Beamforming MISO-OFDM PAPR Reduction: A Space-User Perspective

Kuang Xu, Timothy A. Thomas, Douglas L. Jones

Abstract

A novel space-user approach to multiple-input-single-output (MISO)[1] OFDM peak-to-average-power ratio (PAPR) reduction is examined. We propose two techniques that significantly reduce PAPR for MISO-OFDM at various user-loads. The first exploits the degrees of freedom in the channel's null-space (transmit weight optimization), while the second provides a more robust and higher reduction by further relaxing the constraint on the constellation domain (constellation-beam modification) with modest computational overhead. Both methods can be implemented by low-complexity projection-onto-convex-set (POCS) iterations, have no data rate loss, and are fully compatible with various constellation choices and receiver structure in practical systems. The algorithms proposed achieve up to 6.65 dB reduction in peak power for an 8-antenna MISO transmitter.

Index Terms

Orthogonal frequency multiplexing (OFDM), peak-to-average-power ratio (PAPR), multiple-input-single-output (MISO), beamforming.

I. Introduction

The combination of beamforming multiple-input-single-output (MISO) and orthogonal frequency multiplexing (OFDM) provides a robust communications link in fading environments by maximizing received signal strength via beam-steering, and has been incorporated in standards such as WiMAX and 3GPP Long Term Evolution. A MISO-OFDM system is often characterized by a single base station with multiple antennas transmitting simultaneously to a number of individual users. Like single-input-single-out (SISO) OFDM, MISO-OFDM suffers from a potentially high peak-to-average power ratio (PAPR) at the transmit antennas, caused by the superposition of multiple sub-channel signals in the time domain. A high PAPR requires the power amplifier at the transmitter to be built with a larger back-off, resulting in greater power consumption and energy inefficiency.

A. Related work

A number of PAPR reduction techniques have been proposed for PAPR reduction in SISO OFDM systems [3]. The tone-reservation technique inserts signals into unused subchannels in order to cancel the time-domain peaks in the data-bearing channels, while sacrificing data rates due to the reduction in data-bearing channels [1][2]. In a wireless

[1]The MISO scheme considered in this paper refers to the situation where a single base station transmits simultaneously to multiple mobiles. Technically, some may consider this a multi-user multiple-input-multiple-ouput (MU-MIMO) system but we use the term MISO to emphasize that the mobiles all have a single receive antenna.

setting, however, it is often inefficient to reserve a fixed set of tones due to the time-varying characteristics of the channel. Some methods utilize the pre-known phase knowledge to improve PAPR via encoding: the selected mapping (SLM) method multiplies the original data block by pseudo-random phase-shift sequences, and selects the one with the lowest PAPR [4]. The partial transmit sequence (PTS) method, like SLM, scrambles the data bits across frequencies to avoid combinations with high PAPR [6]. However, SLM and PTS both require side information for the receiver to decode, and an optimal search in PTS can be quite computationally expensive. The active constellation extension (ACE) method allows the outer-most constellation points to be moved outward into several allowable regions of extension, while the constellation can still be correctly detected by a standard maximum likelihood receiver [13][14]. ACE is able to reduce PAPR without sacrificing data rate at the cost of an increased average power due to the extensions in constellations.

There has also been work on PAPR reduction for multiple-input-multiple-output (MIMO) systems. A number of papers have generalized SLM and PTS methods to MIMO-OFDM, obtaining up to 2 to 3 dB PAPR reduction [7][8][9][10]. A mode-reservation algorithm extends the tone-reservation concept to MIMO-OFDM by exploiting unused eigenmodes for PAPR reduction, achieving a PAPR reduction of 3 dB in one iteration when allowing a 10% increase in average transmit power [12]. The ACE scheme was extended to MIMO in [15]. Fewer works have been devoted specifically to MISO-OFDM systems. In [11], an extension of PTS to MISO with antenna selection is proposed, which requires SNR loss to achieve reduction in PAPR.

B. *Our approach*

While many methods have been developed for OFDM PAPR reduction, most lack practicality due to a high computational complexity, data rate loss or the requirement for additional side information, as mentioned in the previous section. In addition, most existing methods fail to address the fact that the number of users served by the base station *varies with time* in a real MISO system (e.g. mobile users joining and leaving the coverage of the base station), which requires the consistency of performance across *all* possible user loads to be guaranteed.

We propose two novel techniques, which exploit the unique space-user characteristics of MISO-OFDM systems, achieving significant and user-load-stable performance with zero data rate loss, low computational complexity, and no modification at the receivers. The main contributions of this paper are as follows:

1) The *transmit weight optimization (TWO)* algorithm for MISO-OFDM is proposed and analyzed, which exploits the degrees of freedom (DoF) in the null-space of the channel matrix. TWO achieves significant PAPR reduction when the number of users is less than the number of antennas (as much as 6.65 dB). An early version of this technique was first reported in [18].
2) Inspired by TWO, we identify a critical *space-user trade-off*, intrinsic to all MISO-OFDM systems. More specifically, the performance of a PAPR reduction algorithm applied to MISO-OFDM can be significantly affected

as the number of users served by a based station changes, which in reality could occur frequently due to a time-varying service load. To address this challenge, we propose the *constellation-beam modification (CBM)* method to overcome TWO's performance decrease as the number of users increases. By simultaneously exploiting degrees of freedom in *both* the channel's null-space and the constellation domains, CBM provides consistently high PAPR reduction at all user-loads with negligible computational overhead.

3) We provide a computationally efficient *smart-gradient-project (SGP)* algorithm for practical implementations of CBM. The SGP algorithm consists of a projection-onto-convex-set (POCS) iteration with a fast step-size search and converges quickly in 3 to 4 iterations.

The remainder of the paper is organized as follows: Section II reviews the structure and terminologies for a beamforming MISO-OFDM system and formally defines the problem of peak-power reduction in MISO-OFDM examined in this paper. Section III presents the method of transmit weight optimization (TWO), where we also briefly analyze the performance of the algorithm in relation to the trade-off between the available spatial freedom in a MISO channel and the number of users served by the base, referred to as the *space-user trade-off*. Based on the observation from Section III, we analyze in Section IV how one could cope with the performance degradation of TWO in high-user-number regimes by exploiting complementary degrees of freedom in the constellation domain, leading to the constellation-beam modification (CBM) technique. We proceed to formally describe the method of CBM in Section V and present two algorithms for CBM in Section VI. Simulation results are presented and discussed in Section VII.

II. Model for Beamforming MISO-OFDM and Problem Formulation

A. Signal Model for Beamforming MISO-OFDM

Consider a beamforming multiple-input-single-output (MISO) system where P transmit antennas are present at the base with B beams to be transmitted, one for each user. Each user is assumed to have only one antenna[2]. The number of users, B, is allowed to vary with time from 1 to P. Note that this time-varying feature of a MISO system distinguishes it from a beamforming MIMO scenario, where the numbers of transmit and receiving antennas are fixed.

