MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Joint Lattice and Subspace Vector Perturbation with PAPR Reduction for Massive MU-MIMO Systems

Koike-Akino, T.; Wang, P.; Orlik, P.V.

TR2018-163 December 07, 2018

Abstract

State-of-the-art base stations can be equipped with a massively large number of antenna elements, often several hundreds of elements, thanks to the rapid advancement of wideband radio-frequency (RF) analog circuits and compact antenna design techniques. With massive antenna systems, a relatively large number of users can be served at the same time by means of analog and digital beamforming and spatial multiplexing. We investigate such a large-scale multi-user multipleinput multiple-output (MU-MIMO) wireless system employing an orthogonal frequency-division multiplexing (OFDM)-based downlink transmission scheme. The use of OFDM causes a high peak-to-average power ratio (PAPR), which usually calls for expensive and power-inefficient RF components at the base station. In this paper, we propose a nullspace vector perturbation (VP) which integrates both nonlinear lattice and linear subspace precoding approaches. By exploiting high degrees of freedom available in massive MU-MIMO OFDM systems, the signal PAPR can be significantly reduced with the proposed method. We also introduce a Gaussian process (GP) regression approach to be robust against the imperfect channel knowledge, which is required for the VP operation, in time-varying fading channels. Our analysis of outage capacity reveals that the proposed VP with GP regression offers a significant improvement in sum-rate spectral efficiency while reducing the PAPR.

IEEE Global Communications Conference (GLOBECOM)

Joint Lattice and Subspace Vector Perturbation with PAPR Reduction for Massive MU-MIMO Systems

Toshiaki Koike-Akino, Pu Wang, Philip V. Orlik
Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA
Email: {koike, pwang, porlik}@merl.com

***Abstract*—State-of-the-art base stations can be equipped with a massively large number of antenna elements, often several hundreds of elements, thanks to the rapid advancement of wideband radio-frequency (RF) analog circuits and compact antenna design techniques. With massive antenna systems, a relatively large number of users can be served at the same time by means of analog and digital beamforming and spatial multiplexing. We investigate such a large-scale multi-user multiple-input multiple-output (MU-MIMO) wireless system employing an orthogonal frequency-division multiplexing (OFDM)-based downlink transmission scheme. The use of OFDM causes a high peak-to-average power ratio (PAPR), which usually calls for expensive and power-inefficient RF components at the base station. In this paper, we propose a nullspace vector perturbation (VP) which integrates both nonlinear lattice and linear subspace precoding approaches. By exploiting high degrees of freedom available in massive MU-MIMO OFDM systems, the signal PAPR can be significantly reduced with the proposed method. We also introduce a Gaussian process (GP) regression approach to be robust against the imperfect channel knowledge, which is required for the VP operation, in time-varying fading channels. Our analysis of outage capacity reveals that the proposed VP with GP regression offers a significant improvement in sum-rate spectral efficiency while reducing the PAPR.**

I. Introduction

The recent hardware technologies have enabled the deployment of very large-scale antenna arrays, so-called massive multiple-input multiple-output (MIMO) wireless systems [1]–[5]. Such large-scale antenna arrays have a potential to accommodate ever-growing mobile data traffic for the coming fifth-generation (5G) networks. For multi-user (MU) scenarios, a large number of antennas at the base station (BS) would serve a large number of users concurrently in the same frequency band by means of analog and digital beamforming techniques. In particular for high carrier frequencies including millimeter waves (mmWave), high beam gain achieved by the massive-antenna BS is of great importance to combat severe propagation loss. To realize practical massive MU-MIMO systems, various configurations have been investigated, e.g., fully digital and digital-analog hybrid subarrays [3]–[5]. While the fully digital configuration yields excellent transmission performance utilizing a large number of degrees of freedom (DoF), the implementation and computational costs can be significantly high. Hence, the digital-analog hybrid configurations have been discussed for the most 5G systems to provide a better tradeoff between the cost and performance.

Most typical massive MU-MIMO systems have much larger number of BS antennas than the number of concurrent users, e.g., hundreds of antennas serving tens of users. Such a large-scale MIMO system can exploit sufficient DoF to suppress inter-user interference (IUI). As one of the simplest ways for IUI mitigation, a zero-forcing (ZF) method [6] has been used to nullify the IUI in prior to transmission by employing a pseudo inverse of the channel matrix. The ZF-based precoding can be slightly improved by using generalized minimum mean-square error (MMSE) precoding in particular when the channel state information (CSI) is imperfectly known at the transmitter (Tx). Those linear precoding methods have been generalized to a block diagonalization (BD) method [7]–[9], which can exploit eigen-beam gains by relaxing the nullspace projection to accept intra-user interference when the user equipment (UE) also have multiple antennas to receive multiple data streams. The BD precoding has been further extended to a block triangularization (BT) [10] and multi-block diagonalization (MBD) [11], which can improve the diversity gain by increasing the available DoF via an additional nonlinear precoding based on Tomlinson–Harashima precoding (THP) [12]–[15] to cancel residual IUI. Furthermore, a generalized variant of THP has been also investigated as a lattice precoding (LP) or vector perturbation (VP) for MU-MIMO wireless systems [16], [17].

The nonlinear VP precoding relies on the principle that an addition of any arbitrary vector on lattice points can be self-canceled through the use of a simple modulo operation. In fact, the VP is useful not only for IUI cancellation but also for reducing a peak-to-average power ratio (PAPR) as discussed in [18]–[21]. The PAPR has been one of major problems in digital communications to implement power-efficient radio-frequency (RF) components, particularly when orthogonal frequency-division multiplexing (OFDM) is used to combat frequency-selective fading. Although the OFDM is an efficient and well-established method in the modern communications systems, it is known to suffer from a high PAPR, which necessitates the use of high-cost and high-power linear RF components to prevent out-of-band radiation and signal distortions. Therefore, it is of importance to reduce the PAPR for OFDM-based large-scale MU-MIMO systems to facilitate low-cost and low-power BS deployment.

