Diversity Fairness in Tomlinson–Harashima Precoded Multiuser MIMO Through Retransmission

Ioannis Krikidis, *Senior Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

Abstract—In this letter, we study the diversity unfairness associated with the conventional Tomlinson–Harashima precoding (THP) in multiuser multiple-input multiple-output downlink transmission. A single-retransmission scheme that combines two THP signals at each user with a complementary multi-user suppression order is investigated. For a system with M antennas at the transmitter and a single antenna at each user, the proposed scheme provides a diversity order $M + 1$ for all users and ensures diversity fairness. We study two retransmission policies, where the users either consider only the current received codeword or combine both codewords for decoding. An asymptotic analysis of the outage probability for both THP retransmission schemes is provided. In addition, a power allocation policy that minimizes the outage probability and accommodates the same coding gain at each user is discussed by formulating a geometric optimization problem.

Index Terms—Diversity, fairness, MIMO downlink, outage probability, THP.

I. INTRODUCTION

THE multiuser multiple-input multiple-output (MU-MIMO) downlink is a fundamental network topology that characterizes the modern cellular systems. The optimal MU-MIMO scheme is based on dirty-paper coding (DPC) which is a multiuser encoding strategy [1]. Due to the high complexity of DPC, the investigation of efficient linear precoders was an active research area over several years i.e., [2]. However, the efficiency of linear precoders requires a high number of users and appropriate user selection policies. For practical scenarios with moderate number of users, the performance difference between linear precoders and DPC becomes critical; the Tomlinson–Harashima precoding (THP) [3] is a non-linear technique that overcomes the limitations associated with the linear precoders and keeps the complexity low in comparison to DPC [4]. In THP, the user performance depends on the multiuser interference (MUI) suppression order and this causes a diversity unfairness to the downlink users; diversity gain/order[1] is an important performance metric at the physical

[1]At high SNRs, if the outage probability behaves like $P \approx c \cdot SNR^{-d}$, then c is the coding gain and d is the diversity order/gain.

layer. For communication scenarios where the downlink users can have equivalent performance requirements (i.e., downlink in LTE [5]), diversity unfairness becomes a critical issue.

In order to overcome this performance unfairness, the authors in [6] investigate a modified zero-forcing (ZF) THP scheme that ensures a diversity order equal to one to all users. However, although this approach balances the achieved diversity gains, it sacrifices the performance of the best-ordered users and does not exploit the available transmitter's spatial diversity. An extension of this scheme is proposed in [7] for a scenario with different quality-of-service requirements at the downlink users. The combination of multiple Tomlinson–Harashima precoded signals could be an efficient solution to support diversity fairness without affecting the best-ordered users. The work in [8] presents a combination of two THP signals in a base-station (BS) cooperation context; this work focuses on the coherent combination of the simultaneous transmissions without discussing fairness diversity issues.

The main target of this letter is to introduce a simple technique that resolves the THP diversity unfairness and jointly exploits the transmitter's spatial diversity. The proposed scheme elaborates a single-retransmission of the generated THP downlink signal by using two complementary MUI suppression orders at the transmitter. In this way, each user receives two copies of the transmitted signal via two "diversity-complementary" channels that keep the total number of diversity degrees constant for all users. We study two fundamental combination techniques at the receivers, which are inspired by the repeat request (ARQ) and the hybrid ARQ with chase combining (H-ARQ CC) schemes [9], and we analyze their performance for high signal-to-noise ratios (SNRs) in terms of outage probability. In addition, a power allocation (PA) scheme that ensures the same coding gain for all users is investigated by solving a generalized geometric program (GPP).

Notation

Vectors and matrices are denoted by boldface letters; $\mathbf{I_M}$ is the $M \times M$ identity matrix and the notation $(\bullet)^H$ denotes Hermitian transposition; $\mathcal{CN}(0, \sigma^2)$ represents a complex Gaussian distribution with zero mean and variance σ^2.

