Abstract — In this paper, we present a joint forward-error-correction (FEC) coding and linear precoding scheme for multiple-input multiple-output (MIMO) and inter-symbol interference (ISI) channels with imperfect channel state information at the transmitter (CSIT). We first study the performance of ideally coded systems. We focus on an average power gain (APG) method that can achieve capacity in the two extreme cases of no CSIT and perfect CSIT. In the more general case, the performance of the APG method improves progressively with the CSIT quality. We then consider the implementation of this APG method in a practically coded system. We propose a unified scheme involving beamforming, water-filling, and diversity coding. The core of the new scheme is a joint FEC coding and linear precoding strategy at the transmitter and an iterative detection process at the receiver. Simulation results demonstrate that the proposed scheme can achieve significant performance gain by efficiently utilizing the available CSIT.

Index Terms—precoder, MIMO ISI channel, iterative linear minimum-mean-square-error (LMMSE) detection.

I. INTRODUCTION

Beamforming, water-filling, and diversity coding are well studied transmission methods for multiple-input multiple-output (MIMO) systems [1][2]. When channel state information at the transmitter (CSIT) is not available, diversity coding is applicable. When CSIT is available, beamforming and water-filling can provide additional performance gain. In the latter case, a traditional approach is first to transform the channel into a set of parallel sub-channels by singular value decomposition (SVD). Then multiple forward-error-correction (FEC) codes are used to adjust the rates and transmission power levels on different sub-channels based on the water-filling principle.

In practice, CSIT may be imperfect, which can be caused by a number of reasons, e.g., quantization effect or channel drift during the feedback process. Imperfect CSIT imposes a serious problem on multiple-code transmission, since the overall performance is dominant by the code with the worst performance.

Linear precoding is an alternative for MIMO systems. Most related work so far has been on un-coded systems [3]-[7]. The recent work in [8] takes into consideration the impact of FEC coding. It is shown that by proper design, linear precoding can achieve near-capacity performance.

In this paper, we first study the system performance assuming ideal coding. We focus on an average power gain (APG) method. This APG method can achieve capacity in the two extreme cases of no CSIT and perfect CSIT. In the general case of imperfect CSIT, the APG performance improves progressively with the CSIT quality. We then consider the implementation techniques of the APG method in a practically coded system. A unified beamforming, water-filling, and diversity coding scheme is developed. The core of the new scheme is a joint FEC coding and linear precoding strategy at the transmitter and an iterative detection process at the receiver. The precoding involves three matrixes, namely, a unitary matrix for beamforming, a diagonal matrix for power allocation among different beams using the water-filling principle, and a second unitary matrix for diversity coding. Design techniques for these matrixes are presented. Simulation results demonstrate that the proposed scheme can achieve significant performance gain by efficiently utilizing the available CSIT.

II. CAPACITY ANALYSIS

A. Channel Model

Assume that the cyclic-prefix and discrete-Fourier-transform techniques [9] are used to transform a MIMO ISI channel into a MIMO OFDM system with J subcarriers modeled as

$$y = Gx + \eta$$

where $y$ is the received signal vector from $N_r$ receive antennas, $G$ a $JN_r$-by-$JN_t$ channel transfer matrix, $x$ the signal vector of length $JN_t$ transmitted over $N_t$ transmit antennas, $\eta$ a $JN_r$-by-1 vector of complex additive white Gaussian noise (AWGN) with variance $\sigma^2$. The matrix $G$ can be further expressed in a block-diagonal form as $G = \text{diag}(G(1), ..., G(J))$, where $G(j)$ is an $N_r$-by-$N_t$ matrix representing the MIMO channel at subcarrier $j$. The transmitted signal $x$ has a zero mean and a total power limit $P$, i.e.

$$E[x] = \mathbf{0} \text{ and } \text{tr}(S) \leq P$$

where $S = E[xx^H]$ is the transmission covariance matrix.

