Abstract—In the uplink of network MIMO systems, in order to cope with backhaul capacity limitations, base stations (BSs) must compress the received baseband data signal and the channel state information (CSI) for communication to the central unit (CU). Assuming ergodic fading, an Estimate-Compress-Forward (ECF) approach is investigated, whereby the BSs perform CSI estimation and forward a compressed version of the CSI to the CU. This approach contrasts with the previously studied Compress-Forward-Estimate (CFE) strategy, whereby CSI estimation is performed at the CU, and is motivated by the information-theoretic optimality of separate estimation and compression. Various ECF schemes are proposed that perform either separate or joint compression of estimated CSI and received baseband signal. Via numerical results, it is shown that a proper design of ECF strategies leads to substantial performance gains compared to the CFE approach.

Index Terms—Uplink network MIMO, limited backhaul, imperfect CSI, compress and forward, distributed compression.

I. INTRODUCTION

In network MIMO systems, multiple base stations (BSs), or remote radio heads, are connected via backhaul links to a central unit (CU) which performs joint decoding in uplink on behalf of all the connected BSs (see [1] and references therein). In the presence of limitations on the backhaul links, a solution that appears to be favored due to its practicality and good theoretical performance is based on compress-and-forward [2]. Accordingly, the BSs compress the received baseband signal and forward it to the CU. Most previous work on the design of backhaul compression strategies for the uplink has implicitly assumed full channel state information (CSI) to be available at the CU [2], [3]. However, in the presence of time-varying (e.g., ergodic) channels, CSI overhead can become significant. It is hence important to properly design the compression of CSI and data from the BSs to the CU. This problem was tackled in [4] by proposing a Compress-Forward-Estimate (CFE) approach, whereby the BSs compress the uplink training signals and the CSI estimate is performed at the CU.

In contrast, motivated by the information-theoretic optimality of separate estimation and compression [5], this work adopts an Estimate-Compress-Forward (ECF) approach. Accordingly, each BS first estimates the CSI and then compresses it for transmission to the CU. Three different strategies are proposed that carry out separate or joint compression of the estimated CSI and the received signal in the data part of the block. The proposed schemes are based on distributed source coding for the received signals following [2], [3]. Numerical results demonstrate the effectiveness of the proposed ECF strategies as compared to the CFE approach.

Notation: The covariance matrix $R_X$ of the random vector $X$ is computed $R_X = \mathbb{E}[XX^\dagger]$, the cross covariance matrix $R_{XY}$ of $X$ and $Y$ is $R_{XY} = \mathbb{E}[XY^\dagger]$, and $R_{XX}$ denotes the conditional covariance matrix of $X$ conditioned on $Y$, i.e., $R_{X|Y} = R_X - R_{XY}R_{YY}^{-1}R_{XY}$. For a subset $S \subseteq \{1, \ldots, n\}$, given matrices $X_1, \ldots, X_n$, we define the matrix $X_S$ by stacking the matrices $X_i$ with $i \in S$ vertically in ascending order, namely $X_S = [X_1^\dagger, \ldots, X_n^\dagger]^T$.

II. SYSTEM MODEL

Consider the uplink of a cellular system consisting of $N_M$ mobile stations (MSs), $N_B$ BSs and a CU, as shown in Fig. 1. We denote the set of all MSs as $\mathcal{N}_M = \{1, \ldots, N_M\}$ and of all BSs as $\mathcal{N}_B = \{1, \ldots, N_B\}$. The MSs, the $i$-th of which has $N_{t,i}$ transmit antennas, communicate in the uplink to the BSs, where the $j$-th BS is equipped with $N_{r,j}$ receive antennas. Each $j$-th BS is connected to the CU via a backhaul link of capacity $C_j$. We define $N_t$ and $N_r$ as the number of total transmit antennas and total receive antennas, that is $N_t = \sum_{i=1}^{N_M} N_{t,i}$ and $N_r = \sum_{j=1}^{N_B} N_{r,j}$, respectively.

