Cascaded Tomlinson Harashima Precoding and Block Diagonalization for Multi-User MIMO

Diwakar Sharma, Sriram N. Kizhakkemadam
Samsung India Software Operations, Bangalore, India
{diwakar, sriram.kn}@samsung.com

Abstract—We propose and evaluate the performance of a cascaded THP and BD system (THP-BD) for a MIMO downlink in which BD is used to mitigate multi-user interference and THP is used to mitigate spatial interference. We show that at moderate to high SNR, the cascaded THP-BD system achieves the same error performance as the system employing THP on a multi-user MIMO (MU-MIMO) downlink. Also, symbol error rate (SER) of the THP-BD precoding technique is at least 3 dB better than a linear precoding scheme when perfect CSIT is available. We also evaluate the performance of this scheme for channels with noisy feedback and with temporal correlation. Significantly, we find that while THP is sensitive to temporal correlation of the channel, THP-BD is more resilient.

Index Terms—Tomlinson Harashima precoding (THP), Multi-user MIMO, block diagonalization,

I. INTRODUCTION

In a MIMO broadcast channel, multiple antennas can be used to transmit independent data streams to spatially separated receivers. The superposition of symbols from multiple antennas causes inter-user interference (IUI) at each receiver (also called UE). Since the interfering signal is known apriori at the transmitter, dirty paper coding (DPC) technique can be used to completely mitigate the interference and achieve the sum capacity of MIMO broadcast channel [1]. Due to the complexity of DPC, simpler precoding techniques that equalize the spatial interference at the transmitter are generally preferred. For single user MIMO point to point link (SU-MIMO), wireless standards like 3GPP and IEEE 802.16m use a linear precoding technique based on the quantized eigen directions of the channel (also called a codebook) that has low channel feedback requirement and is computationally less complex. For multi-user scenarios, knowledge of eigen directions of the channel is insufficient to cancel IUI. Complete knowledge of the channel, is required to cancel IUI. For a MIMO broadcast channel, some of the precoding techniques that require complete knowledge of the channel are THP [2], [3] and BD [4], [5].

THP has been shown to almost achieve the capacity of DPC with the advantage of a realizable transceiver architecture. The THP structure consists of a feedback and feed-forward filter at the transmitter. The feedback filter cancels the inter-stream interference successively while the feedforward filter ensures that the noise at the decision devices is spatially white. A modulo operator at the transmitter and receiver is used to limit the magnitude of the transmitted symbol and this makes THP a non-linear scheme. In BD, a precoding matrix is designed for each receiver such that the received signal is in the null space of the channel of all other UE’s [4]. The MU-MIMO channel will then be reduced to a bank of IUI free SU-MIMO channels with residual multi-antenna interference at each receiver. While using THP for mitigating both IUI and multi-antenna interference is a close to capacity achieving scheme, it is computationally complex. In contrast, using a linear precoder for mitigating inter-stream interference (say SVD) and a linear precoder like BD to mitigate IUI is computationally simpler, but entails a significant penalty in error rate performance.

Simplifying the THP structure for MU-MIMO has garnered significant attention [6], [7]. In this paper, we propose a multi-user spatial multiplexing scheme that uses THP to mitigate multi-antenna interference and BD to mitigate IUI. This scheme has the benefit of orthogonalizing the channels using BD and coming close to capacity of the underlying MIMO channel for each using using THP. The proposed scheme acieves a better error rate performance compared to a complete linear pre-coding scheme at the cost of a modest increase in computational complexity at the receiver. For the case of perfect CSIT, we observe that the error performance of the cascaded THP-BD scheme is close to that of THP and significantly better compared to using a combination of SVD based single user precoding and BD to mitigate IUI.