We assume that partial channel state information (CSI) is available at the transmitter, and full CSI is available at the receiver. Following [1][10], we model $G$ as

$$G = \sqrt{\alpha} G_{\text{mean}} + \sqrt{1 - \alpha} G_{\text{var}}$$

where $G_{\text{mean}} = \text{diag}(G_{\text{mean}}(1), ..., G_{\text{mean}}(J))$ and $G_{\text{var}} = \text{diag}(G_{\text{var}}(1), ..., G_{\text{var}}(J))$ represent, respectively, the known and unknown parts of the channel transfer matrix at the transmitter.

This work was fully supported by a grant from the Research Grant Council of the Hong Kong SAR, China [Project No. CityU 117007].
and $\alpha \in [0,1]$ is a measure on the quality of the CSIT. For simplicity, we further assume that both $G_{\text{mean}}(j)$ and $G_{\text{ua}}(j)$ for each index $j$ have independent and identically distributed (i.i.d.) entries, with each entry distributed as $CN(0,1)$.

### B. Perfect CSIT Case

In the special case of $\alpha = 1$, the CSIT is perfect. Define the following power gain matrix of the channel

$$M = G^H G.$$  \hfill (3)

Denote the eigenvalue decomposition of $M$ by

$$M = V A V^H,$$  \hfill (4)

where $V$ is a unitary matrix and $A$ is a diagonal matrix. The channel capacity $C$ can be found by solving:

$$C = \max \{ J^{-1} \log \det(I + GSG^H) \},$$  \hfill (5a)

where the optimal $S$ takes the form

$$S = VW^H,$$  \hfill (5b)

and $W$ is a diagonal matrix with diagonal entries given by the well-known water-filling coefficients [11].

### C. Imperfect CSIT Case

In the general case of $0 \leq \alpha < 1$, the CSIT is imperfect. Due to the CSIT uncertainty, the achievable rate of the channel is a random variable from the perspective of the transmitter. In this case, the outage probability is a more appropriate measure of system performance. Denote the outage probability $\varepsilon$ given the transmission covariance matrix $S$ and the target rate $R_0$:

$$\varepsilon(S,R_0) = \Pr\{J^{-1} \log \det (I + GSG^H) < R_0\}. $$ \hfill (6)

The outage capacity is defined as

$$C_{\varepsilon} = \max_{\varepsilon(S,R_0)} \{ R_0 : \varepsilon(S,R_0) \leq \varepsilon_0 \}. $$ \hfill (7)

To the best of our knowledge, there is no closed-form solution to (7) for the system in (1) in general.

Since the optimization over all possible $S$ is computational demanding, we now consider a suboptimal but simple choice for $S$. Define the average power gain (APG) matrix of the channel as

$$\bar{M} = E[G^H G] = \alpha G_{\text{mean}}^H G_{\text{mean}} + (1-\alpha) N/\sqrt{J}$$ \hfill (8)

where the expectation $E[\cdot]$ treats $G_{\text{mean}}$ as deterministic since it is known at the transmitter. Denote the eigenvalue decomposition of $\bar{M}$ by

$$\bar{M} = \bar{V} \bar{A} \bar{V}^H.$$ \hfill (9)

The transmission covariance matrix is designed as

$$\tilde{S} = W \tilde{W}^H,$$ \hfill (10)

where the diagonal matrix $\tilde{W}$ is obtained by water-filling over a set of sub-channels with power gain given by the diagonal entries of $\tilde{A}$. The performance of the system can then be evaluated based on (6) by letting $S = \tilde{S}$.

Note that due to the CSIT uncertainty, the actual eigenvectors and eigenvalues of the channel power gain matrix are unknown at the transmitter. In the above APG method, we use their expected values instead.

In the extreme case of no CSIT, substituting $\alpha = 0$ into (8), we obtain $\bar{M} = NJ$. Thus the water-filling solution to the equivalent channel $\tilde{M}$ is given by $\tilde{S} = J^{-1} \tilde{P}$. According to a conjecture in [12], the optimal solution to (7) is given by $S = J^{-1} \tilde{P}$ (assuming a sufficiently small outage probability), which implies that APG can achieve the capacity of the system without CSIT.

In the other extreme case of perfect CSIT, (8) reduces to (3), i.e., $\bar{M} = M$. Hence the APG method is still optimal.

