Transmitter Design for Uplink MIMO Systems With Antenna Correlation

Chongbin Xu, Peng Wang, Member, IEEE, Zhonghao Zhang, and Li Ping, Fellow, IEEE

Abstract—We study the uplink transmission in multiple-input multiple-output (MIMO) systems with antenna correlation. We focus on schemes that require only channel covariance information at the transmitter (CCIT), which involves lower cost than full channel state information at the transmitter (CSIT). We start from mutual information analysis and show that a simple CCIT-based scheme, referred to as statistical water-filling (SWF), can perform close to the optimal full CSIT-based one in MIMO systems with more receive antennas than transmit ones. We then focus on the implementation of SWF in practically coded systems. An iterative linear minimum mean squared error (LMMSE) receiver is assumed and an extrinsic information transfer (EXIT) chart type curve matching technique is developed based on Hadamard precoding techniques. Simulation results show that the proposed scheme can obtain significant performance improvement compared to the conventional equal power transmission. Finally, we show that the proposed scheme is also very efficient in multi-user uplink MIMO systems with distributed channel information.

Index Terms—Uplink MIMO, antenna correlation, channel covariance information.

I. INTRODUCTION

RECENTLY, massive multiple-input multiple-output (MIMO) systems [1]–[3] have attracted extensive research attention. Let us focus on the uplink. Denote by \( N_{\text{BS}} \) and \( N_{\text{MT}} \), respectively, the numbers of antennas at the base station (BS) and at the mobile terminal (MT). The basic assumptions of massive MIMO are \( N_{\text{BS}} \rightarrow \infty \) and \( N_{\text{BS}} \gg N_{\text{MT}} \). These assumptions lead to the so-called favorable propagation effect. Then equal power and rate allocation over all antennas at an MT is nearly optimal provided that these antennas are uncorrelated [2], [3]. This does not require channel state information (CSI) at the transmitter (CSIT) and so avoids the cost related to CSIT acquisition. It is also known that simple maximum ratio combining (MRC) is nearly optimal for massive MIMO [1]–[3]. Therefore, transmission and detection can be greatly simplified when MIMO is “massive”.

However, in practice, the antennas on a MT are typically correlated due to the limited physical size. It is also reasonable to expect that the MIMO size will become large but may not be massive in the near future due to technical difficulties. There are works considering antenna correlation or “not so large” MIMO, but they are mostly on mutual information analysis [4]–[6]. In this paper we focus on practical transmitter design for such systems.

In general, CSIT can be exploited to improve MIMO performance [7]–[18]. However, CSIT acquisition becomes costly when \( N_{\text{BS}} \times N_{\text{MT}} \), the number of channel coefficients, is even just modestly large. A conventional low-cost option is to utilize channel covariance information (CCI) at the transmitter (CCIT) [7], [8], since CCI typically changes more slowly than CSI itself, implying that CCIT requires less updates and so lower cost than CSI.

Mutual information analysis for CCIT channels has been extensively studied [19]–[30]. Optimization techniques have also been developed for precoder design for uncoded systems with CCIT [31]–[34]. Coded systems, however, represent a difficulty. Joint forward-error-control (FEC) coding, adaptive modulation and linear precoding are generally required in MIMO systems to realize the performance promised by capacity analysis. Iterative detection can be applied to improve the performance of such joint schemes. The extrinsic information transfer (EXIT) chart technique [35], [36] is a common design tool for systems with iterative detection. However, analysis for iterative system is highly complicated under CCIT, especially when MIMO size is large. This makes it difficult to generate the EXIT chart functions analytically. Monte Carlo method can be used instead, but it is costly and so not suitable for real time applications.

Detection complexity is another problem. In a MIMO channel with antenna correlation, information rates should be carefully allocated to different channel eigen-directions. Adaptive modulation is a standard approach for this purpose [37], [38]. Denote the number of bits carried by the symbol on the \( n \)-th eigen-direction by \( Q_n \), and its average over all eigen-directions by \( \bar{Q} \). In general, \( Q_n \) varies in the range of \([0,N_{MT} \bar{Q}]\), which increases with the MIMO size. Then the demodulation complexity \( O(2^{Q_n}) \) for some large \( Q_n \) values can be a problem even if the MIMO size is only modestly large.

This paper is concerned with uplink MIMO systems with antenna correlation at the MT side. We start with a statistical water-filling (SWF) technique for channels with CCIT. SWF was mentioned briefly in [23], [29], only as an alternative...
We show that SWF using simple maximum ratio combing (MRC) is actually nearly optimal when $N_{BS}$ is sufficiently large, but it performs poorly when $N_{BS}$ is “not-so-large”. This motivates the joint FEC coding and linear precoding scheme developed below. The main contributions of this paper are as follows.

- We develop a fast optimization technique for joint FEC encoding and linear precoding in MIMO systems with CCIT. This technique is efficient for three reasons. First, simple SWF is used to facilitate precoder design. Second, based on Hadamard precoding [11], [39]–[41], a closed form expression of the EXIT chart function is derived for iterative detection in a channel with CCIT and a sufficiently large $N_{BS}$. It facilitates fast EXIT chart curve matching for system optimization. Excellent performance is demonstrated by simulation results even if $N_{BS}$ is “not-so-large”. Third, the proposed scheme employs a uniform modulation constellation over all signal streams, but can still realize the same water-filling effect of adaptive modulation. This greatly eases the demodulation complexity problem mentioned earlier.

- We extend the discussions to uplink systems in which multiple users transmit in the same time slot and sub-carrier simultaneously. To achieve optimal performance, the transmitters of all users should be designed jointly with global CSIT, which is very costly. We present a very simple solution in which each transmitter is optimized based on its own correlation matrix. We show that this simple scheme provides a low-cost but nearly optimal solution when the ratio of $N_{BS}/(KN_{MT})$ is only moderately large.

The joint FEC coding and linear precoding scheme proposed in this paper represents a good trade-off between cost and performance for modestly large MIMO systems, as confirmed by both analysis and simulation. This property is attractive for practical applications.

II. TRANSMISSION STRATEGY WITH CCIT

A. System Model

Consider an uplink MIMO system. For simplicity, we first assume that the transmissions of multiple users are orthogonal. (We will discuss non-orthogonal multi-user concurrent transmissions in Section V.) Then we can write the received signal as

$$r = H y + \eta$$  \hspace{1cm} (1)

where $r$ is an $N_{BS} \times 1$ signal vector received at BS antennas, $H$ an $N_{BS} \times N_{MT}$ channel transfer matrix, $y$ an $N_{MT} \times 1$ signal vector transmitted from MT antennas, and $\eta$ an $N_{BS} \times 1$ vector of complex additive white Gaussian noise (AWGN) with mean 0 and variance $\sigma^2$. We will assume that $y$ has a zero mean $E[y] = 0$ and a total power constraint $P_t$, i.e.,

$$\text{tr}\{Q\} \leq P_t$$  \hspace{1cm} (2)

where $Q = E[yy^H]$ is the transmission covariance matrix and tr[·] denotes the trace operation. Throughout this paper, we will always assume $E[H] = 0$.

