}\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$ denote an upper matrix and a lower part of the semi-orthogonal matrix. If the term in the left hand side of (9) is substituted for (19), the transmission signal vector $\mathbf{X}$ and the virtual transmission signal vector $\tilde{\mathbf{X}}$ can be obtained by introducing the semi-orthogonal matrix $\mathbf{Q}_{\mathrm{T}}$ into (19) as follows.

$$\mathbf{X} = g_{\mathrm{T}}\mathbf{Q}_{\mathrm{T},1}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\mathbf{R}_{\mathrm{T}}^{-\mathrm{H}}\dot{\mathbf{D}} = g_{\mathrm{T}}\mathbf{Q}_{\mathrm{T},1}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\mathbf{V} \tag{20}$$

$$\tilde{\mathbf{X}} = g_{\mathrm{T}}\mathbf{Q}_{\mathrm{T},2}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\mathbf{R}_{\mathrm{T}}^{-\mathrm{H}}\dot{\mathbf{D}} = g_{\mathrm{T}}\mathbf{Q}_{\mathrm{T},2}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\mathbf{V} \tag{21}$$

In (20) and (21), $\mathbf{V} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ denotes a feedback filter output vector defined in the following.

$$\mathbf{R}_{\mathrm{T}}^{\mathrm{H}}\mathbf{V} = \dot{\mathbf{D}} \tag{22}$$

$\dot{\mathbf{D}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ in (22) represents a modulation signal vector added with a Gaussian integer multiples vector $\mathbf{K}M \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ where $\mathbf{K} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times 1}$ and $M \in \mathbb{Z}$ indicate a Gaussian integer vector and a modulus. In a word, the vector $\dot{\mathbf{D}}$ is defined as $\dot{\mathbf{D}} = \mathbf{D}+\mathbf{K}M$. The entries of the Gaussian integer vector are successfully obtained as is done in the Tomlinson Harashima precoding (THP) [40] [††]. Therefore, the proposed feedback filter is called as THP based on the MMSE even in this paper. Same to the linear precoding, since only the vector $\mathbf{X}$ is actually transmitted, the normalization factor g_{T} is obtained as,

$$g_{\mathrm{T}} = \sqrt{\frac{P_0}{\mathrm{tr}\left[\mathbf{Q}_{\mathrm{T},1}^{\mathrm{H}}\mathbf{Q}_{\mathrm{T},1}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\mathbf{\Phi}_{\mathrm{V}}\mathbf{\Gamma}_{\mathrm{T}}^{-1}\right]}}. \tag{23}$$

In (23), $\mathbf{\Phi}_{\mathrm{V}} \in \mathbb{C}^{(N_{\mathrm{R}}+N_{\mathrm{A}})\times(N_{\mathrm{R}}+N_{\mathrm{A}})}$ represents an autocorrelation matrix of the feedback filter output vector $\mathbf{V}$, which is defined as $\mathbf{\Phi}_{\mathrm{V}} = \mathrm{E}\left[\mathbf{V}\mathbf{V}^{\mathrm{H}}\right]$.

3.3 SNR Performance

Since the transmission performance can be characterized by the signal to noise power ratio (SNR) of detector output signals in any wireless systems, the SNR performance is analyzed for the performance evaluation of the proposed overloaded MIMO spatial multiplexing in the following.

3.3.1 Linear Precoding

When the linear precoding defined in (16) is applied in the

[†]While the QR-decomposition based on the SQRD has been applied in detectors [34], the QR-decomposition can be transferred from a receiver to a transmitter, because our proposed overloaded MIMO spatial multiplexing is regarded as a linear system as described above.

[††]The triangular matrix $\mathbf{R}_{\mathrm{T}}$ obtained by the transform matrix $\mathbf{Q}_{\mathrm{T}}$ makes the feedback filter available at the transmitter.

extended system, a detector input vector $\mathbf{Z} \in \mathbb{C}^{(N_\mathrm{T}+N_\mathrm{B})\times 1}$ is obtained with the transform matrix and the normalization factor, which can be fed to detectors.

$$\mathbf{Z}_\mathrm{R} = g_\mathrm{R}^{-1}\mathbf{Q}_\mathrm{R}^\mathrm{H}\overline{\mathbf{Y}} = \mathbf{\Gamma}_\mathrm{R}\mathbf{R}_\mathrm{R}\mathbf{D} + g_\mathrm{R}^{-1}\mathbf{Q}_\mathrm{R}^\mathrm{H}\overline{\mathbf{N}} \tag{24}$$

Because the matrix $\mathbf{R}_\mathrm{R}$ is upper triangular as shown in the above equation, serial interference cancelers (SICs) can be used to detect the modulation signal vector $\mathbf{D}$ [†]. Though the SNR performance of SICs is influenced by the error propagation, it is not easy to evaluate the error propagation theoretically. As has been done in the SNR performance analysis, we neglect the error propagation in the SNR performance analysis. Let $\underline{\mathbf{N}}_\mathrm{R} \in \mathbb{C}^{(N_\mathrm{T}+N_\mathrm{B})\times 1}$ denote a detector input noise vector fed into the SIC, i.e., $\underline{\mathbf{N}}_\mathrm{R} = g_\mathrm{R}^{-1}\mathbf{Q}_\mathrm{R}^\mathrm{H}\overline{\mathbf{N}}$, the correlation matrix of the detector input noise vector is derived in (25). σ_d^2 and σ^2 indicate the power of the modulation signal and that of the AWGN. The SNR of the mth stream $\rho_\mathrm{R}(m) \in \mathbb{R}$ can be defined as,

$$\rho_\mathrm{R}(m) = \frac{\gamma_\mathrm{R}^2(m)\,\sigma_\mathrm{d}^2}{\mathrm{E}\left[\underline{\mathbf{N}}_\mathrm{R}\,\underline{\mathbf{N}}_\mathrm{R}^\mathrm{H}\right]_{m,m}}. \tag{26}$$

3.3.2 THP Based on MMSE

Even when precoding based on the MMSE is applied to the system, similar as the system with the linear precoding, a detector input vector can be obtained by multiplying the received signal vector with inverse of the normalization factor g_T^{-1} and the transform matrix $\mathbf{S}_\mathrm{T}$ as,

$$\mathbf{Z}_\mathrm{T} = g_\mathrm{T}^{-1}\mathbf{S}_\mathrm{T}^\mathrm{H}\overline{\mathbf{Y}} = \dot{\mathbf{D}} + g_\mathrm{T}^{-1}\mathbf{S}_\mathrm{T}^\mathrm{H}\overline{\mathbf{N}}. \tag{27}$$

As is seen, the region detection can be used to estimate the modulation signal vector. Let a detector input noise vector $\underline{\mathbf{N}}_\mathrm{T} \in \mathbb{C}^{(N_\mathrm{R}+N_\mathrm{A})\times 1}$ be defined as $\underline{\mathbf{N}}_\mathrm{T} = g_\mathrm{T}^{-1}\mathbf{S}_\mathrm{T}^\mathrm{H}\overline{\mathbf{N}}$, the correlation matrix of the detector input noise vector is shown in (28).

When the feedback filter output signals are assumed to be uniformly distributed, all the diagonal elements of the correlation matrix $\mathbf{\Phi}_\mathrm{V}$ are reduced to $\frac{M^2}{6}$. If the elements of the feedback filter output vector are uncorrelated from each other, the matrix $\mathbf{\Phi}_\mathrm{V}$ approximately results in the diagonal matrix $\frac{M^2}{6}\mathbf{I}_{(N_\mathrm{A}+N_\mathrm{R})\times(N_\mathrm{A}+N_\mathrm{R})}$ [††], which is used in the derivation of (28). The SNR of the mth stream $\rho_\mathrm{T}(m)$ can be defined as,

$$\rho_\mathrm{T}(m) = \frac{\sigma_\mathrm{d}^2}{\mathrm{E}\left[\underline{\mathbf{N}}_\mathrm{T}\,\underline{\mathbf{N}}_\mathrm{T}^\mathrm{H}\right]_{m,m}}. \tag{29}$$

3.4 Discussion

As is shown in (25) and (26) as well as (28) and (29), the SNR depends on the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$. Because this is a first step of the research, we would like to start with the simplest example in order to grasp the basic characteristics of the proposed overloaded MIMO spatial multiplexing [†††]. We would like to apply diagonal matrices and null matrices to those matrices in the following sections, as the simplest example.