The channel coherence block, of length $T$ channel uses, is split into a phase for channel training of length $T_p$ channel uses and a phase for data transmission of length $T_d$ channel uses, with $T_p + T_d = T$ as in [4], [6]. The $N_t \times T$ signal $X$ transmitted by all MSs is hence divided into the $N_t \times T_p$ pilot...
signal $X_p$ and the $N_t \times T_d$ data signal $X_d$. For simplicity, we assume equal transmit power allocation for each antenna of all MSs and then that the transmit signal $X$ has a total per-block power constraint $\frac{1}{T} ||X||^2 = P$, and we define $\frac{1}{T} ||X_p||^2 = P_p$ and $\frac{1}{T} ||X_d||^2 = P_d$ as the powers used for training and data. In terms of pilot and data signal powers, then, the power constraint becomes $\frac{1}{T} P_p + \frac{1}{T} P_d = P$.

The $N_{r,j} \times N_t$ channel matrix $H_{j}$ collects all the $N_{r,j} \times N_t$ channel matrix $H_{ji}$ from the $i$-th MS to the $j$-th BS as $H_j = [H_{j1}, \ldots, H_{jN_{u,j}}]$. The channel matrix $H_{ji}$ is modeled as Rician fading with channel gain $\alpha_{ji}$ between the $j$-th BS and the $i$-th MS and hence includes the line-of-sight component $H_{ji}$, which is deterministic, and the scattered component $H_{wi,j}$ with independent and identically distributed (i.i.d.) complex Gaussian variables with zero mean and unit variance, i.e., $CN(0, 1)$, that is $H_{ji} = \sqrt{\alpha_{ji}} (\sqrt{\frac{K}{K+1}} H_{ji} + \sqrt{\frac{1}{K+1}} W_{wi,j})$, where $K$ is the Rician factor. The channel matrix $H_j$ is assumed to be constant during each channel coherence block and to change according to an ergodic process from block to block. As in [6], we assume that coding is performed across multiple channel coherence blocks.

### A. Training Phase

We assume as in [4] that the training signal is $X_p = \sqrt{\frac{P}{N_t}} S_p$ where $S_p$ is a $N_t \times T_p$ matrix of i.i.d. $CN(0, 1)$ entries. The received training signal $Y_{pj}$ at the $j$-th BS is then given as $Y_{pj} = H_{p}X_p + Z_{pj} = \sqrt{\frac{P}{N_t}} H_{p} S_p + Z_{pj}$, where the noise matrix $Z_{pj}$ has i.i.d. $CN(0, 1)$ entries.

By performing the minimum mean square error (MMSE) estimate of $H_{j}$ (see [7]), we can decompose the channel matrix $H_{ji}$ into the estimate $\hat{H}_{ji}$ and the independent estimation error $E_{ji}$, as $H_{ji} = \hat{H}_{ji} + E_{ji}$, where the error $E_{ji}$ has i.i.d. $CN(0, \sigma_{e_{ji}}^2)$ entries with $\sigma_{e_{ji}}^2 = \frac{\alpha_{ji}^2}{T_p P_e + N_t (K + 1)}$. The estimated channel matrix $\hat{H}_{ji}$ has a matrix-variate complex Gaussian distribution with mean matrix $\sqrt{\frac{\alpha_{ji}^2}{K+1}} H_{ji}$ and covariance matrix $\sigma_{e_{ji}}^2 I$, where $\sigma_{e_{ji}}^2 = \frac{\alpha_{ji}^2}{T_p P_e + N_t (K + 1)}$.

The sequence of channel estimates $\hat{H}_{ji}$ for all coherence blocks in the coding block is compressed by the $j$-th BS and forwarded to the CU on the backhaul link. The compressed channel $\tilde{H}_{ji}$ is related to the estimate $\hat{H}_{ji}$ as

$$
\tilde{H}_{ji} = \hat{H}_{ji} + Q_{p,j},
$$

where the $N_{r,j} \times N_t$ quantization noise matrix $Q_{p,j}$ has zero-mean i.i.d. $CN(0, \sigma_{q_{p,j}}^2)$ entries (see, e.g., [8, Ch. 3]) and the compressed estimate $\tilde{H}_{ji}$ is complex Gaussian with mean matrix $\sqrt{\frac{2T_d}{K+1}} H_{ji}$ and covariance matrix $R_{\tilde{H}_{ji}} = \sigma_{q_{p,j}}^2 I$, where $H_{ji} = [\sqrt{\sigma_{e_{ji}}^2} H_{ji}, \ldots, \sqrt{\sigma_{e_{ji}}^2} H_{ji}]$, and $R_{\tilde{H}_{ji}}$ is diagonal matrices with main diagonals given by $[\sigma_{q_{p,j}}^2 I_{N_{u,j}}, \ldots, \sigma_{q_{p,j}}^2 I_{N_{u,j}}]$. The choice of independent entries in $Q_{p,j}$ is dictated by the i.i.d. distribution of the channel entries of the $H_{ji}$. We will discuss later how to relate the quantization noise variance $\sigma_{q_{p,j}}^2$ to the backhaul capacity $C_j$.