B. Kronecker Model

The discussions in the rest of this paper are based on the Kronecker model that has been widely discussed and validated under certain practical channel conditions [7], [8], [19]–[30]. This model is given by [19], [30]

$$H = C_{BS}^{1/2}H_0C_{MT}^{1/2}$$  \hspace{1cm} (3)

where $H_0$ is an $N_{BS} \times N_{MT}$ matrix whose elements are independent and identically distributed (i.i.d.) complex Gaussian random variables with distribution $CN(0,1)$, and $C_{BS}$ and $C_{MT}$ are, respectively, $N_{BS} \times N_{BS}$ and $N_{MT} \times N_{MT}$ Hermitian matrices characterizing the antenna correlation at the receiver and transmitter. We adopt the following normalization

$$\text{tr}\{C_{BS}\} = N_{BS} \text{ and} \text{ tr}\{C_{MT}\} = N_{MT}.$$  \hspace{1cm} (4)

In this paper, we will always assume that the receiver knows full channel state information (given by $H$) via proper channel estimation. We will also assume that the transmitter only knows channel covariance information (given by $C_{BS}$ and $C_{MT}$), which leads to significantly low cost. We will return to this in Section II-E.

C. Statistical Water-Filling (SWF)

Denote by

$$R(H, Q) = \log_2 \det(I_{N_{BS}} + HQH^H/\sigma^2)$$  \hspace{1cm} (5)

the achievable rate of the system in (1) with channel $H$ and transmission covariance matrix $Q$. With full CSIT, the channel ergodic capacity under the power constraint (2) is given by

$$C_{\text{CSIT}} = E[H] \left( \max_{\text{tr}\{Q\} \leq P_t} R(H, Q(H)) \right)$$  \hspace{1cm} (6)

where $Q(H)$ represents that the transmission covariance matrix $Q$ is optimized based on $H$ and the expectation is over the distribution of $H$ conditioned on the known $C_{BS}$ and $C_{MT}$.

Under CCIT assumption, $Q$ is independent of particular channel realization. The related ergodic capacity is given by

$$C_{\text{CCIT}} = \max_{\text{tr}\{Q\} \leq P_t} E[H][R(H, Q)].$$  \hspace{1cm} (7)

The key here is to find $Q$ that maximizes the average mutual information $E[H|R(H, Q)]$. In general, this is a highly complicated problem.

The following is a sub-optimal solution. Let the eigenvalue decompositions of $Q$ and $C_{MT}$ be

$$Q = U_Q D_Q U_Q^H,$$  \hspace{1cm} (8)

$$C_{MT} = U_{MT} D_{MT} U_{MT}^H,$$  \hspace{1cm} (9)
It is known that the optimal $U_Q$ in (8) under the CCIT assumption is [26]–[28]

$$U_Q = U_{MT}. \quad (10)$$

Following [23], [29], we generate a suboptimal $D_Q$ by maximizing the Jensen bound of the average mutual information $E_H[\mathbb{R}(H, Q)]$. Note that

$$\det(I_{NBS} + HQH^H/\sigma^2) = \det(I_{NMT} + QH^H H/\sigma^2),$$

$$E_H[H^H H] = N_{BS} C_{MT}.$$ 

Then we solve the following problem:

$$\max_{D_{Q}} \log_2 \det(I_{NMT} + N_{BS} D_Q D_{MT}/\sigma^2)$$

subject to $\text{tr}(D_Q) \leq P_t$. \quad (11a)

Solving (11) is straightforward; it is equivalent to water-filling over a parallel channel with channel gains given by the diagonal elements of $N_{BS} D_{MT}$. This method is referred to as statistical water-filling (SWF) in [23]. The resultant $Q$, denoted by $Q_{SWF}$, serves as a low-cost and sub-optimal solution to (7). In the sequel, we will assess the performance of SWF and develop its implementation techniques.

**D. Linear Precoding**

The SWF scheme can be realized using a precoding matrix

$$P = U_Q D_Q^{1/2} \mathbf{V} \quad (12)$$

where $U_Q$ and $D_Q$ are defined in (8) and $\mathbf{V}$ is an arbitrary $N_{MT} \times N_{MT}$ unitary matrix that does not affect capacity. (However, $\mathbf{V}$ may affect the performance of a practical system. See Section IV.) Assume that $x$ is a length $N_{MT} \times 1$ coded sequence with elements independently drawn from $CN(0, 1)$ and $y = Px$. Then the required $Q$ is realized as

$$Q = E[yy^H] = PP^H. \quad (13)$$

Substituting $y = Px$ into (1), we have

$$r = HPx + \eta.$$ \quad (14)

In general, $HP$ is not diagonal and so there is inter-stream interference. This is inevitable since the transmitter does not know $H$ perfectly. We will discuss the related issues in detail in Section IV.

**E. Cost Related to Estimating $C_{MT}$**

With SWF, $P$ in (12) can be generated solely based on $C_{MT}$. In a time varying channel, $C_{MT}$ typically changes more slowly than $H$, so it requires less updates. With frequency division duplex (FDD), this clearly reduces the cost related to obtaining $C_{MT}$.

With time division duplex (TDD), $C_{MT}$ can be obtained by sending pilot signals periodically from the BS. As $C_{MT}$ typically changes slowly, the related overhead is relatively low (compared with estimating the full channel matrix $H$ involving fast-changing Rayleigh fading).

Note that an MT needs to estimate channel for data detection but this does not, even with TDD, necessarily provide the full information about $H$. To see this, assume that a linear precoder is used in the downlink. The received signal at an MT can be written as

$$r_{DL} = H_{DL} P_{DL} x_{DL} + \eta_{DL}$$ \quad (15)

where $H_{DL} = H^H$ assuming channel reciprocity, $P_{DL}, P_{DL} x_{DL}$, and $\eta_{DL}$ are defined similarly to their counterparts in the uplink model (14). To detect $x_{DL}$, it suffices to estimate $H_{DL} P_{DL}$ instead of $H_{DL}$. We cannot directly obtain $H_{DL}$ from $H_{DL} P_{DL}$ if $H_{DL} P_{DL}$ has a smaller size than $H_{DL}$ (the number of columns of $P_{DL}$ is given by the number of signal streams in $x_{DL}$). The latter is limited by $\min(N_{MT}, N_{BS})$ that is typically less than the number of columns of $H_{DL}$ (i.e., $N_{BS}$) when $N_{MT} < N_{BS}$.

**III. Mutual Information Analysis**

The low-cost SWF scheme outlined in Section II is suboptimal. In this section, we will show by mutual information analysis that the potential performance loss is actually marginal for the uplink with $N_{BS} > N_{MT}$.

**A. Asymptotic Analysis**

In this subsection, we consider the asymptotic situation of $N_{BS} \to \infty$. We will assume that $C_{BS}$ does not contain dominant eigenvalues when $N_{BS} \to \infty$, i.e.,

$$\lim_{N_{BS} \to \infty} \lambda_n(C_{BS})/\text{tr}(C_{BS}) = 0, \forall n \quad (16)$$

where $\lambda_n(C_{BS})$ is the $n$th eigenvalue of matrix $C_{BS}$. It can be verified that the above assumption holds provided that there is sufficient spacing among BS antennas [42], [43].

Denote by $C_{FCSIT}$ the capacity of the system in (1) with full CSIT (see (6)). Also denote by $R_{SWF}$ the average achievable rate of SWF:

$$R_{SWF} = E[H \mathbb{R}(H, Q_{SWF})] \quad (17)$$

with $Q_{SWF}$ given by (8), (10), and (11). Clearly, $C_{FCSIT} \geq R_{SWF}$.

**Proposition 1:** The SWF scheme is asymptotically optimal, i.e.,

$$\lim_{N_{BS} \to \infty} |C_{FCSIT} - R_{SWF}| = 0, \quad (18)$$

if

(i) $N_{BS} \to \infty$ with $N_{MT}$ fixed and $C_{BS}$ meets (16); or

(ii) $N_{BS} \to \infty$ with $N_{MT}^3/N_{BS} \to 0$, and $C_{BS} = I$.