3.5 Linear Precoding

As is shown in (25), the matrix $\mathbf{J}_1$ should be set to the null matrix in order to reduce the noise power of the 1st to N_Rth streams. Besides, $\mathbf{J}_2 = \mathbf{J}_3 = \sqrt{\gamma_\mathrm{R}}\mathbf{I}_{N_\mathrm{T}}$ looks the simplest where $\gamma_\mathrm{R} \in \mathbb{R}$ represents a constant. However, this setting constrains the number of the spatially multiplexed signal streams to $2N_\mathrm{T}$, because this setting imposes the parameters to satisfy $N_\mathrm{A} = N_\mathrm{B} = N_\mathrm{T}$. In other words, this setting gets rid of the freedom to set the number of the spatially multiplexed signal streams. On the other hand, let γ_R be set as $\gamma_\mathrm{R} = \frac{\sigma^2}{\sigma_\mathrm{d}^2 g^2} = \frac{N_\mathrm{T}\sigma^2}{P_0}$, it has the probability that the following setting can equalize the noise power of all the streams.

$$\mathbf{J}_1 = \mathbf{0}_{N_\mathrm{R}\times N_\mathrm{B}}, \quad (\mathbf{J}_2 \quad \mathbf{J}_3) = \sqrt{\gamma_\mathrm{R}}\mathbf{I}_{N_\mathrm{A}} \tag{30}$$

When the setting written in (30) is used, the parameter N_B is expressed with the other parameters, i.e., $N_\mathrm{B} = N_\mathrm{A} - N_\mathrm{T}$. The number of the spatially multiplexed signal streams N_S gets equal to N_A. In a word, $N_\mathrm{S} = N_\mathrm{A}$.

When the above matrices are applied to the extended channel matrix, an extended channel matrix $\overline{\mathbf{H}}_\mathrm{R} \in \mathbb{C}^{(N_\mathrm{R}+N_\mathrm{A})\times N_\mathrm{A}}$ for the linear precoding is expressed as [††††],

$$\overline{\mathbf{H}}_\mathrm{R} = \begin{pmatrix} \mathbf{H} & \mathbf{0}_{N_\mathrm{R}\times(N_\mathrm{A}-N_\mathrm{T})} \\ \sqrt{\gamma_\mathrm{R}}\mathbf{I}_{N_\mathrm{A}\times N_\mathrm{A}} \end{pmatrix} \tag{31}$$

When the above setting is applied, if the transform matrix is unitary, i.e., $\mathbf{S}_\mathrm{R}^\mathrm{H}\mathbf{S}_\mathrm{R} = \mathbf{I}_{(N_\mathrm{T}+N_\mathrm{B})\times(N_\mathrm{T}+N_\mathrm{B})}$, the correlation in the detector input noise vector is reduced to the following equation.

$$\mathrm{E}\left[\underline{\mathbf{N}}_\mathrm{R}\,\underline{\mathbf{N}}_\mathrm{R}^\mathrm{H}\right] = \frac{N_\mathrm{T}\sigma_\mathrm{d}^2}{P_0}\sigma^2\mathbf{I}_{N_\mathrm{A}} \tag{32}$$

As is described above, all the streams get uncorrelated and the power of them is equalized. By substituting the noise power in (32) for (26), the SNR of the mth stream $\rho_\mathrm{R}(m)$ is rewritten as,

$$\rho_\mathrm{R}(m) = \frac{\gamma_\mathrm{R}^2(m)}{N_\mathrm{T}}\frac{P_0}{\sigma^2}. \tag{33}$$

While the SNR is described in (33) when the transform matrix $\mathbf{S}_\mathrm{R}$ is unitary, the SNR is dependent on not only the

[†] The transform matrix $\mathbf{Q}_\mathrm{R}$ transforms the extended channel matrix into the triangular matrix $\mathbf{R}_\mathrm{R}$, which makes it possible to apply the SIC at the receiver.

[††] The correlation matrix is given with the assumption that all the streams are added with the Gaussian integer multiples. The derivation on the assumption has been shown in [40].

[†††] The optimization of the those matrices is definitely one of our important future works.

[††††] The extended channel matrix $\overline{\mathbf{H}}_\mathrm{R}$ is identical to the matrix $\overline{\mathbf{H}}$. However, because the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$ are designed suitable for the linear precoding, the extended matrix dares to be named as $\overline{\mathbf{H}}_\mathrm{R}$.

$$\mathrm{E}\left[\underline{\mathbf{N}}_{\mathrm{R}}\,\underline{\mathbf{N}}_{\mathrm{R}}^{\mathrm{H}}\right] = g_{\mathrm{R}}^{-2}\mathbf{Q}_{\mathrm{R}}^{\mathrm{H}}\mathrm{E}\left[\overline{\mathbf{N}}\,\overline{\mathbf{N}}^{\mathrm{H}}\right]\mathbf{Q}_{\mathrm{R}} = \mathbf{Q}_{\mathrm{R}}^{\mathrm{H}}\begin{pmatrix} \frac{\sigma^2}{g_{\mathrm{R}}^2}\mathbf{I}_{N_{\mathrm{R}}} + \sigma_{\mathrm{d}}^2\mathbf{J}_1\mathbf{S}_{\mathrm{R},2}\mathbf{S}_{\mathrm{R},2}^{\mathrm{H}}\mathbf{J}_1^{\mathrm{H}} & \sigma_{\mathrm{d}}^2\mathbf{J}_1\mathbf{S}_{\mathrm{R},2}\left(\mathbf{J}_2\mathbf{S}_{\mathrm{R},1}+\mathbf{J}_3\mathbf{S}_{\mathrm{R},2}\right)^{\mathrm{H}} \\ \sigma_{\mathrm{d}}^2\left(\mathbf{J}_2\mathbf{S}_{\mathrm{R},1}+\mathbf{J}_3\mathbf{S}_{\mathrm{R},2}\right)\left(\mathbf{J}_1\mathbf{S}_{\mathrm{R},2}\right)^{\mathrm{H}} & \sigma_{\mathrm{d}}^2\left(\mathbf{J}_2\mathbf{S}_{\mathrm{R},1}+\mathbf{J}_3\mathbf{S}_{\mathrm{R},2}\right)\left(\mathbf{J}_2\mathbf{S}_{\mathrm{R},1}+\mathbf{J}_3\mathbf{S}_{\mathrm{R},2}\right)^{\mathrm{H}} \end{pmatrix}\mathbf{Q}_{\mathrm{R}} \quad (25)$$

$$\begin{aligned}\mathrm{E}\left[\underline{\mathbf{N}}_{\mathrm{T}}\,\underline{\mathbf{N}}_{\mathrm{T}}^{\mathrm{H}}\right] &= g_{\mathrm{T}}^{-2}\mathbf{S}_{\mathrm{T}}^{\mathrm{H}}\mathrm{E}\left[\overline{\mathbf{N}}\,\overline{\mathbf{N}}^{\mathrm{H}}\right]\mathbf{S}_{\mathrm{T}} = \mathbf{S}_{\mathrm{T}}^{\mathrm{H}}\begin{pmatrix} \frac{\sigma^2}{g_{\mathrm{T}}^2}\mathbf{I}_{N_{\mathrm{R}}} + \mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\Phi_{\mathrm{V}}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\mathbf{Q}_{\mathrm{T},2}^{\mathrm{H}}\mathbf{J}_1^{\mathrm{H}} & \mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\Phi_{\mathrm{V}}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} \\ \left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\Phi_{\mathrm{V}}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\left(\mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} & \left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\Phi_{\mathrm{V}}\boldsymbol{\Gamma}_{\mathrm{T}}^{-1}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} \end{pmatrix}\mathbf{S}_{\mathrm{T}} \\ &= \mathbf{S}_{\mathrm{T}}^{\mathrm{H}}\begin{pmatrix} \frac{\sigma^2}{g_{\mathrm{T}}^2}\mathbf{I}_{N_{\mathrm{R}}} + \frac{M^2}{6}\mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\mathbf{Q}_{\mathrm{T},2}^{\mathrm{H}}\mathbf{J}_1^{\mathrm{H}} & \frac{M^2}{6}\mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} \\ \frac{M^2}{6}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\left(\mathbf{J}_1\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} & \frac{M^2}{6}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\left(\mathbf{J}_2\mathbf{Q}_{\mathrm{T},1}+\mathbf{J}_3\mathbf{Q}_{\mathrm{T},2}\right)^{\mathrm{H}} \end{pmatrix}\mathbf{S}_{\mathrm{T}} \end{aligned} \quad (28)$$

diagonal elements $\gamma_{\mathrm{R}}^2(m)$ and the noise power but also the characteristics of the transform matrix, when the transform matrix is not unitary. Therefore, the SNR performance is evaluated by computer simulation in Sec.4.