The PAPR reduction for OFDM signals has been investigated based on various other methods, e.g., selected mapping (SM) [23], [24], partial transmit sequence (PTS) [25], active constellation extension (ACE) [26], [27], and tone reservation (TR) [28], [29]. In [22], a simple linear subspace precoding was proposed for MU-MIMO OFDM systems, where we can

exploit the excessive DoF underlying the massive antennas in comparison to the number of serving users. It was shown that the convex optimization of subspace precoding offers significant reduction of PAPR without sacrificing IUI penalty, when compared to the conventional PAPR reduction methods, i.e., SM, PTS, ACE, and TR.

In this paper, we investigate the joint use of nonlinear lattice precoding [18]–[21] and linear subspace precoding [22] to reduce the PAPR while suppressing the IUI for a massive MU-MIMO OFDM system. The proposed precoding method involves relatively intensive signal processing only at the BS having massive antennas while keeping the low-complexity UE operations. Although the hybrid use of linear and nonlinear precoding has been well studied for IUI cancellation, there are few investigation in terms of PAPR reduction for massive MU-MIMO systems to be best of authors' knowledge.

Our contributions can be summarized as follows:

- We investigate the joint use of nonlinear lattice precoding and linear nullspace precoding to reduce the PAPR for large-scale MU-MIMO OFDM systems.
- We develop an alternating optimization algorithm, which successively designs the lattice perturbation vector and nullspace perturbation vector in an iterative fashion.
- We introduce the Gaussian process (GP) regression [31] to be robust against the CSI knowledge imperfection. The regression confidence value is exploited to modify the precoding matrix.
- We also consider a low-complexity subspace tracking method [30] to adapt the precoding matrix over the GP-predicted time-varying fading channels.
- We present numerical simulation results to demonstrate the capabilities of the proposed MU-MIMO OFDM downlink transmission scheme.

Throughout the paper, boldface uppercase letters denote matrices and boldface lowercase letters denote vectors. An operator $[\cdot]^{\mathrm{T}}$ denotes transpose and $[\cdot]^{\dagger}$ complex-conjugation, i.e., Hermitian. Unless otherwise specified, all vectors are assumed to be column vectors. We use $\mathbb{Z}_q$, $\mathbb{C}^{n\times m}$, and $\mathbb{E}[\cdot]$ to respectively denote the set of non-negative integers less than q, the set of complex-valued matrices of size $n \times m$, and the expectation operation. The $\mathcal{L}_p$-norm of a vector $\boldsymbol{x}$ is expressed as $\|\boldsymbol{x}\|_p$. We denote $\mathbf{0}$ and $\boldsymbol{I}$ as all-zero matrix and identity matrix, respectively, with a proper dimension.

II. Massive Multi-User MIMO Systems

A. System Description

Fig. 1 depicts the MU-MIMO wireless downlink systems having massive antenna elements at the BS to serve multiple UEs concurrently. For example, the BS having $N_{\mathrm{t}} = 1024$ antenna elements is configured with digital-analog hybrid arrays constituting of digitally-controlled dual-polarization (4×8)-dimensional subarrays, each of which further implements $N_{\mathrm{a}} = 4 \times 4$ analog-controlled antenna elements (thus, $2\times4\times8\times4\times4 = 1024$). Most typically, the analog-controlled subarray antennas use a bulk phase shifting for beam staring towards a target UE, and hence relatively high side-lobe interference occurs towards the other UEs. The remaining $N_{\mathrm{d}} = N_{\mathrm{t}}/N_{\mathrm{a}}$ antennas, which are digitally-controlled, can deal with the residual IUI more precisely in real time.

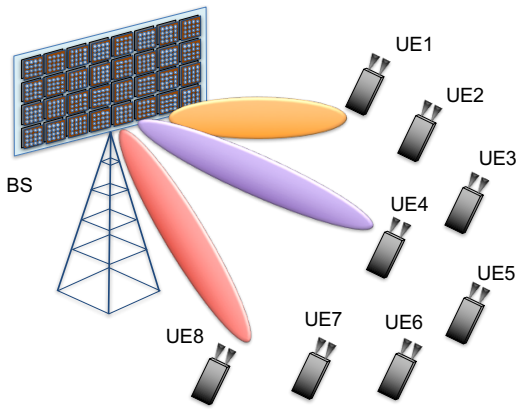


Fig. 1. Massive MU-MIMO downlink systems having $N_{\mathrm{t}} = 1024$ antenna elements (e.g., 2 polarization, 32 subarrays, and 16 elements) for serving $N_{\mathrm{u}} = 8$ UEs having $N_{\mathrm{r}} = 2$ antenna elements each.

Consider serving N_{u} concurrent UEs, equipped with N_{r} receiving (Rx) antennas each. We assume much more DoF is available for digital-controlled antennas than the total number of Rx antennas; specifically, $N_{\mathrm{d}} \gg N_{\mathrm{u}}N_{\mathrm{r}}$. For example in Fig. 1, the total number of Rx antennas are 16 for $N_{\mathrm{u}} = 8$ UEs with $N_{\mathrm{r}} = 2$ elements, whereas the number of digital Tx antennas is $N_{\mathrm{d}} = 64$ for the BS with dual-polarized 32 subarrays. For such a case, we can utilize the excessive DoF (i.e., $N_{\mathrm{d}} - N_{\mathrm{u}}N_{\mathrm{r}}$) to manipulate the transmission signal, e.g., such that the PAPR is minimized as studied in [22].