**Proof:** This can be proved using large random matrix analysis [44], [45]. For details, see Appendix A.

Proposition 1 reveals that the full CSIT capacity can be approached by simple SWF under the given two conditions. Recall that SWF only requires the knowledge of $C_{MT}$ at the transmitter. Full CSIT provides no further performance improvement. This indicates that SWF is an attractive option for
uplink massive MIMO systems. For practical considerations, it is also interesting to examine the implication of Proposition 1 in systems with limited size. The numerical results below show that SWF is still attractive in the latter case.

B. Numerical Results

We now present numerical results to verify the efficiency of the SWF scheme. For simplicity, we adopt the exponential model [45], [46] to characterize the antenna correlation, i.e.,

\[
C_{BS}(m,n) = \rho^{|m-n|} e^{i(m-n)\theta_{BS}} \quad (19a)
\]

\[
C_{MT}(m,n) = \rho^{|m-n|} e^{i(m-n)\theta_{MT}} \quad (19b)
\]

where \( i \triangleq \sqrt{-1}, \ \rho_{BS} \) and \( \rho_{MT} \in [0, 1] \) are respectively receive and transmit correlation factors, and \( \theta_{BS} \) and \( \theta_{MT} \) are uniformly distributed over \([0, 2\pi)\).

Fig. 1 compares the SWF performance with the full CSIT (FCSIT) capacity. The latter serves as a performance upper bound. We can see that SWF is nearly optimal even for modestly large \( N_{BS} \). This verifies the asymptotic analysis in Proposition 1. The gap between the two schemes reduces when the transmit correlation factor \( \rho_{MT} \) increases. Intuitively, this is because a larger \( \rho_{MT} \) implies a lower degree of freedom, which has similar effect as a smaller \( N_{MT} \) (and so a larger \( N_{BS}/N_{MT} \) ratio).

Fig. 2 provides the performance when \( N_{MT} \) varies while the ratio \( N_{BS}/N_{MT} \) is fixed. For reference, we also include the performance of systems of no precoding (NP), in which we select \( Q = P_{t}/N_{MT}I \). We can see that SWF can obtain performance close to the FCSIT upper bound. Its gain over the system of no precoding is noticeable, especially when \( N_{MT} \) is large. This indicates the necessity to introduce proper precoding in the uplink when MTs are equipped with multiple antennas. Comparing Fig. 2(a) and 2(b), we can see that the advantage of SWF over NP increases when the transmit correlation factor \( \rho_{MT} \) increases. This is expected, since the related gain is 0 in the extreme case of \( \rho_{MT} = 0 \) (and so \( C_{MT} = I \)) where equal power allocation is optimal.

Condition (ii) in Proposition 1 requires \( N_{MT}^3/N_{BS} \to 0 \), which is a relatively strong requirement. In practice, this may not hold and then the terms \( N_{MT} \log_2(1 + 2\sqrt{N_{MT}/N_{BS}}) \) in (46) and (47) (see Appendix A) is not negligible. This effect can be seen in Fig. 2: the gap between the FCSIT bound and SWF performance increases slightly as \( N_{MT} \) increases.

IV. TRANSMITTER AND RECEIVER DESIGN

We now discuss the implementation of SWF. The mutual information analysis in Section III shows that SWF can obtain a good performance even if \( N_{BS} \) is “not so large”. However, careful system design is still required for practical systems. In this section, we will show the following.
A simple maximum ratio combining (MRC) receiver performs well if \( N_{BS} \rightarrow \infty \), which is consistent with mutual information analysis. However, it may perform poorly for a “not so large” \( N_{BS} \) value due to the residual interference problem when the channel is not perfectly diagonalized.

Iterative detection can be used to suppress interference and provides a possible way to implement the potential benefit of SWF for a modest large \( N_{BS} \) value. The transmitter can be optimized by a curve matching technique carefully for “not so large” MIMO.

B. Hadamard Precoding

In what follows, we develop an improved precoding scheme based on iterative detection. First, let \( V = V_{\text{Hadamard}} \) in (12), where \( V_{\text{Hadamard}} \) is a Hadamard matrix with a proper size. (See Fig. 4.) For convenience, we will call

\[
P = U_Q \Phi^{1/2} V_{\text{Hadamard}}
\] (20)

a Hadamard precoder. Similar precoder structures have been previously discussed for channels without CSIT in [39] and for channels with full CSIT in [11], [40]. In this paper, our focus is its application in SWF. As we will see below, the use of \( V_{\text{Hadamard}} \) in (20) facilitates the EXIT chart type curve matching for SWF.

C. Iterative LMMSE Detection

Substituting (20) into (14), the received signal is given by (see Fig. 4)

\[
r = H U_Q \Phi^{1/2} V_{\text{Hadamard}} x + \eta.
\] (21)

Applying the standard LMMSE detection to \( r \) in (21), we have [47]

\[
\dot{x} = E[x] + v(\tilde{H} V_{\text{Hadamard}})^H r^{-1} (r - \tilde{H} V_{\text{Hadamard}} E[x])
\] (22a)

where \( \tilde{H} = H U_Q \Phi^{1/2} \), \( E[x] \) is the \textit{a priori} mean of \( x \), \( vI \) is the \textit{a priori} covariance of \( x \) and

\[
R \Delta \text{cov}(r, r^H) = v(\tilde{H} V_{\text{Hadamard}})(\tilde{H} V_{\text{Hadamard}})^H + \sigma^2 I.
\] (22b)

Initially, if we do not have any information on \( x \), we set \( E[x] = 0 \) and \( vI = I \) (assuming that the symbols in \( x \) have average power of 1). We rewrite (22a) into a symbol-wise form as

\[
\hat{x}(i) = v \Omega(i, i) x(i) + \xi(i)
\] (23a)

where \( \Omega(i, i) \) is the \( i \)-th diagonal element of the following matrix

\[
\Omega \triangleq ((\tilde{H} V_{\text{Hadamard}})^H R^{-1} (\tilde{H} V_{\text{Hadamard}})_{\text{diag}}
\] (23b)

and \( \xi(i) \) is an interference-plus-noise term. Following the treatments in [48], we approximate \( \xi(i) \) by an AWGN sample with variance (c.f. (18) in [49])

\[
\text{Var}(\xi(i)) = v \Omega(i, i)(1 - v \Omega(i, i)) v.
\] (23c)

We can then estimate \( x(i) \) using the symbol-wise model in (23a).

The estimate of \( x \) is then processed by a forward-error-control (FEC) decoder (denoted by DEC in Fig. 4) using \textit{a posteriori} probability (APP) decoding. The decoding output can be used to update the values of \( E[x] \) and \( v \) for the sequence \( x \). Then LMMSE detection can be performed again. This process continues iteratively.
D. Demodulation Complexity

We now briefly discuss the complexity issue. Let x be divided into sections; each section contains \( N_{\text{MT}} \) symbols transmitted over each time use of the channel. Assume that these \( N_{\text{MT}} \) symbols carry a total of \( N_{\text{MT}} \times \bar{Q} \) bits. For simplicity, let \( \bar{Q} \) be an integer.

For the Hadamard precoding scheme in (21), each symbol in x carries \( \bar{Q} \) bits based on a constellation of size \( 2^\bar{Q} \). The demodulation complexity involved in (23) is fixed at \( O(2^\bar{Q}) \).

As a comparison, consider an alternative method based on adaptive modulation, in which \( N_{\text{MT}}\bar{Q} \) bits are allocated among \( N_{\text{MT}} \) symbols in a section of x. Denote by \( Q_n \) the number of bits carried by the symbol on the \( n \)th eigen-direction, which can vary in the range of \([0,N_{\text{MT}}\bar{Q}]\). Note that \( N_{\text{MT}}\bar{Q} \) can be quite large for a relatively large MIMO. The demodulation complexity \( O(2^{Q_n}) \) for some large \( Q_n \) values then becomes a serious problem.

