Modified Tomlinson-Harashima Precoding for Downlink MU-MIMO Channel with Arbitrary Precoder

Hamid Farmanbar and Hadi Baligh
Ottawa R&D Centre
Huawei Technologies Canada
Ottawa, ON, Canada
{hamid.farmanbar, hadi.baligh}@huawei.com

Abstract—**Tomlinson-Harashima precoding (THP) for downlink MU-MIMO channel can be modeled as user signal perturbation (constellation expansion) followed by linear precoding. The *effective linear precoder* in standard THP is merely a function of the MU-MIMO channel, which does not provide the flexibility for user power allocation. A modified THP algorithm is proposed so that the effective linear precoder can be arbitrary. As a special case, zero-forcing THP (ZF-THP) with arbitrary user power allocation is considered for which optimal user power allocation is derived to maximize the weighted sum-rate of users.**

Keywords-MIMO broadcast channel; Tomlinson-Harashima precoding; vector perturbation; multi-user power allocation

I. INTRODUCTION

The capacity region of multiple-input multiple-output broadcast channel (MIMO-BC), also known as downlink multi-user MIMO (DL MU-MIMO) channel, is achievable by dirty paper coding (DPC) [1]. However, DPC is too complex for implementation. Tomlinson-Harashima Precoding (THP) [2] and vector perturbation (VP) [3] are lower-complexity nonlinear precoding techniques for DL MU-MIMO channel, which offer performance improvement over linear precoding techniques.

In VP, a linear precoder is used to transmit the signal of users. However, the vector containing users' signals is perturbed prior to linear precoding to minimize instantaneous transmit power. It can be shown that THP can be formulized as VP. In other words, in THP as well, effectively, the vector containing users' signals is perturbed and then an *effective linear precoder* is applied. The idea behind THP and VP is to decrease the average transmit power while maintaining the same SNR levels at the receivers as their linear counterparts which do not apply perturbation (ignoring modulo loss at the receivers).

Although VP lends itself to using arbitrary linear precoder, which can be an advantage as it provides more freedom to improve performance, it has two major drawbacks. The first drawback is that the calculation of perturbed vector to minimize transmit power requires sphere decoding algorithm [4], which is NP-Hard. The second drawback is the calculation of perturbed vector covariance matrix. The covariance matrix of perturbed signal is required for the calculation of transmit power normalization factor. The brute-force calculation of the covariance matrix involves sphere decoding for all possible user signal vectors and then averaging out over all possibilities, which is prohibitively complex. In order to estimate the transmit power normalization factor, averaging over transmissions over some intervals has been proposed [5]. However, the estimated power by this approach is not accurate and could cause performance degradation.

THP is less complex than VP in terms of both precoding and perturbed signal covariance calculations. However, in its standard form, it does not lend itself to using arbitrary linear precoding. For example, in ZF-THP, the power allocation on the columns of the effective linear precoder is fixed and cannot be optimized. In this paper, we modify the standard THP algorithm such that it accepts arbitrary precoder. As a special case, we focus on our proposed modified zero-forcing THP algorithm with arbitrary user power allocation and optimize user power allocation such that the weighted sum-rate of users is maximized. We obtain closed-form power allocation formula in the case that high-SNR approximation of Shannon capacity is used as user rates.

We use the following notations and operators. Boldface upper-case symbols represent a matrix (e.g., $\mathbf{A}$), and boldface lower-case symbols represent a vector (e.g., $\mathbf{a}$). The notation $\mathrm{diag}(\mathbf{A})$ represents a diagonal matrix made up of the diagonal elements of $\mathbf{A}$. $[\mathbf{A}]_{ij}$ denotes the i-th row and j-th column element of $\mathbf{A}$. tr(.) denotes the trace of a matrix, E[.] denotes expectation, $(.)^T$ denotes transpose, $(.)^H$ denotes conjugate transpose, and $(.)^{-H}$ denotes (pseudo) inverse conjugate transpose.

The rest of the paper is organized as follows. In Section II, we introduce the system model. In Section III, we propose modified THP with arbitrary effective linear precoder. In Section IV, we consider modified ZF-THP with arbitrary user power assignment and we obtain optimal user power allocation to maximize weighted sum-rate of users. We present simulation results in Section V. Section VI contains summary and conclusion.