In general, the APG method is suboptimal since counter-examples of its optimality can be constructed. However, as will be seen later, both analytical and simulation results show that the APG method can exploit the available CSIT efficiently and achieve significant performance gains.

### D. Alternatives for the Imperfect CSIT Case

For comparison with the APG method, we consider two alternatives for the imperfect CSIT case.

The first is the equal power (EP) scheme, in which we select $\bar{W} = J^{-1/2} \bar{P}$ regardless of the CSIT quality. The EP scheme provides a good diversity gain and is optimal in the case of no CSIT. However, its performance remains the same regardless of the CSIT quality, and so it cannot take advantage of the CSIT. Its performance is considerably worse than that of the APG method when CSIT is available, as shown in Fig. 1.

The second is the conventional multiple-code water-filling (MC-WF) scheme, in which the water-filling principle is applied to $G_{\text{mean}}$ (the known part of the channel) to allocate power and rate between different eigenmodes. In the case of perfect CSIT, the MC-WF scheme is optimal and can provide additional water-filling gain compared with the EP one. However, the MC-WF scheme is sensitive to the CSIT error for the following reason. In the MC-WF method, multiple codes are transmitted in parallel. In the presence of CSIT error, the true channel condition can be better or worse than the condition predicted by the available CSIT (i.e., $G_{\text{mean}}$) for the channel seen by each code, which may result in transmission failure. If water-filling is performed exactly according to the available CSIT, the outage probability will be dominant by the worst case performance (related to the worst performed one among the multiple codes transmitted in parallel). Intuitively, a better strategy is to code over all eigenmodes jointly. In this way, the overall performance will be roughly determined by the averaged channel condition. Techniques to implement this intuitive idea are presented in Section III.

Furthermore, the MC-WF scheme relies on the orthogonality of different eigenmodes. In the presence of CSIT error, such orthogonality cannot be maintained at the reciever. Different codes may interfere with each other, which may worsen the situation.

### E. Numerical Examples

Performances of the APG, EP, and MC-WF schemes in a MIMO OFDM system are illustrated in Fig. 1. We can see that the EP scheme has a good performance when $\alpha = 0$. However it cannot utilize the available CSIT. The MC-WF scheme is optimal in the case of perfect CSIT but it is sensitive to the CSIT error. The APG method achieves the same performance as the EP scheme and the MC-WF scheme respectively at $\alpha = 0$ and $\alpha = 1$. In general, it outperforms both EP and MC-WF when...
0 < \alpha < 1. This means that the APG method is a robust solution and can utilize the available CSIT efficiently.

For the MC-WF scheme, the impact of interference effect is difficult to model. (Joint detection may be required). In obtaining the MC-WF curve in Fig.1, we simply ignore this effect. The simulation setting is as follows. Assume that there are N eigenmodes at the transmitter and the i\text{th} eigenmode transmits code C_i with power p_i. At the receiver, each C_i is detected by ignoring the interference of all the other codes C_j, j \neq i (i.e., by artificially setting p_j = 0). This provides an optimistic bound of the MC-WF method. However, as seen in Fig.1, the MC-WF performance degrades severely even after ignoring the interference effect, which is caused by the channel fluctuation experienced by individual codes.

![Fig. 1. Performances of different transmission schemes with imperfect CSIT.

The throughput is 2 bits per channel use. N_t = 4, N_r = 2. L = 4, J = 32. For the MC-WF and APG schemes, the power is averaged with respect to the distribution of G.

III. PRECODING BASED ON ITERATIVE LMMSE DETECTION

A. Precoder Structure

Let \( c \) be a coded vector generated by a generic encoder (consisting of an FEC encoder followed by random interleaving). Without loss of generality, the average power per entry of \( c \) is normalized to 1. We will treat \( c \) as a sequence of i.i.d. symbols, which is approximately ensured by random interleaving [13] [14].

\[
\begin{align*}
\begin{bmatrix}
P & W & V
\end{bmatrix}
\end{align*}
\]

Fig. 2. Precoding Implementation.