However, the above discussion assumes that $N_{\mathrm{B}}\,(=N_{\mathrm{A}}-N_{\mathrm{T}})$ is non-negative, i.e., $N_{\mathrm{A}} \geq N_{\mathrm{T}}$. If the N_{A} is set to be smaller than N_{T}, i.e., $N_{\mathrm{A}} < N_{\mathrm{T}}$, N_{B} gets to be zero, $N_{\mathrm{B}} = 0$, and the number of the spatially multiplexed signal streams is fixed to N_{T}. In fact, the matrices $\mathbf{J}_1$ and $\mathbf{J}_3$ become null, and the matrix $\mathbf{J}_2$ is reduced as,

$$\mathbf{J}_1 = \mathbf{J}_3 = \varnothing, \quad \mathbf{J}_2 = \sqrt{\gamma_{\mathrm{R}}}\left(\mathbf{I}_{N_{\mathrm{A}}} \quad \mathbf{0}_{N_{\mathrm{A}}\times(N_{\mathrm{T}}-N_{\mathrm{A}})}\right). \quad (34)$$

3.6 THP Based on MMSE

We would like to apply the above successful discussion in this section for the THP. However, the settings described in the previous section can not be directly applied to the THP based on the MMSE, because the setting does not meet the requirement in (10). As is described in (9) and (10), the Hermite transpose of the extended channel matrix should be slim to apply it to the THP. Since the extended channel matrix in (31) is slim, it is a good start point that the matrix is used as an Hermite transpose of the extended channel matrix for the THP. However, the channel matrix $\mathbf{H}$ needs to be replaced with its Hermite transpose. In addition, $\mathbf{J}_1$ and $\mathbf{J}_2$ are renamed as $\mathbf{J}_2$ and $\mathbf{J}_1$ in the matrix to be content with the definition of the extended channel matrix. If the setting is applied, the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$ are written as,

$$\mathbf{J}_2 = \mathbf{0}_{N_{\mathrm{A}}\times N_{\mathrm{T}}}, \quad \begin{pmatrix}\mathbf{J}_1^{\mathrm{T}} & \mathbf{J}_3^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}} = \sqrt{\gamma_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{B}}} \quad (35)$$

In (35), the sizes of the channel matrices are adjusted to be consistent with the definition of the extended channel matrix, i.e., $N_{\mathrm{B}} = N_{\mathrm{A}} + N_{\mathrm{R}}$. In a word, the number of the spatially multiplexed signal streams N_{S} becomes equal to $N_{\mathrm{B}}\,(=N_{\mathrm{A}}+N_{\mathrm{R}})$.

If the above replacement is used, consequently, the extended channel matrix $\mathbf{H}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{B}}\times(N_{\mathrm{T}}+N_{\mathrm{B}})}$ for the THP based on the MMSE can be derived as,

$$\overline{\mathbf{H}}_{\mathrm{T}}^{\mathrm{H}} = \begin{pmatrix}\mathbf{H}^{\mathrm{H}} & \mathbf{J}_2^{\mathrm{H}} \\ \mathbf{J}_1^{\mathrm{H}} & \mathbf{J}_3^{\mathrm{H}}\end{pmatrix} = \begin{pmatrix}\mathbf{H}^{\mathrm{H}} & \mathbf{0}_{N_{\mathrm{T}}\times(N_{\mathrm{B}}-N_{\mathrm{R}})} \\ \sqrt{\gamma_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{B}}\times N_{\mathrm{B}}}\end{pmatrix}. \quad (36)$$

Because the parameter setting $\gamma_{\mathrm{T}} = \frac{N_{\mathrm{R}}\sigma^2}{P_0}$ optimizes the transmission performance of THPs [40], we borrow the parameter setting for the proposed THP based on the MMSE [†]. Let the transform matrix $\mathbf{S}_{\mathrm{T}}$ be decomposed as $\mathbf{S}_{\mathrm{T}} = \left(\mathbf{S}_{\mathrm{T},1}^{\mathrm{T}},\, \mathbf{S}_{\mathrm{T},2}^{\mathrm{T}}\right)^{\mathrm{T}}$ where $\mathbf{S}_{\mathrm{T},1} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{B}}}$ and $\mathbf{S}_{\mathrm{T},2} \in \mathbb{C}^{N_{\mathrm{A}}\times N_{\mathrm{B}}}$ represent submatrices of the transform matrix, as a result, the noise correlation is reduced to,

$$\begin{aligned}\mathrm{E}\left[\underline{\mathbf{N}}_{\mathrm{T}}\,\underline{\mathbf{N}}_{\mathrm{T}}^{\mathrm{H}}\right] = \frac{\sigma^2}{g^2}\Big\{&\mathbf{S}_{\mathrm{T},1}^{\mathrm{H}}\mathbf{S}_{\mathrm{T},1} + \Big(\mathbf{S}_{\mathrm{T},1}^{\mathrm{H}}\mathbf{Q}_{\mathrm{T},2,1}\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\mathbf{Q}_{\mathrm{T},2,1}^{\mathrm{H}}\mathbf{S}_{\mathrm{T},1} \\ &+ \mathbf{S}_{\mathrm{T},2}^{\mathrm{H}}\mathbf{Q}_{\mathrm{T},2,2}\boldsymbol{\Gamma}_{\mathrm{T}}^{-2}\mathbf{Q}_{\mathrm{T},2,2}^{\mathrm{H}}\mathbf{S}_{\mathrm{T},2}\Big)\Big\}. \end{aligned} \quad (37)$$

In (37), $\mathbf{Q}_{\mathrm{T},2,1} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{B}}}$ and $\mathbf{Q}_{\mathrm{T},2,2} \in \mathbb{C}^{N_{\mathrm{A}}\times N_{\mathrm{B}}}$ denote submatrices of the matrix $\mathbf{Q}_{\mathrm{T},2}$, which are defined as $\mathbf{Q}_{\mathrm{T},2} = \begin{pmatrix}\mathbf{Q}_{\mathrm{T},2,1}^{\mathrm{T}} & \mathbf{Q}_{\mathrm{T},2,2}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$. Nevertheless, the above analysis is available if $N_{\mathrm{A}}\,(=N_{\mathrm{B}}-N_{\mathrm{R}})$ is non-negative, i.e., $N_{\mathrm{B}} \geq N_{\mathrm{R}}$.

When N_{B} is set to be less than N_{R}, i.e., $N_{\mathrm{B}} < N_{\mathrm{R}}$, N_{A} is imposed to be set to zero, $N_{\mathrm{A}} = 0$, and the number of the spatially multiplexed signal streams is fixed to N_{R}. As a result, the matrices $\mathbf{J}_2$ and $\mathbf{J}_3$ become null, and the matrix $\mathbf{J}_1$ is reduced as,

$$\mathbf{J}_2 = \mathbf{J}_3 = \varnothing, \quad \mathbf{J}_1 = \sqrt{\gamma_{\mathrm{T}}}\left(\mathbf{I}_{N_{\mathrm{B}}} \quad \mathbf{0}_{N_{\mathrm{B}}\times(N_{\mathrm{R}}-N_{\mathrm{B}})}\right)^{\mathrm{T}}. \quad (38)$$

The SNR of the m stream $\rho_{\mathrm{T}}(m)$ is rewritten by substituting the noise variance in (37) for (29). Compared with the SNR performance of the linear precoding, that of the THP is a little bit more complex.