B. Signal Model

Let $\boldsymbol{x}_k \in \mathbb{C}^{N_{\mathrm{r}}}$ be the Tx modulated streams vector for the kth UE, at a certain subcarrier (subcarrier index is omitted for convenience throughout the paper). We consider 256-ary quadrature-amplitude modulation (QAM) format for the constellation of $\boldsymbol{x}_k$ as an example. The data streams $\boldsymbol{x}_k$ are fed into RF components to transmit through massive subarrays via a linear precoding matrix $\boldsymbol{B}_k \in \mathbb{C}^{N_{\mathrm{d}}\times N_{\mathrm{r}}}$ as follows: $\boldsymbol{B}_k\boldsymbol{x}_k$. Correspondingly, the BS carriers out a spatial multiplexing for all data streams $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_{\mathrm{u}}}]^{\mathrm{T}}$, through the N_{d} subarrays, using a compound linear precoding matrix $\boldsymbol{B} = [\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_{N_{\mathrm{u}}}] \in \mathbb{C}^{N_{\mathrm{d}}\times N_{\mathrm{u}}N_{\mathrm{r}}}$. Accordingly, the transmission signal vector $\boldsymbol{u} \in \mathbb{C}^{N_{\mathrm{d}}}$ across the BS subarrays is expressed as follows:

$$\boldsymbol{u} = \sum_{k=1}^{N_{\mathrm{u}}} \boldsymbol{B}_k\boldsymbol{x}_k = \boldsymbol{B}\boldsymbol{x}. \tag{1}$$

Note that the time-domain Tx sequence is obtained by taking an inverse Fourier transform across all subcarriers, denoted as $\mathcal{F}^{-1}[\boldsymbol{u}]$ for convenience. We suppose that the peak envelope shall not surpass a predefined power truncation limit,

(a) BD (b) BT (c) BBD

Fig. 2. Linear BD, BT, BBD precoding for MU-MIMO systems.

such that the infinity norm lies under a threshold with a high probability as follows:

$$\left\|\mathcal{F}^{-1}[\boldsymbol{u}]\right\|_{\infty}^{2} \leq E_{\mathrm{th}}, \tag{2}$$

where E_{th} denotes the peak power limitation. To prevent nonlinear distortion due to a low-cost power amplifier under the peak power limitation, the average power of Tx signal is also constrained within a backoff margin as

$$\mathbb{E}\left[\left\|\mathcal{F}^{-1}[\boldsymbol{u}]\right\|_{2}^{2}\right] \leq \tau E_{\mathrm{th}}, \tag{3}$$

where τ is a constant value determining the backoff (typically a few dB less than peak power).

The multiplexed data streams $\boldsymbol{u}$ are propagated after N_{a}-element analog beamforming towards a receiver UE through wireless fading channels. At the kth UE, the Rx baseband signal is modeled as follows:

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{u} + \boldsymbol{z}_k = \sum_{j=1}^{N_{\mathrm{u}}} \boldsymbol{H}_k\boldsymbol{B}_j\boldsymbol{x}_j + \boldsymbol{z}_k, \tag{4}$$

where $\boldsymbol{y}_k \in \mathbb{C}^{N_{\mathrm{r}}}$ is the Rx signal vector, $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{d}}}$ denotes the subcarrier channel matrix, and $\boldsymbol{z}_k \in \mathbb{C}^{N_{\mathrm{r}}}$ is the additive white Gaussian noise (AWGN) vector following the complex Gaussian distribution of $\mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ with a noise variance of σ^2. Note that handling cyclic prefix (CP) and Fourier transform are assumed appropriately.

C. Block-Diagonalization (BD) Precoding

As shown in (4), the Rx signal at the kth UE may contain undesired signals $\boldsymbol{x}_j$, sent for the other users $j \neq k$. The main purpose employing the linear precoding $\boldsymbol{B}$ is to suppress such an IUI. To do so, we perform a block-wise diagonalization [7]–[9] for the compound channel matrix $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{u}}N_{\mathrm{r}}\times N_{\mathrm{d}}}$ such that

$$\boldsymbol{H}\boldsymbol{B} = \begin{bmatrix}\boldsymbol{H}_1\\ \boldsymbol{H}_2\\ \vdots\\ \boldsymbol{H}_{N_{\mathrm{u}}}\end{bmatrix}\boldsymbol{B} = \begin{bmatrix}\boldsymbol{G}_1 & & & \boldsymbol{0}\\ & \boldsymbol{G}_2 & & \\ & & \ddots & \\ \boldsymbol{0} & & & \boldsymbol{G}_{N_{\mathrm{u}}}\end{bmatrix}, \tag{5}$$

where $\boldsymbol{G}_k \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{r}}}$ is an effective channel matrix after precoding for the kth UE. This BD precoding enforces nullifying all off-diagonal blocks to avoid IUI at the other UEs, i.e., $\boldsymbol{H}_k\boldsymbol{B}_j = \boldsymbol{0}$ for $k \neq j$, as shown in Fig. 2(a). This constraint can often lead to a limited diversity gain in particular when the excessive antenna DoF is small. The use of MMSE can relax the constraint by accepting small IUI leakage comparable to a noise level at off-diagonal blocks.

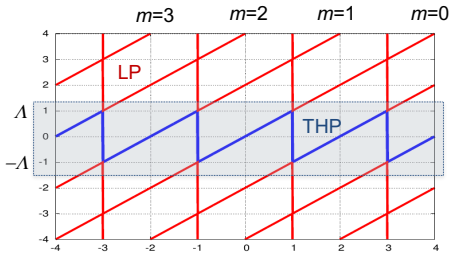


Fig. 3. Modulo operation for THP and its generalization with LP/VP.