We adopt the scheme illustrated in Fig. 2. The precoding is defined by a linear transform

\[
x = VWpc
\]

where \( x \) is the transmitted signal vector over all subcarriers in (1). The linear transform \( V, W, \) and \( P \) implement the operation of beamforming, water-filling, and diversity coding, respectively. Their functions are detailed as follows.

The matrix \( V \) is unitary. We will call each signal transmitted along a column of matrix \( V \) a beam and call \( V \) the beamforming matrix\(^1\). In the case of perfect CSIT, \( V \) can be obtained from (4). Denote the SVD of \( G \) in (1) by

\[
G = UDV^H.
\]

At the receiver, \( U \) is applied to rotate the received signal \( y \) (see Fig. 2). Then we have

\[
U^HV = U^HVDV^H = D.
\]

The channel gains of these beams are the corresponding singular values of channel matrix \( G \). Furthermore, since different beams are orthogonal to each other at the receiver, there is no interference between them.

For imperfect CSIT, the actual \( V \) is unknown at the transmitter. We simply replace \( V \) by \( \hat{V} \) in (11). Then (12b) becomes

\[
U^HG\hat{V} = U^HVDV^H = DV^H\hat{V}.
\]

With (12c), the orthogonality of different beams is not guaranteed at the receiver.

The digonal matrix \( W \) is designed to adjust the power levels on different beams. It can be obtained from the discussion in the ideal coding case. (See (10) in Section II.C). Such an option of \( W \) may not be optimal for practical coding. However, as will be shown in Fig.3, the related performance loss is marginal.

The matrix \( P \) is for diversity coding. In this paper, we select \( P \) to be a Hadamard matrix for the following reasons. First, Hadamard matrix is unitary and hence of full-rank. No capacity loss is caused by the use of \( P \). Second, for MIMO channels, different beams have different channel gains in general. The use of a Hadamard matrix can efficiently remove this fluctuation effect, which results in a diversity advantage. We will return to this issue in Section III.F. Third, fast Hadamard transform (FHT) [15] makes the calculation involving matrix \( P \) much simpler, which is discussed in Section III.C.

B. Detection Principles

The use of the precoder in (11) causes inter-symbol interference in general. We can apply iterative LMMSE detection [16] to suppress this interference. A key in the discussion below is to take into account the feedback information from the decoder. This brings about significant performance improvement.

Combining (1), (11) and the rotation \( U^H \) at the receiver, we have

\[
y' = A c + U^H \eta
\]

(13a)

\[
A = U^H \tilde{G} V WP = D V^H \tilde{V} WP
\]

(13b)

The optimal estimation of \( c \) usually involves prohibitive complexity. We take a suboptimal alternative involving two local operators, namely, the elementary signal estimator (ESE) and the decoder (DEC). The coding constraint is ignored in the ESE and the impact of the channel is ignored in the DEC, which results in reduced complexity at the cost of certain performance loss. To compensate for the performance loss, an iterative procedure is applied to the two local operators, following the renowned turbo principle.