Table 1 Simulation parameters

Modulation	QPSK
Channel model	IID, Keyhole based on Rayleigh fading
$(N_{\mathrm{T}}, N_{\mathrm{R}})$	$(2,\ 2)$
Channel estimation	Perfect
Lattice reduction	LLL algorithm
Overloading Ratio	2.0, 3.0
Error correction coding	N.A.

3.7 Characteristics of Proposed Scheme and Other Issues

As is described in Sec. 1, many overloaded MIMO techniques have been proposed before. Some of them achieve

[†]The optimization of the parameter γ_{T} for the proposed THP based on the MMSE is one of our future works.

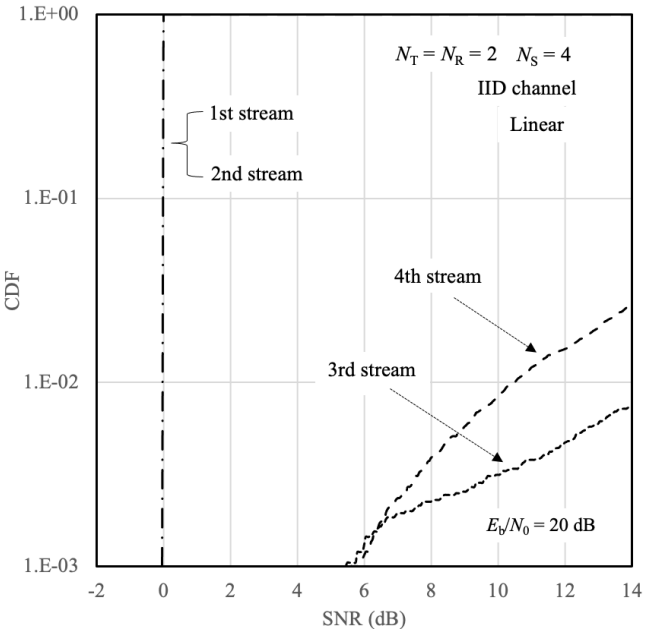


Fig. 2 SNR Distribution of Linear Precoding with SQRD

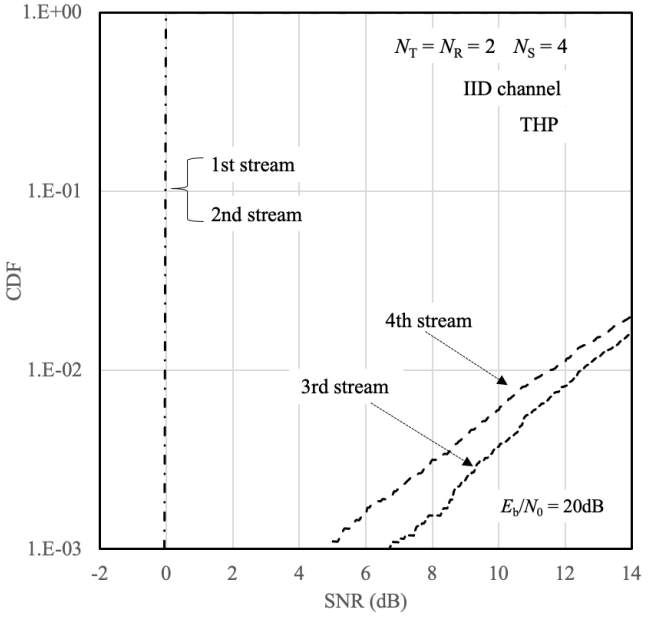


Fig. 3 SNR Distribution of THP with SQRD

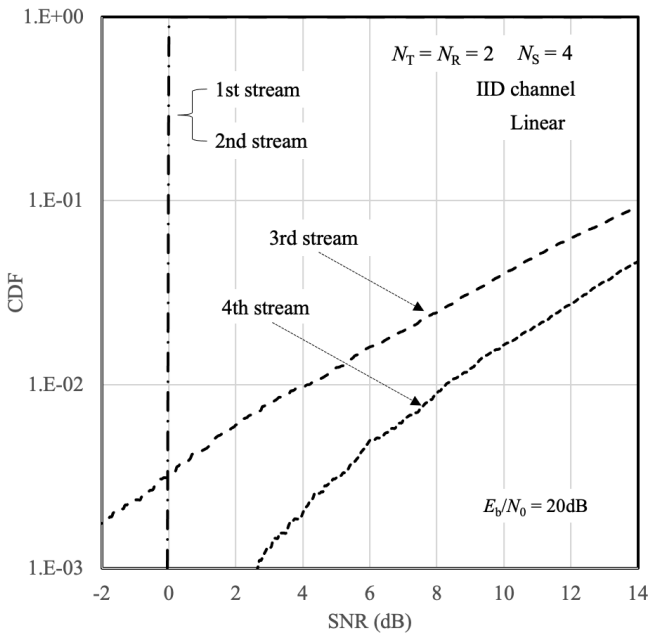


Fig. 4 SNR Distribution of Linear Precoding with Lattice Reduction

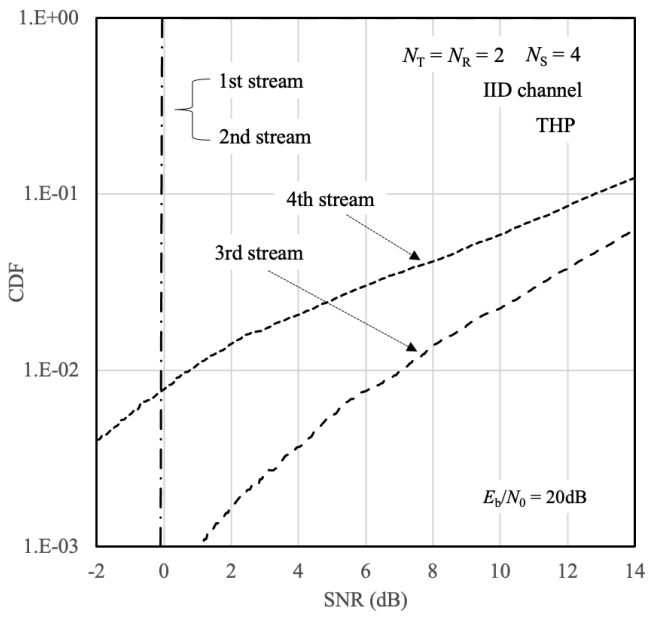


Fig. 5 SNR Distribution of THP with Lattice Reduction

fair transmission performance with relatively high computational complexity. For instance, the overloaded MIMO with the virtual channels needs sequence estimation like the maximum likelihood sequence estimation (MLSE) to achieve such performance [30]. Another overloaded MIMO with the virtual channels requires a complex non-linear optimization every packet [31]. In those techniques, the number of the spatially multiplexed signal streams is limited by that of the antennas on the transmitter. On the other hand, the proposed overloaded MIMO spatial multiplexing can raise the number of the spatially multiplexed signal streams to more than that of the antennas on the receiver or the transmitter, which characterizes the proposed multiplexing. Because such overloaded spatial multiplexing can not be implemented other than the proposed overloaded MIMO, it is impossible to compare the proposed overloaded MIMO spatial multiplexing with the conventional techniques in terms of the transmission performance and the computational complexity.

