Interference Alignment with Analog Channel State Feedback

Omar El Ayach and Robert W. Heath, Jr.

Abstract

Interference alignment (IA) is a multiplexing gain optimal transmission strategy for the interference channel with an arbitrary number of users. While the achieved sum rate with IA is much higher than previously thought possible, the improvement comes at the cost of requiring network channel state information at the transmitters. This can be achieved by explicit feedback, a flexible yet costly approach that incurs large overhead and limits throughput. We propose using analog feedback as an alternative to limited feedback or reciprocity based alignment. We show that the full multiplexing gain observed with perfect channel knowledge is preserved by analog feedback and the mean loss in sum rate is bounded by a constant when signal-to-noise ratio is comparable in both forward and feedback channels. When such feedback quality is not quite possible, a fraction of the degrees of freedom is achieved. We consider the overhead of training and feedback and use this framework to optimize the system’s effective throughput. We present simulation results to demonstrate the performance of IA with analog feedback, verify our theoretical analysis, and extend our conclusions on optimal training and feedback length.

I. INTRODUCTION

The interference channel is an information theoretic concept that models wireless networks in which several transmitters simultaneously communicate data to their paired receivers. Communication in such channels has often relied on orthogonal resource allocation, resulting in decaying per-user data rates as networks grow. Recent work on the interference channel, however, has shown that networks can support sum rates that, under certain assumptions, scale linearly...
with the number of users at high signal-to-noise ratios (SNR) [2]. This multiplexing gain can be achieved by a transmission strategy known as interference alignment (IA) [3].

Interference alignment is a linear precoding technique that attempts to align interfering signals in time, frequency, or space. In multiple-input multiple-output (MIMO) networks, IA utilizes the spatial dimension offered by multiple antennas for alignment. By aligning interference at all receivers, IA reduces the number of discernible interfering signals allowing users to cancel interference and decode their desired signals interference free. When coding over infinitely many dimensions, IA allows users to communicate at approximately half their interference free rate [2]. While such a general result has not been proven for the constant MIMO channel, multiple antennas can provide a practical source of dimensionality and have been shown to perform close to the optimal solution [4], [5]. While various methods of calculating MIMO IA precoders have been proposed [4], [6]–[12], all solutions assume some form of channel state information (CSI) at the transmitter obtained either implicitly, through channel reciprocity, or explicitly through feedback. As a result, an efficient method to transfer CSI back to the transmitters must be devised to realize the sum capacities promised by interference alignment.

Interference alignment was introduced in the seminal work in [2], [3] where it was used to prove the achievability of the maximum degrees of freedom in a time varying single-input single-output (SISO) channel. Since then, IA has been studied further [5] and several methods of finding precoders both algorithmically [4], [6]–[10] and in closed form [11], [12] have been proposed. While these solutions differ in the tools used to solve IA’s system of equations, they all require the same amount of channel information obtained via feedback or reciprocity. Reciprocity, however, can not be assumed in general as it does not apply to frequency division duplexed links, and further it requires accurate calibration. Therefore, low overhead feedback strategies that preserve the channel’s optimal multiplexing gain must be used to satisfy the CSI requirement. Existing work on tackling the overhead problem is limited to network design strategies such as partitioning to reduce the required CSI [13]. An alternative approach to reduce feedback requirements is to employ channel state quantization. For example, [14] uses Grassmannian codebooks to directly quantize and feedback the channel coefficients in single antenna systems with frequency extensions. This was later extended in [15] to multiple-input single-output systems; MIMO was only briefly considered. In both [14] and [15], multiplexing gain is preserved by scaling the number of feedback bits, and effectively codebook size, with SNR. The overhead of quantized
feedback, however, increases with codebook size. Large Grassmannian codebooks are hard to design, encode, and can incur large feedback delays.

In this paper we propose using IA with a low overhead CSI feedback strategy based on analog feedback [16]. Instead of quantizing the channel state information, analog feedback, in its simplest form, directly transmits the channel matrix elements as uncoded quadrature amplitude modulated symbols. In the proposed strategy, receivers estimate the forward channels, train the feedback channels, and then feedback the forward channel matrices using analog feedback.