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

II. System Model

We consider a transmitter with N antennas and K single-antenna users ($N \geq K$). The MU-MIMO channel $\mathbf{H}_{K\times N}$ is the channel from the transmitter to all users and is assumed to be known at the transmitter. Let $\mathbf{u}_{K\times 1}$ be the vector of users' signals chosen from a unit power constellation.

A. Vector Perturbation

In VP, the transmit signal is given by

$$\begin{aligned}\mathbf{x} &= \mathbf{P}\hat{\mathbf{u}} \\ &= \mathbf{P}(\mathbf{u}+\tau\mathbf{l})\end{aligned} \tag{1}$$

where $\mathbf{P}_{N\times K}$ is a precoder matrix which is calculated based on $\mathbf{H}_{K\times N}$, τ is the constellation width plus twice the minimum distance of the constellation, and $\mathbf{l}$ is a complex integer vector that minimizes the instantaneous transmit power $\|\mathbf{x}\|^2$. The perturbed vector $\hat{\mathbf{u}}$ takes values in an infinite extended constellation.

We denote the covariance matrix of $\hat{\mathbf{u}}$ by $\mathbf{R}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}$. $\mathbf{R}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}$ is a function of precoder $\mathbf{P}$ which in turn is a function of the MU-MIMO channel $\mathbf{H}$. The average transmit power is given by

$$\begin{aligned}\mathrm{E}[\mathbf{x}^H\mathbf{x}] &= \mathrm{E}[\mathrm{tr}(\mathbf{x}\mathbf{x}^H)] \\ &= \mathrm{E}[\mathrm{tr}(\mathbf{P}\hat{\mathbf{u}}\hat{\mathbf{u}}^H\mathbf{P}^H)] \\ &= \mathrm{tr}(\mathbf{P}\,\mathrm{E}[\hat{\mathbf{u}}\hat{\mathbf{u}}^H]\mathbf{P}^H) \\ &= \mathrm{tr}(\mathbf{P}\mathbf{R}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}\mathbf{P}^H)\end{aligned} \tag{2}$$

The transmit signal (1) requires normalization such that the average transmit power is fixed at some value. From (2), the calculation of power normalization factor requires the calculation of $\mathbf{R}_{\hat{\mathbf{u}}\hat{\mathbf{u}}}$, which is prohibitively complex.

In zero-forcing VP (ZF-VP), the precoder is given by

$$\mathbf{P} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{D} \tag{3}$$

where $\mathbf{D}_{K\times K}$ is an arbitrary diagonal matrix which can provide the flexibility of user power allocation.

B. Tomlinson-Harashima Precoding

In standard ZF-THP, QR decomposition is applied to $\mathbf{H}^H$ as follows

$$\mathbf{H}^H = \mathbf{Q}_{N\times K}\mathbf{R}_{K\times K} \tag{4}$$

where $\mathbf{Q}$ has orthonormal columns and $\mathbf{R}$ is an upper triangular matrix with real entries on the diagonals. The first step of ZF-THP is given by

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= u_2 - \frac{r_{12}^*}{r_{22}}v_1 + \tau l_2 \\ &\vdots \\ v_K &= u_K - \frac{r_{(K-1)K}^*}{r_{KK}}v_{K-1} - \cdots - \frac{r_{1K}^*}{r_{KK}}v_1 + \tau l_K \end{aligned} \tag{5}$$

where u_k s are the elements of signal vector $\mathbf{u}$ and r_{ij} s are the elements $\mathbf{R}$ and $l_2,\ldots,l_K$ are complex integers chosen such that elements of the nonlinearly precoded vector $\mathbf{v}$ remain within the original constellation boundaries. The second step of standard THP is given by

$$\mathbf{x} = \mathbf{Q}\mathbf{v} \tag{6}$$

where $\mathbf{x}$ is the transmit signal.

For both VP and THP, the vector of received signals by users is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{7}$$

where $\mathbf{z} = [z_1 \cdots z_K]^T$ is the noise vector at receivers with independent elements.