In order to drastically increase the DoF, some relaxed variants including BT [10] and block bi-diagonalization (BBD) [11] allow IUI at partial off-diagonal blocks as shown in Figs. 2(b) and (c). The admitted IUI are then canceled by introducing a nonlinear lattice precoding, which is based on THP [12]–[15] or more generally VP [16], [17]. Nevertheless, the BD and BBD precoding are advantageous over the BD only for the case when there is a limited antenna DoF. In this paper, we consider a sufficiently large number of antenna elements (e.g., $N_{\mathrm{d}} = 64$ for $N_{\mathrm{u}} = 8$ and $N_{\mathrm{r}} = 2$, thus having 4-times more DoF), and therefore will focus on the BD precoding for the IUI suppression.

Note that there are usually more than a unique solution of $\boldsymbol{B}$ which can nullify off-diagonal blocks. In order to achieve the highest eigen-beam gain, the effective channel matrix $\boldsymbol{G}_k$ should have the identity matrix for its right singular matrix after precoding. More practically, once we obtained $\boldsymbol{G}_k$ after a particular block diagonalization process, we should multiply the right singular matrix from the right of precoding matrix $\boldsymbol{B}_k$. Specifically, with a singular-value decomposition of $\boldsymbol{G}_k = \boldsymbol{S}_k\boldsymbol{V}_k\boldsymbol{D}_k^{\dagger}$, the BD precoding matrix $\boldsymbol{B}_k$ shall be replaced with $\boldsymbol{B}_k\boldsymbol{D}_k$.

III. Joint Lattice & Subspace VP

A. Nonlinear Lattice VP

Even without IUI by using the BP precoding, the VP technique is still useful for reducing the PAPR. Because of the central limit theorem, OFDM signals multiplexing QAM constellations over a number of subcarriers tend to follow the Gaussian distribution, resulting into a significantly high PAPR in general. The high PAPR in turn necessitates high-power and high-cost RF components for all the massive antenna subarrays in MU-MIMO systems.

The lattice VP uses a modulo operation, $\mathcal{M}_{\Lambda}[x] = x - 2\Lambda\lfloor x/2\Lambda\rceil$, where Λ is a modulo constant and $\lfloor x\rceil$ denotes the rounding operation returning the closest integer to x. This modulo operator used for the conventional THP limits the amplitude within Λ, as shown in Fig. 3. The VP uses a transparency feature through the modulo operation even after modifying the signal x by adding $2m\Lambda$ as $\mathcal{M}_{\Lambda}[x+2m\Lambda] = x$ for any auxiliary integer $m \in \mathbb{Z}$.

At the BS, we can modify the OFDM signal $\boldsymbol{x}_k$ by adding any arbitrary lattice vector $\boldsymbol{\lambda}_k$, whose element is a product of 2Λ and a certain integer (specifically, complex-valued square

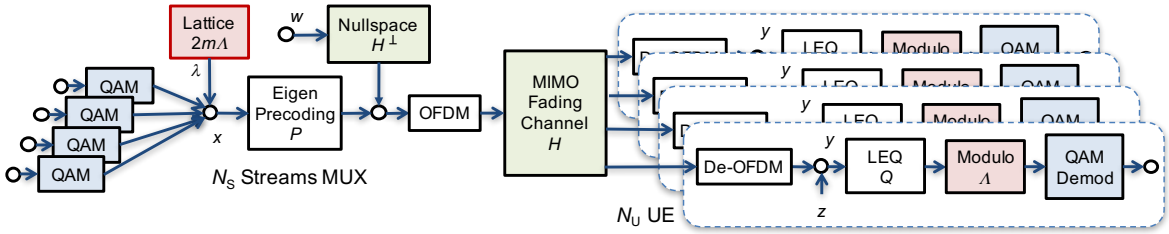


Fig. 4. Joint lattice and subspace vector perturbation for MU-MIMO.

grid). Assuming that IUI is nullified by the BD precoding, the kth Rx signal in (4) is rewritten with lattice vector perturbation $\boldsymbol{\lambda}_k$ as follows:

$$\boldsymbol{y}_k = \boldsymbol{G}_k(\boldsymbol{x}_k + \boldsymbol{\lambda}_k) + \boldsymbol{z}_k. \tag{6}$$

The desired signal $\boldsymbol{x}_k$ can be easily retrieved at the Rx by employing the modulo operation as $\mathcal{M}_\Lambda[\boldsymbol{G}_k^{-1}\boldsymbol{y}_k]$ even without knowledge of lattice perturbation vector $\boldsymbol{\lambda}_k$. The auxiliary vector addition provides a great opportunity to manipulate the precoded OFDM signals $\boldsymbol{u}$ such that the PAPR is minimized. As we will discuss later, we use a simplified version of the p-sphere encoder [19] to obtain optimized lattice perturbation vector $\boldsymbol{\lambda}_k$.