We show that, under mild assumptions on feedback quality, performing IA on the channel estimates obtained via analog feedback incurs no loss in multiplexing gain. This first result was presented in part in our previous work [1]. Specifically, using a cooperative analog feedback strategy, we show that when the SNR on both forward and reverse channels is comparable, the loss in sum rate achieved by a linear zero-forcing receiver is upper bounded by a constant. Constant loss is a slightly stronger result than the preservation of multiplexing gain. We extend the performance analysis, both in theory and simulation, to demonstrate that other analog feedback strategies, which assume no cooperation, perform similarly. Using analog feedback allows IA to approach the performance predicted in [17]. Moreover, we show that when such symmetry is not possible, the system still achieves a fraction of its degrees of freedom. Using the derived bounds on loss in sum rate due to imperfect CSI, we consider the system’s effective throughput, accounting for the overhead incurred by training and feedback. We use our constructed framework to optimize training and feedback lengths to find the optimal operating point on the trade off curve between overhead and sum rate.

This paper extends the analog feedback framework that was originally proposed for the MISO broadcast channel [16], [18]–[21], to the MIMO interference channel. We study the performance of analog feedback in interference channels. We show that analog feedback is a viable alternative to quantization based schemes presented in [14], [15]. As a further advantage, our strategy does not suffer from the same exploding complexity at high SNR as in [14] and [15], and the required feedback scaling will come naturally in most wireless ad hoc networks. We also study the effect of feedback and training overhead on IA’s performance, a concept which, with the exception of [13], has been mostly neglected.

This paper is organized as follows. Section II introduces the MIMO interference channel system model in the presence of interference for the forward and reverse channels and reviews the
concept of interference alignment. Section III introduces the analog feedback scheme proposed and proves the preservation of the achievable multiplexing gain with imperfect CSI. Section IV introduces the concept of overhead and presents results on the optimal training and feedback lengths. We then present simulation results to verify the claims made throughout the paper in Section V and conclude with Section VI.

Throughout this paper we use the following notation: \( A \) is a matrix; \( a \) is a vector; and \( a \) is a scalar; \( A^* \) and \( a^* \) denote the conjugate transpose of \( A \) and \( a \) respectively; \( \|A\|_F \) is the Frobenius norm of \( A \); \( \|a\|_p \) is the \( p \)-norm of \( a \); \( I_N \) is the \( N \times N \) identity matrix; \( \mathbb{C}^N \) is the \( N \)-dimensional complex space; \( \mathcal{CN}(a, A) \) is a complex Gaussian random vector with mean \( a \) and covariance matrix \( A \). Other notation is defined when needed.

II. SYSTEM MODEL AND BACKGROUND

In this section we introduce the MIMO interference channel model under consideration. Then we review the concept of interference alignment, based on perfect channel state information, with a focus on the equations and properties that will be used in our analysis of analog feedback in later sections.

A. Interference Channel Model

Consider the narrowband MIMO interference channel with \( K \) users shown in Fig. 1 in which each source node \( i \) communicates with its sink node \( i \) and interferes with all other sink nodes, \( \ell \neq i \). We refrain from using the term transmitter-receiver since all nodes will be involved in both transmission and reception in either payload transmission or feedback intervals. Instead, we use the source-sink terminology, motivated by the direction of the payload data flow. For simplicity of exposition, we consider the case of the homogeneous network where each source and sink is equipped with \( N_t \) and \( N_r \) antennas respectively. The results can be generalized to a network with a different number of antennas per node, as long as IA remains feasible [5]. Therefore, node \( i \) transmits \( d_i \leq \min(N_t, N_r) \) independent spatial streams to its corresponding sink.

We consider a block fading channel model in which channels are drawn from the same zero-mean continuous distribution independently across all users and antennas and remain fixed for
the interval of interest. We assume perfect time and frequency synchronization when expressing received baseband signals.

Under our assumptions, the received signal at sink node \( i \) can be written as

\[
y_i = \sqrt{\frac{P_d}{d_i}} \mathbf{H}_{i,i} \mathbf{F}_i \mathbf{s}_i + \sum_{\ell \neq i} \sqrt{\frac{P_d}{d_\ell}} \mathbf{H}_{i,\ell} \mathbf{F}_\ell \mathbf{s}_\ell + \mathbf{v}_i,
\]

where \( y_i \) is the \( N_r \times 1 \) received signal vector, \( \mathbf{H}_{i,\ell} \) is the \( N_r \times N_t \) channel matrix from source \( \ell \) to sink \( i \), \( \mathbf{F}_i \) is the \( N_t \times d_i \) unitary precoding matrix used at source \( i \), \( \mathbf{s}_i \) is the \( d_i \times 1 \) transmitted symbol vector at node \( i \), with unit norm elements, i.e. \( \mathbb{E} [\| \mathbf{s}_i \|^2] = d_i \), and \( \mathbf{v}_i \) is a complex vector of i.i.d. circularly symmetric white Gaussian noise with covariance matrix \( \sigma^2 \mathbf{I}_{N_r} \).