The average transmit power for THP is given by

$$\begin{aligned}\mathrm{E}[\mathbf{x}^H\mathbf{x}] &= \mathrm{E}[\mathrm{tr}(\mathbf{x}\mathbf{x}^H)] \\ &= \mathrm{E}[\mathrm{tr}(\mathbf{Q}\mathbf{v}\mathbf{v}^H\mathbf{Q}^H)] \\ &= \mathrm{tr}(\mathbf{Q}\,\mathrm{E}[\mathbf{v}\mathbf{v}^H]\mathbf{Q}^H) \\ &= \mathrm{tr}(\mathbf{Q}\mathbf{R}_{\mathbf{v}\mathbf{v}}\mathbf{Q}^H) \\ &\cong \mathrm{tr}(\mathbf{Q}\mathbf{Q}^H) \\ &= K\end{aligned} \tag{8}$$

as $\mathbf{R}_{\mathbf{v}\mathbf{v}}$ can be approximated as identity matrix [6]. From (8), the normalization of transmit signal in THP is much easier than in VP.

C. Standard THP Formulation as VP

Eqn. (5) can be rewritten as

$$\left(\mathrm{diag}(\mathbf{R})\right)^{-1}\mathbf{R}^H\mathbf{v} = \mathbf{u} + \tau\mathbf{l} \tag{9}$$

where $\mathbf{l} = [0, l_2, \ldots, l_K]^T$. The transmit signal for ZF-THP can be written as

$$\begin{aligned}\mathbf{x} &= \mathbf{Q}\mathbf{v} \\ &= \mathbf{Q}\mathbf{R}^{-H}\,\mathrm{diag}(\mathbf{R})(\mathbf{u}+\tau\mathbf{l}) \\ &= \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\,\mathrm{diag}(\mathbf{R})(\mathbf{u}+\tau\mathbf{l}) \\ &= \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\,\mathrm{diag}(\mathbf{R})\hat{\mathbf{u}}.\end{aligned} \tag{10}$$

It can be seen that it has the form of ZF-VP with the *effective linear precoder*

$$\mathbf{P}_{eff} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\,\mathrm{diag}(\mathbf{R}). \tag{11}$$

It should be noted that in ZF-THP, the perturbation vector $\mathbf{l}$ is obtained from successive interference cancellation at the transmitter (see (5)) and may not be the same as the optimal perturbation obtained from sphere decoding in ZF-VP.

III. Modified THP with Arbitrary Effective Linear Precoder

If standard ZF-THP is used, the received signal at user k is given by

$$y_k = \left[\mathrm{diag}(\mathbf{R})\right]_{kk}(u_k + \tau l_k) + z_k. \tag{12}$$

The perturbation on u_k is removed by modulo operation at receiver. The received power of UE k is then given by $[\text{diag}(\mathbf{R})]_{kk}^2$. Therefore, power allocation is fixed and cannot be optimized. We propose a modified THP algorithm such that it can accept arbitrary effective linear precoder.

Let $\mathbf{B}_{K\times K}$ and $\mathbf{F}_{N\times K}$ be the THP feedback and feedforward filters [6], respectively, where $\mathbf{B}$ is a unit triangular matrix (a triangular matrix with unit elements on the main diagonal). The transmitted signal is given by

$$\mathbf{x} = \mathbf{F}\mathbf{B}^{-1}\hat{\mathbf{u}}. \tag{13}$$

Suppose we want the effective linear precoder to be $\mathbf{P}$, where $\mathbf{P}$ is an arbitrary $N\times K$ matrix for which a pseudo inverse exists. In other words, we must decompose $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{F}\mathbf{B}^{-1} \tag{14}$$

where $\mathbf{B}$ is unit triangular matrix. We put no restriction on $\mathbf{F}$. Let's apply QR decomposition on $\mathbf{P}^{-H} = \mathbf{P}(\mathbf{P}^H\mathbf{P})^{-1}$ as follows

$$\mathbf{P}^{-H} = \mathbf{Q}_1\mathbf{R}_1. \tag{15}$$

We choose

$$\mathbf{B} = \mathbf{R}_1^H\left(\text{diag}(\mathbf{R}_1)\right)^{-1}, \tag{16}$$

and

$$\mathbf{F} = \mathbf{Q}_1\left(\text{diag}(\mathbf{R}_1)\right)^{-1}. \tag{17}$$