B. Linear Nullspace VP

When there is an excessive DoF, any subspace projection on kernel can be auto-canceled over the MIMO channels. Let $\boldsymbol{H}^\perp \in \mathbb{C}^{N_\mathrm{d}\times N_\mathrm{e}}$ be the nullspace projection matrix of the channel matrix $\boldsymbol{H}$ such that $\boldsymbol{H}\boldsymbol{H}^\perp = \boldsymbol{0}$, where N_e is the excess DoF which is at most $N_\mathrm{d} - N_\mathrm{u}N_\mathrm{r}$. Consequently, we can further modify the precoded OFDM signal $\boldsymbol{u}$ in (1) by adding any arbitrary perturbation vector $\boldsymbol{w} \in \mathbb{C}^{N_\mathrm{e}}$ over the nullspace projection as follows:

$$\boldsymbol{u} = \boldsymbol{B}(\boldsymbol{x} + \boldsymbol{\lambda}) + \boldsymbol{H}^\perp\boldsymbol{w}, \tag{7}$$

as the nullspace vector $\boldsymbol{w}$ is automatically canceled:

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{H}\boldsymbol{u} + \boldsymbol{z}\\ &= \boldsymbol{H}\boldsymbol{B}(\boldsymbol{x}+\boldsymbol{\lambda}) + \underbrace{\boldsymbol{H}\boldsymbol{H}^\perp\boldsymbol{w}}_{\boldsymbol{0}} + \boldsymbol{z}\\ &= \boldsymbol{H}\boldsymbol{B}(\boldsymbol{x}+\boldsymbol{\lambda}) + \boldsymbol{z}.\end{aligned} \tag{8}$$

Because of the transparency, the subspace VP can modify the Tx signal statistics without disturbing the Rx detection. The subspace perturbation vector $\boldsymbol{w}$ which reduces the PAPR can be obtained by a gradient descent method based on fast iterative truncation algorithm (FITRA) [22]. In this paper, we investigate a joint use of nonlinear lattice VP and linear nullspace VP, whose schematic is illustrated in Fig. 4.

C. Alternating Minimization for PAPR Reduction

For the joint vector & subspace VP, we need to optimize the two perturbation vectors $\boldsymbol{\lambda}$ and $\boldsymbol{w}$ given a channel realization $\boldsymbol{H}$ and transmitting data streams $\boldsymbol{x}$ as well as the BD precoding $\boldsymbol{B}$. Under the constraint of average Tx power in (3), we wish to minimize the peak power in (2), or simply

$$\min_{\boldsymbol{\lambda},\boldsymbol{w},\boldsymbol{B}} \left\| \mathcal{F}^{-1}\left[\boldsymbol{B}(\boldsymbol{x}+\boldsymbol{\lambda}) + \boldsymbol{H}^\perp\boldsymbol{w}\right]\right\|_\infty^2. \tag{9}$$

However, minimizing PAPR does not always improve the sum-rate spectral efficiency because the desired signal term of $\boldsymbol{B}\boldsymbol{x}$ can be less power if too much energy are assigned for those perturbation vectors. To simply the problem, we impose another constraint such that the desired signal power cannot be lower than a threshold as

$$\mathbb{E}\left[\left\|\mathcal{F}^{-1}\left[\boldsymbol{B}\boldsymbol{x}\right]\right\|_2^2\right] \geq \beta\tau E_\mathrm{th}, \tag{10}$$

where β is a power allocation factor no greater than 1.

We use an alternating minimization method, which successively determines the best vectors $\boldsymbol{\lambda}$ and $\boldsymbol{w}$ to minimize the peak power while keeping the other power constraints in an iterative fashion. To find the best lattice vector $\boldsymbol{\lambda}$ given the nullspace vector $\boldsymbol{w}$, we employ the p-sphere encoder [19] with a few number of survivors to minimize the infinity norm of $\boldsymbol{u}$ according to (2). Then the Tx power is scaled to satisfy the constraints of (3) and (10). We next use the FITRA algorithm [22] to find the nullspace vector $\boldsymbol{w}$ (implicitly, via $\boldsymbol{u}$ to provide the minimum infinity norm) given the lattice vector $\boldsymbol{\lambda}$. Note that the precoding constraint in the original FITRA should be replaced from the ZF precoding to the BD precoding as follows:

$$\boldsymbol{H}\boldsymbol{u} = \boldsymbol{G}(\boldsymbol{x}+\boldsymbol{\lambda}), \tag{11}$$

in other words, the precoded channel matrix $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{B}$ is no longer identity. Since the individual steps are well-established for the lattice VP and nullspace VP in the other literature, the iteration process is rather straightforward with minor modifications.

D. Computational Complexity

We briefly discuss the computational complexity of the proposed joint nonlinear and linear VP. The p-sphere encoder [19] tries to minimize the $\mathcal{L}_\infty$-norm in (10) subject to the $\mathcal{L}_2$-norm bound in (3) so that PAPR is reduced given $\boldsymbol{x}$ and $\boldsymbol{w}$. In order to constrain the computational complexity low, we employ the M-algorithm where the constant number of surviving candidates is retained to search small $\mathcal{L}_2$-norm. Since the Fourier transform is invariant under $\mathcal{L}_2$-norm, we can

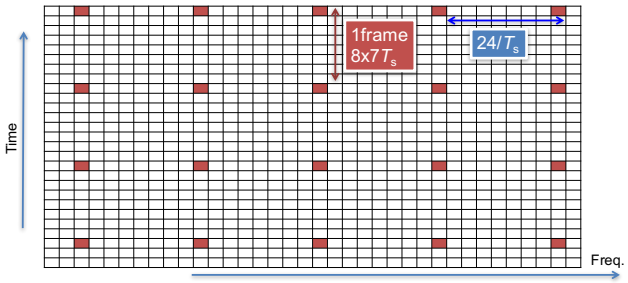


Fig. 5. Distributed pilots in OFDM transmission.

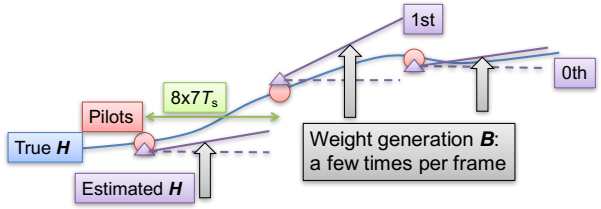


Fig. 6. GP regression to interpolate and extrapolate CSI estimation.