We place no assumption on the reciprocity of the forward and reverse channel. This is similar to a frequency division duplexed (FDD) system model, for example, in which the forward and reverse channels are uncorrelated. For the reverse or feedback channel we write the received signal at source node \( i \) as

\[
\mathbf{\tilde{y}}_i = \sqrt{\frac{P_f}{N_t}} \mathbf{G}_{i,i} \mathbf{\tilde{x}}_i + \sum_{\ell \neq i} \sqrt{\frac{P_f}{N_t}} \mathbf{G}_{i,\ell} \mathbf{\tilde{x}}_\ell + \nu_i,
\]

(1)

where \( P_f \) is the transmit power used to transmit pilot and feedback symbols, \( \mathbf{G}_{i,\ell} \) is the \( N_t \times N_r \) reverse channel between sink node \( i \) and source node \( \ell \) with i.i.d. \( \mathcal{CN}(0,1) \) elements\(^1\), \( \mathbf{\tilde{x}}_i \) is the symbol vector with unit norm elements sent by sink \( i \), and \( \nu_i \) is a complex vector of i.i.d. circularly symmetric white Gaussian noise with covariance matrix \( \sigma^2 \mathbf{I}_{N_t} \).

B. Interference Alignment

Interference alignment for the MIMO interference channel is a linear precoding technique that can achieve the maximum multiplexing gain, or degrees of freedom defined as \( \lim_{P \to \infty} R_{\text{sum}} \log_2 P \), which in this case is \( \frac{K N_r}{2} \), when coding over infinitely many channel extensions [11]. While the maximum multiplexing gain may not be achieved without time extensions, interference alignment for the constant MIMO channel can still provide an increase in sum rate [22]. To do so, given global channel knowledge, interference alignment computes the transmit precoders \( \mathbf{F}_i \) to align interference at all receivers in a strict subspace of the received signal space, thus

\(^1\text{Note the effective reversal in the indexing of the channel, while the indexing still has the form “sink,source” the transmit receive roles have been switched.}\)
leaving interference free dimensions for the desired signal. While interference alignment is only one of the many precoding strategies for the interference channel [4], [23], [24], some of which marginally outperform it at low SNR [4], it is analytically tractable. Its complete interference suppression properties make it especially amenable to the study of performance with feedback and imperfect CSI.

To express the conditions for alignment, consider the \( K \)-user interference channel with precoding presented in Section II-A and any corresponding achievable degree of freedom allocation vector \( \mathbf{d} = [d_1\ d_2\ \ldots\ d_K] \). Source node \( i \) sends its \( d_i \) spatial streams along the columns \( f_i^\ell \) of the precoder \( \mathbf{F}_i \), resulting in a transmitted symbol

\[
x_i = \frac{1}{\sqrt{d_i}} \mathbf{F}_is_i = \frac{1}{\sqrt{d_i}} \sum_{\ell=1}^{d_i} f_i^\ell s_i^\ell \quad i = 1, \ldots, K
\]

where we note that \( \|f_i^\ell\|^2 = 1 \) and \( \|s_i^\ell\|^2 = 1 \) to satisfy the total power constraint with equality. We assume equal power allocation since the gain observed from water-filling is at most a constant and thus is negligible at high SNR [25].

While in general interference alignment can be used with any receiver design, the discussion and proofs in this paper assume a linear zero-forcing receiver in which sink node \( i \) projects its received signal on to the columns, \( \mathbf{w}_i^\ell \), of the \( N_r \times d_i \) combiner \( \mathbf{W}_i \). Simulations in Section V indicate that the same performance can be expected from an optimal receiver.