\[1\] Let \{D(i,j)\} be the diagonal entries of matrix \( D \) (i.e., the singular values of matrix \( G \)). In the special case of multiple-input single-output (MISO) channel, \( G \) reduces to a vector \( g \). Then only one of \{D(i,j)\}, say \( D(1,1) \) with corresponding right-singular vector \( v_1 \), is non-zero. If the MISO channel has a single line-of-sight, \( v_1 \) determines the actual angle of incidence. This is why we call \( V \) the beamforming matrix and the columns of \( V \) the directions of beams.
For the ESE, even after ignoring the coding constraint, the optimal detection can still be difficult task for a discrete $c$. To overcome the difficulty, we follow the two-step approach first proposed in [13]. In the first step, we estimate $c$ as approximately Gaussian distributed. We assume that the \textit{a priori} mean (denoted by $E[c]$) and \textit{a priori} auto-covariance of $c$ are available. To simplify the problem, we further assume that the \textit{a priori} auto-correlation of $c$ can be written as $\nu I$. The \textit{a priori} information is initialized to $E[c] = 0$ and $\nu = 1$, and updated using the feedbacks from the DEC during the iterative process. (We will return to this later in III.E.) Then the standard LMMSE estimator of $c$ is given by [17]
$$
\hat{c} = E[c] + \nu A^H \tilde{R}^{-1}(y' - AE[c])
$$
(14a)
where $R = \text{Cov}(y', y') = \nu A^H \tilde{R}^H + \sigma^2 I$. In the second step, we write the $i$th entry of $\hat{c}$ in the following form
$$
(\hat{c}(i) = \phi(i)c(i) + \tilde{\xi}(i)).
$$
(14b)
In the above,
$$
\phi(i) = v \tilde{Q}(i, i),
$$
(15a)
$$
\tilde{\xi}(i) = E[c(i)] + \nu v^{(i)}(i)^{R}(y' - AE[c]) - v \tilde{Q}(i, i) \phi(i)
$$
(15b)
where $\phi(i)$ is the $i$th row of $\phi$, and $\tilde{Q}(i, i)$ is the $i$th diagonal entry in the following matrix
$$
\tilde{Q} = (A^H \tilde{R}^{-1} A)_{\text{diag}}.
$$
(15c)
In (15b), $\tilde{\xi}(i)$ represents an unknown interference-plus-noise term that is independent of $c(i)$. We treat $\tilde{\xi}(i)$ as an AWGN sample, with the variance given by [18, Eqn. (18)]
$$
\text{Var}(\tilde{\xi}(i)) = \phi(i)(1 - \phi(i))\nu.
$$
(15d)
Then we can detect $c(i)$ based on (14b) in a symbol-by-symbol manner.

\textbf{C. Complexity}

We now briefly discuss the complexity of the above approach. Substituting $A = DV^{1/2} \tilde{V}$ into (14) and noting that $P$ is unitary, we have
$$
R = \nu D V^{1/2} \tilde{V} W A^H \tilde{V}^H V^{1/2} \nu I + \sigma^2 I
$$
(16)
From the discussion above (2), $G$ is block-diagonal. All the matrices in (16) are also block-diagonal, and so $R^{-1}$ can be computed in a block-by-block manner with complexity $O(N(R^3)N(N, N))$ [19].

Furthermore, let $\tilde{Q} = D \tilde{V}^{1/2} \tilde{V} W$ and so the equivalent channel $A$ can be expressed as $A = QP$. When $P$ is a Hadamard matrix, following a similar approach as in [20] (cf., Appendix B in [20]), it can be shown that (15c) can be approximated as
$$
\tilde{Q} \approx \omega I
$$
(17a)
where
$$
\omega = \frac{1}{L} \int_{-\nu}^{\nu} \frac{1}{\nu} \frac{d\xi}{|\xi|^2 + \sigma^2} p(\xi)
$$
(17b)
with $p(\xi)$ being the empirical density function of the singular values of $Q$. We omit the detailed derivation due to space limitation.

All the other operations in (14) and (15) involving $A = D \tilde{V}^{1/2} \tilde{V} WP$ can be efficiently implemented using FHT or in a block-by-block manner. Thus the complexity involved in (14) and (15) is $O(N(N, N)N \log (N))$, which is modest from a practical point of view.

\textbf{D. SINR-Variance Transfer Function}

The performance of the ESE can be characterized using the output SINR. The output SINR, denoted by $\rho(j)$, is defined as the ratio of the signal power to the interference-plus-noise power in the ESE output (14b). Based on (15a) and (15d), we have
$$
\rho(j) = \frac{\nu\phi(j)^2}{\text{Var}(\tilde{\xi}(j))} = \frac{\Omega(j, j)}{1 - \nu \Omega(j, j)}.
$$
(18a)
From (17a), all $\{\phi(j)\}$ are equal and so they can be given by the same constant $\rho$ defined as
$$
\rho = \omega(1 - \nu \rho).
$$
(18b)
Define $\phi(\nu) = \omega(1 - \nu \rho)$. The behavior of the ESE can then be characterized by a univariant transfer function $\rho = \phi(\nu)$.