Complete CSIT is vital to accrue the benefit of both THP and BD. Channel errors due to estimation errors, quantization errors and outdated CSIT limit the gains due to precoding. In TDD systems, the channel can be known by channel reciprocity. However, outdated CSIT may pose a problem, especially at high mobility (or equivalently at high Doppler frequencies). For FDD systems, complete knowledge of the channel at the transmitter is often prohibitive. Hence precoding techniques for MU-MIMO often necessitate impact of finite rate feedback. For BD, the feedback rate has to scale linearly with the SNR to preserve the multiplexing gain of the broadcast channel [8]. For determining the practical viability of the cascaded THP-BD scheme, we numerically evaluate its performance for

- Noisy estimate of the channel known at transmitter
- Channel with Dynamic CSIT and finite rate feedback

In both cases, we observe that the THP-BD scheme has performance close to that of using THP for mitigating multi-antenna interference and IUI and at the same time is significantly better
compared to BD based linear precoding schemes. The effect is more visible at higher SNR’s which is the regime of interest for spatial multiplexing.

This paper is organized as follows. In section II, we describe the system model and follow it with describing our proposed cascaded THP-BD scheme in section III. In section IV, we describe channel models for partial CSIT. In section V, we evaluate numerically the performance of THP-BD scheme with complete and partial CSIT in comparison with THP and BD precoding schemes.

II. System Model

We consider a MIMO downlink wherein a base station (BS) equipped with $N_t$ transmit antennas serves $K$ users. The $k^{th}$ user has provision for $N_r$ receive antennas. We consider multi-user spatial multiplexing and the BS allocates $L_k$ transmit antennas for $k^{th}$ user. The channel between the BS and the $k^{th}$ user is denoted by $\mathbf{H}_k \in \mathbb{C}^{N_r \times L_k}$. We assume $\mathbf{H}_k$ to be flat fading with elements $h_{ij}$ taken i.i.d from a zero mean unit variance Gaussian distribution. The channel is assumed to be block fading with independent fading across each block. We denote by $s_k$ the output of a $M - QAM$ modulator that is used to transmit the signal information for the $k^{th}$ user at the BS. We employ THP on $s_k$ to obtain $x_k$. If we denote by $\mathbf{T}_k$, a linear pre-coder for the $k^{th}$ user, the received signal at the $n^{th}$ time instant is

$$y^k[n] = \mathbf{H}_k[n] \sum_{i}^{K} T_i[n]x_i[n] + w_k[n]$$

(1)

where the received vector is corrupted by additive complex Gaussian noise $w_k$ with elements distributed i.i.d according to $\mathcal{CN}(0, \sigma_w^2)$. We also denote by $\mathcal{H}_k \triangleq \{\mathbf{H}_k : i = 1, 2, \ldots, K; i \neq k\}$, the complex conjugate transpose is denoted by *. Throughout, each receiver is assumed to have perfect information of its respective channel. Also, the channel is assumed to be wide sense stationary. Without loss of generality, we describe the design of precoders for the $k^{th}$ user.

The transmitter for each user is assumed to be power limited to $P$ Watts with a total power constraint of $KP$. Accordingly, the signal to noise ratio SNR is defined as $SNR = P/\sigma_w^2$ for each user. For perfect CSIT, waterfilling achieves the capacity of a MIMO channel. However, since our interest for spatial multiplexing in this paper is for medium to high SNR where equal power loading and waterfilling achieve almost the same capacity, we consider equal power loading at the transmitter for each user. In the following section, we assume that the transmitter has perfect CSIT and knows the channel non-causally. In Section III, we assume that the transmit and receive processing is done on a per-symbol basis. Hence, we drop the time index $n$ for the following section, but use it in Section IV.

III. Cascaded BD THP

We now describe the design of precoders for cascaded BD-THP operation. Since both BD and THP are well described schemes, we describe them only briefly. In our scheme, the precoder $T_k$ designed using BD for the $k^{th}$ user (say) is a function of $\mathcal{H}_k^k$ while the precoder designed using THP is a function of $\mathbf{H}_k$ and $\mathbf{T}_k$. Hence, we first describe BD and then summarily describe THP. A block diagram of the cascaded BD-THP system is given in Fig. 1