The key here is that the worst-case complexity can be very high if variable constellation sizes are involved. The precoding scheme in (21) employs a uniform constellation size for all symbols, which avoids the problem. Note that here the Hadamard precoding scheme is still adaptive to channel condition, which is achieved by controlling the power allocation matrix \( D_0 \) in (21).

E. Transfer Function for LMMSE Detection

Note that in the symbol-wise demodulation in (23), the interference among different symbols in x is treated as additive noise (included in \( \xi(i) \)) for simplicity. The residual interference may still affect performance noticeably if not treated properly. In the following subsections, we outline an optimization procedure to minimize the interference effect, which is equivalent to maximize the signal to interference-plus-noise ratio (SINR), during the iterative process.

Recall that \( \mathbf{Q} \) in (23) characterizes the SINR achieved by LMMSE estimation. In general, such SINR can fluctuate for different symbols with different index \( i \). We can smoothen out such fluctuation by a transmission technique involving an augmented channel model. This is detailed in Appendix B where the following approximation is derived:

\[
\mathbf{Q} \approx \omega \mathbf{I} \tag{24}
\]

with \( \omega = N_{\text{MT}}^{-1} \text{tr} \{ \mathbf{D}_0 \} \). The use of the Hadamard matrix \( V_{\text{Had}} \) in (21) is the key to this approximation. Then, using (24) and (23c), we can compute the SINR in \( \hat{x}(i) \) in (23a) as

\[
\rho(i) = \left| \mathbf{Q}(i,i) \right|^2 / \text{Var} \{ \xi(i) \} = \omega / (1 - \nu \omega), \forall i. \tag{25a}
\]

Then the LMMSE module can be characterized by a function

\[
\rho = \phi(v) \triangleq \omega / (1 - \nu \omega). \tag{25b}
\]

where

\[
\phi(v) = \frac{1}{N_{\text{MT}}} \sum_n D_{\text{MT}}(n,n)D_Q(n,n) + \sigma^2/N_{\text{BS}} \tag{26b}
\]

if the two conditions in Proposition 1 hold.

Proof: See Appendix C.

Proposition 2 shows that \( \phi(v) \) is asymptotically determined by \( \mathbf{C}_{\text{MT}} \), since both \( D_{\text{MT}} \) and \( D_Q \) are functions of \( \mathbf{C}_{\text{MT}} \). (See (9) and (11) in Section II.) This is consistent with Proposition 1 that the system performance is determined by \( \mathbf{C}_{\text{MT}} \), except here a practical iterative LMMSE receiver is considered.

Recall from Section IV-A that MRC performance requires a relatively large \( N_{\text{BS}} \) value to converge. On the contrary, the convergence speed for LMMSE detection related to Proposition 2 is much faster. This is illustrated in Fig. 5 by a numerical example. The \( \phi(v) \) curves for the actual channel \( \mathbf{H} \) and the artificial parallel channel \( N_{\text{BS}}D_{\text{MT}} \) are provided. Noting the term \( \sigma^2/N_{\text{BS}} \) in the denominator in (26b), we set \( E_b/N_0 = -10 \log_{10}(N_{\text{BS}}) \) in Fig. 5 for different \( N_{\text{BS}} \) values for a fair comparison. We can see that the curves are quite close for all the \( N_{\text{BS}} \) values considered. This property is crucial for our discussions in the following subsections.

F. Transmitter Optimization

The performance of the FEC decoder is determined by its input SINR (denoted by \( \rho \)) and the quality of its output can be measured by the variance (denoted by \( \nu \)). This can be characterized by a function \( \nu = \psi(\rho) \) that can be produced by pre-simulation [50]. This function is the counterpart of \( \rho = \phi(v) \) discussed earlier.
Similar to [35], [36], we can show that the performance of the iterative receiver in Fig. 4 is characterized by a fixed point of the two transfer functions $\rho = \phi(v)$ and $\psi(\rho)$. Given $\rho = \phi(v)$, we can carefully design an irregular LDPC code such that $\psi(\rho)$ matches $\phi(v)$. This EXIT chart technique has been discussed in [51], [52] and details will be omitted.

The above design strategy requires knowing $C_{\text{MT}}$, but not $H$, at the transmitter. This leads to a key conclusion of this paper that for an uplink MIMO system with $N_{\text{BS}} > N_{\text{MT}}$, an efficient transmitter can be designed without the explicit knowledge of the complete channel coefficients. Knowing $C_{\text{MT}}$ is sufficient for this purpose. This fact greatly relieves the burden of channel state information acquiring at the transmitter.

### G. Variable and Fixed FEC Code Structures

We now consider two situations that may arrive in practice. We first assume FEC code optimization is allowed. In this case, the SWF based design can be used as initial values for the precoder (see Section II-C for details) that provides $\phi(v)$ and $\psi(\rho)$ matches $\phi(v)$. This EXIT chart technique has been discussed in [51], [52] and details will be omitted.

The above design strategy requires knowing $C_{\text{MT}}$, but not $H$, at the transmitter. This leads to a key conclusion of this paper that for an uplink MIMO system with $N_{\text{BS}} > N_{\text{MT}}$, an efficient transmitter can be designed without the explicit knowledge of the complete channel coefficients. Knowing $C_{\text{MT}}$ is sufficient for this purpose. This fact greatly relieves the burden of channel state information acquiring at the transmitter.

In the above, when $C_{\text{MT}}$ changes, a different LDPC code is optimized. The transmitter needs to inform the receiver about the structure of this code. This incurs considerable complexity overhead. Alternatively, we may fix the FEC code and optimize the precoder. We assume the receiver can acquire the overall equivalent channel including the physical channel and the new precoder through proper channel estimation. See (15) and the related discussions. This avoids the overhead mentioned above. An example for such situation can be found in Fig. 8.

### H. Numerical Results

We now demonstrate the effectiveness of transmitter optimization using numerical results in Fig. 6. For reference, we consider equal power (EP) allocation (in which the precoder $P = (P/N_{\text{MT}})^{1/2}V_{\text{Had}}$) and a regular LDPC code for the case without CSIT. Specifically, in Fig. 6(a), a rate-0.25 $(3, 4)$ LDPC code with length 32768 is adopted, followed by QPSK modulation. In Fig. 6(a), we can observe an early crossing point between $\phi(v)$ for EP precoder and $\psi(\rho)$ for the $(3, 4)$ LDPC code. This leads to considerable performance deterioration, as will be shown in Fig. 7.

Note that EP in the above still involves a precoder $P = (P/N_{\text{MT}})^{1/2}V_{\text{Had}}$, which is different from NP (i.e., no precoding) in Fig. 2 in which $P = (P/N_{\text{MT}})^{1/2}I$. For mutual information analysis, EP and NP are equivalent since they lead to the same transmission covariance matrix. However, for a practically coded system, the use of a Hadamard matrix in EP can provide a better diversity gain as discussed in [39].

In Fig. 6(b), we first generate $\phi(v)$ using the SWF technique described in Section II. We then generate a rate-0.25 irregular LDPC code with a $\psi(\rho)$ function that best matches $\phi(v)$. (Variable and check nodes degree distributions are given, respectively, by $\lambda(x) = 0.570146x^4 + 0.047378x^2 + 0.193282x^7 + 0.035182x^8 + 0.154016x^{34}$ and $\rho(x) = x^3$.)

Note that we can further match the two functions by alternating between optimizing $\phi(v)$ for a given $\psi(\rho)$ and optimizing $\psi(\rho)$ for a given $\phi(v)$). However, we observed only limited gain in this way.