Writing the per stream input-output relation after projection gives

\[
(\mathbf{w}_i^m)^* \mathbf{y}_i = (\mathbf{w}_i^m)^* \sqrt{P} \mathbf{H}_{i,i} f_i^m s_i^m + \sum_{\ell \neq m} (\mathbf{w}_i^m)^* \sqrt{P} \mathbf{H}_{i,i} f_i^\ell s_i^\ell + \sum_{k \neq i} \sum_{\ell = 1}^{d_k} (\mathbf{w}_i^m)^* \sqrt{P} \mathbf{H}_{i,k} f_k^\ell s_k^\ell + (\mathbf{w}_i^m)^* \mathbf{v}_i,
\]

for \( m \in \{1, \ldots, d_i\} \) and \( i \in \{1, \ldots, K\} \), where \( \|\mathbf{w}_i^m\|^2 = 1 \). At the output of these linear receivers \( \mathbf{w}_i^\ell \), the conditions for perfect interference alignment can be restated as

\[
(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} f_i^\ell = 0, \quad \forall i, \ell \neq m
\]

\[
(\mathbf{w}_i^m)^* \mathbf{H}_{i,k} f_k^\ell = 0, \quad \forall k \neq i, \text{ and } \forall m, \ell
\]

\[
| (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} f_i^m | \geq c > 0, \quad \forall i, m
\]

where interference alignment is guaranteed by the first two conditions, and the third ensures the decodability of the \( d_i \) desired streams.
The suboptimal sum rate achieved by the linear zero-forcing receiver, assuming Gaussian input signals and treating interference as noise, is

\[ R_{\text{sum}} = \sum_{i=1}^{K} \sum_{m=1}^{d_i} R^m_i = \sum_{i=1}^{K} \sum_{m=1}^{d_i} \log_2 \left( 1 + \frac{P}{\mathcal{I}_{i,m}^1 + \mathcal{I}_{i,m}^2 + \sigma^2} \| (w^m_i)^* H_{i,i} f^m_i \|^2 \right) , \]

where \( \mathcal{I}_{i,m}^1 \) and \( \mathcal{I}_{i,m}^2 \) are the inter-stream and inter-user interference, respectively. These sum interference terms are given by

\[ \mathcal{I}_{i,m}^1 = \sum_{\ell \neq m} \frac{P}{d_i} \| (w^m_i)^* H_{i,i} f^\ell_i \|^2 , \]
\[ \mathcal{I}_{i,m}^2 = \sum_{k \neq i} d_k \sum_{\ell = 1}^{d_k} \frac{P}{d_k} \| (w^m_i)^* H_{i,k} f^\ell_k \|^2 . \]

In the presence of perfect channel knowledge, and for an achievable degree of freedom vector \( \mathbf{d} = [d_1, d_2, \ldots, d_K] \), equations (2), (3), and (4) can be satisfied with probability one and thus \( \mathcal{I}_{i,m}^1 = \mathcal{I}_{i,m}^2 = 0 \). This gives

\[ \lim_{P \to \infty} \frac{R_{\text{sum}}}{\log_2 P} = \lim_{P \to \infty} \frac{\sum_{i,m} \log_2 \left( 1 + \frac{P}{\| (w^m_i)^* H_{i,i} f^m_i \|^2} \right)}{\log_2 P} = \sum_{i=1}^{K} d_i \leq \frac{KN_r}{2} . \]

It is not immediately clear, however, if the same sum rate scaling behavior can be expected from a network with only imperfect knowledge of the channel derived from noisy feedback. Results on single user MIMO prove an acceptable constant loss in sum rate due to imperfect channel state information [26]. In multi-user scenarios, however, the cost of imperfect channel knowledge may be much higher, potentially resulting in the loss of the channel’s multiplexing gain [27] which saturates achieved sum rate at high SNR [28]. In Section III, we show that such performance can be expected from a realistic system via interference alignment provided that the quality of channel knowledge scales sufficiently with transmit power, or effectively the forward channel’s SNR. This is similar to the results presented in [14] and [15] for SISO channels where the feedback scaling is in terms of codebook size. We discuss several advantages of using the feedback scheme proposed in this paper over limited feedback quantization.
III. INTERFERENCE ALIGNMENT WITH ANALOG FEEDBACK

In this section we propose a feedback and transmission strategy based on analog feedback and interference alignment, which uses the estimated channels as if they were the true propagation channels.

A. Analog Feedback

To feedback the forward channel matrices $H_{i,\ell}$ reliably across the feedback channel given in (1), we propose dividing the feedback stage into two phases. First, each source learns its reverse channels. Second, the forward channels are fed back and estimated. We neglect the initial training phase in which sinks learn the forward channels and, thus, assume they have been estimated perfectly. This is since imperfect CSI at the receiver will adversely affect performance of all feedback schemes and is not exclusive to analog feedback. In fact, any error in estimating the forward channels will only add an error term to the forward channels in (7), which also decays with transmit power, and thus similar results can be shown.