To implement them, one of the overloaded MIMO systems with the virtual channels optimizes the precoding weights and feeds them back to the precoders at the trans-

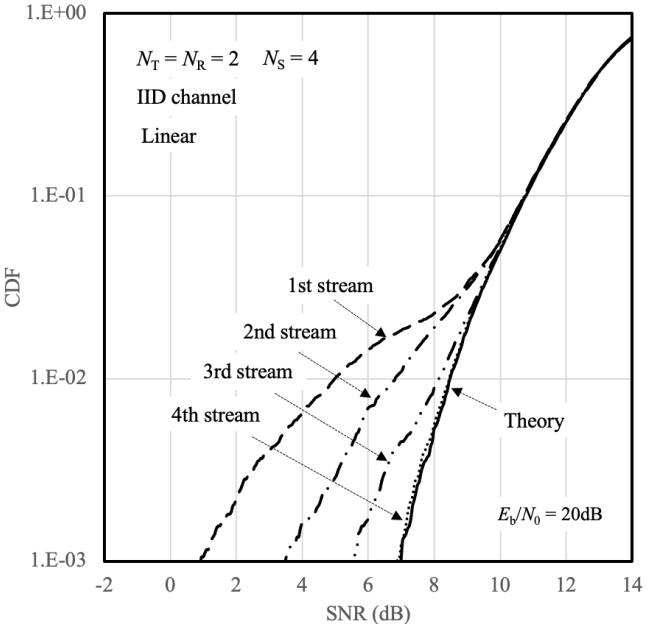


Fig. 6 SNR Distribution of Linear Precoding with Equal Gain Transform

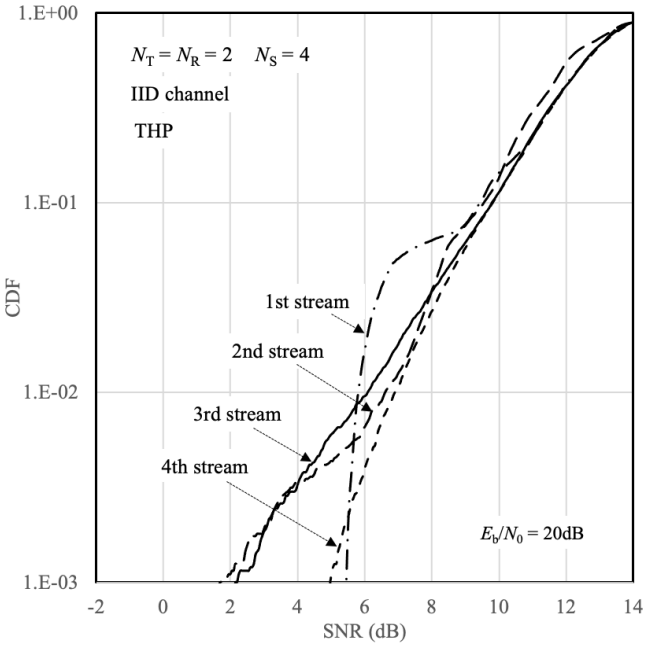


Fig. 7 SNR Distribution of THP with Equal Gain Transform

mitter, every slot [31]. While no information needs to be fed back in the MIMO system proposed in the paper [30], the packet length should be small if the computational complexity is taken into account. On the other hand, the proposed overloaded MIMO spatial multiplexing should share the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$ among all the receiver and the transmitter before the communication starts. However, the matrices are products of the identity matrices and the scalar value as defined in (30), (34), (35), and (38). While the scalar value needs to be set based on the noise variance, those identity matrices can be set in the transmitter and the receiver when they are configured at factories. If the identity matrices are not set at factories, it is enough to broadcast them for the receiver and the transmitter once when they are turned on. In a word, the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$ don't need to be frequently exchanged between the transmitter and the receiver. The noise variance has to be sent not only in the proposed multiplexing but also in systems with the MMSE precoding. The amount of the information to send the noise variance is negligible small compared with that of the information bits to send. This information exchange before the communication does not cause any serious problem.

On the other hand, the proposed overloaded MIMO spatial multiplexing needs the spatial filters at the receiver as shown in (24) and (27). The need for the spatial filters at the receiver restricts the proposed overloaded MIMO spatial multiplexing to single-user MIMO systems.

4. Computer Simulation

The performance of the proposed overloaded MIMO spatial multiplexing is evaluated by computer simulation. The modulation scheme is quaternary phase shift keying (QPSK) [41]. The proposed spatial multiplexing enables the number of the spatially multiplexed signal streams to be greater than the maximum number of the antennas in the system, i.e., $\max[N_T, N_R]$. This means that the number of the spatially multiplexed signal streams exceeds that of the eigenvalues in the channel. To confirm how the number of the eigenvalues affects the transmission performance, we evaluate the transmission performance in not only independent and identically distributed (IID) Rayleigh fading channels but also Keyhole channels based on the Jakes' model [42]. The number of the transmit antennas N_T and that of the received antennas N_R are all set to 2. The number of the spatially multiplexed signal streams N_S is increased to 4 or 6, which correspond with the overloading ratio of 2.0 or 3.0, respectively. The LLL algorithm is used for the lattice reduction. The simulation parameters are listed in Table 1. The following performance comparison is performed on the same assumption, for instance, the error correction coding is not applied as listed in the Table. Only the QR-decomposition makes the difference in the performance as shown in the following sections.

4.1 SNR performance

The signal power to noise power ratio (SNR) distribution of the proposed overloaded spatial multiplexing is analyzed to grasp the transmission performance. The number of the spatially multiplexed signal streams N_S is set to 4 in Sec. 4.1. This subsection shows some figures where the abscissa is the SNR in dB and the ordinate is the cumulative distribution function. The E_b/N_0 is set to 20 dB in this section. While the SNR performance are analyzed above theoretically, the SNR performance is evaluated by means of computer simulations in this section. Since the detector input signals are defined in (24) and (27), the SNR measurement

depends on the precoding schemes. When the linear precoding is employed, the noise power can be measured as $\sigma_{\mathrm{R}}^2(m) = \mathrm{E}\left[\left|\frac{1}{\gamma_{\mathrm{R}}} z_{\mathrm{R}}(m) - d(m) - \sum_{k=N_{\mathrm{T}}}^{m-1} r_{\mathrm{R}}(m,k)\,\bar{d}(k)\right|^2\right]$ where $\sigma_{\mathrm{T}}^2(m)$, $z_{\mathrm{T}}(m)$, $d(m)$, and $\bar{d}(k)$ denote a noise power in an mth detector input signal, an mth entry of the detector input signal $\mathbf{Z}_{\mathrm{T}}$, that of the modulation signal vector $\mathbf{D}$, and a kth estimated modulation signal, respectively. When the THP is used. on the other hand, the noise power can be measured as $\sigma_{\mathrm{T}}^2(m) = \mathrm{E}\left[|z_{\mathrm{T}}(m) - d(m)|^2\right]$ where $\sigma_{\mathrm{T}}^2(m)$ and $z_{\mathrm{T}}(m)$ represent a noise power in the mth detector input signal and an mth entry of the detector input signal $\mathbf{Z}_{\mathrm{T}}$, respectively. Hence, the SNR $\rho_\Omega(m)$ can be calculated as $\rho_\Omega(m) = \frac{\sigma_{\mathrm{d}}^2}{\sigma_\Omega^2(m)}$ where Ω takes R or T.

Fig. 2 and Fig. 3 show the distribution functions of the SNR of all the streams when the linear precoding and the THP with the SQRD are applied to the proposed multiplexing, respectively. While the probability that the SNR distribution of 3rd and 4th streams gets less than 5 dB is about 10^{-3}, the SNR of the 1st and the 2nd streams is fixed about 0 dB. The performance is greatly dependent on the distribution of the diagonal elements of the matrix $\mathbf{\Gamma}_\Omega$ $\Omega =$ R or T. While the number of the eigenvalues in the channel is equal to 2 in the setting of the section, the number of the eigenvalues in the extended channel matrices is the same to that of the signal streams, i.e., 4 in spite of the type of the precoding. Also, the diagonal matrix $\mathbf{\Gamma}_\Omega$ has 4 diagonal non-zero elements. On the other hand, when the E_{b}/N_0 is high enough, $\sqrt{\gamma_\Omega}$ is smaller than all the eigenvalues in the channel matrix $\mathbf{H}$. Even if the SQRD is applied, the eigenvalues are concentrated into only 2 ($= N_{\mathrm{R}} = N_{\mathrm{T}}$) diagonal elements, while the other diagonal elements are almost same to $\sqrt{\gamma_\Omega}$, which causes that the SNR of the two streams gets much less than that of the others. This causes half of the detector input signals to be almost zero, while the other input signals are non-zero values. In addition, the non-diagonal elements in the 1st and 2nd row of the upper triangular matrix $\mathbf{R}_{\mathrm{R}}$ become zero [†]. This causes the noise power $\sigma_{\mathrm{R}}(m)$ rewritten as $\sigma_{\mathrm{R}}^2(m) = \mathrm{E}\left[\left|\frac{1}{\gamma_{\mathrm{R}}} z_{\mathrm{R}}(m) - d(m)\right|^2\right]$ for $m = 1, 2$. Hence, the noise power can be reduced to $\sigma_{\mathrm{R}}(m) = \mathrm{E}[|d(m)|^2] = \sigma_{\mathrm{d}}^2$. Eventually, the SNR of the 1st and the 2nd streams is reduced to 0dB. Even when the THP is applied, the similar performance can be seen due to the reason described above, as long as the SQRD is applied. Therefore, the SNR of the 1st and the 2nd streams is fixed to 0dB.