Fig. 7 shows the simulated performance for the design examples in Fig. 6. We can see that though the LDPC code is designed based on the asymptotic analysis, it works well in systems even for $N_{\text{BS}} = 8$. The proposed scheme obtains a significant improvement from the EP performance (about 5.0 dB for $N_{\text{BS}} = 64$ and 5.4 dB for $N_{\text{BS}} = 8$). The threshold given by the EXIT chart type analysis is less than 0.5 dB away from the FCSIT capacity limit given in Section III.

![Fig. 6. SINR-variance transfer curves. $N_{\text{MT}} = 8$, $N_{\text{BS}} = 64$, $\rho_{\text{MT}} = 0.8$, and $\rho_{\text{BS}} = 0.5$. (a) $E_b/N_0 = -14.6$ dB, (b) $E_b/N_0 = -20.4$ dB.](image)

![Fig. 7. Simulation performance of the proposed scheme. $N_{\text{MT}} = 8$, $\rho_{\text{MT}} = 0.8$, $\rho_{\text{BS}} = 0.5$, and rate = 4 bits/symbol.](image)
An alternative is to fix coding for all realizations of \( C_{MT} \) and employ adaptive precoding (by optimizing \( D_Q \) in (21)) for different \( C_{MT} \). In this way, the transmitter does not need to inform the receiver for the following reason. For data detection, the receiver can estimate \( HP \) (instead of \( H \)), where \( HP \) is the equivalent channel formed by the physical channel \( H \) and the precoder \( P \). This avoids the additional cost.

Fig. 8 shows an example of a fixed LDPC code and an adaptively optimized precoder, as discussed in Section IV-G. We assume that \( C_{MT} \) varies with \( \rho_{MT} \in [0.7, 0.9] \) with the code optimized at \( \rho_{MT} = 0.8 \). We can again see significant performance gain in this case.

V. EXTENSION TO MULTI-USER SYSTEMS

In a multi-user MIMO system, global CSIT can lead to significant performance gain [53]. However, this requires feeding back CSI for all users together with centralized optimization [54], [55]. It is very costly in terms of computation and feedback bandwidth. In this section, we consider an individual CCIT scheme, in which each user individually designs transmitter based on its own CCIT. We will show that this low-cost option can still achieve near-optimal performance.

A. System Model

The system model in (1) is rewritten as follows

\[
r = \sum_{k=1}^{K} \sqrt{\gamma_{k}} H_k y_k + \eta
\]

(27)

where \( \gamma_{k} \) is the large scale fading factor of user \( k \), \( H_k \) is the Rayleigh-fading channel matrix modeled as (3), and \( y_k \) the transmitted signal vector of user \( k \) with zero mean and a power constraint \( P_k \), i.e.,

\[
E[y_k] = 0 \quad \text{and} \quad \text{tr}\{Q_k\} \leq P_k
\]

(28)

with \( Q_k = E[y_k y_k^H] \) the transmission covariance matrix of user \( k \). Following (3) and (19), we denote by \( C_{BS,k} \) and \( C_{MT,k} \) the receive and transmit correlation matrices of user \( k \) modeled by

\[
C_{BS,k} \{m,n\} = \rho_{BS,k}^{m-n} e^{i(m-n)\theta_{BS,k}}, \quad (29a)
\]

\[
C_{MT,k} \{m,n\} = \rho_{MT,k}^{m-n} e^{i(m-n)\theta_{MT,k}}. \quad (29b)
\]

And we also assume that the channels of different users are independent.

B. Mutual Information Performance

Denote by \( R_{\text{sum}}(S) \) the achievable sum rate of user subset \( S \subseteq \{1, \ldots, K\} \), i.e.,

\[
R_{\text{sum}}(S) = \log_2 \det \left( I_{N_{BS}} + \sum_{k \in S} \gamma_k H_k Q_k H_k^H / \sigma^2 \right).
\]

(30a)

The achievable rate region of the system in (27) is given by

\[
R(\{H_k\}, \{Q_k\}) = \left\{ (r_1, \ldots, r_K) \middle| \sum_{k \in S} r_k \leq R_{\text{sum}}(S), \forall S \right\}.
\]

(30b)

Denote by \( C_{\text{FCSIT}} \) the full CSIT capacity region that can be obtained by optimizing \( \{Q_k\} \) in (30). Denote by \( R_{\text{SWF}} \) the achievable rate region of individual SWF (I-SWF) scheme in which each user individually performs SWF precoding and ignores the presence of other users. We now show that \( R_{\text{SWF}} \) approaches \( C_{\text{FCSIT}} \) asymptotically when \( N_{BS} \rightarrow \infty \).

Proposition 3: With \( K \) and \( N_{MT} \) fixed, when \( N_{BS} \rightarrow \infty \) and (16) holds for each \( C_{BS,k} \), we have \( \forall (r_1, \ldots, r_K) \in C_{\text{FCSIT}} \), \( \exists (r_1', \ldots, r_K') \in R_{\text{SWF}} \),

\[
\lim_{N_{BS} \rightarrow \infty} \max_k |r_k - r_k'| = 0.
\]

(31)

Proof: See Appendix D.

Fig. 9 shows the I-SWF performance. We assume that \( K \) MTs are uniformly distributed in a hexagon region with sides of 1. The large scale fading factor \( \gamma_{k} \) consists of path loss with decay order 4 and lognormal fading with standard derivation 8 dB, and \( \rho_{MT,k} \) is uniformly distributed over [0.6, 0.9]. The channel capacity based on global CSIT and no CSIT are included for reference. We can see from Fig. 9 that I-SWF can obtain

Fig. 8. Simulation performance of the proposed scheme with varying \( \rho_{MT} \). \( N_{MT} = 8, N_{BS} = 64, \rho_{BS} = 0.5 \).

Fig. 9. Mutual information performance in a \( K \)-user system. \( K = 4, N_{MT} = 8, \rho_{BS} = 0.5, P_S = P_k/k, k = 1, \ldots, K \).
performance close to the global CSIT upper bound. It offers significant performance improvement compared to systems of no precoding (NP).

C. Implementation in Practical Systems

To implement I-SWF, we consider the similar scheme as in Section IV except that multi-user detection is adopted. We outline the main detection and analysis process as follows.

Substituting (20) into (27), the received signal is given by

\[ \mathbf{r} = \sum_{k=1}^{K} \sqrt{\gamma_k} \mathbf{H}_k \mathbf{U}_{Q,k} \mathbf{D}_{Q,k}^{1/2} \mathbf{V}_{\text{H}} \mathbf{x}_k + \eta. \]  

(32)

With LMMSE detection, we have

\[ \hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k] + v_k (\hat{\mathbf{H}}_k \mathbf{V}_{\text{H}}) \mathbf{R}^{-1} \left( \mathbf{r} - \sum_{k=1}^{K} \hat{\mathbf{H}}_k \mathbf{V}_{\text{H}} \mathbb{E}[\mathbf{x}_k] \right) \]  

(33)

where \( \hat{\mathbf{H}}_k = \sqrt{\gamma_k} \mathbf{H}_k \mathbf{U}_{Q,k} \mathbf{D}_{Q,k}^{1/2} \), and \( \mathbf{R} = \sum_{k=1}^{K} v_k \mathbf{D}_{Q,k} \mathbf{R} + \sigma^2 \mathbf{I} \).