II. SYSTEM MODEL

We assume a MU-MIMO downlink where one BS, equipped with M antennas, transmits M independent data flows to M single-antenna users U_i with $i = 1, \ldots, M$. The signal received by the M users can be expressed in matrix form by an equivalent MIMO channel as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}, \qquad (1)$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$ denotes the received signal vector, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmitted symbol vector, and $\mathbf{n} \in \mathbb{C}^{M \times 1}$ is
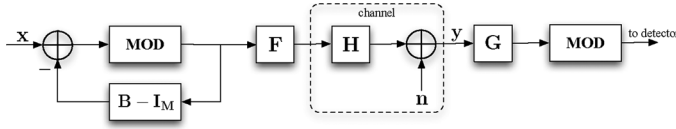
Fig. 1. Transmission/reception chain for THP.

the vector of the additive white Gaussian noise (AWGN) with entries $n_i \sim \mathcal{CN}(0,1)$, where $i = 1, \ldots, M$; $\mathbf{H} \in \mathbb{C}^{M \times M}$ denotes the channel matrix with entries $h_{m,n} \sim \mathcal{CN}(0,1)$, where the (m,n)-th element $h_{m,n}$ represents the flat fading co-efficient between the m-th transmit antenna and the n-th user. The channel remains constant during a single codeword and the fading changes independently from one codeword to another (block fading channel). This channel model can refer to a single subcarrier in orthogonal frequency-division multiplexing (OFDM)-based systems and is well-known in the literature i.e., [10]. The total available transmitted power is equal to P and p_i is the power allocated to the i-th data flow with $\sum_{i=1}^{M} p_i \leq P$; the required spectral efficiency for each user is r in bits per channel use (BPCU). We assume perfect channel state information (CSI) at the BS, which enables the design of precoding techniques for the downlink; perfect CSI simplifies the analysis and provides useful performance bounds for practical scenarios with non-ideal channel estimation.

III. CONVENTIONAL ZF-THP SCHEME AND DIVERSITY UNFAIRNESS

The ZF-THP is a well-known non-linear technique for the downlink that diagonalizes the MIMO wireless channel and allows simple scalar detection at the receivers. It is an efficient implementation of the DPC and consists of a feed-forward matrix $\mathbf{F}$, a feed-back matrix $(\mathbf{B} - \mathbf{I}_M)$ (where $\mathbf{B}$ is a unit triangular matrix) and a modulo device, as it is depicted in Fig. 1. The matrix $\mathbf{F}$ is used in order to make the MIMO channel spatially causal while the feed-back matrix $(\mathbf{B} - \mathbf{I}_M)$ eliminates the causal interference [3], [6]; these two matrices are based on the QR factorization of the MIMO channel matrix with $\mathbf{H}^H = \mathbf{QR}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R} = [r_{m,n}]$ is an upper triangular matrix. The modulo device constraints the transmitted power and can be seen as a replication of the used constellation in order to adapt the transmitted symbols to the MUI. The THP scheme is summarized as follows

$$\mathbf{B} = \mathbf{GR}^H, \quad \mathbf{F} = \mathbf{Q}, \qquad (2)$$

where the diagonal matrix $\mathbf{G} = [g_{m,n}] = \mathrm{diag}(r_{1,1}^{-1}, r_{2,2}^{-1}, \ldots, r_{M,M}^{-1})$ ensures causality in the MUI suppression process and distributed detection.

At the reception side, the i-th user multiples the received signal with the i-th diagonal element of the matrix $\mathbf{G}$ and performs a modulo operation in order to create the detector's input; the SNR for the link between the BS and the i-th user is equal to $\rho_i = p_i |r_{i,i}|^2$. By using the properties of the QR decomposition and basic statistical theory [6, Lemma 1], the terms $|r_{i,i}|^2$ follow a chi-square distribution with $2\bar{i}$ degrees of freedom (DoF), where $\bar{i} \triangleq M - i + 1$. The outage probability for the i-th user is expressed as

$$\mathrm{P}_{\mathrm{THP}}^{(i)}(p_i) = \mathbb{P}\left\{\log_2\left(1 + p_i|r_{i,i}|^2\right) < r\right\}$$
$$\approx \Phi\left(\frac{C_0}{p_i}, \bar{i}\right) \propto P^{-\bar{i}}, \qquad (\text{for } P \to \infty), \quad (3)$$

where $C_0 = 2^r - 1$ and the asymptotic expression $\Phi(x, n) = x^n/n\Gamma(n)$ is given in [9, Eq. (6)] (it approximates the cumulative distribution function (cdf) of a chi-square random variable with n DoF as $x \to 0$). The asymptotic expression in (3) shows that the conventional THP provides a diversity order $\bar{i}$ to the i-th user and therefore suffers from a diversity unfairness. For this scheme, we assume a symmetric PA with $p_i = P/M$ and thus further PA issues are not discussed (PA cannot resolve the diversity unfairness).