Let the number of total OFDM subcarriers be N. For the k-th subcarrier, the *frequency-antenna-domain* symbol to be transmitted at the p-th antenna, $\mathbf{z}_{p,k}$, can be represented by

$$\mathbf{z}_{p,k} = \mathbf{w}_{p,k}^T \mathbf{x}_k, \tag{1}$$

where $\mathbf{x}_k$ is the $B \times 1$ OFDM *frequency-user-domain* data vector, and $\mathbf{w}_{p,k}$ is the $B \times 1$ antenna weight for the p-th antenna. In matrix form, the $P \times 1$ transmit vector at all P antennas for subcarrier k can be written as

$$\mathbf{z}_k = \mathbf{W}_k^T \mathbf{x}_k, \tag{2}$$

[2]Our method does not preclude the use of more than one receiver antenna, but the weights have to be designed to just one of the antennas.

where $\mathbf{W}_k$ is the $P \times B$ *transmit weight matrix* for subcarrier k, and $\mathbf{W}_k = [\mathbf{w}_{1,k}\ \mathbf{w}_{2,k}\ \ldots\ \mathbf{w}_{P,k}]$.

Assuming the base station has perfect channel knowledge, which is represented by the set of $B \times P$ channel matrices for all N subcarriers $\{\mathbf{H}_k, 1 \leq k \leq N\}$, the received signal at the b-th user, $1 \leq b \leq B$, can be written as

$$y_{k,b} = (\mathbf{H}_k \mathbf{W}_k^T \mathbf{x}_k + \mathbf{n}_k)_b, \tag{3}$$

where $\mathbf{n}_k$ denotes the $B \times 1$ noise vector introduced by the channel at subcarrier k, and by $(\mathbf{x})_b$ we mean the b-th element of vector $\mathbf{x}$.

Note that the zero-crosstalk constraint is enforced in Tx-SDMA and MRT (i.e. $B = 1$) systems, or equivalently,

$$\mathbf{I} = \mathbf{H}_k \mathbf{W}_k^T, \forall k \in \{1, 2..., N\}, \tag{4}$$

where $\mathbf{I}$ is the $B \times B$ identity matrix. This can be achieved by having the transmit weight matrix take the form

$$\mathbf{W}_k = \frac{(\mathbf{H}_k^H \mathbf{H}_k)^{-1} \mathbf{H}_k^H}{\sqrt{\text{trace}\left((\mathbf{H}_k^H \mathbf{H}_k)^{-1}\right)}} \tag{5}$$

where the denominator of (5) is a normalizing constant [18].

B. Peak-power Reduction for Beamforming MISO-OFDM

The peak-to-average-power ratio (PAPR) is widely used in measuring the degree to which peaks in time-domain signals occur. Since we are concerned about reducing the analog time-domain peaks at *all* antennas, PAPR is defined as follows within the context of beamforming MISO-OFDM:

Definition 1. *Let $\mathbf{a}_p$ be the $(NL) \times 1$ time-domain signal at the p-th antenna, i.e.*

$$\mathbf{a}_p = \mathit{interp}_L\left(\mathit{IFFT}\left(\mathbf{z}_p\right)\right), \tag{6}$$

where $\mathit{interp}_L(\cdot)$ is an interpolation that up-samples the time-domain signal by a factor of L in order to approximate the continuous waveforms at the antenna, and $\mathbf{z}_p$ is the frequency-domain transmit signal at the p-th antenna. The peak-to-average-power ratio (PAPR) is defined by:

$$\mathit{PAPR}(\boldsymbol{a}) = \frac{\max_{p,n} |\mathbf{a}_p[n]|^2}{\frac{1}{NLP} \sum_p \sum_n |\mathbf{a}_p[n]|^2}, \tag{7}$$

where $\boldsymbol{a}$ is the vector of time-antenna-domain signals from all antennas, $\boldsymbol{a} = \left[\mathbf{a}_1^T\ \mathbf{a}_2^T\ \ldots\ \mathbf{a}_P^T\right]^T$.

Reduction in PAPR versus reduction in peak power: While the value of PAPR for a given signal is widely adopted as a standard metric in the literature, it is often dangerous to overly trust the reduction in PAPR. When a PAPR reduction algorithm increases the average power of the signal, which is common to many existing techniques [3], the reduction of the actual *peak power* is often well below the reported reduction in PAPR values. Since the reduction of peak power is what truly reduces the power consumption of an amplifier, we choose to measure only the *reduction in peak power* when evaluating algorithms, which is more conservative yet realistic.

C. Problem Formulation

Following the discussion above regarding PAPR versus peak power, we introduce the quantity *peak-to-original-average-power ratio (POAPR)*, defined as follows:

Definition 2. *Let* $\mathbf{a}$ *and* $\hat{\mathbf{a}}$ *be the original and the modified* $NLP \times 1$ *time-antenna-domain signals, respectively. The peak-to-original-average-power ratio (POAPR) between* $\mathbf{a}$ *and* $\hat{\mathbf{a}}$ *is defined by the function:*

$$\text{\textit{POAPR}}\left(\hat{\mathbf{a}}, \mathbf{a}\right) = \frac{\max_{p,n} \left|\hat{\mathbf{a}}_p\left[n\right]\right|^2}{\frac{1}{NLP}\sum_p \sum_n \left|\mathbf{a}_p\left[n\right]\right|^2} \tag{8}$$

By comparing (7) and (8), one sees that the only difference lies in that POAPR is defined with respect to the average power of the original signal, which stays constant. Hence any reduction in the value of POAPR is essentially equivalent to the reduction of peak power, scaled by a constant. We now define our goal as solving a general optimization problem to minimize the value of POAPR between the modified signal and the original signal.

Problem formulation

Given an OFDM signal block to be transmitted, $\mathbf{a} = \left[\mathbf{a}_1^T \ \mathbf{a}_2^T \ ... \ \mathbf{a}_P^T\right]^T$. Let $\boldsymbol{\Delta}\mathbf{a}$ be an additive change to the original signal $\mathbf{a}$. We solve for:

$$\boldsymbol{\Delta}\mathbf{a}^* = \arg\min_{\boldsymbol{\Delta}\mathbf{a}} \text{POAPR}\left(\mathbf{a} + \boldsymbol{\Delta}\mathbf{a}, \mathbf{a}\right), \tag{9}$$

with the constraint that **no increase in bit-error-rate (BER)** is permitted.

Comments:

- The constraint of no bit-error-rate increase prevents trivial solutions where peak power can be reduced by decreasing overall transmit power or introducing signal distortions.
- The constraint in the problem formulation does not include *average power*, which could become an area of concern for applications where the allowed average radiation power is confined by regulatory constraints. The issue of average radiation power will be discussed and examined in Section VII.

III. Method of Transmit Weight Optimization

We now present our first result, the transmit weight optimization (TWO) method for peak-power reduction. Consider again a beamforming MISO-OFDM system with P base antennas transmitting to B users. Observe that when the number of users B is less than the number of antennas P, the $B \times P$ channel matrix for each subcarrier has a $(P - B)-$dimensional null-space, in which the transmit power is completely eliminated by the channel and not received by any users. In other words, when $B < P$ the set of all possible transmit weights W lies in a $(P - B)-$dimensional subspace.