### B. Data Phase

During the data phase, the MSs transmit signals $X_d$, which are i.i.d. $CN(0, P_d/N_t)$ variables. The BSs compress the signal $Y_{d,j}$ and sends it to the CU on the backhaul link. The CU recovers the sequence of quantized data signals $\hat{Y}_{d,j}$ in (2) and of quantized channel estimates $\tilde{H}_{ji}$ in (1) from the information received on the backhaul link. Separating the desired signal and the noise, the received signal $\hat{Y}_{d,j}$ from the $j$-th BS can be expressed as

$$
\hat{Y}_{d,j} = Y_{d,j} + Q_{d,j} = \tilde{H}_{ji} X_d + N_{d,j},
$$

where the equivalent noise $N_{d,j} = (Q_{p,j} + E_{j}) X_d + Z_{d,j} + Q_{d,j}$ is not Gaussian distributed and is not independent of $X_d$ (see also [6]); $E_{j}$ collects all estimation errors as $E_{j} = \{E_{j1}, \ldots, E_{jN_{u,j}}\}$; the noise matrix $Z_{d,j}$ has i.i.d. $CN(0, 1)$ entries; and the quantization noise matrix $Q_{d,j}$ is independent of $Y_{d,j}$ and is zero-mean complex Gaussian with covariance matrix $E\{\text{vec}(Q_{d,j}) \text{vec}(Q_{d,j})^H\} = R_{d,j} \otimes I_{p,e}$. The role of the covariance matrix $R_{d,j}$ and its relationship with the backhaul capacity will be clarified in the next sections.

### III. Joint Signal and CSI Compression

In this section, we discuss how to calculate the quantization noises statistics, namely $\sigma_{q_{p,j}}^2$ for the estimated CSI in (1) and $R_{d,j}$ for the data in (2) with the aim of maximizing the ergodic achievable sum-rate. We adopt distributed source coding for the compression of the received data signals, as implemented via successive compression following [3]. Accordingly, the CU decodes the data signals in a given order $\pi$ by treating all previously decoded signals as side information.

### A. Problem Definition

Here we define the optimization problem and the proposed sequential solution. We recall that we need to optimize the compression parameters $(\sigma_{p,j}, R_{d,j})$ for all $j \in N_B$ and that each BS uses the test channel (1) for the training phase and (2) for the data phase. The goal is to maximize the ergodic achievable sum-rate $\frac{1}{T} I(X_d; Y_{d,\pi(j)}, \tilde{H}_{j})$, where $Y_{d}$ is the data signal received at the CU from all BSs and $\tilde{H}_{j}$ collects all compressed channel matrices. Using the bounding technique in [9] and by the chain rule of mutual information, we can write the ergodic achievable sum-rate as the given $\frac{1}{T} \sum_{j=1}^{N_B} I(X_d; Y_{d,\pi(j)}, \tilde{H}_{j}) \geq \frac{1}{T} \sum_{j=1}^{N_B} R_{j}$, where

$$
R_{j} = \frac{T_d}{T} E [\log_2 \det (I + \hat{H}_{\pi(j)} R_{X|\tilde{Y}_{\tilde{H}_{ji}}})],
$$

In (3), $\pi$ is a given ordering of the BSs; $\sigma_{p,j}^2 = \text{tr} (R_{X|\tilde{Y}_{\tilde{H}_{ji}}}) (\sigma_{q_{p,j}}^2 + \sigma_{e_{j}}^2)^{-1}$, where $\sigma_{j} = \sum_{i=1}^{N_t} \sigma_{e_{ji}}^2$, and the conditional correlation matrix $R_{X|\tilde{Y}_{\tilde{H}_{ji}}}$ is calculated as

$$
R_{X|\tilde{Y}_{\tilde{H}_{ji}}} = \frac{P_d}{N_t} I_{N_t} \tilde{H}_{ji}^H \tilde{H}_{ji} + \frac{P_d}{N_t} \tilde{H}_{ji}^H \tilde{H}_{ji} + \frac{P_d}{N_t} \tilde{H}_{ji}^H \tilde{H}_{ji} + N_t R_{p,\pi(j)} + 2 \epsilon_{j}.
$$