Then $\mathbf{F}\mathbf{B}^{-1} = \mathbf{Q}_1\mathbf{R}_1^{-H}$. We just need to show $\mathbf{P} = \mathbf{Q}_1\mathbf{R}_1^{-H}$. By taking pseudo inverse Hermitian of both sides of (15), we obtain

$$\begin{aligned}\mathbf{P} &= \mathbf{Q}_1^{-H}\mathbf{R}_1^{-H}\\ &= \mathbf{Q}_1\mathbf{R}_1^{-H}\end{aligned} \tag{18}$$

as $\mathbf{Q}_1^{-H} = \mathbf{Q}_1$.

In summary, the first step of modified THP is given by

$$\begin{aligned}v_1 &= u_1\\ v_2 &= u_2 - b_{21}v_1 + \tau l_2\\ &\vdots\\ v_K &= u_K - b_{K(K-1)}v_{K-1} - \cdots - b_{K1}v_1 + \tau l_K\end{aligned} \tag{19}$$

and the second step is given by

$$\mathbf{x} = \mathbf{F}\mathbf{v} \tag{20}$$

where $\mathbf{F}$ is given by (17).

IV. Optimal Power Allocation for Modified ZF-THP

In Section III, we designed a modified THP algorithm that can accept arbitrary precoder. As a special case, we consider the popular case of ZF precoder with arbitrary power allocation. We optimize power allocation to maximize weighted sum rate of users. Let

$$\mathbf{P} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{D} \tag{21}$$

be the desired effective ZF precoder, where $\mathbf{D}_{K\times K}$ is a diagonal matrix with real diagonal entries $d_1,\ldots,d_K$. Then

$$\begin{aligned}\mathbf{P}^{-H} &= \mathbf{P}\left(\mathbf{P}^H\mathbf{P}\right)^{-1}\\ &= \mathbf{H}^H\mathbf{D}^{-1}\end{aligned} \tag{22}$$

Let $\mathbf{Q}_1$ and $\mathbf{R}_1$ be the unitary and the upper triangular matrix obtained from QR decomposition of $\mathbf{P}^{-H}$. Also, let $\mathbf{Q}$ and $\mathbf{R}$ be the unitary and the upper triangular matrix obtained from QR decomposition of $\mathbf{H}^H$. Since $\mathbf{D}$ is diagonal, we have $\mathbf{Q}_1 = \mathbf{Q}$ and

$$\mathbf{R}_1 = \mathbf{R}\mathbf{D}^{-1} \tag{23}$$

The transmit signal is given by

$$\begin{aligned}\mathbf{x} &= \mathbf{F}\mathbf{v}\\ &= \mathbf{Q}_1\left(\text{diag}(\mathbf{R}_1)\right)^{-1}\mathbf{v}\end{aligned} \tag{24}$$

Therefore, the average transmit power is given by

$$\begin{aligned}P_t &= \text{tr}\left(\text{E}\left[\mathbf{x}\mathbf{x}^H\right]\right)\\ &= \text{tr}\left(\mathbf{Q}_1\left(\text{diag}(\mathbf{R}_1)\right)^{-2}\mathbf{Q}_1^H\right)\\ &= \text{tr}\left(\mathbf{Q}_1^H\mathbf{Q}_1\left(\text{diag}(\mathbf{R}_1)\right)^{-2}\right)\\ &= \text{tr}\left(\left(\text{diag}(\mathbf{R}_1)\right)^{-2}\right)\\ &= \text{tr}\left(\mathbf{D}^2\left(\text{diag}(\mathbf{R})\right)^{-2}\right)\\ &= \sum_{k=1}^{K} a_k\delta_k\end{aligned} \tag{25}$$

where $\delta_k = d_k^2$ and $a_k = [\mathbf{R}]_{kk}^{-2}$. The received SNR of user *k* is given by