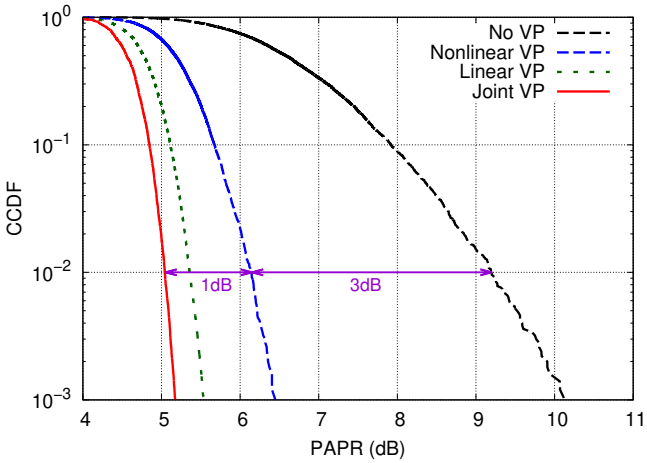


Fig. 7. PAPR performance with/without nonlinear lattice VP and linear nullspace VP for massive MU-MIMO OFDM systems.

independently perform the sphere decoding for each subcarrier. For each subcarrier, we employ the quadrature-residue (QR) factorization for the precoder matrix $\boldsymbol{B}$ so that efficient tree-based M-algorithm is feasible. The computational complexity for the QR decomposition is in the order of $\mathcal{O}[N_{\rm d}^3 N_{\rm s}]$ where $N_{\rm s}$ denotes the number of subcarrier tones. The M-algorithm having M survivors requires $\mathcal{O}[M N_{\rm s} N_{\rm d}]$ complexity to search for small $\mathcal{L}_2$-norm candidates. Among the best M survivors, we then determines the PAPR minimizing solution which offers the smallest $\mathcal{L}_\infty$-to-$\mathcal{L}_2$ norm ratio, which requires $\mathcal{O}[M N_{\rm d} N_{\rm s}(1+\log_2 N_{\rm s})]$ complexity due to the inverse Fourier transform.

Given the optimized lattice vector $\boldsymbol{\lambda}$, the FITRA [22] updates the subspace vector $\boldsymbol{w}$ via an iterative gradient and truncation process based on a proximal map. The gradient computation requires the Fourier transform of complexity order $\mathcal{O}[N_{\rm d} N_{\rm s} \log_2 N_{\rm s}]$ to convert back and forth between frequency and time representations of precoded signal $\boldsymbol{u}$ in (7). In addition, each subcarrier needs channel projections of complexity order $\mathcal{O}[N_{\rm s} N_{\rm r}^2 N_{\rm u}^2 N_{\rm d}^2]$ in total. The proximal map uses an element-wise truncation of the complexity order of $\mathcal{O}[N_{\rm s} N_{\rm d}]$.

E. Precoder Adaptation over Time-Varying Fading Channels

Supposed that the VP vectors can be perfectly canceled, the spectral efficiency is upper-bounded by

$$R = \log_2 \det\Big[\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{B}\boldsymbol{B}^{\dagger}\boldsymbol{H}^{\dagger}\Big] = \sum_{k=1}^{N_{\rm u}} \sum_{i=1}^{N_{\rm r}} \log_2\Big(1 + \frac{v_{k,i}^2}{\sigma^2}\Big), \tag{12}$$

where $v_{k,i}$ is the ith singular value of $\boldsymbol{G}_k$, which depends on the power allocation factor β. One major challenge of the VP technique lies in the requirement of accurate CSI feedback from the Rx UEs to the Tx BS before applying the BD precoding and VP. When the imperfect CSI is provided, the undesired IUI can considerably degrade the spectral efficiency as residual IUI plays an effectively colored noise source.

The CSI imperfection can be more serious when pilots tones are sparsely distributed over time and frequencies as in Fig. 5. To mitigate the CSI error, we introduce the GP interpolation and extrapolation between pilots. Fig. 6 illustrates the GP regression [31], where CSI interpolation and extrapolation are performed given available estimates of CSI assuming that the stochastic process is based on the Gaussian kernel. The GP regression provides not only the CSI estimates but also the estimation mean-square error (MSE), which is useful to use MMSE-based BD precoding for higher tolerance of the CSI mismatch. Because the CSI can slowly vary over time and frequency, we can employ the recursive projection approximation subspace tracking (PAST) algorithm [30] to reduce the computational complexity at the BS.

IV. Performance Analysis

We use the lattice factor of $\Lambda = 1.5$ for lattice vector search based on the p-norm sphere encoder [19] with $M = 16$ survivors. To search for the nullspace vector, we use the FITRA [22] with a maximum number of 2000 iterations and a regularization parameter of 0.25. Unless explicitly stated otherwise, all simulation results are for an MU-MIMO OFDM 256QAM system having $N_{\rm t} = 1024$, $N_{\rm a} = 16$, and $N_{\rm d} = 64$ antennas at the BS and serving $N_{\rm u} = 8$ UEs with $N_{\rm r} = 2$ antennas. We employ OFDM with $N_{\rm s} = 64$ subcarrier tones.