(4)
Greedy algorithm for the multi-BS case

\[
C_{d,j} = \frac{1}{T} I \left( \mathbf{Y}_{d,j}; \hat{\mathbf{Y}}_{d,j} \big| \hat{\mathbf{Y}}_{d,S_j}, \hat{\mathbf{H}}_{S_j} \right) \leq \frac{T_d}{T} \log_2 \det \left( I_{N_{r,j}} + \frac{E \left[ \mathbf{H}_j \mathbf{R}_{X|\mathbf{Y}_{S_j},\mathbf{H}} \mathbf{H}_j + \left( \sigma_{p,j}^2 + \sigma_{d,j}^2 \right) \mathbf{R}_{X|\mathbf{Y}_{S_j},\mathbf{H}} + I_{N_{r,j}} \right]}{\sigma_{d,j}^2} \right). \quad (7)
\]

**Algorithm 1** Greedy algorithm for the multi-BS case

1. Initialize set \( S \) to be an empty set, i.e., \( S_0 = \emptyset \).
2. for \( n = 1 \) to \( N_B \) do
3. \[ j^* = \arg \max_{j \in N_B \setminus \bigcup_{n=1}^{n-1} S_n} R_j \] where \( R_j \) is the optimal value of the problem
4. maximize \( R_j \) in (3) \hspace{1cm} (5a)
5. s.t. backhaul constraint (see Sec. III-B, C, D) \( (5b) \)
6. Update the set \( S_n = S_{n-1} \cup \{j^*\} \) and \( \pi^* (n) = j^* \).
7. Assign a solution of (5) for \( j = j^* \) to the optimal \( \sigma_{p,j^*}^2 \) and \( R_{d,j^*} \).
8. end for
9. retur \( \pi^*, \{\sigma_{p,1}^2, \ldots, \sigma_{p,N_B}^2\} \), and \{\( R_{d,1}, \ldots, R_{d,N_B} \)\}

with \( R_{d,S_j}, R_{p,S_j} \) and \( R_{d,\pi} \) being diagonal matrices with main diagonals given by \( [R_{d,\pi}(1), \ldots, R_{d,\pi}(j-1), \sigma_{p,\pi}^2(1), \ldots, \sigma_{p,\pi}(j-1)] \) and \( [\sigma_{p,\pi}^2(1), \ldots, \sigma_{p,\pi}(j-1)] \), respectively. We remark that the rate \( R_j \) can be interpreted as the contribution of \( j \)-th BS to the ergodic achievable sum-rate.

The proposed approach to the optimization of the ergodic achievable sum-rate with respect to the order \( \pi \) and the compression parameters \( \sigma_{p,j}^2 \) and \( R_{d,j} \) for all \( j \in N_B \) is summarized in Algorithm 1. The rate maximization step in (5) is discussed in the next sections.

**B. Separate Compression of Channel and Received Data Signal**

In this subsection, we solve the problem (5) for a given \( j \)-th BS assuming separate compression of CSI and received data signal. Due to the identical distribution of the entries of \( \mathbf{Y}_{d,j} \), here we choose \( R_{d,j} = \sigma_{d,j}^2 I_{N_{r,j}} \) in (2) and hence the optimization is over the pair \( (\sigma_{p,j}^2, \sigma_{d,j}^2) \). To this end, let us denote as \( C_{p,j} \) and \( C_{d,j} \) the backhaul rate used for transmitting the channel estimates (1) and the received signals (2) from the \( j \)-th BS to the CU, respectively. The backhaul constraint in (5) is thus \( C_{p,j} + C_{d,j} = C_j \). Moreover, from rate-distortion considerations, we can relate the compression noise power \( \sigma_{p,j}^2 \) with the backhaul capacity needed for the transmission of the sequence of channel estimates \( \hat{\mathbf{H}} \). This leads to the condition \( C_{p,j} = \frac{1}{T} I(\mathbf{H}, \hat{\mathbf{H}}) \), namely