IV. THP WITH RETRANSMISSION

The main objective of the proposed THP-based scheme is to ensure the same diversity gain at each user without affecting the performance of the best-ordered users. In contrast to the previous approaches that balance the diversity unfairness by accommodating a diversity gain equal to one at each user [6], our motivation is to allocate a diversity gain $(M + 1)$ to all users. The proposed scheme employs a single retransmission of the M data flows with a complementary (inverse) MUI suppression order, followed by a combination of the two transmitted copies at each receiver. More specifically, assume that the BS performs a conventional THP scheme and transmits the M data flows by ordering[2] the users as $(U_1, U_2, \ldots, U_M)$, without loss of generality; this means that the i-th user achieves a diversity $\bar{i}$ according to our previous discussion. The proposed scheme introduces a retransmission procedure and after the first data transmission, the BS retransmits the same M data flows using THP but with the inverse MUI order i.e., $(U_M, U_{M-1}, \ldots, U_1)$ which means that the i-th user achieves a diversity gain equal to i. In order to have a fair comparison with the conventional single-transmission THP scheme, we assume that the total transmitted power for the two phases of the protocol is constrained to P; if $p_i^{(1)}$ and $p_i^{(2)}$ denote the power allocated to the i-th data flow for the first and the second phase of the protocol, respectively, we have $\sum_{m=1}^{2} \sum_{i=1}^{M} p_i^{(m)} \leq P$.

At the reception side, each user combines the two received signals in order to decide about the transmitted symbols. Our analysis is independent of particular coding/decoding technique and concerns random coding and typical set decoding. We consider two combination/decoding approaches that refer to systems with different complexity/computation constraints. The first approach assumes that each user considers only the current received signal in decoding of its data flow; this requires that the BS employs the same codebook for both transmissions (a similar coding/decoding operation is performed in the ARQ policy [9, Sec. II. A]). In the second approach, we assume that each user considers both received signals in decoding its current data flow. If the generated codebooks are independent and identically distributed according to a given probability density function, the combination process becomes similar to the H-ARQ with CC [9, Sec. II. B]. It is worth noting that the retransmission policy with combination can be seen as a space-time block coding technique, where the information is transmitted over two different time instances with a complementary interference suppression order.

---

[2]The user ordering can improve the performance (coding gain) in conventional THP systems [6]; however, given that the main objective of this work is to study the diversity order, further user ordering issues are beyond the scope of this letter.

V. ASYMPTOTIC OUTAGE ANALYSIS

We focus on the high SNR regime ($P \to \infty$) and we provide asymptotic closed-form expressions for the achieved outage probability performance.

A. Retransmission Without Combination (THP-R)

In the THP-R case, each receiver decodes its data flow based on the current transmission without combining the two transmitted copies. This scheme does not require data buffers or signal combiners at the receivers and keeps the implementation complexity low. The associated outage probability for the i-th user is written as

$$
\begin{aligned}
\mathrm{P}_{\mathrm{THP\text{-}R}}^{(i)}\left(p_i^{(1)}, p_i^{(2)}\right) &= \mathbb{P}\left\{\log_2\left(1 + p_i^{(1)}|r_{i,i}^{(1)}|^2\right) < 2r\right\} \\
&\times \mathbb{P}\left\{\log_2\left(1 + p_i^{(2)}|r_{\bar{i},\bar{i}}^{(2)}|^2\right) < 2r\right\} \\
&\approx \Phi\left(\frac{C}{p_i^{(1)}}, \bar{i}\right)\Phi\left(\frac{C}{p_i^{(2)}}, i\right) \qquad (4) \\
&= \frac{C^{M+1}}{\zeta(i)\left(p_i^{(1)}\right)^{\bar{i}}\left(p_i^{(2)}\right)^{i}} \propto P^{-(M+1)}, \quad (5)
\end{aligned}
$$