Similarly as in the single-user case, we can use the following matrix to characterize the SINR of LMMSE estimation in (33)

\[ \mathbf{\Omega}_k \overset{\Delta}{=} \left( (\hat{\mathbf{H}}_k \mathbf{V}_{\text{H}})^{\mathbf{R}^{-1}} (\hat{\mathbf{H}}_k \mathbf{V}_{\text{H}}) \right)_{\text{diag}}. \]  

(34)

With an augmented channel model, we can show that \( \mathbf{\Omega}_k \approx \omega_k \mathbf{I} \) with \( \omega_k = N_{\text{MT}}^{-1} \text{tr}(\mathbf{\Omega}_k) \). Hence, the LMMSE module of user \( k \) can be characterized by a function

\[ \rho_k = \phi_k(v_k) = \omega_k / (1 - v_k \omega_k). \]  

(35)

When \( N_{\text{BS}} \rightarrow \infty \), the transfer function in (35) can be simplified as given below.

Proposition 4: With \( K \) and \( N_{\text{MT}} \) fixed, and \( \mathbf{U}_{Q,k} = \mathbf{U}_{\text{MT},k} \), when \( N_{\text{BS}} \rightarrow \infty \) and (16) holds for each \( \mathbf{C}_{\text{BS},k} \), we have

\[ \phi_k(v_k) = \omega_{\text{m},k}(v_k) / (1 - v_k \omega_{\text{m},k}(v_k)) \]  

(36a)

where

\[ \omega_{\text{m},k}(v) = \frac{1}{N_{\text{MT}}} \sum_{n} D_{\text{MT},k}(n,n) D_{Q,k}(n,n) / (v_k D_{\text{MT},k}(n,n) D_{Q,k}(n,n) + \sigma^2 / N_{\text{BS}}). \]  

(36b)

Proof: See Appendix E.

Proposition 4 shows that \( \phi_k(v_k) \) is asymptotically determined by \( \mathbf{C}_{\text{MT},k} \) ignoring the presence of other users when \( N_{\text{BS}} \) is sufficiently large. Later, we will show by numerical results that good performance can be obtained even if \( N_{\text{BS}} / (K N_{\text{MT}}) \) is only moderately large.

Fig. 10 provides the simulation results. Again EP and regular LDPC code are used for systems without CSIT. The irregular LDPC code is fixed to that used in Fig. 7 (i.e., optimized according to \( \rho_{\text{MT}} = 0.8 \)). The precoder for each user is optimized according to its own transmit correlation matrix. We can see that the simple individual CCIT based scheme can offer significant performance improvement compared to EP even if \( N_{\text{BS}} / (K N_{\text{MT}}) \) is only moderately large (\( N_{\text{BS}} / (K N_{\text{MT}}) = 1 \) and 2, respectively, for \( N_{\text{BS}} = 32 \) and 64 in Fig. 10).

VI. CONCLUSION

We have studied a simple SWF scheme that requires only the correlation matrix \( \mathbf{C}_{\text{MT}} \) at the transmitter in uplink MIMO systems. This scheme can efficiently relieve the high cost related to CSIT acquisition in MIMO systems. We show by mutual information analysis that the performance of SWF is potentially close to that with full CSIT when \( N_{\text{BS}} > N_{\text{MT}} \). We also show that iterative LMMSE detection can be used to handle the interference problem related to SWF and transmitter optimization in this system can also be accomplished using only \( \mathbf{C}_{\text{MT}} \). Significant performance improvement is demonstrated by simulation results.

APPENDIX

A. Proof of Proposition 1

To prove Proposition 1(i), we first show that the achievable rate \( R(\mathbf{H}, \mathbf{Q}) \) can be asymptotically characterized by \( \mathbf{C}_{\text{MT}} \) and \( \mathbf{Q} \) regardless of channel realization \( \mathbf{H} \).

Remark 1: When \( N_{\text{BS}} \rightarrow \infty \) with \( N_{\text{MT}} \) fixed, and \( \mathbf{C}_{\text{BS}} \) meets the condition in (16), we have

\[ \lim_{N_{\text{BS}} \rightarrow \infty} R(\mathbf{H}, \mathbf{Q}) = \log_2 \det \left( \mathbf{I}_{\text{MT}} + \mathbf{C}_{\text{MT}}^{1/2} \mathbf{Q} \mathbf{C}_{\text{MT}}^{1/2} / N_{\text{BS}} / \sigma^2 \right). \]  

(37)

To show (37), we first rewrite \( R(\mathbf{H}, \mathbf{Q}) \) in (5) as follows

\[ R(\mathbf{H}, \mathbf{Q}) = \log_2 \det \left( \mathbf{I}_{\text{MT}} + \mathbf{C}_{\text{MT}}^{1/2} \mathbf{Q} \mathbf{C}_{\text{MT}}^{1/2} \mathbf{H}_{\text{BS}}^{H} \mathbf{C}_{\text{BS}} \mathbf{H}_{\text{w}} / \sigma^2 \right). \]  

(38)

Recall that the elements of \( \mathbf{H}_{\text{w}} \) are i.i.d. complex Gaussian random variables \( \mathcal{CN}(0,1) \). From the law of large numbers and the assumption in (16), we can show that

\[ \lim_{N_{\text{BS}} \rightarrow \infty} \mathbb{E}[\mathbf{H}_{\text{w}}] \cdot \mathbf{C}_{\text{BS}} \cdot \mathbf{H}_{\text{w}} = \mathbb{E}[\mathbf{C}_{\text{BS}}] \cdot \mathbb{E}[\mathbf{I}_{\text{MT}}] = N_{\text{BS}} \mathbf{I}_{\text{MT}}. \]  

(39)

Combining (38) and (39), we obtain (37).
With Remark 1, we can obtain an upper bound for the full CSIT capacity as detailed below.

Remark 2: When \( N_{BS} \to \infty \) with \( N_{MT} \) fixed, and \( C_{BS} \) meets the condition in (16), we have

\[
\lim_{N_{BS} \to \infty} C_{CSIT} \leq \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q_{SWF} C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right). \tag{40}
\]

Recall that the full CSIT capacity is defined in (6) by

\[
C_{CSIT} = E_H \left[ \max_{Q} R(H, Q(H)) \right]. \tag{41}
\]

Denote the optimized \( Q \) for the channel realization \( H \) by \( Q_H \). From (37), we have

\[
C_{CSIT} = E \left[ \log_2 \det \left( I_{N_{BS}} + H Q_H H^H / \sigma^2 \right) \right]
= E \left[ \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q H C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right) \right]. \tag{42}
\]

Note that for the channel characterized by \( C_{MT}^{1/2}, Q_{SWF} \) is optimal. Then we have

\[
E \left[ \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q_{SWF} C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right) \right]
\leq \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q_{SWF} C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right). \tag{43}
\]

Combining (42) and (43), we obtain (40).

On the other hand, by setting \( Q = Q_{SWF} \) in (37) and taking expectation over \( H \), we can show that the upper bound in Remark 2 can be asymptotically approached by SWF as detailed below.

Remark 3: When \( N_{BS} \to \infty \) with \( N_{MT} \) fixed, and \( C_{BS} \) meets the condition in (16), we have

\[
\lim_{N_{BS} \to \infty} R_{SWF} = \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q_{SWF} C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right). \tag{44}
\]

Combining Remarks 2 and 3 and noting that \( R_{SWF} \leq C_{CSIT} \), we obtain Proposition 1(i).

Proof of Proposition 1(ii): Similar to Remark 1, the achievable rate \( R(H, Q) \) in case (ii) can be characterized by \( C_{MT} \) and \( Q \) regardless of the channel realization \( H \).

Remark 4: When \( N_{BS}, N_{MT} \to \infty \) with \( N_{MT} / N_{BS} \to 0 \), and \( C_{BS} = I \), we have

\[
\lim_{N_{BS} \to \infty} R(H, Q) = \log_2 \det \left( I_{N_{MT}} + C_{MT}^{1/2} Q C_{MT}^{1/2} \cdot N_{BS} / \sigma^2 \right) \tag{45}
\]

To prove (45), we first introduce the following result.