1) Reverse Link Training: To learn all reverse links, the $K$ sink nodes must transmit pilot symbols over a period $\tau_p \geq KN_r$. Similar to the analysis done in [16], we let the sinks collectively transmit a $KN_r \times \tau_p$ unitary matrix of pilots $\Phi$ shown to be optimal in [29]. Such reverse channel training requires no assumptions other than training sequences be known to all sink nodes, to guarantee orthogonality. Of course, if we enforce orthogonality in time, this assumption reduces to synchronization, and exact sequences need not be known to all sinks.

Let $\tilde{Y}_i = [\tilde{y}_i[1], \ldots, \tilde{y}_i[\tau_p]]$ be the $N_t \times \tau_p$ received training matrix at source node $i$, and let $\tilde{Y}_p = [\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_K]$ be the composite received training matrix. We write the received training as

$$\tilde{Y}_p = \sqrt{\frac{\tau_p P_f}{N_r}} G \Phi + V,$$

where $G$ is the $KN_t \times KN_r$ composite reverse channel matrix between all sinks and sources and $V$ is a $KN_t \times \tau_p$ matrix with i.i.d $\mathcal{CN}(0, \sigma^2 I_{N_r})$ elements. Using the received training, each source derives an MMSE estimate of its channels all of which can be together written as

$$\hat{G} = \frac{\sqrt{\tau_p P_f}}{\sigma^2 + \frac{\tau_p P_f}{N_r}} \tilde{Y}_p \Phi^*.$$ (6)
Since $\hat{G}$ is an MMSE estimate of Gaussian random variables corrupted by Gaussian noise, this results in $\hat{G} \sim \mathcal{CN} \left( 0, \frac{\tau p_f}{N_r} \right)$ and $G = G - \hat{G}$ with i.i.d. $\mathcal{CN} \left( 0, \frac{\sigma^2}{\sigma^2 + \frac{\tau p_f}{N_r}} \right)$ elements. No collaboration between sources is needed in (6); each node estimates its channels based on its own received signals which we have concatenated in (6) and written in terms of the composite received training matrix, $\bar{Y}_{p}$, to conserve space.

2) Analog CSI Feedback: After training the reverse channels, each sink node $i$ directly sends its unquantized uncoded estimates of $H_{i,\ell} \forall \ell$ over a period $\tau_c$. To have the sink nodes feedback their CSI simultaneously while maintaining orthogonality, each node post multiplies its $N_r \times K N_t$ feedback matrix $[H_{i,1} \ldots H_{i,K}]$ with a unitary $K N_t \times \tau_c$ matrix $\Psi_i^*$ such that $\Psi_i^* \Psi_i = I_{K N_r} \delta_{i,\ell}$, a general orthogonal feedback structure that can capture the case of orthogonality in time [16]. This requires $\tau_c \geq K N_t N_r$. The transmitted $N_r \times \tau_c$ feedback matrix $\bar{X}_i$ from node $i$ can be written as

$$\bar{X}_i = \sqrt{\frac{\tau_c P_f}{K N_t N_r}} [H_{i,1} \ldots H_{i,K}] \Psi_i^*.$$

The concatenated, or composite received feedback $K N_t \times \tau_c$ matrix is then given by

$$\bar{Y}_c = \sqrt{\frac{\tau_c P_f}{K N_t N_r}} \sum_{i=1}^{K} \begin{bmatrix} G_{i,1} \\ \vdots \\ G_{i,K} \end{bmatrix} [H_{i,1} \ldots H_{i,K}] \Psi_i^* + V$$

where $V$ is now a $K N_t \times \tau_c$ noise matrix.