Fig. 4 and Fig. 5 show the distribution functions of the SNR of all the streams when the linear precoding and the THP with the lattice reduction are applied to the proposed multiplexing, respectively. Similar as the performance with the SQRD, the SNR of the 1st and the 2nd is about 0 dB, while that of the 3rd and the 4th streams is much bigger than 0 dB. Even when the lattice reduction is applied for QR-decomposition, the 1st and 2nd diagonal elements of the diagonal matrix $\mathbf{\Gamma}_\Omega$ is much smaller than the 3rd and 4th diagonal elements, which causes the SNR performance to degrade as shown in Fig. 4 when the linear precoding is applied. Even if the THP is used, the similar SNR performance can be obtained as shown in Fig. 5.

Fig. 6 and Fig. 7 show the SNR distribution functions when the linear precoding and the THP with the equal gain transform are applied to the proposed multiplexing, respectively. "Theory" in Fig. 6 shows the theoretical performance written in (33) [††]. The theoretical SNR distributions of all the streams are same when the linear precoding is used. Since the equal gain transform equalizes the diagonal elements of the diagonal matrix $\mathbf{\Gamma}_\Omega$, the SNR of all the streams is equalized, which agrees with the simulation result shown in Fig. 6. Actually, the SNR distribution of the signal streams is different from each other because of the error propagation. While the SNR distribution of the 4th stream is the worst, that of the 1st stream is the best, which is the same to the theoretical SNR performance, because the propagation error does not happen in the first stream. Even when the THP is employed, the similar performance is expected, because of the performance of the equal gain transform. In fact, the SNR distributions of all the streams are only a little bit deviated as shown in Fig. 7. As is shown in the figure, the probability that the SNR is less than 10 dB is about 10^{-1}. The SNR is called a required SNR for 10% outage probability. On the other hand, when the linear precoding is applied, the required SNR for 10% outage probability is improved to about 11 dB as shown in Fig. 6. In a word, the linear precoding achieves about 1.0 dB better SNR performance than the THP at the 10% outage probability.

4.2 BER Performance

Since the three QR-decomposition techniques are considered, those techniques are compared in terms of the BER performance in this section. The number of the spatially multiplexed signal streams is fixed to 4, which means that the overloading ratio is 2.0. The channel model is the IID based on the Jake's model.

4.2.1 Performance VS. QR-Decomposition Techniques

Fig. 8 shows the BER performances of the proposed overloaded MIMO spatial multiplexing when the linear precoding is applied. In the figure, the three QR-decomposition techniques are compared in terms of the BER. When the LLL and the SQRD are used, the proposed spatial multiplexing can't reduce the irreducible error to less than 0.3, which agrees with the SNR distributions shown in Fig. 2 and Fig. 4. On the other hand, the equal gain transform makes the proposed spatial multiplexing achieve superior performance,

[†]The theoretical performance analysis of those phenomena is one of our future works.

[††]While the performance can be regarded as an upper bound, we dare to name the performance as "Theory", because we intend to emphasize that the performance is derived theoretically, and to show the performance in comparison with the simulation results.

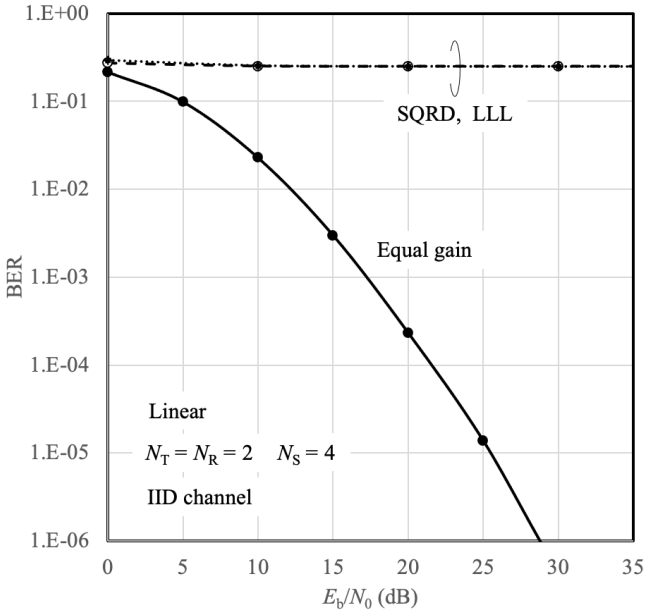


Fig. 8 BER Performance of Linear Precoding with QR-decomposition techniques

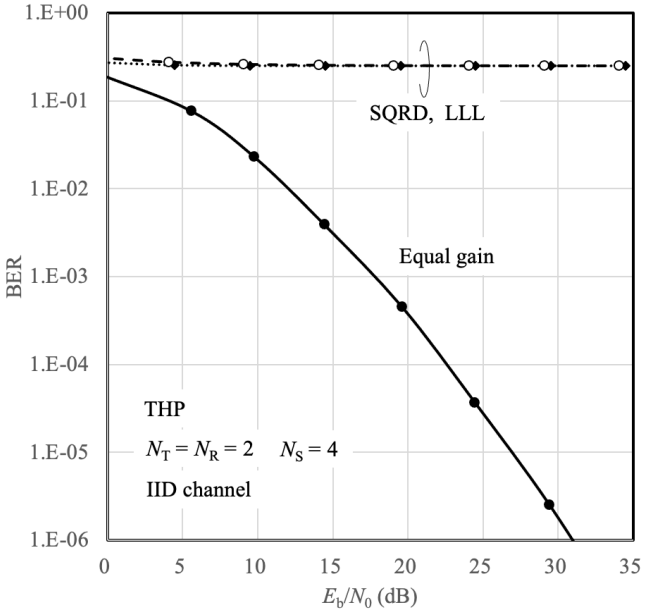


Fig. 9 BER Performance of THP with QR-decomposition techniques

which also agrees with the SNR distribution shown in Fig. 6.

Fig. 9 shows the BER performance when the THP is used. Also, the three QR-decomposition techniques are employed. Similar as Fig. 8, the irreducible error appears at about 0.3 when the LLL and the SQRD are used for the QR-decomposition. On the other hand, the equal gain transform enables the proposed overloaded MIMO spatial multiplexing to attain the superior performance. Besides, the linear precoding achieves about 1.0 dB better BER performance than the THP at the BER of 10^{-5}, which agrees with the SNR performance comparison described at the end of Sec.4.1.

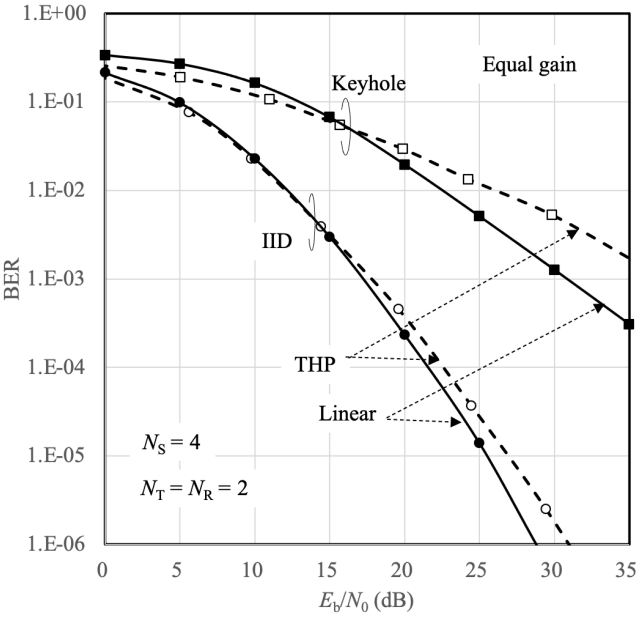


Fig. 10 BER Performance in Keyhole Channels

4.2.2 BER Performance in Keyhole Channel

As is shown in the previous sections, the proposed overloaded MIMO spatial multiplexing achieves superior performance, even though the number of the spatially multiplexed signal streams is twice as many as that of the eigenvalues in the channel. This section evaluates the BER performance in the channels with less than 2 eigenvalues. We apply a keyhole channel for the performance evaluation. Let $\mathbf{h}_R \in \mathbb{C}^{N_R \times 1}$ and $\mathbf{h}_T \in \mathbb{C}^{1 \times N_T}$ denote vectors defined as $\mathbf{h}_R = (h_R(1) \cdots h_R(N_R))^T$ and $\mathbf{h}_T = (h_T(1) \cdots h_T(N_T))$ where $h_\Omega(m) \in \mathbb{C}$ represents a channel gain, Ω = R or T, the keyhole channel is defined as $\mathbf{H} = \mathbf{h}_R \mathbf{h}_T$. In the performance evaluation, the channel gains are generated based on the Jake's model for fair performance comparison. While this keyhole channel is a kind of Rayleigh fading channels, the number of the eigenvalues is reduced to 1.