\textbf{E. Iterative Process}

Recall that the \textit{a priori} information is assumed in III.B. Reliable \textit{a priori} information is crucial to the proposed scheme. Now we consider the following iterative process to refine the \textit{a priori} information. We forward the ESE output (i.e., $\hat{c}$ in (14a)) to the DEC. After decoding, the DEC produces updated $E(c)$ and $\nu$. (The details on computing $E(c)$ and $\nu$ can be found, e.g., in [16].) The ESE is then executed again for refining the system performance. This process continues iteratively.

We can view $E(c)$ as an estimate of $c$ and $\nu$ as the reliability of $E[c]$. Also, $\rho$ in (18) can be viewed as a reliability measure for the input to the DEC. Thus we can characterize the DEC operation by a transfer function $\nu = \varphi(\rho)$. With $\varphi(\nu)$ and $\varphi(\rho)$ available, we can predict the performance of the iterative process using a recursion of $\rho$ and $\nu$. For the details, see [16].

\textbf{F. Matrix $P$ Again}

Recall that $\rho$ is a reliability measure for the input to the DEC. From (18), we can see that all the symbols of sequence $c$ have the same reliability in each iteration. This is the consequence of the use of Hadamard matrix in (17). Intuitively, with a Hadamard transform, each symbol in sequence $c$ is spread uniformly (in the sense of energy) over all directions. Consequently, every symbol undergoes the same channel condition. This explains the diversity advantage related to the use of $P$.

\textbf{IV. Numerical Results}

We now present some numerical examples to demonstrate the efficiency of the proposed scheme. We assume quasi-static Rayleigh-fading MIMO ISI channels with $N_t$ transmit antennas, $N_r$ receive antennas, and $L$ delay taps. The total gain over the $L$ paths between each transmit-receive antenna pair is normalized to 1. The channel remains constant during the transmission of each coded frame, but changes independently from frame to frame. The FEC code is a rate 1/2 convolutional code with generator polynomials $(7, 5)\nu$ followed by random interleaving and QPSK modulation. The frame length is set to 4096 information bits.

2 In contrast to a conventional spreading system, $Pc$ has the same length as $c$ since $P$ is square. There is no rate loss here. The proposed scheme is applicable to high rate transmission, as seen in Figs. 3 and 4.
A. Perfect CSIT

Fig. 3 shows the BER performance of the proposed scheme in systems with perfect CSIT. Performances of both the optimized \( W \) (obtained by searching based on the evolution analysis for practical coding [20]) and approximate \( W \) (obtained by water-filling under the assumption of ideal coding) are included. The channel capacities are included for reference. There is a gap of about 4 dB between the simulated performance and its corresponding capacity. This gap is due to the use of the convolutional code. (The convolutional code used in the simulation has a gap of about 6 dB from capacity in an AWGN channel.)

About 3 dB power gain is obtained when the dimension of the MIMO channels increases from 2×2 to 4×4, which is expected from the capacity analysis [1][12]. It is simple to construct \( W \) by assuming ideal coding. Though suboptimal, it provides reasonably good performance, as seen in Fig. 3. Further optimization based on practical coding only gives marginal performance gain.

![Fig. 3. Performances of the proposed scheme in the MIMO ISI channels with various \( N_t \) and \( N_r \) and \( L = 128 \). The throughput is 2 bits per channel use for the 2×2 MIMO system, and 4 bits per channel use for the 4×4 MIMO system. An outage probability of \( \epsilon = 0.01 \) is allowed.](image)

B. Imperfect CSIT

Fig. 4 shows the simulated FER performance of the proposed method as the CSIT quality varies. We can see that the system performance improves progressively as the CSIT quality improves. Compared with no CSIT, a considerable performance gain of about 3.7 dB (at FER = 10^-3) is obtained in the case of perfect CSIT. The performance of ideal coding (See Section II.C) is also included in Fig. 4 for reference.

V. CONCLUSIONS

In this paper, a robust APG method is proposed for MIMO OFDM systems with imperfect CSIT. A unified beamforming, water-filling, and diversity coding scheme is developed to implement the APG method in practically coded systems. Both analytical and simulation results demonstrate that the proposed method can achieve significant performance gains by efficiently utilizing the available CSIT.

REFERENCES