Proposition 5: (c.f. Example 2.50 in [44]) Let \( H \) be an \( N_{MT} \times N_{BS} \) random matrix whose elements are zero-mean i.i.d. Gaussian random variables with variance \( 1 / \sqrt{N_{BS}/N_{MT}} \) and denote by \( \sqrt{N_{BS}/N_{MT}} = \varsigma \). It can be shown that as \( N_{MT}, N_{BS} \to \infty \) with \( N_{MT} / N_{BS} \to 0 \), the asymptotic spectrum of the matrix \( HH^H - \varsigma \sqrt{N_{MT}} I \) is the semicircle law, i.e.,

\[
w(\lambda) = \begin{cases} \frac{1}{\pi} \sqrt{4 - \lambda^2} & |\lambda| \leq 2; \\ 0 & |\lambda| > 2. \end{cases}
\]

Now we consider the \( N_{MT} \times N_{BS} \) random matrix \( H_w \) in (3) whose elements are i.i.d. random variables \( CN(0,1) \). From Proposition 5, as \( N_{MT}, N_{BS} \to \infty \) with \( N_{MT} / N_{BS} \to 0 \), the asymptotic spectrum of the matrix \( (H_w^H H_w - N_{BS} I_{N_{MT}}) / \sqrt{N_{BS}/N_{MT}} \) obeys the semicircle law. Hence the eigenvalue \( \lambda \) of \( H_w^H H_w \) meets the following condition with probability 1

\[
|\lambda - N_{BS}| / \sqrt{N_{BS}/N_{MT}} \leq 2.
\]

This means

\[
N_{BS} \left( 1 - 2 \sqrt{N_{MT}/N_{BS}} \right) \leq \lambda \leq N_{BS} \left( 1 + 2 \sqrt{N_{MT}/N_{BS}} \right).
\]

Then we can obtain

\[
R(H, Q) \leq \log_2 \det \left( I_{N_{MT}} + N_{BS} C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
= \log_2 \det \left( I_{N_{MT}} + H_w^H H_w C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
\geq \log_2 \det \left( I_{N_{MT}} + H_w^H H_w + 2 \sqrt{N_{BS}/N_{MT}} C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
\geq \log_2 \det \left( I_{N_{MT}} + N_{BS} C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
+ N_{MT} \log_2 \left( 1 + 2 \sqrt{N_{MT}/N_{BS}} \right). \tag{46}
\]

Similarly, we can show that

\[
R(H, Q) \leq \log_2 \det \left( I_{N_{MT}} + N_{BS} C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
\leq \log_2 \det \left( I_{N_{MT}} + H_w^H H_w + 2 \sqrt{N_{BS}/N_{MT}} C_{MT}^{1/2} Q C_{MT}^{1/2} / \sigma^2 \right)
\]

\[
+ N_{MT} \log_2 \left( 1 + 2 \sqrt{N_{MT}/N_{BS}} \right). \tag{47}
\]

Note that \( N_{MT}/N_{BS} \to 0 \) and \( \ln(1 + x) \to x \) as \( x \to 0 \), we obtain (45).

With (45) available, Proposition 1(ii) can be proved similarly as Proposition 1(i).

B. Justification of Equation (24)

In the discussions in Section IV-E, we assume \( \Omega \approx \omega I \). This can be justified for the case of full CSIT, as discussed in [40]. In more general cases of partial CSIT (as discussed in Section II), an augmented channel model can be adopted to ensure (24). We outline the main results as follows.

Denote by \( J \) the number of samples of \( H \) in a coded frame. They can be distributed over different OFDM subcarriers or different time slots. Let the received signals be (see (1))

\[
r(j) = H(j) y(j) + \eta(j), \quad j = 1, \ldots, J. \tag{48}
\]

Denote by \( r^{aug} = [r(1)^T, \ldots, r(J)^T]^T \), \( H^{aug} = \text{diag}(H(1), \ldots, H(J)) \), \( y^{aug} = [y(1)^T, \ldots, y(J)^T]^T \), and \( \eta^{aug} = [\eta(1)^T, \ldots, \eta(J)^T]^T \) with superscript "aug" for augmentation. The system (48) can be rewritten as

\[
r^{aug} = H^{aug} y^{aug} + \eta^{aug}. \tag{49}
\]
Assume that $\mathbf{x}$ is a coded sequence. Following [41], [56], [57], we adopt the precoding technique as follows

$$\mathbf{y}^{\text{aug}} = \mathbf{U}^{\text{aug}} \mathbf{D}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{x}$$

(50)

where $\mathbf{V}^{\text{aug}} = \mathbf{J} \mathbf{N} \mathbf{N}^T$ is a Hadamard matrix, and $\mathbf{U}^{\text{aug}} = \text{diag}(\mathbf{U}(1), \ldots, \mathbf{U}(J))$ and $\mathbf{D}^{\text{aug}} = \text{diag}(\mathbf{D}(1), \ldots, \mathbf{D}(J))$ are, respectively, augmentations of the matrices $\mathbf{U}$ and $\mathbf{D}$ in (20).

Similar to (21), from (49) and (50), we can write the precoded system as

$$\mathbf{r}^{\text{aug}} = \mathbf{H}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{x} + \mathbf{v}^{\text{aug}}$$

(51)

where $\mathbf{H}^{\text{aug}} = \text{diag}(\mathbf{H}(1), \ldots, \mathbf{H}(J))$. Note that $\mathbf{H}^{\text{aug}}$ is block-diagonal and so is $\mathbf{R}$. Denote by $\mathbf{R} = \text{diag}(\mathbf{R}(1), \ldots, \mathbf{R}(J))$. The matrix $\mathbf{\Omega}$ can be calculated in a block-diagonal form as follows

$$\mathbf{\Omega} = \mathbf{\Omega}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{V}^{\text{aug}} \mathbf{V}^{\text{aug}}$$

(52)

where $\mathbf{V}^{\text{aug}}$ denotes the Kronecker product, and

$$\mathbf{\Omega}^{\text{aug}} = \mathbf{J}^{-1} \sum_{j=1}^{J} (\mathbf{\hat{H}}(j))^H (\mathbf{R}(j))^{-1} (\mathbf{\hat{H}}(j))$$

(53)

Note that $\mathbf{\Omega}(j)$ in (53) is a Hermitian matrix. Its diagonal elements are real and non-negative while its off-diagonal elements are complex with phase determined by $\mathbf{\hat{H}}(j)$. When $\{\mathbf{\hat{H}}(j)\}$ in (53) are randomly generated, the off-diagonal elements of $\mathbf{\Omega}(j)$ will have phases varying with $j$, and the summary over $j$ can be constructive or deconstructive. Note that the diagonal elements always sum constructively, we have

$$\mathbf{\Omega}^{\text{aug}} \approx (\mathbf{\Omega}^{\text{aug}})^{\text{diag}}$$

(54)

When $\{\mathbf{\hat{H}}(j)\}$ in (53) are quasi-static, additional i.i.d. permutation matrices $\{\mathbf{\Pi}(j)\}$ can be applied to form equivalent channel $\{\mathbf{\hat{H}}(j)\mathbf{\Pi}(j)\}$ to introduce random phase and obtain (54) [41].

Combining (52) and (54), and using the conclusion in the full CSIT case, we obtain (24).

### C. Proof of Proposition 2

We first rewrite (24) as follows

$$\omega = N_{\text{MT}}^{-1} \text{tr} \left( \mathbf{\hat{H}}^H (\mathbf{\hat{H}}^H + \sigma^2 \mathbf{I})^{-1} \mathbf{\hat{H}} \right)$$

$$= N_{\text{MT}}^{-1} \sum_{n=1}^{N_{\text{BS}}} \lambda_n (\mathbf{\hat{H}}^H) / (v \lambda_n (\mathbf{\hat{H}}^H) + \sigma^2)$$

(55)

where $\lambda_n(A)$ denotes the $n$th eigenvalue of $A$.