Fig. 10 shows the BER performance of the proposed overloaded MIMO spatial multiplexing in the Keyhole channel. In the figure, the equal gain transform is employed for the QR-decomposition. The performances of not only the linear precoding but also the THP are illustrated in the figure. The performance in the IID channel is added as a reference in the figure. Whereas the performance in the keyhole channel is much worse than that in the IID channel, however, the irreducible error does not appear up to the BER of 10^{-3}, This means that the decrease in the number of eigenvalues in channels does not cause the irreducible error up to 10^{-3}, even though the performance is degraded as the number of the eigenvalues in channels is decreased. The linear precoding is able to make the proposed overloaded MIMO spatial multiplexing achieve better BER performance than the THP even in the keyhole channel.

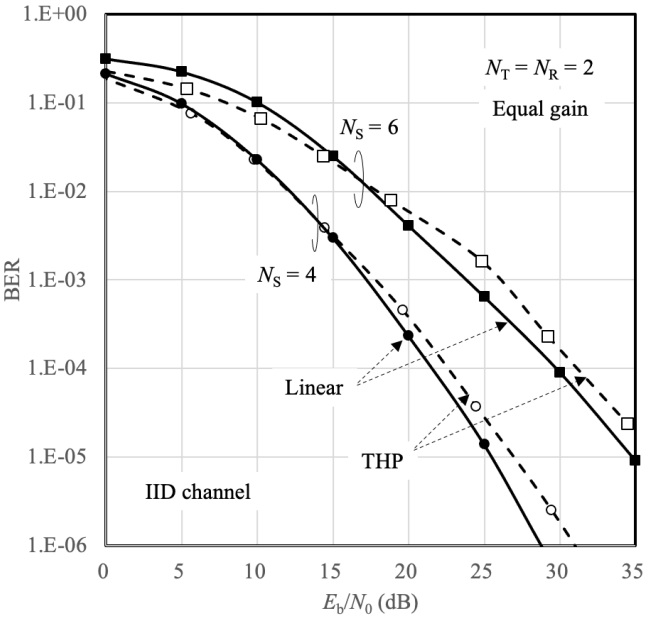


Fig. 11 BER Performance of Linear precoding in overloaded MIMO channels

4.3 Overloading Ratio

Fig. 11 compares the BER performance with the overloading ratio of 2.0 and that of 3.0 in the IID channel. In the figure, the equal gain transform is used. The number of the receive antennas and that of the transmit antennas are all fixed to 2. The overloading ratio is raised by changing the extended channel matrices. For example, the overloading ratios of 2.0 and 3.0 are implemented with setting the parameter N_A to 4 and 6, respectively, when the linear precoding is applied. When the THP is used, the parameter N_B is also set to 4 and 6 for the overloading ratio of 2.0 and 3.0,respectively. The proposed overloaded MIMO spatial multiplexing with the THP outperforms that with the linear filter in the E_b/N_0 of less than 15dB, while the performance of the proposed spatial multiplexing with the THP is inferior to that with the linear filter in the other E_b/N_0 region [†]. The proposed overloaded MIMO spatial multiplexing achieves superior transmission performance even if the overloading ratio is raised to 3.0. The proposed overloaded MIMO spatial multiplexing with the linear precoding achieves about 1.0 dB better BER performance than that with the THP, despite of the overloading ratio.

5. Conclusion

This paper has proposed overloaded MIMO spatial multiplexing that can increase the number of spatially multiplexed signal streams in spite of antenna settings on a receiver or a transmitter. The extension of a channel matrix with some matrices has been proposed to implement such overloaded signal transmission. Three tapes of QR-decomposition techniques such as the SQRD, the lattice reduction with the LLL algorithm, and the equal gain transform have been considered for precoding in the proposed overloaded MIMO spatial multiplexing. We apply linear precoding and the THP for the precoding, which can be used complementary. This paper shows some examples for the extended channel matrices and analyze the performance theoretically. The performance is confirmed by computer simulation. The linear precoding achieves better performance that the THP in the system with the example of the extended channel matrices. The equal gain transform makes the proposed overloaded MIMO spatial multiplexing achieve the best transmission performance in the three types of the QR-decomposition techniques. The proposed overloaded spatial multiplexing with the linear precoding based on the equal gain transform achieves 6 spatially multiplexed signal streams transmission with superior transmission performance, even though only two antennas are employed at both the receiver and the transmitter. In a word, the proposed overloaded MIMO spatial multiplexing is able to increase the number of the spatially multiplexed signal streams despite of the number of the antennas on both the receiver and the transmitter.

[†]Such performances probably depends on the the choice of the matrices $\mathbf{J}_1$, $\mathbf{J}_2$, and $\mathbf{J}_3$. However, the performance analysis is one of our future works.

Acknowledgment

The work has been supported by JSPS KAKENHI JP21K04061 and 24K07475, the support center for advanced telecommunications technology research (SCAT), and Softbank Co. ltd.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol.6, no.3, pp.311—335, 1998.
[2] I. E. Telater, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp.585–595. 1999.
[3] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys & Tutorials*, vol.17, no.4, pp.1941–1988, Fourthquarter 2015.
[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Lab. Tech. J.*, vol.1, no.2, pp.41—59, 1996.
[5] P. W. Wolniansky, G. J. Foschini; G. D. Golden, R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," Proc. IEEE ISSSE-98, Pisa, Italy, 30 September 1998.
[6] Q. H. Spencer, A. L. Swindlehurst, M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol.52, no.2, pp.461–471, 2004.
[7] T. Abe, S. Tomisato, and T. Matsumoto, "A MIMO turbo equalizer for frequency-selective channels with unknown interference," *IEEE Trans. Veh. Technol.*, vol.53, no.3, pp.476–482, 2003.
[8] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine,* vol. 52, no. 2, pp.186–195, Feb. 2014.

[9] M. Sakai, K. Kamohara, H. Iura, H. Nishimoto, K. Ishioka, Y. Murata, M. Yamamoto, A. Okazaki, N. Nonaka, S. Suyama, J. Mashino, A. Okamura, and Y. Okumura, "Experimental field trials on MU-MIMO transmissions for high SHF wide-band massive MIMO in 5G," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp.2196–2207, April 2020.

[10] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Select. Topics in Signal Process.*, vol.8, no.5, pp.742–758, Oct. 2014.

[11] P. Som, T. Datta, A. Chockalingam, and B. S. Rajan, "Improved large-MIMO detection based on damped belief propagation," *IEEE Information Theory Workshop on Information Theory* (ITW), 2010.

[12] W. Fukuda, T. Abiko, T. Nishimura, T. Ohgane, Y. Ogawa, Y. Ohwatari, and Y. Kishiyama, "Low-complexity detection based on belief propagation in a massive MIMO system," *IEEE 77th Veh. Technol. Conf.* (VTC Spring), 2013.

[13] T. Takahashi, S. Ibi, and S. Sampei, "On normalization of matched filter belief in GaBP for large MIMO detection," *IEEE 84th Veh. Technol. Conf.* (VTC-Fall), 2016.

[14] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, Vol.56, No4, pp.1616-1626, April 2008.

[15] R. Stoica, G. Abreu, Z. Liu, T. Hara, and K. Ishibashi, "Massively concurrent non-orthogonal multiple access for 5G networks and beyond" *IEEE Access*, Vol.7, pp.82080-82100, June 2019.