where $C = 2^{2r} - 1$, $\zeta(i) = i\bar{i}\Gamma(i)\Gamma(\bar{i})$ and $r_{i,i}^{(m)}$ refers to the channel's QR factorization for the m-th transmission. It is worth noting that the required spectral efficiency becomes $2r$ for the retransmission-based schemes as the channel is used twice in order to deliver users' data. From (5), we can see that the diversity order of the system using THP-R retransmission is $(M+1)$ for all users. On the other hand, a close observation of (5) shows that the coding gain of each user depends on the parameter $\zeta(i)$ and therefore can be different. Specifically, a symmetric PA that allocates the same power to all transmissions i.e., $p_i^{(1)} = p_i^{(2)} = P/2M$ does not ensure the same coding gain at each user; given that $\zeta(i) = \zeta(\bar{i})$, the users U_i and $U_{\bar{i}}$ achieve the same coding gain that is equal to $(2MC)^{M+1}/\zeta(i)$.

In order to overcome this coding gain unfairness, we design an optimal PA that minimizes the outage performance by jointly providing the same coding gain to all users. This optimal PA is formulated as a GGP optimization problem (in posynomial form) [11, Sec. 4.5.2] as follows

$$
\begin{aligned}
&\left(p_i^{(1)*}, p_i^{(2)*}\right) \\
&= \arg\min_{p_i^{(1)}, p_i^{(2)}} \max_{i \in [1,\ldots,M]}\left\{\frac{1}{\zeta(i)}\left(p_i^{(1)}\right)^{-\bar{i}}\left(p_i^{(2)}\right)^{-i}\right\} \\
&\text{s.t. } \frac{1}{P}\sum_{m=1}^{2}\sum_{i=1}^{M} p_i^{(m)} \le 1, \; p_i^{(1)} \ge 0, \; p_i^{(2)} \ge 0, \quad \forall i. \quad (6)
\end{aligned}
$$

The formulation in (6) is a GGP that can be transformed into a convex optimization problem of the socalled log-sum-exponential form; it then can be solved using standard computational algorithms such as the interior point method [11].

B. Retransmission With Combination (THP-RC)

In the THP-RC scheme, each user decodes its data flow by combining the two transmitted copies. It refers to a higher implementation complexity, as it requires a storage of the first re-ceived signal as well as a signal combiner at each receiver. The outage probability for the i-th user is written as

$$
\begin{aligned}
&\mathrm{P}_{\mathrm{THP\text{-}RC}}^{(i)}\left(p_i^{(1)}, p_i^{(2)}\right) \\
&= \mathbb{P}\left\{\log_2\left(1 + p_i^{(1)}|r_{i,i}|^2 + p_i^{(2)}|r_{\bar{i},\bar{i}}|^2\right) < 2r\right\} \\
&\approx V\left(p_i^{(1)}, p_i^{(2)}, \bar{i}, i, C\right) \\
&= \frac{C^{M+1}}{(M+1)\Gamma(M+1)\left(p_i^{(1)}\right)^{\bar{i}}\left(p_i^{(2)}\right)^{i}} \propto P^{-(M+1)}, \quad (7)
\end{aligned}
$$

where the function $V(\cdot)$ is given in the Appendix. As it can be seen from (7), the THP-RC scheme provides a full diversity order $(M+1)$ to all users. In addition, the expression in (7) shows that a symmetric PA with $p_i^{(1)} = p_i^{(2)} = P/2M$ provides the same coding gain to all users, which is equal to $(2MC)^{M+1}/[(M+1)\Gamma(M+1)]$ (in contrast to the THP-R scheme where the symmetric PA suffers from a coding gain unfairness). Although the symmetric PA is an efficient technique from a fairness standpoint, we study an optimal PA that further improves the coding gain of the users. We formulate the following GGP optimization problem that gives the optimal PA for the THP-RC case

$$
\left(p_i^{(1)*}, p_i^{(2)*}\right) = \arg\min_{p_i^{(1)}, p_i^{(2)}} \max_{i \in [1,\ldots,M]}\left\{\left(p_i^{(1)}\right)^{-\bar{i}}\left(p_i^{(2)}\right)^{-i}\right\}
$$
$$
\text{s.t. } \frac{1}{P}\sum_{m=1}^{2}\sum_{i=1}^{M} p_i^{(m)} \le 1, \; p_i^{(1)} \ge 0, \; p_i^{(2)} \ge 0, \quad \forall i. \quad (8)
$$

The above GGP formulation can be transformed to a convex GP optimization problem and solved via standard convex optimization tools [11].