The idea of TWO is to find an optimal transmit weight matrix which minimizes the peak power at the base antennas. Because such modifications of transmit weight lie strictly in the null-space of the channel matrix, changes at the base will have no impact on the received signals at the users, hence the technique carries no BER increase.

Peak-power reduction using TWO is accomplished by successively clipping the time-domain signals and projecting the modifications in the transmit weight matrix to the linear subspace of all beamformers that enforce zero-crosstalk among users. Since both the peak-limited time-domain signals and the linear subspace for all possible transmit weight matrices are convex, such an algorithm is guaranteed to converge to a solution if it exists, based on the theory of projection-onto-convex-sets (POCS) [5]. Relating to the optimization formulation described in Section II-C, TWO essentially minimizes the peak power over the set of zero-forcing beamformers defined as follows:

Definition 3. ***(TWO Constraints)*** *Let* $\Delta\mathbf{z} = FFT\left(\mathbf{\Delta a}\right)$. *TWO solves the optimization problem in Section II-C over the constraint set* S_T*, defined by:*

$$S_T = \left\{\mathbf{\Delta a} \in \mathbb{C}^{NLP} : \mathbf{H}_k \Delta\mathbf{z}_k = 0, \forall 1 \leq k \leq N\right\} \tag{10}$$

Note that for notational consistency with later parts of the paper, updates are expressed as additive changes to the transmitted frequency-antenna-domain signals z, which is equivalent to modifying the transmit weight matrix W. The outline of the algorithm is presented as follows.

POCS Algorithm for TWO

For each subcarrier k, let $\mathbf{W}_k$ be the $P \times B$ transmit weight matrix and $\mathbf{H}_k$ be the channel matrix, as defined in Section II-A.

1) Assign frequency-user-domain constellation points, $\mathbf{x}_k^i$, for all B users, where i is the iteration number. Construct the initial $P \times 1$ frequency-antenna-domain data by $\mathbf{z}_k^0 = \mathbf{W}_k \mathbf{x}_k^0$. Set $i = 0$.
2) Reconstruct the time-antenna-domain data $\mathbf{a^i} = \left[\mathbf{a}_1^{i\,T}\ \mathbf{a}_2^{i\,T}\ ...\ \mathbf{a}_P^{i\,T}\right]^T$ by applying a zero-padded inverse FFT on each antenna,
$$\mathbf{a}_p^i = \text{interp}\left(\text{IFFT}\left(\mathbf{z}_p^i\right)\right), \forall p \in \{1, 2, ..., P\}. \tag{11}$$
3) Clip the magnitudes of samples in $\mathbf{a^i}$ that exceed a predetermined limit, G_{max}, to get $\hat{\mathbf{a}}^i$.
4) Project the changes in the frequency-antenna-domain data onto the null-space of the channel matrix $\mathbf{H}_k$ and update z,
$$\mathbf{z}_k^{i+1} = \left(\mathbf{I} - \mathbf{W}_k \mathbf{W}_k^{pinv}\right)\left(\hat{\mathbf{z}}_k^i - \mathbf{z}_k^i\right) + \mathbf{z}_k^i, \tag{12}$$

where $\mathbf{W}_k^{pinv}$ is a pseudo-inverse of $\mathbf{W}_k$ and $\hat{\mathbf{z}}_k^i$ is the frequency-antenna-domain data after clipping, i.e.

$$\hat{\mathbf{z}}_k^i = \text{FFT}\left(\hat{\mathbf{a}}^i\right) \tag{13}$$

5) Let $i = i + 1$, return to Step 2 and repeat until a minimum PAPR is achieved or the number of iterations is exhausted.

Value of G_{max}: The clipping limit G_{max} in Step (3) of the algorithm can be either set to a fixed value, or adjusted between the POCS iterations. The problem with a fixed threshold is that one does not know beforehand what an optimal threshold might be: setting G_{max} to be too large will limit the reduction of peak value (The POCS algorithm will end after the peak magnitude falls below G_{max}), while setting G_{max} to be too small may cause the algorithm to never converge, because there simply may not exist such a modified signal that satisfies the TWO constraint *and* has a peak magnitude less that G_{max}. Hence, we choose to use an adjusted G_{max} which is automatically increased whenever an increase in peak power is observed after a POCS iteration. The rules we used for updating G_{max} will be presented in detail in Section VII.

Performance of TWO: Figure 2(a) in Section VII plots the reduction in peak power from TWO for an 8-antenna base transmitting to 1 through 8 users. Although TWO is very efficient in reducing peak power when the number of users is low, achieving a 6.65 dB reduction (14 iterations) at $B = 1$ (MRT), its performance degrades quickly as the number of users increases, due to the decreasing degrees of freedom in the channel's null-space. Finally, TWO fails to yield any reduction when $P = B$.

IV. Space-User Trade-off

A. Space-User Trade-off of MISO-OFDM

The dependency of TWO's performance on the number of users reflects a critical difference between MISO-OFDM and other wireless OFDM systems such as SISO and MIMO: in MISO-OFDM, the spatial dimensionality of the channel *changes* as the number of users served by the base varies. Relating to the problem of peak-power reduction, since the physical properties of amplifiers cannot simply change when the number of users served by a base changes, this means that the usefulness of any peak-power reduction algorithm is in some sense limited by its *worst-case* performance across *all* user-loads.

While TWO yields impressive reduction results when the number of users is low, we would like to take a step further and find a simple algorithm that achieves significant peak-power reduction for *all* numbers of users. Since the performance degradation of TWO essentially results from diminishing degrees of freedom in the constraints, a natural solution is to seek a second constraint space in which the degrees of freedom *increase* as the number of users increases. The hope is that if the optimization is performed over both constraint sets simultaneously, we achieve not only more reduction for a given user number, due to the relaxed constraint, but more importantly, a consistently high reduction performance across all numbers of users thanks to the complementary degrees of freedom of the constraint sets.

B. Solution: Relaxing ACE Constraints

Observe that the amount of constellation symbols in the frequency-user domain increases linearly with the number of users. Hence, we chose to incorporate the additional degrees of freedom in *constellations*, which is utilized by the active constellation extension (ACE) methods previously developed for PAPR reduction in SISO and MIMO systems [13], [14], [15]. The idea of ACE is to allow the outer-most constellation points to be extended *outward*, demonstrated graphically in Figure 1. It has been shown that the allowable regions of extension in ACE are so structured that the modified constellations can in fact result in a *reduced* bit-error-rate with a proper OFDM receiver design [16]. This can be explained by the fact that a constellation point, if moved, becomes further away any other constellation points than when in its original position, hence increasing the probability of correct detection by a standard maximum likelihood receiver. Therefore, the addition of the constellation modifications would satisfy our initial constraint of no bit-error-rate increase.