[16] S. M. A. Kazmi, N. H. Tran, T. M. Ho, D. Niyato, and C. S. Hong, "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks," *IEEE Access*, Vol.6, pp.39860-39875, June 2018.

[17] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, Vol.E98-B, No.3, pp.403–414, March 2015.

[18] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp.2347–2381, Dec. 2017.

[19] H. Yang, X. Fang, Y. Liu, X. Li, Y. Luo, and D. Chen, "Impact of overloading on link-level performance for sparse code multiple access," *25th Wireless and Optical Communication Conference* (WOCC), 2016.

[20] M. Anan, M. Sawahashi, and Y. Kishiyama, "BLER performance of windowed-OFDM using faster-than-Nyquist signaling with 16QAM," *21st International Symposium on Wireless Personal Multimedia Communications* (WPMC), 2018.

[21] K. K. Wong, A. Paulraj, and R. D. Murch, "Efficient high-performance decoding for overloaded MIMO antenna systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp.1833–1843 ,May 2007.

[22] N. Surajudeen-Bakinde, X. Zhu, J. Gao, and A. K. Nandi, "Improved signal detection approach using genetic algorithm for overloaded MIMO systems," *4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008.

[23] I. Shubhi and Y. Sanada, "Joint turbo decoding for overloaded MIMO-OFDM systems," *IEEE Trans. Vehicular Technology*, vol.66, no.1, pp.433–442, Jan. 2017.

[24] X. Lian and D. Li, "On the application of sphere decoding algorithm in overload MIMO systems," *IEEE International Conference on Information Theory and Information Security*, Dec. 2010.

[25] S. Yoshikawa, S. Denno, and M. Morikura, "Complexity reduced lattice-reduction-aided MIMO receiver with virtual channel detection," *IEICE Trans. Commun.*, vol. E96-B, no.1, 263–270, 2013.

[26] R. Hayakawa, K. Hayashi, and M. Kaneko, "Lattice reduction-aided detection for overloaded MIMO using slab decoding," *IEICE Trans. Commun.*, vol. E99.B, no.8, pp.1697–1705, Jan. 2016.

[27] R. Hayakawa and K. Hayashi, "Convex optimization-based signal detection for massive overloaded MIMO systems," *IEEE Trans. Wireless Commun.*, vol.16, no.11, pp.7080–7091, Nov. 2017.

[28] R. Shioji, T. Imammura, and, Y. Sanada, "Overloaded MIMO detection based on two-stage belief propagation with MMSE pre-cancellation," *IEEE Vehicular Technol. Conf.*(VTC2021-Fall), Sept.27–30, 2021.

[29] T. Takahashi, S. Ibi, A. Tölli, and S. Sampei, "Subspace marginalized belief propagation for mmWave overloaded MIMO signal detection," *IEEE Intern. Conf. Commun.*(ICC2020), June 7–11, 2020.

[30] D. K. C. So and Y. Lan, "Virtual receive antenna for overloaded MIMO layered space-time system," *IEEE Trans. Commun.*, vol.60, no.6, pp.1610–1620, June 2012.

[31] G. Yang, Y. Zhou, and W. Xia, "Performance optimization for overloaded MIMO systems with virtual channel approach," *Hindawi Wireless Commun. Mobile Comput.*, vol.2018, Article ID:9651378, pp.1–6, 2018.

[32] S. Denno, Y. Kawaguchi, H. Murata, and D. Umehara, "An iterative noise cancelling receiver with soft-output LR-aided detection for collaborative reception," *the 19th International Symposium on Wireless Personal Multimedia Communications* (WPMC 2016), Shenzhen China, Nov. 14–16, 2016.

[33] S. Denno, T. Inoue, T. Fujiwara, and Y. Hou, "Low complexity soft input decoding in an iterative linear receiver for overloaded MIMO," *IEICE Trans. Commun.*, vol. E103-B, no.5, pp.600–608, May 2020.

[34] D. Wübben, R. Böhnke, V. Kühn, and K.-D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," *IEEE 58th Veh. Technol. Conf.* (VTC2003-Fall), Orlando Fl, USA, Oct. 6–9, 2003.

[35] D. Wübben, D. Seethaler, J. Jalden, and G. Mats, "Lattice reduction," *IEEE Signal Process. Mag.*, vol.28, No.3, pp.70–91, 2011.

[36] A. K. Lenstra, H. W. Lenstra, Jr, and L. Lovász, "Factoring polynomials with rational coefficients," Mathematische Annalen, vol. 261, no.4, pp.515–534, 1982.

[37] S. Denno, Y. Kawaguchi, T. Inoue, and Y. Hou, "A novel low xomplexity lattice eeduction-aided iterative receiver for overloaded MIMO," *IEICE Trans. Commun.*, vol.E102-B, No.5, pp.1045–1054, May. 2019.

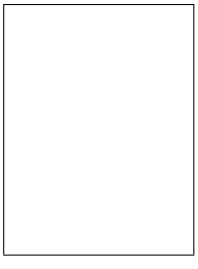
[38] J.-K. Zhang, A. Kavcic, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inform. Theory*, vol.51, no.1, pp.154–172, 2005.

[39] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol.53, no.8, pp.2700–2712, 2005.

[40] K. Kusume, M. Joham, W. Utschick, and G. Bauch, "Cholesky factorization with symmetric permutation applied to detecting and precoding spatially multiplexed data streams," *IEEE Trans. Wireless. Commun.*, vol.55, no.6, pp.3089–3103, 2007.

[41] J. G. Proakis and M. Salehi, "Digital communications,-5th ed.-," McGraw-Hill, 2008.

[42] W. C. Jakes, "Microwave mobile communications," IEEE Press, 1994.



Satoshi Denno received the M.E. and Ph.D degrees from Kyoto University, Kyoto, Japan in 1988 and 2000, respectively. He joined NTT radio communications systems labs, Yokosuka, Japan, in 1988. He was seconded to ATR adaptive communications research laboratories, Kyoto, Japan in 1997. From 2000 to 2002, he worked for NTT DOCOMO, Yokosuka, Japan. In 2002, he moved to DOCOMO communications laboratories Europe GmbH, Germany.

From 2004 to 2011, he worked as an associate professor at Kyoto University. Since 2011, he is a full professor at Okayama University. From the beginning of his research career, he has been engaged in the research and development of digital mobile radio communications. In particular, he has considerable interests in channel equalization, array signal processing, Space time codes, spatial multiplexing, and multimode reception. He won the Best paper award of the 19th international symposium on wireless personal multimedia communications (WPMC2016), and the outstanding paper award of the 23rd international conference on advanced communications technology (ICAT2021). He received the excellent paper award from the IEICE in 1995 and the best paper award from the IEICE communication society in 2020, respectively.

Takumi Sugimoto received B.S. and M.S. degrees from Okayama University, Japan, in 2022 and 2024, respectively. He joined with Murata Manufacturing Co. ltd., in 2024. His research interests include signal processing, wireless communication systems, and overloaded MIMO systems.

Koki Matoba received B.S. degree from Okayama University, Japan, in 2023. He is a master course student in Graduate School of Environmental, life, Natural Science and Technology, Okayama University. His research interests include signal processing, wireless communication systems, and overloaded MIMO systems.

Yafei Hou received his Ph.D. degrees from Fudan University, China and Kochi University of Technology (KUT), Japan in 2007. He was a post-doctoral research fellow at Ryukoku University, Japan from August 2007 to September 2010. He was a research scientist at Wave Engineering Laboratories, ATR Institute International, Japan from October 2010 to March 2014. He was an Assistant Professor at the Graduate School of Information Science, Nara Institute of Science and Technology, Japan from April 2014 to March 2017. He became an assistant professor at Okayama University, Japan from April 2017. He is a guest research scientist at Wave Engineering Laboratories, ATR Institute International, Japan from October 2016. His research interest are communication systems, wireless networks, and signal processing. He received IEICE (the Institute of Electronics, Information and Communication Engineers) Communications Society Best Paper Award in 2016, 2020, and Best Tutorial Paper Award in 2017. Dr. Hou is a senior member of IEEE and IEICE.