A. PAPR Reduction

To compare the PAPR characteristics of different precoding schemes, we evaluate the complementary cumulative distribution function (CCDF) and the 1%-outage PAPR that is valid at most for 99% of all simulated OFDM symbols. Fig. 7 shows the CCDF of PAPR performance for the BD precoding with/without the linear nullspace VP and the nonlinear lattice

VP. It is seen that the nonlinear VP significantly reduce the 1%-outage PAPR performance by at least 3 dB. The linear VP offers an additional 0.3 dB gain because the excessive DoF is relatively large, $N_{\mathrm{e}} = 48$ compared to the number of multiplexing streams $N_{\mathrm{u}}N_{\mathrm{r}} = 16$ per symbol. In addition, the nullspace vector $\boldsymbol{w} \in \mathrm{C}^{N_{\mathrm{e}}}$ can take any complex-valued numbers, whereas the lattice vector $\boldsymbol{\lambda} \in 2\Lambda \cdot \mathbb{Z}^{N_{\mathrm{u}}N_{\mathrm{r}}}$ can take from lattice points with smaller dimension. Nevertheless, this additional DoF over nonlinear lattice space on top of the linear subspace can improve the PAPR performance slightly more when both VPs are jointly used. Consequently, the joint VP achieves a total of 4 dB PAPR reduction. This PAPR reduction can directly improve the power efficiency of the RF components, e.g., the backoff for power amplifier.

B. Spectral Efficiency

We then evaluate the achievable spectral efficiency of the proposed joint VP. Fig. 8 shows the outage sum rates at a channel SNR of 20 and 30 dB for different fading speed with a normalized maximum Doppler frequency of $f_{\mathrm{D}}T_{\mathrm{s}} = 0.002, 0.003, 0.004, 0.005$. This figure also presents the BD precoding with the perfect CSI knowledge at the BS as a benchmark. We can see that the rapid fading degrades the achievable spectral efficiency in particular at high SNR regimes. For example, the penalty due to the imperfect CSI is about 22% loss in spectral efficiency at an outage probability of 0.1 for an SNR of 30 dB and $f_{\mathrm{D}}T_{\mathrm{s}} = 0.0002$. This penalty will be more significant at faster fading of $f_{\mathrm{D}}T_{\mathrm{s}} = 0.005$; specifically, 65% reduction in outage spectral efficiency is observed. It is nonetheless confirmed that the GP regression offers considerable tolerance against time-varying fading. In fact, the decreased spectral efficiency will be raised up by 130% with the joint VP and GP. Note that the joint VP can improve the sum rates by approximately 10 bps/Hz, because of the improved power efficiency, compared to the BD precoding without VP.

V. Conclusion

We investigated large-scale MU-MIMO OFDM wireless downlink systems employing a nullspace VP with both nonlinear lattice and linear subspace precoding to achieve low PAPR while nullifying IUI. The proposed method exploits excessive degrees of freedom offered in massive antennas. We also introduced the GP regression approach to be robust against the imperfect CSI, which is required for the VP operation. In addition, we introduced a low-complexity subspace tracking method for time-varying fading channels. Our analysis of outage capacity revealed that the proposed VP with GP regression offers a significant improvement in sum-rate spectral efficiency while reducing the PAPR by 4 dB.

Acknowledgment

The authors would like to thank Dr. H. Nishimoto, A. Taira, and A. Okazaki for valuable discussion.

References

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

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[4] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Ant. Prop.*, vol. 61, no. 7, pp. 3814–3827, 2013.

[5] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, and K. Cheun, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, 2014.

[6] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 461–471, 2004.

[7] M. Rim, "Multi-user downlink beamforming with multiple transmit and receive antennas," *Electron. Lett.*, vol. 38, no. 25, pp. 1725–1726, 2002.

[8] K.-K. Wong, R. D. Murch, and K. B. Letaief, "Performance enhancement of multiuser MIMO wireless communication systems," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1960–1970, 2002.

[9] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, 2004.

[10] E. C. Y. Peh and Y.-C. Liang, "Power and modulo loss tradeoff with expanded soft demapper for LDPC coded GMD-THP MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 714–724, 2009.

[11] H. Nishimoto, H. Iura, A. Taira, A. Okazaki, and A. Okamura, "Block lower multi-diagonalization for multiuser MIMO downlink," *IEEE WCNC*, pp. 303–308, Doha, Qatar, Apr. 2016.

[12] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5, pp. 138–139, 1971.

[13] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, 1972.

[14] C. Masouros, M. Sellathurai, and T. Rantarajah, "Interference optimization for transmit power reduction in Tomlinson-Harashima precoded MIMO downlinks," *IEEE Trans. Sig. Proc.*, vol. 60, no. 5, pp. 2470–2481, May 2012.

[15] A. Garcia and C. Masouros, "Power-efficient Tomlinson-Harashima precoding for the downlink of multi-user MISO systems," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1884–1896, Jun. 2014.

[16] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, 2005.

[17] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, 2005.