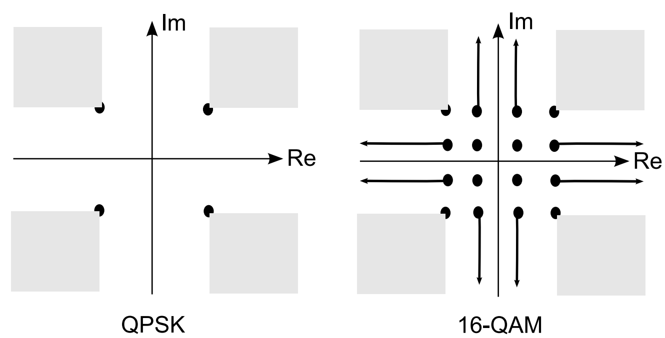


Fig. 1. Allowable regions of extensions in QPSK and 16-QAM constellations, represented by the gray areas and arrows.

The constraint set over which ACE optimizes is defined as follows:

Definition 4. (ACE Constraints) *ACE solves the optimization problem in Section II-C over the constraint set S_A, which is the set of all time-antenna-domain modifications $\Delta\mathbf{a} \in \mathbb{C}^{NPL}$, of which the corresponding constellation modifications fall under the allowable regions of extension on the constellation domain, for which the complete definition can be found in [13].*

Since the regions of extension as defined in [13] form a convex set, ACE can also be efficiently implemented using an iterative POCS algorithm similar to that of TWO. Here we extend the ACE-POCS algorithm developed for SISO [13] and MIMO [15] to MISO systems, accomplished by incorporating the user-antenna transformation, expressed by the antenna weight matrices $\mathbf{W}_k$.

POCS Algorithm for ACE

For each subcarrier k, let $\mathbf{W}_k$ be the $P \times B$ transmit weight matrix and $\mathbf{H}_k$ be the channel matrix, as defined in Section II-A.

1)-3) Same as that of the POCS algorithm for TWO in Section III.

4) Reconstruct the frequency-user-domain data $\hat{\mathbf{x}}_k^i$ by inverting the beamforming transformation:

$$\hat{\mathbf{x}}_k^i = \mathbf{W}_k^{pinv} \hat{\mathbf{z}}_k^i \tag{14}$$

where $\mathbf{W}_k^{pinv}$ is a pseudo-inverse of $\mathbf{W}_k$ and $\hat{\mathbf{z}}_k^i$ is the frequency-antenna-domain data after clipping, i.e.

$$\hat{\mathbf{z}}_k^i = \text{FFT}\left(\hat{\mathbf{a}}^i\right) \tag{15}$$

5) Apply active constellation constraints to $\hat{\mathbf{x}}_k^i$ on all OFDM frequencies and all antennas, by preserving the modified constellations that stay within the allowable regions of extensions, and projecting those which do not onto the boundaries of the regions [13]. Label the resulting user-frequency constellations as $\mathbf{x}_k^{i+1}$.

6) Apply the transmit weight matrix to the updated constellation to obtain the new frequency-antenna-domain signal:

$$\mathbf{z}_k^{i+1} = \mathbf{W}_k \mathbf{x}_k^{i+1} \tag{16}$$

7) Let $i = i + 1$, return to Step 2 and repeat until a minimum PAPR is achieved or the number of iterations is exhausted.

Performance of ACE. Figure 2(a) in Section VII shows the performance of the ACE algorithm applied to MISO-OFDM. It can be observed that the reduction performance does increase as the number of users increases from 1 to 8 users, which agrees with our intuition of the additional degrees of freedom due to more constellation symbols. There is however a performance degradation when the number of users reaches 8. We conjecture such a phenomenon is due to the high condition numbers of square matrices with i.i.d. Gaussian entries. This is, however, beyond the scope of this paper.

V. Method of Constellation-Beam Modification

The interesting complementary behavior of TWO and ACE as the number of users changes motivates us to optimize *jointly* over the union of the constraint sets from both TWO and ACE, which we call the constellation-beam modification (CBM) method.

Definition 5. (CBM Constraints) *CBM solves the optimization problem in Section II-C over the constraint set* S_C*,*

$$S_C = S_T \cup S_A \tag{17}$$

where S_A *and* S_T *are defined in Definition 3 and Definition 4, respectively.*

It is important to note that the TWO and ACE constraint sets are independent, so a joint modification preserves an undistorted signal reception at the receivers. This orthogonality between the constraint sets allows for a combined algorithm to be implemented very efficiently, since any combined adjustment can be perceived as *one* additive change to the original signal that simultaneously encapsulates the effects of TWO and ACE.

For any given user number, the CBM algorithm essentially optimizes over a higher-dimensional convex set with looser constraints, thus offering better peak-power reduction than either ACE or TWO alone.

VI. Efficient Algorithms for Constellation-Beam Modification

Similar to TWO and ACE, it was shown in Section V that CBM can also be framed as an optimization over a convex set. In this section, we first present a POCS implementation for CBM. Since the TWO and ACE constraints are essentially orthogonal, the constraint enforcement for CBM becomes two simple projections onto the ACE and TWO constraint sets which can be performed sequentially in the frequency domain. We then describe a more efficient smart-gradient-project (SGP) algorithm, which incorporates a fast stepsize search in each step, and converges quickly in only about three iterations (see Section VII).

The performance of CBM is deferred to the next section for more in-depth analysis.

A. Projection-onto-convex-sets (POCS)

The POCS for CBM is similar to that of TWO and ACE, described as below.

POCS Algorithm for CBM

1)-3) Same as that of the POCS algorithm for TWO in Section III.

4) Reconstruct the frequency-user-domain data $\hat{\mathbf{x}}_k^i$ by inverting the beamforming transformation:

$$\hat{\mathbf{x}}_k^i = \mathbf{W}_k^{pinv}\hat{\mathbf{z}}_k^i \tag{18}$$

where $\mathbf{W}_k^{pinv}$ is a pseudo-inverse of $\mathbf{W}_k$ and $\hat{\mathbf{z}}_k^i$ is the frequency-antenna-domain data after clipping, i.e.

$$\hat{\mathbf{z}}_k^i = \text{FFT}\left(\hat{\mathbf{a}}^i\right) \tag{19}$$

5) Apply active constellation constraints to $\hat{\mathbf{x}}_k^i$ on all OFDM frequencies and all antennas. Label the resulting user-frequency constellation as $\mathbf{x}_k^{i+}$. Denote the *frequency-user-domain descent vector* by $\Delta\mathbf{x}_k^i = \mathbf{x}_k^{i+} - \mathbf{x}_k^i$.