\[
C_{p,j} = \frac{N_{r,j}}{T} \log_2 \left( \prod_{i=1}^{N_{r,j}} \frac{\sigma_{b,ij}^2}{\left( \sigma_{p,j}^2 + \sigma_{d,j}^2 \right)^{N_{r,j}}} \right). \quad (6)
\]

where we have used the test channel defined by (1). Using the well-known Wyner-Ziv theorem (see, e.g., [8, Sec. 11.3]), the rate \( C_{d,j} \) needed to compress the data received signal \( \mathbf{Y}_{d,j} \) given the side information \( \mathbf{Y}_{d,S_j}, \hat{\mathbf{H}}_{S_j} \) available at the CU about the previously processed BSs according to order \( \pi \) is given by (7). The optimization (5) thus requires a one-dimensional search over \( C_{p,j} \) or \( C_{d,j} \).

**C. Joint Compression of Channel and Received Data Signal**

Here we propose a more sophisticated method to convey the sequence of the channel estimates \( \hat{\mathbf{H}}_j \) in (1) and of received data signals \( \mathbf{Y}_{d,j} \) in (2) over the backhaul link. This method leverages the fact that channel estimates \( \hat{\mathbf{H}}_j \) and received signals \( \mathbf{Y}_{d,j} \) are correlated. As in Sec. III-B, we assume an uncorrelated compression covariance \( R_{d,j} = \sigma_{d,j}^2 I_{N_{r,j}} \) in the test channel (2) and hence we are interested in optimizing the pair \( (\sigma_{p,j}^2, \sigma_{d,j}^2) \). Given side information \( \hat{\mathbf{Y}}_{d,S_j} \) and \( \hat{\mathbf{H}}_{S_j} \), the backhaul constraint (5b) can be formulated as

\[
C_j = \frac{1}{T} I \left( \mathbf{Y}_{d,j}, \hat{\mathbf{H}}_j, \hat{\mathbf{Y}}_{d,j}, \hat{\mathbf{H}}_j, \hat{\mathbf{Y}}_{d,S_j}, \hat{\mathbf{H}}_{S_j} \right) \geq \frac{T_d}{T} \left[ \log_2 \det \left( \mathbf{H}_j \mathbf{R}_{X|\mathbf{Y}_{S_j},\mathbf{H}} \mathbf{H}_j + \left( \sigma_{p,j}^2 + \sigma_{d,j}^2 \right) \mathbf{R}_{X|\mathbf{Y}_{S_j},\mathbf{H}} + I_{N_{r,j}} \right) \right] - N_{r,j} \log_2 (\sigma_{d,j}^2) + C_{p,j}, \quad (8)
\]

where the second inequality can be shown similar to the derivations in [7, Appendix A]. The problem (5) then requires an optimization over parameters \( (\sigma_{p,j}^2, \sigma_{d,j}^2) \).

**D. Joint Adaptive Compression of Channel and Received Data Signal**

In this section, we introduce an improved method for joint compression of channel and received data signal. The main idea is that of adapting the covariance matrix \( R_{d,j} \) of the compression noise added to the data signal (see (2)) to the channel estimate in each channel coherence block. We start by observing that (8) suggests that joint compression can be performed in two steps: (i) first, the channel estimate sequence in compressed with required backhaul rate \( \frac{1}{T} I(\mathbf{H}_j; \hat{\mathbf{H}}_j) \); (ii) then, given that the sequence of channel estimates \( \hat{\mathbf{H}}_j \) for all coherence blocks is known at both the \( j \)-th BS and the CU, the \( j \)-th BS uses a different compression strategy for the quantization of \( \mathbf{Y}_{d,j} \) depending on the value of \( \mathbf{H}_j \). Based on this observation, we propose here to adapt the choice of matrix \( R_{d,j} \) to the current estimate \( \mathbf{H}_j \) and the estimates \( \hat{\mathbf{H}}_{S_j} \) of the previously selected BSs for each coherence block. To emphasize this fact, we use the notation \( R_{d,j}(\hat{\mathbf{H}}_{S_j}, \mathbf{H}(j)) \).