$$\begin{aligned}\rho_k &= \frac{d_k^2}{N_k}\\ &= \frac{\delta_k}{N_k}\end{aligned} \tag{26}$$

where N_k is the noise power at user *k*, *k= 1,…,K*. In the above SNR formula, we have ignored modulo loss at the receivers as it can be modeled as a constant reduction from received SNRs. The weighted sum-rate of users is given by

$$\begin{aligned}r(\boldsymbol{\delta}) &= \sum_{k=1}^{K} w_k\log\left(1+\rho_k\right)\\ &= \sum_{k=1}^{K} w_k\log\left(1+\frac{\delta_k}{N_k}\right)\end{aligned} \tag{27}$$

where w_k is the weight associated to user $k, k = 1,\ldots,K$. The user weights could come from MU-MIMO scheduler in a wireless network. For example, in a proportional fair scheduler, the weights are the inverse of average rate of users. The problem is to maximize the objective function in (27) subject to unit average transmit power (25), i.e.,

$$\max_{\boldsymbol{\delta}} \; r(\boldsymbol{\delta}) = \sum_{k=1}^{K} w_k \log\left(1+\frac{\delta_k}{N_k}\right)$$

s.t.

$$\sum_{k=1}^{K} a_k \delta_k \leq 1 \tag{28}$$

The above optimization problem is convex and can be solved using dual decomposition method [7]. If we use high-SNR approximation of Shannon rate, i.e., $\log(1+SNR) \approx \log(SNR)$, the optimization problem (28) reduces to

$$\max_{\boldsymbol{\delta}} \sum_{k=1}^{K} w_k \log(\delta_k)$$

s.t.

$$\sum_{k=1}^{K} a_k \delta_k \leq 1 \tag{29}$$

for which a closed-form solution is obtained by applying K.K.T. conditions. The optimal solution for the optimization problem (29) is given by

$$d_k = [\mathbf{R}]_{kk} \sqrt{\frac{w_k}{\sum_j w_j}} \tag{30}$$

The effective linear precoder with the above optimal power allocation can then be written as

$$\mathbf{P} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1} \operatorname{diag}(\mathbf{R})\mathbf{W} \tag{31}$$

where $\mathbf{W}$ is a diagonal matrix with diagonal entries $\sqrt{\frac{w_k}{\sum_j w_j}}, k = 1, \ldots, K.$

It would be insightful to compare the effective linear precoder with optimal power allocation (31) with effective linear precoder of standard ZF-THP (11). For the special case that all weights are equal, i.e., $w_1 = \cdots = w_K$, the two effective precoders are the same, which means that standard THP is optimal in terms of maximizing sum rate.

V. Simulation Results

We have conducted link-level simulations to compare the performances of standard ZF-THP and modified ZF-THP with optimal user power assignment. We consider downlink MU-MIMO scenario with K=4 single-antenna users and N=4 transmit antennas. The maximum transmit power is one and noise variance is fixed at $N_k = 0.05, \forall k$. We have used user weights $w_1 = 0.1$, $w_2 = 0.3$, $w_3 = 0.4$, and $w_4 = 0.2$. We have generated 10000 i.i.d. MU-MIMO complex Gaussian channel matrix samples with unit average power entries for each of which we have calculated the weighted sum-rate with standard ZF-THP and modified ZF-THP with optimal power assignment (30) (high-SNR approximation) as well as modified ZF-THP with optimal power assignment (optimal solution of (28)). We have plotted the cumulative distribution function (CDF) of the three weighted sum rates in Figure 1. Modified ZF-THP increases the weighted sum rate by about 0.19 b/s/Hz on the average compared to standard ZF-THP.

We have also conducted system-level simulations to compare the performances of standard ZF-THP and our proposed modified ZF-THP algorithm with optimal user power assignment. We consider a cellular system with 19 sites with 3 cells per site. We use Long Term Evolution (LTE) simulation methodology [8]. We consider intra-site cooperative multiple point (CoMP) transmission. There are 30 users per site (10 users per cell) from which the site "super" base station picks K users for MU-MIMO transmission. The "super" base stations, which are composed of three cell base station in a site, have a total number of 12 transmit antennas. A proportional fair (PF) scheduler is used for assigning resource blocks (RBs) to users. In the PF scheduler, the weights of users are the inverse of their current average received rates. The total cell throughput vs. cell edge (5 percentile) average throughput is plotted in Figure 2 for both standard and modified THP. The numbers shown on the figure represent the maximum allowed paired users for MU-MIMO transmission. As it can be seen, the modified THP algorithm offers both aggregate and cell-edge throughput gains over standard ZF-THP. It is worth mentioning that these gains are obtained at almost no added complexity.