6) Project the frequency-antenna-domain clipping $(\hat{\mathbf{z}}_k^i - \mathbf{z}_k^i)$ onto the null-space of the channel matrix. Combine the result with the frequency-user-domain descent vector to form the frequency-antenna-domain descent vector:

$$\Delta\mathbf{z}_k^i = \left(I - \mathbf{W}_k\mathbf{W}_k^{pinv}\right)(\hat{\mathbf{z}}_k^i - \mathbf{z}_k^i) + \mathbf{W}_k\Delta\mathbf{x}_k^i \tag{20}$$

7) Update both frequency-antenna and frequency-user-domain signals:

$$\mathbf{z}_k^{i+1} = \mathbf{z}_k^i + \Delta\mathbf{z}_k^i$$
$$\mathbf{x}_k^{i+1} = \mathbf{x}_k^i + \Delta\mathbf{x}_k^i$$

8) Let $i = i + 1$, return to Step 2 and repeat until a minimum PAPR is achieved or the number of iterations is exhausted.

Computational complexity of POCS: The computational costs of the POCS algorithm for CBM, ACE and TWO are the same in the asymptotic sense with respect to the OFDM block size: the major cost is induced by two FFT/IFFT operations (per antenna) performed in each POCS iteration with a complexity of $O(N \log N)$, where N is the size of the OFDM block, while the rest of the operations, including peak clipping, enforcement of ACE constraints and projection onto the null-space, all have a complexity $O(N)$. As N grows large, the overall complexity becomes $O(N \log N)$.

B. Smart gradient project (SGP)

If we consider the frequency-antenna-domain update $\Delta \mathbf{z}_k$ in Step 6 of the CBM-POCS algorithm as a descent direction, the POCS algorithm is essentially a projected gradient descent algorithm with a constant unit stepsize [14]. Naturally, we can expect to improve the rate of convergence significantly with an optimal stepsize search along the gradient direction in each iteration. To do so, two questions need to be answered: 1) how can the stepsize search be done in a computationally efficient manner and 2) how to ensure the constraints are not violated when a stepsize > 1 is taken.

1) Fast Stepsize Search with Smart Gradient Project: We address the first challenge by generalizing to a MISO setting the technique of smart-gradient-project (SGP), which was first proposed for PAPR reduction in SISO-OFDM [13].

Let $\mathbf{\Delta a}$ be an interpolated time-antenna-domain descent vector across *all* antennas, obtained from a CBM-POCS iteration:

$$\mathbf{\Delta a} = \text{interp}_L \left[IFFT(\Delta \mathbf{z}) \right], \tag{21}$$

based on which an optimal stepsize μ^* can be expressed as:

$$\mu^* = \arg\min_{\mu} \left\| \mathbf{a} + \mu \mathbf{\Delta a} \right\|, \tag{22}$$

where $\mathbf{a}$ is the current time-domain signal across all antennas.

Let $a_{max}(\mu) = \max_n |\mathbf{a}[n] + \mu \mathbf{\Delta a}[n]|$. Note that as we increase the value of μ starting from $\mu = 0$, $a_{max}(\mu)$ decreases while the magnitudes of some other samples in $|\mathbf{a}[\mathbf{n}] + \mu \mathbf{\Delta a}|$ increase. At some point the a_{max} will reach a balancing with that of the second-largest sample(s), and the peak magnitude of the given block is minimized. To see why, note that further increasing μ by any positive amount will break the balancing by increasing the magnitude of the second-largest sample, resulting in a higher peak level.

As was shown in [13], solving (22) exactly involves computing the solutions of $NLP(NLP-1)$ quadratic equations for the optimal balancing that minimizes the peak magnitude among all possible combinations of two time-domain samples. This could be prohibitively expensive as the OFDM block size N becomes large. SGP greatly reduces the amount of computation by making the following two approximations:

- It is assumed that the optimal balancing will involve the sample n_{max} that has the largest initial magnitude, namely $n_{max} = \arg\max_n |\mathbf{a}[n]|$, hence reducing the number of comparisons from $NLP(NLP-1)$ down to $NLP-1$. This is a reasonable assumption based on the analysis of $a_{max}(\mu)$ given in the previous paragraph.
- To find the stepsize that will lead to a balancing between two time-domain samples i, j, one needs to solve for μ in the quadratic equation:

$$|\mathbf{a}[i] + \mu \boldsymbol{\Delta}\mathbf{a}[i]|^2 = |\mathbf{a}[j] + \mu \boldsymbol{\Delta}\mathbf{a}[j]|^2 \,, \tag{23}$$

  Instead, in SGP each sample in $\boldsymbol{\Delta}\mathbf{a}$ is projected onto the *phase angle* of $\mathbf{a}$ to avoid solving quadratic equations. This linearization reduces the computation of μ^* to simply (assuming $|\mathbf{a}[i]| > |\mathbf{a}[j]|$):

$$\mu = \frac{|\mathbf{a}[i]| - |\mathbf{a}[j]|}{\boldsymbol{\Delta}\mathbf{a}_{\mathbf{proj}}[j] - \boldsymbol{\Delta}\mathbf{a}_{\mathbf{proj}}[i]}, \tag{24}$$

  where

$$\boldsymbol{\Delta}\mathbf{a}_{\mathbf{proj}}[i] = \frac{Re\left\{\mathbf{a}[i]\boldsymbol{\Delta}\mathbf{a}[i]^*\right\}}{|\mathbf{a}[i]|}. \tag{25}$$

2) Constraints on Descent Vector: Regarding the concern of constraint enforcement with an arbitrary stepsize, note that the descent direction $\Delta \mathbf{z}_k^i$ consists of two components: an update that lies in the channel's null-space, $\left(I - \mathbf{W}_k \mathbf{W}_k^{pinv}\right)(\hat{\mathbf{z}}_k^i - \mathbf{z}_k^i)$, and an update that modifies the constellation, $\mathbf{W}_k \Delta \mathbf{x}_k^i$. The null-space update is not of concern, since any scaled version of the vector will still lie in the same linear subspace of the channel as before. The constellation update needs to be carefully handled, by making sure that the real and imaginary parts of the modification vector $\Delta \mathbf{x}_k^i$ have the same signs as that of the original constellation. In other words, the constellation points are now only allowed to be extended *outward* in *every* iteration, so that any positive scaling will keep the points within the allowable regions of extensions.

Note that in POCS, constellation points are allowed to move *inward* in any given iteration, so long as the resulting position still conform with the constraints. Theoretically this implies that POCS could explore the constellation domain more thoroughly than SGP. However, in actual simulations, we will see that SGP's efficiency gained by the stepsize search greatly overrules the small loss of freedom in the constellation domain.