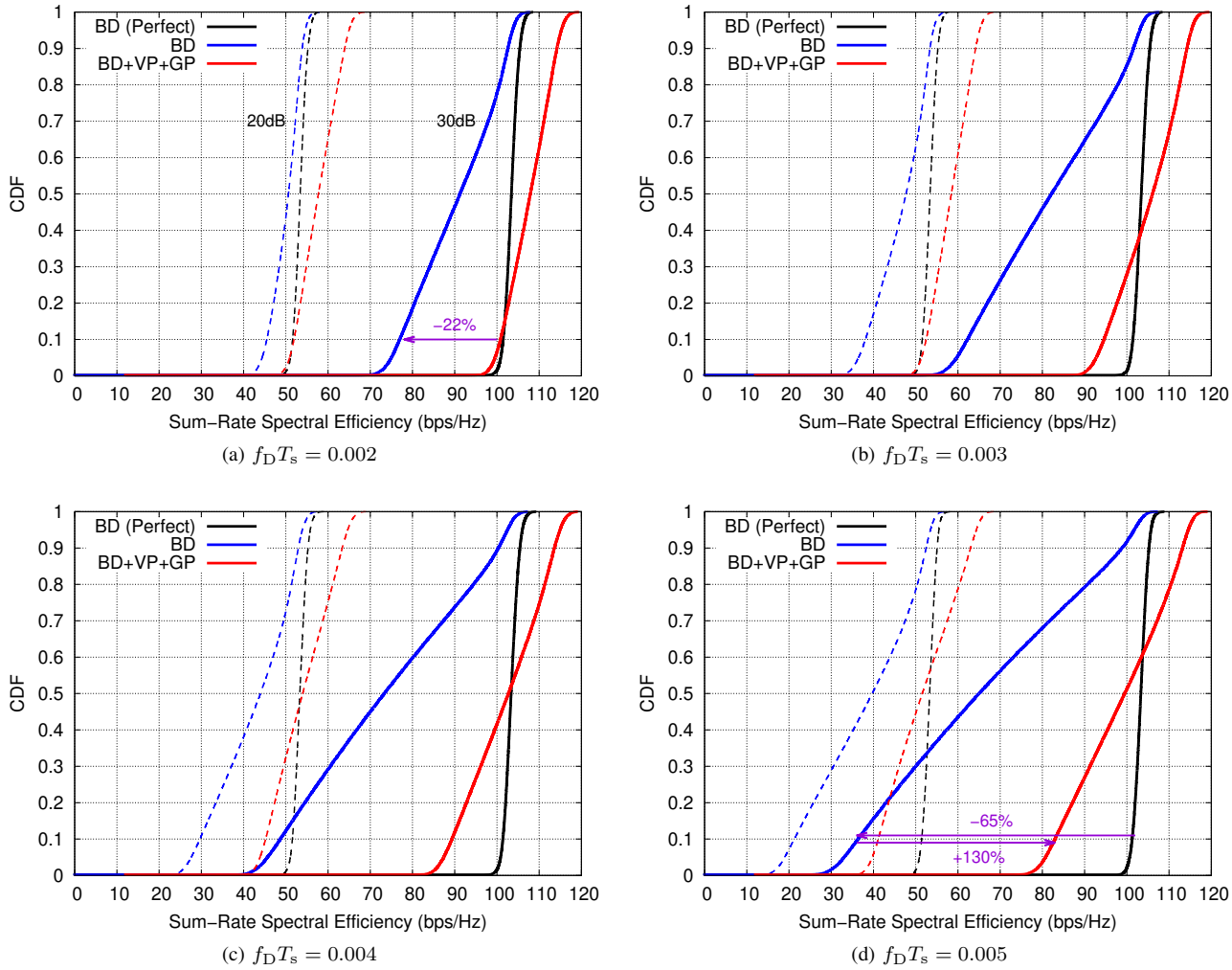
[18] L. Yang and E. and Alsusa, "Selective vector perturbation precoding and peak to average power ratio reduction for OFDM systems," *IEEE GLOBECOM*, Nov. 2008.

[19] F. Boccardi and G. Caire, "The p-sphere encoder: Peak-power reduction by lattice precoding for the MIMO Gaussian broadcast channel," *IEEE trans. Commun.*, vol. 54, no. 11, pp. 2085–2091, 2006.

[20] M. Mazrouei-Sebdani and W. A. Krzymien, "Sum rate of p-sphere encoding for MIMO broadcast channels with reduced peak power," *IEEE VTC-Spring*, May 2012.

[21] C. Siegl and R. F. H. Fischer, "Selected basis for PAR reduction in multi-user downlink scenarios using lattice-reduction-aided precoding," *EURASIP J. Advanced Sig. Proc.*, vol. 17, pp. 1–11, July 2011.

[22] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 303–313. 2013.



(a) $f_{\mathrm{D}}T_{\mathrm{s}} = 0.002$ (b) $f_{\mathrm{D}}T_{\mathrm{s}} = 0.003$

(c) $f_{\mathrm{D}}T_{\mathrm{s}} = 0.004$ (d) $f_{\mathrm{D}}T_{\mathrm{s}} = 0.005$

Fig. 8. Sum-rate spectral efficiency of BD precoding with/without VP and GP regression in time-varying fading channels at an SNR of 20 and 30 dB.

[23] R. W. Bäuml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electron. Lett.*, vol. 32, no. 22, pp. 2056–2057, Oct. 1996.

[24] R. F. H. Fischer and M. Hoch, "Directed selected mapping for peak-to-average power ratio reduction in MIMO OFDM," *Electron. Lett.*, vol. 42, no. 2, pp. 1289–1290, Oct. 2006.

[25] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electron. Lett.*, vol. 33, no. 5, pp. 368–369, Feb. 1997.

[26] T. Tsiligkaridis and D. L. Jones, "PAPR reduction performance by active constellation extension for diversity MIMO-OFDM systems," *J. Electrical and Computer Eng.*, no. 930368, 2010.

[27] B. S. Krongold and D. L. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE ICASSP*, pp. 525–528, Hong Kong, China, Apr. 2003.

[28] —, "An active-set approach for OFDM PAR reduction via tone reservation," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 495–509, Feb. 2004.

[29] J. Illic and T. Strohmer, "PAPR reduction in OFDM using Kashin's representation," *IEEE SPAWC*, pp. 444–448, Perugia, Italy, June 2009.

[30] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Sig. Proc.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[31] F. Pérez-Cruz, S. Van Vaerenbergh, J. J. Murillo-Fuentes, M. Lázaro-Gredilla, and I. Santamaria, "Gaussian processes for nonlinear signal processing: An overview of recent advances," *IEEE Sig. Proc. Mag.* vol. 30, no. 4, pp. 40–50, July 2013.