A. Mitigating multi-user interference using Block Diagonalization

From (1), the received signal for the $k^{th}$ user is

$$y_k = \mathbf{H}_k T_k x_k + \mathbf{H}_k \sum_{i \neq k} T_i x_i + w_k$$

(2)

where $\mathbf{H}_k \sum_{i \neq k} T_i x_i$ is the interference at the $k^{th}$ receiver due to the superposition of transmitted codewords of $K - 1$ users. To mitigate this interference, the design of the precoder $T_k$ for the $k^{th}$ user should satisfy $[\mathbf{H}_k^T \mathbf{H}_k^T \ldots \mathbf{H}_k^T \mathbf{H}_{k+1}^T \ldots \mathbf{H}_K^T]^T T_k = 0$. If we denote by $\overline{\mathbf{H}}_k = [\mathbf{H}_1^T \mathbf{H}_2^T \ldots \mathbf{H}_{k-1}^T \mathbf{H}_k^T \mathbf{H}_{k+1}^T \ldots \mathbf{H}_K^T]^T$, we can equivalently write $\overline{\mathbf{H}}_k^T T_k = 0$ where 0 is a $(K - 1)N_r$ dimensional vector of 0’s. Therefore, if $T_k$ is in the null space of its corresponding aggregated channel matrix $\overline{\mathbf{H}}_k^T$, inter user interference for the $k^{th}$ user will be mitigated. A QR decomposition of $\overline{\mathbf{H}}_k$ to obtain its null-space yields

$$\overline{\mathbf{H}}_k = \mathbf{R}_k^T \mathbf{Q}_k = [\mathbf{R}_k^T \mathbf{0}_{N_r \times (L_k - L_k)}] [\mathbf{Q}_k^T \mathbf{Q}_k^T]$$

(3)

where $\mathbf{R}_k$ is a $N_r - L_k$ dimensional square right triangular matrix. The unitary matrices $\mathbf{Q}_{k_1}$ and $\mathbf{Q}_{k_2}$ are of dimension $N_t \times N_t - L_k$ and $N_t \times L_k$ respectively. From (3), we see that $\mathbf{Q}_{k_2}$ is in the null space of $\overline{\mathbf{H}}_k$ and hence we choose $T_k = \mathbf{Q}_{k_2}$. Note that the column vectors of $\mathbf{Q}_{k_2}$ are the $N_t - L_k + 1$ to $L_k^{th}$ column vectors of $\mathbf{Q}_k$. A sufficient condition for the existence of a non-zero precoding matrix is $L_k > 0$ [4].

B. THP to mitigate inter-stream interference

After canceling the inter-user interference for each user, the effective channel for each user is $\mathbf{H}_k^k T_k^k = \mathbf{H}_k^k \mathbf{Q}_{k_2}$. This can be decomposed as $[\mathbf{H}_k T_k] = \mathbf{F}_k \mathbf{U}_k^H$, where $\mathbf{F}_k$ is a lower triangular matrix and $\mathbf{U}_k$ is a unitary matrix. The lower triangular characteristic of $\mathbf{F}_k$ can be exploited for successive interference cancellation.

- If we pre-multiply $x_k$ by $\mathbf{U}_k$, the effective channel will be the lower triangular matrix $\mathbf{F}_k$. The superposition of the signals from multiple transmit antennas at each receive antenna will then be due to the lower triangular elements of $\mathbf{F}_k$. Consequently, the symbol at each receive antenna (say $l^{th}$) is a weighted sum of the transmitted symbols from the first to the $l^{th}$ antenna.
the feed forward matrix \( U \) (or equivalently at low Doppler frequency), we can exploit

\[ \text{For further details, the interested reader is referred to [2], [3].} \]

encoder and decoder to cancel the inter-stream interference.

\[ \text{A. Dynamic CSIT} \]

A wireless channel has a varying channel response in time and frequency domain. Temporal variation of the channel is due to the motion of the transmitter, receiver or scatterers in the channel. If the coherence time \( T_c \) of the channel is large (or equivalently at low Doppler frequency), we can exploit the temporal correlation of the channel to send feedback less frequently. In the sequel, we follow the exposition of [9] to describe partial CSIT in a Rayleigh fading channel with temporal correlation.