The algorithm for stepsize search using SGP is described as the following (starting from Step 6 of CBM-POCS):

Algorithm for SGP Stepsize Search

---

1) Construct the time-antenna-domain descent vector and find the index, n_{max}, of the peak sample:

$$\boldsymbol{\Delta}\mathbf{a} = \text{interp}_L \left[IFFT(\Delta \mathbf{z})\right],$$
$$n_{max} = \arg\max_n |\mathbf{a}[n] + \mu \boldsymbol{\Delta}\mathbf{a}[n]| \,. \tag{26}$$

2) For every sample n_t in $\mathbf{\Delta a}$, including $n_t = n_{max}$, compute the projection of the descent vector along the phase angle of $\mathbf{a}[n_t]$:

$$\mathbf{\Delta a_{proj}}[n_t] = \frac{Re\left\{\mathbf{a}[n_t]\mathbf{\Delta a}[n_t]^*\right\}}{|\mathbf{a}[n_t]|}.. \tag{27}$$

3) If $\mathbf{\Delta a_{proj}}[n_{max}] > 0$, terminate the algorithm, since any positive stepsize would increase the current peak level. Otherwise, for every $\mathbf{\Delta a_{proj}}[n_t] > 0, n_t \neq n_{max}$, calculate the stepsize, $\mu[n_t]$, which will result in a balancing between $\mathbf{\Delta a}[n_t]$ and $\mathbf{\Delta a}[n_{max}]$:

$$\mu[n_t] = \frac{|\mathbf{a}[n_{max}]| - |\mathbf{a}[n_t]|}{\mathbf{\Delta a_{proj}}[n_t] - \mathbf{\Delta a_{proj}}[n_{max}]}, \tag{28}$$

4) Select the minimum among all positive $\mu[n]$ as the the optimal stepsize μ^*, i.e.:

$$\mu^* = \min_{1 \leq n \leq NLP, n \neq n_{max}, \mu[n] > 0} \mu[n] \tag{29}$$

5) Update the frequency-antenna, frequency-user, and time-antenna-domain data:

$$\begin{aligned} \mathbf{z}_k^{i+1} &= \mathbf{z}_k^i + \mu^* \Delta \mathbf{z}_k^i \\ \mathbf{x}_k^{i+1} &= \mathbf{x}_k^i + \mu^* \Delta \mathbf{x}_k^i \\ \mathbf{a}^{i+1} &= \mathbf{a}^i + \mu^* \mathbf{\Delta a} \end{aligned} \tag{30}$$

Computational overhead of SGP: Since the SGP algorithm examines each sample of the time-domain descent vector *exactly once*, the search for the optimal stepsize is achieved in $O(N)$, where N is the OFDM block size. Compared to the dominant computational costs of two FFT/IFFT operations inside each POCS iteration, the additional expense of SGP is small, especially when the size of the data block N becomes large. The reader may wonder why the IFFT in Step 1 of the SGP algorithm is not taken into account as part of the analysis. This is because in the next iteration we will no longer need an IFFT before clipping the peak, essentially replacing the IFFT in Step 2 of a POCS iteration.

VII. Simulation Results and Discussions

A. Simulation Parameters

All simulations in this section are performed over 10000 OFDM blocks with 128 subchannels, 8 antennas, and 1 through 8 users with no unused subchannels. While QPSK is used as the main constellation setting when comparing across different peak-power-reduction schemes, simulations are also conducted for 16-QAM and 64-QAM using CBM. The analog waveform at the transmit antennas is approximated using an up-sampling factor L = 4. All POAPR reduction values reported are measured at 0.1% clipping probability.

We use two channel models to capture the extreme conditions to give some sense of performance limits for the algorithms in reality. Let $h_k^{i,j}$ denote an entry of the channel matrix for subfrequency k, $\mathbf{H}_k$. The *Random Gaussian* model generates every element $h_k^{i,j}$ across all subfrequencies as i.i.d. complex Gaussian random variables, such that $\forall i, j$:

$$Re\left(h_k^{i,j}\right), Im\left(h_k^{i,j}\right) \sim \mathcal{N}(0,1), \forall k.$$

In the *Flat Gaussian* model, we first generate *one* random Gaussian channel matrix, and apply it to *all* subfrequencies, such that $\forall i, j$:

$$Re\left(h_k^{i,j}\right), Im\left(h_k^{i,j}\right) \sim \mathcal{N}(0,1), k=1,$$
$$h_k^{i,j} = h_1^{i,j}, \forall k \geq 2.$$

The Random Gaussian model is meant to simulate a rich-scattering environment with varying channel gains across frequencies, while the Flat Gaussian captures the opposite case where all channels have highly correlated gains; the reality should fall somewhere in between.

The clipping threshold G_{max} is expressed as a ratio of the peak of the current time-domain signal,

$$G_{max} = \beta \max_n |a[n]| \quad (31)$$

In order to ensure the convergence of the POCS algorithms, we increase β (with a decreasing increment) if and only if the previous iteration resulted in a peak growth. Specifically, β is updated according to:

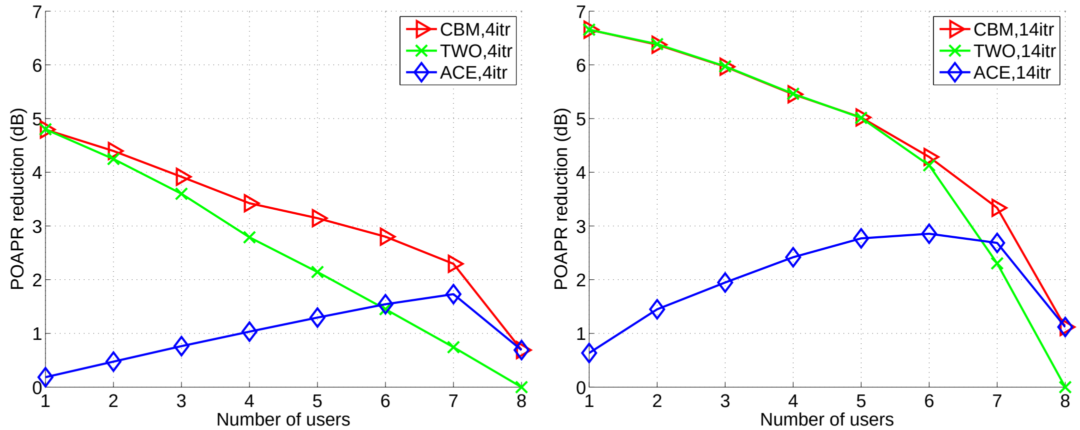
$$\beta^0 = 0.7, c^0 = 0.5,$$
$$\beta^{i+1} = \beta^i + c^i(1-\beta^i), c^{i+1} = c^i - 0.05$$

B. POAPR Reduction Performance with QPSK Constellations

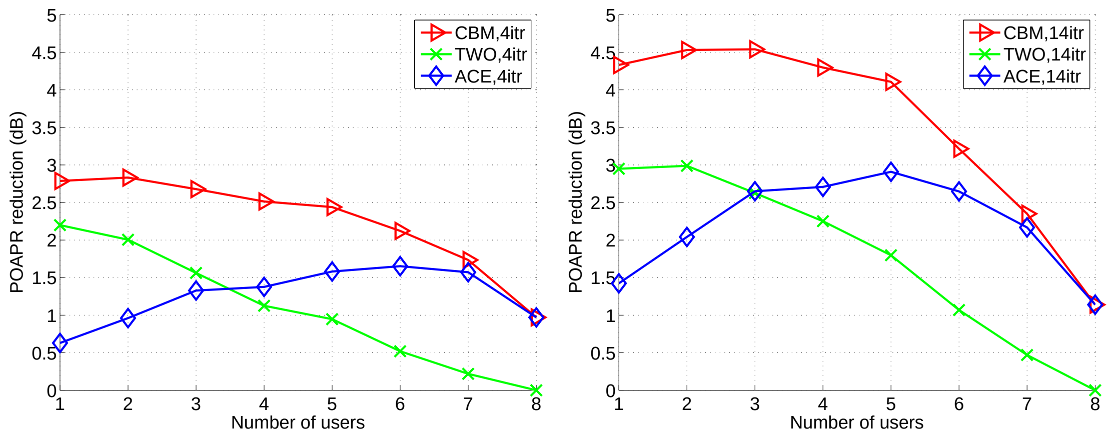
1) Channel Model I - Random Gaussian: Figure 2(a) illustrates the performance for CBM, TWO and ACE using QPSK constellations and 8 base antennas using POCS.