VI. NUMERICAL RESULTS

Computer simulations are carried out in order to demonstrate the performance of the proposed schemes. We use CVX toolbox [12] to solve the GGP problems. Fig. 2 plots the outage probability of the investigated THP approaches versus P for a simulation setup with $r = 1$ BPCU and $M = 2$ users. The first main observation is that the conventional THP provides a diversity gain equal to one for the first user and equal to two for the second user; thus it suffers from a diversity unfairness. On the other hand, the proposed retransmission schemes accommodate a diversity gain equal to three for both users and ensure diversity fairness. A comparison between the retransmission policies shows that the THP-RC scheme outperforms the THP-R scheme with a coding gain 1.5 dB at high SNRs. In addition, it can be seen that the optimal PA slightly improves the achieved performance in comparison to the symmetric PA with a gain 0.25 dB; we note that for the considered simulation setup ($M = 2$ users), the symmetric PA provides the same coding gain at both users in the THP-R scheme. As for the theoretical approximations, it can be seen that the asymptotic expressions given in (5), (7) match with the simulations results at high SNRs and validate our analysis.

In Fig. 3, we depict the outage probability performance versus P for a simulation setting with $r = 1$ BPCU and $M = 4$ users.
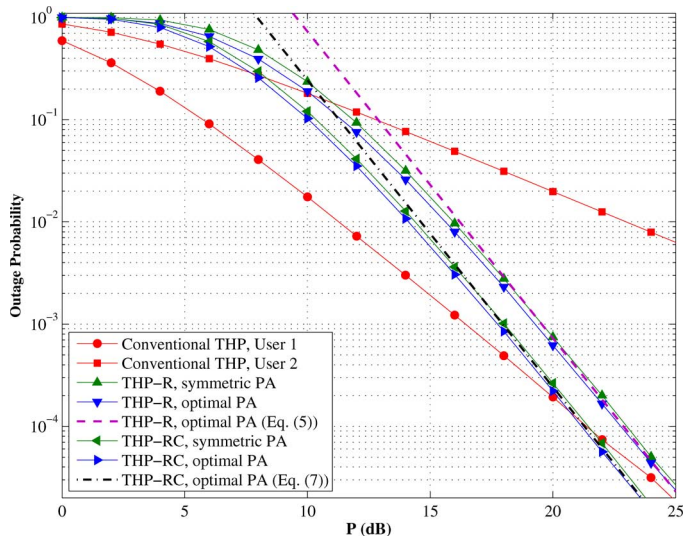
Fig. 2. Outage probability performance versus the transmitted power P; $r = 1$ BPCU and $M = 2$ users.
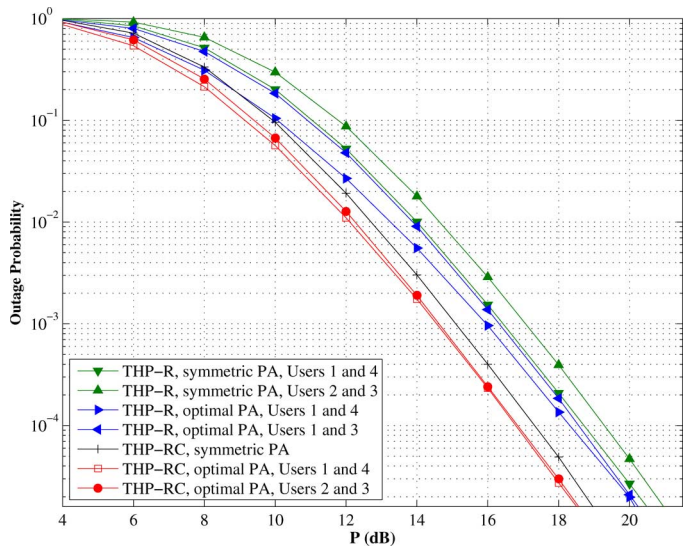

Fig. 3. Outage probability performance versus the transmitted power P; $r = 1$ BPCU and $M = 4$ users.

The main observations are in line with the above simulation example and validate our previous conclusions. In addition, for this case, it can be seen that the symmetric PA provides a different coding gain for the users (U_1, U_4) and (U_2, U_3) in the THP-R scheme. The optimal PA ensures the same power gain and significantly improves the outage probability performance for all users at high SNRs. For the THP-RC scheme, it can be seen that the optimal PA improves the achieved performance with a gain 0.5 dB in comparison to the symmetric PA.