From (3), (9), (10) and (22), we have

$$\mathbf{\hat{H}}^H \mathbf{\hat{H}} = \mathbf{D}_Q^{1/2} \mathbf{D}_\text{MT}^{1/2} \mathbf{U}_\text{MT}^H \mathbf{H}_\text{w}^H \mathbf{C}_\text{BS} \mathbf{H}_\text{w} \mathbf{U}_\text{MT} \mathbf{D}_\text{MT}^{1/2} \mathbf{D}_Q^{1/2}$$

On the other hand, when the two conditions in Proposition 1 is satisfied, from the proof of Proposition 1, we can show that

$$\lim_{N_{\text{BS}} \to \infty} \mathbf{\hat{H}}^H \mathbf{\hat{H}} = \mathbf{D}_Q \mathbf{D}_\text{MT} \cdot N_{\text{BS}} \mathbf{I}_{N_{\text{MT}}}.$$ (56)

Substituting (56) into (55), we complete the proof.

### D. Proof of Proposition 3

To prove Proposition 3, we first show that the achievable rate region $\mathcal{R}(\{\mathbf{H}_k\}, \{\mathbf{Q}_k\})$ can be asymptotically characterized by $\{\mathbf{C}_{\text{MT}, k}\}$ and $\{\mathbf{Q}_k\}$ regardless of the channel realizations $\{\mathbf{H}_k\}$.

**Remark 5:** When $N_{\text{BS}} \to \infty$ with $N_{\text{MT}}$ and $K$ fixed, and (16) holds for each $C_{\text{BS}, k}, \forall S = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, K\}$, we have,

$$\lim_{N_{\text{BS}} \to \infty} \log_2 \det \left( \mathbf{I}_{N_{\text{BS}}} + N_{\text{BS}} \mathbf{C}_{\text{MT}, S} \mathbf{Q}_S \right) / \sigma^2$$

(57)

where

$$\mathbf{C}_{\text{MT}, S} = \text{diag}(\mathbf{C}_{\text{MT}, i_1}, \ldots, \mathbf{C}_{\text{MT}, i_k})$$

To show (57), we rewrite the expression in (57) as follows

$$\log_2 \det \left( \mathbf{I}_{N_{\text{BS}}} + N_{\text{BS}} \mathbf{C}_{\text{MT}, S} \mathbf{Q}_S \right) / \sigma^2$$

(58)

Recall that the elements of $\mathbf{H}_w, k$ are i.i.d. complex Gaussian random variables $CN(0,1)$. From the law of large numbers and the assumption in (16), we can show that

$$\lim_{N_{\text{BS}} \to \infty} \mathbf{H}_w^H \mathbf{H}_w = N_{\text{BS}} \mathbf{C}_{\text{MT}, S}.$$ (59)

Combining (58) and (59), we obtain (57).

Based on Remark 5, Proposition 3 can be proved similarly as Proposition 1.

### E. Proof of Proposition 4

From (34), we have

$$\omega_k = N_{\text{MT}}^{-1} \text{tr} \left( \mathbf{H}_k^H \left( \mathbf{H}_k^H + \sigma^2 \mathbf{I} + \sum_{i=1}^{K} v_i \mathbf{H}_i \mathbf{H}_i^H \right)^{-1} \mathbf{H}_k \right)$$

$$= N_{\text{MT}}^{-1} \text{tr} \left( \sigma^{-2} \left( \mathbf{I} + \sum_{i=1}^{K} \sigma^{-2} v_i \mathbf{H}_i \mathbf{H}_i^H \right)^{-1} \mathbf{H}_k \mathbf{H}_k^H \right)$$

(60)
where $\mathbf{R} = \sum_{k=1}^{K} \sigma^{-2} v_k \mathbf{H}_k \mathbf{H}_k^H$. From (59), we have

$$
\lim_{N_{BS} \to \infty} \left( \mathbf{R} \right)^n \mathbf{H}_k \mathbf{H}_k^H = \lim_{N_{BS} \to \infty} \left( \sum_{k=1}^{K} \sigma^{-2} v_k \mathbf{H}_k \mathbf{H}_k^H \right)^n \mathbf{H}_k \mathbf{H}_k^H = \left( \sigma^{-2} v_k \mathbf{H}_k \mathbf{H}_k^H \right)^n \mathbf{H}_k \mathbf{H}_k^H .
$$

Hence

$$
\lim_{N_{BS} \to \infty} \sigma^{-2} (1 + \mathbf{R})^{-1} \mathbf{H}_k \mathbf{H}_k^H = \lim_{N_{BS} \to \infty} \sigma^{-2} (1 - \mathbf{R} + \mathbf{R}^2 - \mathbf{R}^3 + \ldots) \mathbf{H}_k \mathbf{H}_k^H = \lim_{N_{BS} \to \infty} \sigma^{-2} (1 - (\sigma^{-2} v_k \mathbf{H}_k \mathbf{H}_k^H) + \ldots) \mathbf{H}_k \mathbf{H}_k^H = \left( \sigma^{-2} I + v_k \mathbf{H}_k \mathbf{H}_k^H \right)^{-1} \mathbf{H}_k \mathbf{H}_k^H .
$$

Then Proposition 4 can be proved similarly as Proposition 2.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

REFERENCES


Chongbin Xu received the B.S. degree in information engineering from Xi’an Jiaotong University in 2005 and the Ph.D. degree in information and communication engineering from Tsinghua University in 2012. He is now a Research Fellow in the Department of Electronic Engineering, City University of Hong Kong. His research interests are in the areas of signal processing and communication theory, including linear precoding, iterative detection, and random access techniques.

Peng Wang (S’05–M’10) received the B. Eng. degree in telecommunication engineering and the M. Eng. degree in information engineering, from Xidian University, Xi’an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong SAR, in 2010. He was a Research Fellow with the City University of Hong Kong and a visiting Post-Doctor Research Fellow with the Chinese University of Hong Kong, Hong Kong SAR, both from 2010 to 2012. Since 2012, he has been with the Centre of Excellence in Telecommunications, School of Electrical and Information Engineering, the University of Sydney, Australia, where he is currently a Research Fellow. His research interests include channel and network coding, information theory, iterative multi-user detection, MIMO techniques and millimetre-wave communications. He has published over 40 peer-reviewed research papers in the leading international journals and conferences, and received the Best Paper Award at the IEEE International Conference on Communications (ICC) in 2014. He has also served on a number of technical programs for international conferences such as ICC and WCNC.

Zhonghao Zhang received the B.S. degree and M.S. degree in electronic and information engineering from University of Electronic Science and Technology of China in 2005 and 2008, respectively. He received the Ph.D. degree in electronic engineering from City University of Hong Kong in 2014. He is now with the information technology group in the Department of Electronic Engineering at City University of Hong Kong. His research interests are in the areas of signal processing in communication systems, including spatial coupling transmission and iterative detection.

Li Ping (S’87–M’91–SM’06–F’10) received the Ph.D. degree at Glasgow University in 1990. He lectured in the Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and worked as a member of research staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a Chair Professor. He received a British Telecom-Royal Society Fellowship in 1986, the IEE J J Thomson premium in 1993, the Croucher Foundation Award in 2005, and a British Royal Academy of Engineering Distinguished Visiting Fellowship in 2010. He served as a member of Board of Governors of IEEE Information Theory Society from 2010 to 2012.