We report CBM's reduction with 14 POCS iterations as our primary result, while the reduction results with 4 iterations are also included for comparison. At 14 iterations, CBM yields a maximum POAPR reduction of 6.65 dB with 1 user, and a minimum reduction of 1.06 dB with 8 users. CBM's advantages over both ACE and TWO are evident at 5 to 7 users. In particular, CBM in the 6-user case yields an additional 0.97 dB reduction compared to any of the other methods alone. When the number of users is close to 1 or 8, CBM offers little benefits, due to the diminishing degrees of freedom in either TWO or ACE, respectively.

It is also interesting to notice the pattern of the evolution of performance: CBM shows a significant advantage over both TWO and ACE when the number of iteration is low (4 itr), indicating very efficient reduction during the initial iterations thanks to the increased degrees of freedom. CBM is caught up by TWO as the iteration number increases.



(a) Random Gaussian channel model



(b) Flat Gaussian channel model

Fig. 2. POAPR reduction of CBM, ACE and TWO using QPSK with 8 antennas (POCS). The left plots show the algorithms with 4 iterations and the right plots show the algorithms with 14 iterations.

In the convergent case (14 itr), TWO performs almost as well as CBM in low-user-number regimes, although CBM's additional degrees of freedom still yields an additional 1.05 dB reduction over TWO when the number of users is 7.

2) Channel Model II - Flat Gaussian: Based on the results from the Random Gaussian channel, it seems that compared to TWO, ACE makes limited, although noticeable, contribution to CBM's POAPR reduction performance. Before drawing this conclusion, however, it is important to realize that TWO's exploitation of the channel's spatial freedom may be favored by the Random Gaussian model, where every single entry of the channel matrix for *all* subfrequencies are generated independently, providing ample spatial freedom for TWO.

Figure 2(b) illustrates POAPR reduction using the Flat Gaussian model, where the freedom in the channel's null-space is reduced by having the same channel matrix for all subcarriers. In both plots, CBM shows significant advantage over both ACE and TWO. More importantly, while both ACE and TWO's POAPR reduction varies drastically as the number of users changes, CBM's performance remains stable across all user numbers. Also, contrary to the Random Gaussian case, the relative performance among CBM, TWO and ACE remain almost identical as the number of iterations varies.

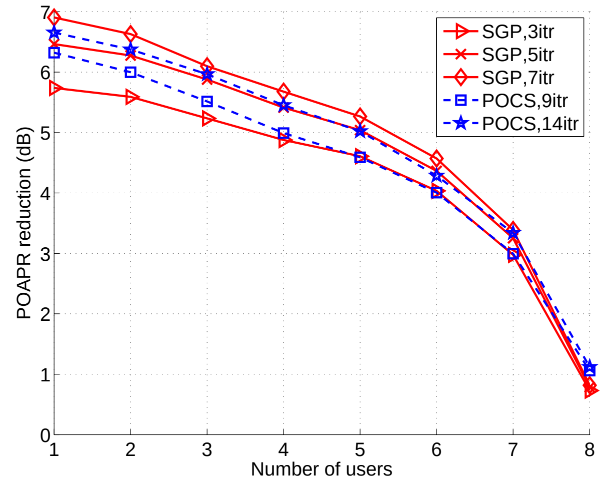


Fig. 3. POAPR comparison between CBM-POCS and CBM-SGP with QPSK constellations.

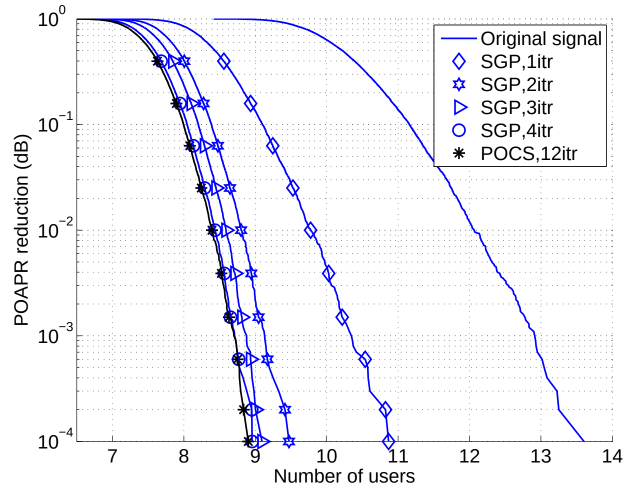


Fig. 4. Convergence comparison between CBM-POCS and CBM-SGP for 6 users with QPSK constellations.

C. Convergence: POCS vs. SGP

Implementing CBM using SGP shows outstanding convergence performance compared to using POCS. It can be seen from Figure 3 that on average SGP is able to achieve the same POAPR reduction with only one third the number of iterations of POCS.

Figure 4 plots the convergence comparison for the six-user case. The first two iterations of SGP shows impressive performance: the POAPR reduction achieved in the first two iterations ($\sim$ 4 dB) accounts for 89% of the total reduction ($\sim$ 4.5 dB).

D. Extensions to Higher Constellations

When CBM is used for larger constellations, ACE contributes less to the total reduction, since the outer-most points now constitute a smaller portion of the entire constellation. Figure 5 compares CBM's performance across three constellation settings: 16-QAM and 64-QAM show similar performance to QPSK, while the reduction with 5 to 8

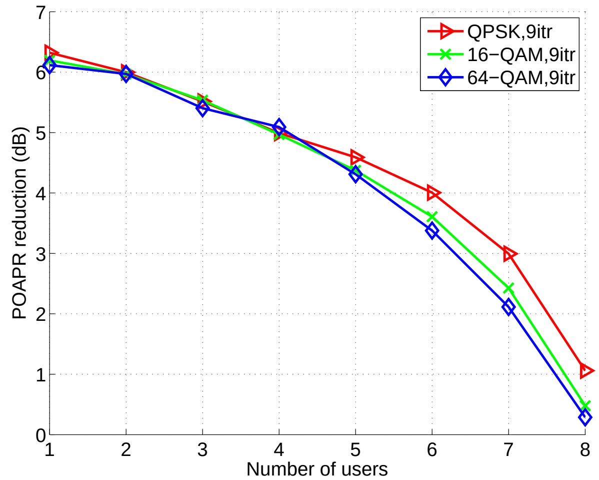


Fig. 5. Performance comparison of CBM using QPSK,16-QAM and 64-QAM constellations.