VII. Conclusion

In this letter, we investigated a single-retransmission scheme that ensures diversity fairness in THP MIMO downlink. The proposed scheme involves the redunduncy of two THP signals that are generated by using a complementary MUI suppression order at the transmitter. We studied two different retransmission policies that enable data decoding with or without combination of the two transmissions. The outage probability for both retransmission schemes has been derived in closed-form for high SNRs. In order to further ensure coding gain fairness, a

PA policy has been studied by solving a geometric optimization problem. An extension of this work is to study the impact of the MUI order on the achieved performance.

Appendix

Let X_1, X_2 two central chi-squared random variables with $2m_1$ and $2m_2$ DoF, respectively; the cdf of X_i is approximated as $F_{X_i}(x) = x^{m_i}/m_i\Gamma(m_i)$ as $x \to 0$ [9, Eq. (6)] with a corresponding probability density function $f_{X_i}(x) = \vartheta F_{X_i}(x)/\vartheta x = x^{m_i-1}/\Gamma(m_i)$. We assume the random variable $Z \triangleq \alpha X_1 + \beta X_2$ where $\alpha, \beta > 0$; an approximation of the cdf of Z (for small X_1, X_2) is given by

$$V(\alpha, \beta, m_1, m_2, z)$$
$$= \mathbb{P}\{Z < z\} = \mathbb{P}\left\{X_1 < \frac{z - \beta X_2}{\alpha}\right\}$$
$$= \int_0^{z/\beta} \frac{1}{m_1\Gamma(m_1)} \left(\frac{z - \beta y}{\alpha}\right)^{m_1} \frac{y^{m_2-1}}{\Gamma(m_2)} dy$$
$$= \frac{1}{m_1\Gamma(m_1)\Gamma(m_2)} \left(\frac{\beta}{\alpha}\right)^{m_1} \int_0^u (u - y)^{m_1} y^{m_2-1} dy \quad (9)$$
$$= \frac{B(m_1 + 1, m_2)}{m_1\Gamma(m_1)\Gamma(m_2)} \left(\frac{\beta}{\alpha}\right)^{m_1} u^{m_1+m_2} \quad (10)$$
$$= \frac{1}{(m_1 + m_2)\Gamma(m_1 + m_2)\alpha^{m_1}\beta^{m_2}} z^{m_1+m_2}, \quad (11)$$

where $u = z/\beta$ in (9), $B(x, y) = \int_0^1 t^{x-1}(1 - t)^{y-1} dt$ denotes the beta function and (10) follows from the integral [13, 3.191.1].

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 4, pp. 1691–1706, Jul. 2003.

[2] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, pp. 3857–3868, Oct. 2005.

[3] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1305–1316, Jul. 2004.

[4] A. Tomasoni, M. Ferrari, S. Bellini, and G. Caire, "Effective beamforming and efficient scheduling for multi-user MIMO broadcast channels," in *Proc. IEEE Inf. Theory Workshop*, Taormina, Italy, Oct. 2009, pp. 520–523.

[5] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Process Lett.*, vol. 16, pp. 461–464, Jun. 2009.

[6] J. Liu and W. A. Krzymien, "A novel nonlinear precoding algorithm for the downlink of multiple antenna multi-user systems," in *Proc. IEEE Veh. Tech. Conf.*, Stockholm, Sweden, May 2005, pp. 887–891.

[7] L. Sanguinetti and M. Morelli, "Non-linear pre-coding for multiple-antenna multi-user downlink transmissions with different QoS requirements," *IEEE Trans. Wireless Commun.*, pp. 852–856, Mar. 2007.

[8] B. Wang, B. Li, and M. Liu, "A novel precoding method for joint processing in CoMP," in *Proc. IEEE Int. Conf. Netw. Comp. Inf. Sec.*, Guilin, China, May 2011, pp. 126–129.

[9] P. Hesami and J. N. Laneman, "Incremental use of multiple transmitters for low-complexity diversity transmission in wireless systems," *IEEE Trans. Commun.*, vol. 60, pp. 2522–2533, Sep. 2012.

[10] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2845–2866, Jun. 2010.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[12] M. Grant and S. Boyd, CVX Users' Guide for CVX Ver. 1.22, 2012.

[13] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York, NY, USA: Academic, 2007.