To estimate the forward channels $H_{i,\ell}$, the source nodes first isolate the training from sink node $i$ by post multiplying their received training by $\Psi_i$ which gives

$$\bar{Y}_c \Psi_i = \sqrt{\frac{\tau_c P_f}{K N_t N_r}} \begin{bmatrix} G_{i,1} \\ \vdots \\ G_{i,K} \end{bmatrix} [H_{i,1} \ldots H_{i,K}] + V \Psi_i.$$  \hspace{1cm} (7)

To simplify the analysis in Section III-B, we assume sources share the complete matrix $\bar{Y}_c$, and
effectively compute a common least squares estimate \( \hat{H}_i \) of \( H_i \) given by

\[
\hat{H}_i = \sqrt{\frac{K N_i N_r}{\tau_c P_f}} \left( \tilde{G}_i^* \tilde{G}_i \right)^{-1} \tilde{G}_i^* \tilde{Y}_c \Psi_i,
\]


which makes it clear that the error in the estimate consists of two error terms: the first due to noisy feedback and the second due to a noisy estimate of the feedback channel. To quantify the effect of the error on the achieved sum rate, we derive the variance of the error term introduced by analog feedback. Recall that the elements of \( H_{i,\ell} \) are \( CN(0,1) \), those of \( V \) are \( CN(0,\sigma^2) \), and those of \( \tilde{G}_i \) are \( CN(0,\frac{\sigma^2}{\sigma^2 + \frac{\tau_p P_f}{N_r}}) \). As a result, the error term \( \tilde{G}_i H_i \) due to the reverse channel estimation has independent elements with a variance of \( \frac{N_r \sigma^2}{\sigma^2 + \frac{\tau_p P_f}{N_r}} \). Similarly to [16] we see that the covariance of each columns of \( \tilde{H}_i \) denoted \( \tilde{H}_i^{(\ell)} \), conditioned on \( \tilde{G}_i \) is

\[
\text{Cov}(\tilde{H}_i^{(\ell)}|\tilde{G}_i) = \left( \frac{K N_i N_r \sigma^2}{\tau_c P_f} + \frac{N_r \sigma^2}{\sigma^2 + \frac{\tau_p P_f}{N_r}} \right) \left( \tilde{G}_i^* \tilde{G}_i \right)^{-1}.
\]

Since the elements of the MMSE estimate \( \tilde{G}_i \) are Gaussian and uncorrelated, the diagonal elements of \( \left( \tilde{G}_i^* \tilde{G}_i \right)^{-1} \) are reciprocals of scaled chi-squared random variables with \( 2(K N_i - N_R + 1) \) degrees of freedom [16]. As a result, the mean square error, \( \sigma_f^2 \), in the elements of \( \tilde{H}_{i,\ell} \) is given by

\[
\sigma_f^2 = \frac{\sigma^2}{(K N_i - N_r) P_f} \left( \frac{N_r^2}{\tau_p} + \frac{K N_i N_r}{\tau_c} \left( 1 + \frac{N_r \sigma^2}{\tau_p P_f} \right) \right).
\]

At high SNR this gives

\[
\sigma_f^2 \approx \frac{\sigma^2 \left( \frac{N_r^2}{\tau_p} + \frac{K N_i N_r}{\tau_c} \right)}{(K N_i - N_r) P_f}.
\]
Having computed feedback error, we return to the assumption on node cooperation. Source cooperation simplifies analysis by effectively making a common \( \hat{H} \) known to all users, however, this cannot be assumed in practical systems. The first alternative to cooperation is to have one user calculate precoders and combiners, and feed them forward to the other users as in [30]. The other alternative is for each node to calculate its precoder and combiner based on the feedback only it receives. We refer to this as the “distributed processing” approach. Due to independent noise across users in the feedback stage, distributed processing implies that users will calculate vectors based on mismatched information, the effect of which can also be bounded. We elaborate on this after presenting our results in Sections III-B and V and show that the cost of no cooperation is limited.

### B. Multiplexing Gain with Analog Feedback

To characterize the performance of interference alignment with analog feedback, we examine the mean loss in sum-rate [27] incurred by naive interference alignment where the channel estimates obtained at the sources are used, as if they were the true channels, to calculate the columns of the precoders, \( f_i^m \forall i, m \) and combiners \( w_i^m \forall i, m \). Transmit and receive vectors are calculated to satisfy (2), (3), and (4) using the estimated channels.

The mean loss in sum-rate is defined as

\[
\Delta R_{sum} \triangleq E[H]R_{sum} - E[H]\hat{R}_{sum}, \tag{10}
\]

where \( E[H]R_{sum} \) is the average sum rate from interference alignment with perfect CSI, with instantaneous rate given in (5), and \( E[H]\hat{R}_{sum} \) is the average rate with CSI obtained via feedback.