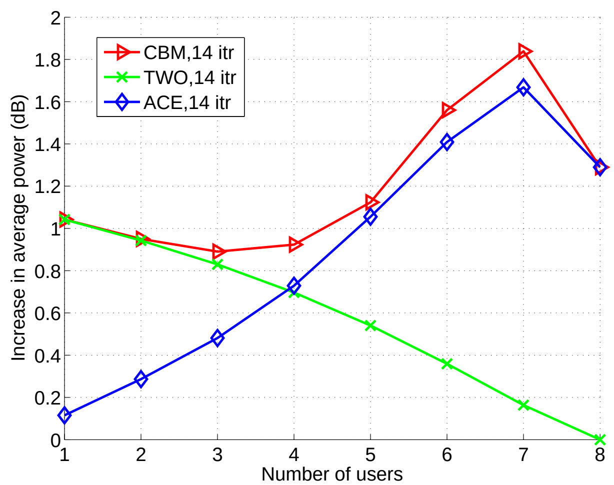


Fig. 6. Increase in average power with QPSK constellations.

users decreases by an average of 0.8 dB.

E. Issue of Average Power

The average transmit power does increase in all three methods (Figure 6). In TWO, additional energy is injected into the channel null space, which leads to a increase in average power, while in ACE it is mainly due to the extension of constellations. Both of these effects contribute jointly to the average power increase in CBM.

For the simulated QPSK system with an up-sampling factor $L = 4$, CBM increases the average power by a minimum of 0.89 dB (3 users) and a maximum of 1.84 dB (7 users). However, since the power consumption will be dominated by the maximum power capacity in systems with linear amplifiers, the effects of such an increased average power are more than offset by the benefits of a significantly reduced peak power.

VIII. Conclusion

We presented in this paper two efficient peak-power reduction schemes for beamforming MISO systems, the transmit weight optimization (TWO) and constellation-beam modification (CBM). TWO utilizes the degrees of freedom in the channel's null-space, which performs extremely well when the number of users is small compared to the number of base antennas, but has a degraded performance as the number of users increases. The performance variation of TWO is characterized as an intrinsic space-user trade-off of MISO-OFDM, where the spatial freedom in the channel necessarily decreases as the service load increases. This leads to CBM, which builds upon TWO by exploiting adjustments in the constellation domain, compensating the loss in spatial freedom with the increasing number of constellation points to be modified.

Based on the simulations, CBM offers a significant advantage over TWO in high-user-number regimes. When the number of users is small, CBM offers limited benefits compared to TWO in a rich scattering environment (modeled by the Random Gaussian channel). However, when the channel has limited spatial diversity among subcarriers (modeled by the Flat Guassian channel), CBM provides significantly higher and more robust reduction performance compared to ACE and TWO across different numbers of users.

Both TWO and CBM can be implemented with an iterative projection-onto-convex-set (POCS) algorithm with $N \log(N)$ complexity, which is due to the FFT/IFFT operation during each iteration. We also provided an $O(N)$ step-search algorithm (SGP) to hasten the speed of convergence. While both POCS and SGP have similar limiting performance, SGP is able to achieve an approximately three-fold speed-up, which makes it very attractive for practical implementations of CBM.

References

[1] A. Gatherer and M. Polley, "Controlling Clipping Probability in DMT Transmission," *Proceedings 31st Asilomar Conference on Signals, Systems, and Computers*, pp. 578–584, 1997.

[2] J. Tellado, "Peak to Average Power Reduction for Multicarrier Modulation," *Ph.D. Dissertation, Stanford University*, 2000.

[3] S.H. Han and J.H. Lee, "An Overview of Peak-To-Average Power Ratio Reduction Techniques for Multicarrier Transmission," *IEEE Wireless Communications*, vol. 12, pp. 56–65, April 2005.

[4] R.W. Beaumi, R.F.H. Fischer, and J.B. Huber, "Reducing the Peak-to-Average Power Ratio of Multicarrer Modulation by Selected Mapping," *Electronics Letters*, vol. 32, pp. 2056–2057, October 1996.

[5] L. M. Bregman, "The Method of Successive Projection for Finding a Common Point of Convex Sets," *Soviet Mathematics - Doklady*, vol. 6, pp. 688–692, 1965.

[6] S. Mueller and J. Huber, "OFDM with Reduced Peak-to-Average Power Ratio by Optimum Combination of Partial Transmit Sequences," *Electronics Letters*, vol. 33, pp. 368–369, February 1997.

[7] T.C.W. Schenk, P.F.M. Smulders, and E.R. Fledderus, “Peak-to-average Power Reduction in Space Division Multiplexing Based OFDM Systems Through Spatial Shifting,” *Electronics Letters*, vol. 41, no. 15, pp. 860–861, 2005.

[8] T.C.W. Schenk, P.F.M. Smulders, and E.R. Fledderus, “The Application of Spatial Shifting for Peak-to-Average Power Ratio Reduction in MIMO OFDM Systems,” *IEEE Vehicular Technology Conference*, 2006.

[9] Y. Lee, Y. You, W. Jeon, J. Paik, and H. Song, “Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission,” *IEEE Communications Letters*, vol. 7, no. 12, pp. 575–577, December 2003.

[10] R.F.H. Fischer and M. Hoch, “Peak-to-Average Power Ratio Reduction in MIMO OFDM,” *IEEE International Communication Conference*, pp. 762–767, 2007.

[11] J. Joung, E.R. Jeong, and Y.H. Lee, “Beamforming and PAPR Reduction for MISO-OFDM systems,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007.

[12] H. Zhang and D.L. Goeckel, “Peak Power Reduction in Closed-Loop MIMO-OFDM Systems via Mode Reservation,” *IEEE Communications Letters*, vol. 11, no. 7, pp. 583–585, 2007.

[13] B.S. Krongold and D.L. Jones, “Peak Power Reduction via Active Constellation Extension,” *IEEE Transactions on Broadcasting*, pp. 258–268, September 2003.

[14] D.L. Jones, “Peak Power Reduction in OFDM and DMT via Active Channel Modification,” *Proceedings 33rd Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, California), pp. 1076–1079, October 1999.

[15] G. Woo and D.L. Jones, “Peak Power Reduction in MIMO OFDM via Active Channel Extension,” *IEEE International Conference on Communications*, pp. 2636–2639, May 2005.

[16] T.F Detwiler and D.L. Jones, “OFDM Receiver Design for Active Constellation Extension,” *Proceedings 39th Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, California), pp. 1485–1489, October 2005.

[17] J.K. Carvers, “Single-User and Multiuser Adaptive Maximal Ratio Transmission for Rayleigh Channels,” *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, pp. 2043-2050, November 2000.

[18] T.A. Thomas and D.L. Jones, “PAPR Reduction for Beamforming OFDM Transmitters,” *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2006.