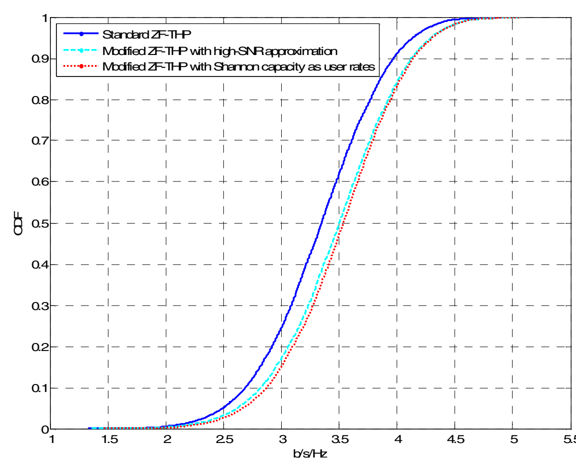

Figure 1. CDF of weighted sum rate comparison: standard ZF-THP vs. modified ZF-THP with optimal user power allocation.

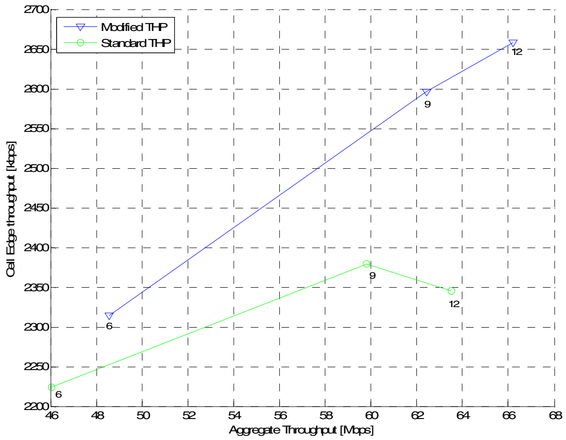

Figure 2. ZF-THP vs. modified ZF-THP with optimal user power assignment.

VI. CONCLUSION

Standard THP algorithm for downlink MU-MIMO does not provide the flexibility for user power allocation. We proposed a modified THP algorithm which accepts arbitrary effective linear precoder. We studied the special case of zero-forcing THP (ZF-THP) with arbitrary user power allocation and derived optimal user power allocation to maximize the weighted sum-rate of users.

REFERENCES

[1] H. Weingarten, Y. Steinberg, and S. Shamai, “Capacity region of the MIMO broadcast channel,” *IEEE Trans. Information Theory*, vol. 52, no. 9, pp. 3936-3964, 2006.

[2] D. A. Schmidt, M. Joham, and W. Utschick, “Minimum mean square error vector precoding,” *European Trans. Telecom.*, vol. 19, pp. 219-231, 2008.

[3] Hochwald, B.M.; Peel, C.B.; Swindlehurst, A.L.; , "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537- 544, March 2005.

[4] U. Fincke and M. Pohst, “Improved methods for calculating vectors of short lengths in a lattice, including a complexity analysis,” *Math. Computat.*, vol. 44, pp. 463–471, Apr. 1985.

[5] Vetter, H.; Sandell, M.; , “A New Dedicated Pilot Allocation Scheme in Multiuser MIMO OFDM with Vector Perturbation," *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd* , vol., no., pp.1-5, 15-18 May 2011.

[6] J. Liu and W. Krzymien, “Improved Tomlinson-Harashima precoding for the downlink of multi-user MIMO systems,” Can. J. Elect. Comput. Eng., vol. 32, no. 3, 2007, pp. 133-144.

[7] D. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, Aug. 2006, pp. 1439-1451.

[8] 3GPP TR36.814, Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9).