The backhaul constraint in (5) is given by (8), which, following similar steps as in [7, Appendix A], leads to

\[
C_{p,j} \geq C_{p,j} + \frac{T_d}{T} E \left[ \log_2 \det \left( I_{N_{r,j}} + \mathbf{R}_{Y_j|\mathbf{Y}_{S_j},\mathbf{H}_j} \mathbf{R}_{Y_j|\mathbf{Y}_{S_j},\mathbf{H}_j}^{-1} \mathbf{I}_{N_{r,j}} \right) \right], \quad (10)
\]

where we have \( \mathbf{R}_{Y_j|\mathbf{Y}_{S_j},\mathbf{H}_j} = \mathbf{H}_j \mathbf{R}_{X|\mathbf{Y}_{S_j},\mathbf{H}_j} \mathbf{H}_j + \sigma_{p,j}^2 \mathbf{I} \). Problem (5) then requires an optimization over \( \sigma_{p,j}^2 \) and \( R_{d,j}(\hat{\mathbf{H}}_{S_j}, \mathbf{H}(j)) \). Optimization over \( R_{d,j}(\hat{\mathbf{H}}_{S_j}, \mathbf{H}(j)) \) for fixed \( \sigma_{p,j}^2 \) can be done in closed form, as reported in [7, Proposition 2] following the analysis in [3].
In this section, we evaluate the performance of the proposed compression strategies in the uplink of a multi-cell system consisting of $N_B = 2$ BSs and $N_M = 2$ MSs. The BSs and the MSs have two transmit and receive antennas, respectively. We compare the three proposed ECF compression methods and also consider the more conventional compression method based on the CFE strategy in terms of ergodic achievable sum-rate. As detailed in [7], the CFE scheme is also implemented by assuming distributed source coding and successive compression for quantization of the data signals and differs only in the way CSI is quantized. Specifically, with CFE, the received pilot signal $Y_{p,j}$ (see Sec. II-A) is quantized and MMSE channel estimation is performed at the CU. Throughout, we assume that each BS has the same backhaul capacity $C_j = C$ for $j \in N_B$. Moreover, we set $H_j = 1_{N_{t,j} \times N_{r,i}}$. We optimize over the power allocation $(P_p, P_d)$ and we set $T_p = N_t$ as in [6].

Fig. 2 plots the ergodic sum-rate obtained as a function of the constrained backhaul capacity for coherence time $T = 10$, power $P = 20dB$, channel gain $\alpha_{ji} = 1$ for all $j \in N_B$, $i \in N_M$, and Rayleigh fading channel, i.e., $K = 0$. Fig. 2 shows that the ECF strategy is generally advantageous over the CFE. Moreover, it is seen that progressively more complex ECF schemes have better performance, with the joint adaptive strategy outperforming the joint approach and the separate strategy. Finally, we observe that the advantages obtained by more complex compression strategies are especially pronounced in the region of moderate backhaul capacity, in which the backhaul capacity is at a premium and should be used efficiently. We show the impact of the Rician factor $K$ in Fig. 3 with backhaul capacity $C = 6$, power $P = 20dB$, coherence time $T = 20$ and channel gain $\alpha_{ji} = 1$ for all $j \in N_B$, $i \in N_M$. We observe that the performance of the joint adaptive compression method approaches that of the joint compression method as the Rician factor $K$ increases. This is because the joint adaptive compression scheme is based on an optimization of the compression strategy that adapts the quantization error on the data signal to the channel estimates for each coherence block. Therefore, in the presence of reduced channel variations due to a larger Rician factor $K$, the performance gain of the adaptive joint approach is decreased.

V. CONCLUSION

In this paper, we have studied the design of the backhaul compression strategies for the uplink of network MIMO systems by accounting for both CSI and data transfer from the BSs to the CU. Motivated by the information-theoretic optimization of separate estimation and compression, we have adopted an ECF approach with distributed source coding, whereby the BSs first estimate the CSI and then forward the compressed CSI to the CU. From numerical results, we have observed that the ECF approaches outperform the CFE approach, and that more complex joint compression strategies have significant advantages in the regime of intermediate backhaul capacity, in which the backhaul capacity should be used efficiently.

REFERENCES

[9] M. Medard, “The effect upon channel capacity in wireless communica-

878