For a zero mean Rayleigh fading channel, the channel auto-covariance characterizes how rapidly the channel decorrelates with time. Let us denote by \( \hat{h}_k[n] \), the vectorized form of the channel \( H_k[n] \). The elements of \( h_k[n] \) can be indexed as \( \hat{h}_k[n] = [\hat{h}_k(1), \ldots, \hat{h}_k(L_k N_c)] \). The temporal auto covariance of the channel between two samples at symbol instances \( n \) and \( n+m \) can be written as

\[ C_k[n] = \mathbb{E} \left[ h_k[n] h_k^* [n + m] \right] \]

The matrix \( C_k \) is a \( L_k \times N_c \) dimensional square matrix and gives the covariance of each co-efficient of the channel with all other channel co-efficients (including itself). Note that due to the assumption of \( i.i.d \) and stationarity in section II, the auto-covariance depends only on the time difference. If the channel temporal statistics is the same for all transmit and receiver antenna pairs, we can assume that the temporal correlation is homogeneous and identical for any element of the channel. If we denote the correlation co-efficient by \( \rho_{k}[m] = \mathbb{E} \left[ \hat{h}_k[n] \hat{h}_k^*[n + m] \right] \), then for the homogeneous temporal correlation channel [9]

\[ C_k[n] = \rho_{k}[m] C_k[0] \]

So, instead of feeding back the channel co-efficients, the receiver can feed back the correlation coefficient at regular intervals of time. The Fourier transform of the correlation co-efficient gives the power spectral response of the channel co-efficients. From the power spectral response of the correlation co-efficients, we get the Doppler spread of the correlation co-efficient. A popular model for the correlation co-efficient is the Clark’s spectrum given by

\[ \rho[m] = J_o \left( 2\pi f_d m \Delta t \right) \]

where \( f_d \) is the Doppler spread, \( J_o \) is the zeroth order Bessel function of the first kind and \( \Delta t \) is the sampling interval.

If \( \hat{H}_k \) is the estimate of the channel at the transmitter based on the receiver’s feedback, a CSIT model can be written as [9]

\[ H_k[n] = \hat{H}_k[n] + E_k[n] \]
where $E_k[n]$ is the estimation error for the $k^{th}$ user’s channel. Assuming unbiased estimates, we can model $E_k[n]$ as a stationary zero mean Gaussian random process. If the initial measurement of the channel at say, time 0 is known perfectly at the transmitter, from MMSE estimation theory we can obtain the estimated channel at time $n$ as,

$$
\hat{h}_k[n] = E[h_k[n]h_k[0]]
\approx C_k[n]C_k^{-1}[0]h_k[0]
$$

(9)

For the homogeneously temporal correlated channel, by substituting (6) in (9), the estimate of the channel becomes

$$
\tilde{H}_k[n] = \rho_k[n]H_k[0]
$$

(10)

The error covariance $C_k[n]$ associated with this estimation will be $C_k[n] = (1 - \rho_k^2[n])C_k[0]$. Thus, as the correlation factor tends to 1, the estimate of the channel becomes the initial estimate $H_k[0]$ and the error covariance approaches 0. As the correlation factor tends to 0, the estimate of the channel tends to its mean (zero) and the error covariance approaches the channel variance $C_k[0]$. Thus, periodic feedback of the initial estimate $H_k[0]$ and the channel correlation factor $\rho_k$ is sufficient. If $B$ bits are available to feedback $\rho_k$ and the initial estimate of the channel $H_k[0]$, we can choose an optimal quantizer based on the Lloyds-Max algorithm [11]. For a $B$ bit representation, the Lloyds-Max algorithm finds the optimal partition and codebook that can represent the variable with a particular distribution with minimal quantization error.

B. Noisy CSIT

In certain scenarios, complete knowledge of the channel of all users may be known at the transmitter, albeit corrupted by noise. This model is particularly appealing for TDD systems where channel reciprocity helps in knowing the channel at the transmitter. A model for the noisy CSIT case can therefore be written as

$$
\tilde{H}_k[n] = H[n] + z[n]
$$

(11)

where $z[n]$ is a zero mean complex Gaussian random variable with variance $\sigma_z^2$. Accordingly, we can define the channel-to-noise ratio (CNR) as $\text{CNR} = 1/\sigma_z^2$.

V. NUMERICAL ANALYSIS

We now present numerical evaluation of THP-BD. For comparison, we also plot results for a linear precoding scheme based on SVD and a non-linear MU precoding scheme based on THP. We assume QPSK modulation and a transmitter with 4 transmit antennas. We assume a 2 user system each of which has 2 receive antennas. For THP-BD, we design the filters as given in Section III. When using THP for both multi-antenna interference and IUI mitigation, we design the filters as given in [2]. We give below the symbol error rate (SER) performance for various channel models.

---

![Fig. 2: Plot of SER for THP, THP-BD and Linear Precoding as a function of $E_b/N_0$ when the channel of both users is known perfectly and non-causally at the transmitter](image-url)

A. Complete and Non-Causal CSIT

In our first comparison, we assume that CSIT is available perfectly and non-causally for all transceiver implementations. From Fig. 2, we cannot distinguish the difference in the performance of THP and THP-BD. In contrast, Linear precoding based on SVD performs poorer. This case of complete and non-causal CSIT models the maximum achievable performance of all the schemes and can hence serve as a benchmark.

B. Noisy estimate of channel

In Fig 3, we plot the SER for THP, THP-BD and linear precoding scheme when a noisy estimate of the channel is available at the transmitter. For CNR below 28 dB, the error performance of all three schemes is poor. As the CNR increases beyond 28 dB, the error performance of all the schemes improves. We also observe that the gap in performance between linear precoding scheme and the THP-BD scheme increases as the CNR improves. The curves for CNR of 100 dB is effectively the no noise estimate for the channel at the transmitter. We observe that asymptotically the gap between linear pre-coding scheme and the proposed THP-BD scheme is 5 dB. Likewise, with improvement in CNR, the performance of the THP-BD scheme approaches that of THP scheme.

C. Dynamic CSIT

In Fig 4, we plot the SER for the three schemes for a correlated Rayleigh fading channel whose autocorrelation coefficient follow the Clark’s spectrum of (7). We assume that updates are sent in intervals of 1 ms and that there are 14 symbols in 1 ms (following the frame structure of LTE-A). The receiver estimates the correlation of the 14 taps, computes the average of the correlation coefficient for the 14 symbols and feeds it back to the transmitter. Also, we assume that the
transmitter has perfect estimate of the channel matrix $H_k$ for all the $K$ users at the beginning of each 1 ms period. We assume a Doppler of 5 Hz as our simulations indicate that for Doppler spread beyond 10 Hz, all the schemes have a high SER. From the figure, we observe that at high SNR, THP-BD outperforms even THP in error rate performance. We are currently investigating the theoretical reasoning for this phenomenon. We chose to average the co-efficients instead of sending a quantized estimate of the channel correlation co-efficients to operate within the uplink feedback budget. Note that each receiver has to transmit 16 real co-efficients (corresponding to 8 channel elements) to the transmitter in addition to the channel correlation coefficient.

VI. CONCLUSION

In this paper we considered using BD to orthogonalize the channels in a MIMO broadcast channel and subsequently use THP to mitigate the effect of multi-antenna interference in the underlying MIMO link channel. Both BD and THP require complete knowledge of the channel for their operation. Our numerical results indicate that with perfect CSIT, the symbol error rate performance of the cascaded THP-BD system approaches that of THP and at the same time significantly outperforms a SVD based linear precoding scheme. The gap in performance between THP-BD and linear precoding is about 5 dB. When we introduce imperfections in the channel in the form of either a noisy estimate of the channel or temporal correlations in the channel, the performance of THP, THP-BD and linear precoding all worsen. While THP-BD follows the trend of the THP scheme, linear precoding continues to perform poorer in comparison to THP and THP-BD. We note that for a noisy estimate of the channel, a CNR less than 40 dB decreases the performance of all the schemes significantly. Likewise, with temporal correlation, if the Doppler spread is greater than 10 Hz, the SER performance of all the precoding schemes worsen. However, we do note that THP-BD performs better than THP based precoding at low Doppler frequencies. Further work on theoretically establishing thresholds for the relative merit of each of these schemes and investigation of the impact of practical finite feedback rate on THP-BD is underway.

REFERENCES