**Theorem 1:** Interference alignment on the \( K \)-user \( N_r \times N_t \) interference channel with imperfect channel state information obtained via the analog feedback strategy described in Section III-A achieves the same average sum-rate scaling observed with perfect interference alignment as long as the feedback power \( P_f \) scales with the transmit power \( P \). Thus, the original multiplexing gain is preserved. Moreover, the mean loss in sum rate \( \Delta R_{sum} \) is \( O(1) \).

**Proof:** Let the \( K \)-user \( N_r \times N_t \) interference channel use the analog feedback scheme presented to achieve a vector of multiplexing gains \( d \). Let the transmit precoding and receive
combining vectors be calculated to satisfy

\[(\hat{w}_i^m)^* \hat{H}_{i,i} \hat{f}_i^m = 0, \quad \forall i, \ell \neq m \]  \hspace{1cm} (11)

\[(\hat{w}_i^m)^* \hat{H}_{i,k} \hat{f}_{k}^m = 0, \quad \forall k \neq i, \text{ and } \forall m, \ell \]  \hspace{1cm} (12)

\[| (\hat{w}_i^m)^* \hat{f}_m | \geq c > 0, \quad \forall i, m. \]  \hspace{1cm} (13)

Using these precoding and combining vectors, the input-output relationship at the output of a linear zero-forcing receiver is

\[(\hat{w}_i^m)^* y_i = (\hat{w}_i^m)^* \sqrt{P} \hat{H}_{i,i} \hat{f}_i^m s_i^m + \sum_{\ell \neq m} (\hat{w}_i^m)^* \sqrt{P} \hat{H}_{i,i} \hat{f}_{k}^m s_{\ell}^m + \sum_{k \neq i} \sum_{\ell = 1}^{d_k} (\hat{w}_i^m)^* \sqrt{P} \hat{H}_{i,k} \hat{f}_{k}^m s_{\ell}^m + (\hat{w}_i^m)^* v_i. \]  \hspace{1cm} (14)

Using the received signal in (2), the instantaneous rate expression in (5), and the sum rate loss defined in (10), this gives the following upper bound on mean loss in sum rate:

\[
\Delta R_{\text{sum}} = \mathbb{E}_{\mathbf{H}} \sum_{i,m} \log_2 \left( 1 + \frac{P}{d_i} | (w_i^m)^* H_{i,i} f_i^m |^2 \right) - \mathbb{E}_{\mathbf{H},\hat{H}} \sum_{i,m} \log_2 \left( 1 + \frac{P}{d_i} | (\hat{w}_i^m)^* \hat{H}_{i,i} \hat{f}_i^m |^2 \right)
\]

\[
= \mathbb{E}_{\mathbf{H}} \sum_{i,m} \log_2 \left( 1 + \frac{P}{d_i} | (w_i^m)^* H_{i,i} f_i^m |^2 \right)
\]

\[- \mathbb{E}_{\mathbf{H},\hat{H}} \sum_{i,m} \log_2 \left( 1 + \frac{P}{d_i} | (\hat{w}_i^m)^* \hat{H}_{i,i} \hat{f}_i^m |^2 \right)
\]

\[+ \mathbb{E}_{\mathbf{H},\hat{H}} \sum_{i,m} \log_2 \left( 1 + \frac{I_{i,m}}{\sigma^2} \right)
\]

\[\leq \mathbb{E}_{\mathbf{HH}} \sum_{i,m} \log_2 \left( 1 + \frac{I_{i,m}}{\sigma^2} \right)
\]

where (a) is due to the fact that \(w_i^m, \hat{w}_i^m, f_i^m, \text{ and } \hat{f}_i^m\) are independent of \(H_{i,i}\) and therefore \(\frac{P}{d_i} | (w_i^m)^* H_{i,i} f_i^m |^2\) and \(\frac{P}{d_i} | (\hat{w}_i^m)^* \hat{H}_{i,i} \hat{f}_i^m |^2\) are identically distributed. As a result \(\frac{P}{d_i} | (w_i^m)^* H_{i,i} f_i^m |^2\) + \(I_{i,m}\) stochastically dominates \(\frac{P}{d_i} | (\hat{w}_i^m)^* \hat{H}_{i,i} \hat{f}_i^m |^2\) [27]. We now apply Jensen’s inequality to the
upper bound in (a) to get
\[ \Delta R_{\text{sum}} \leq \sum_{i,m} \log_2 \left( 1 + \frac{\mathbb{E}_{H,\tilde{H}} I_{i,m}}{\sigma^2} \right). \] (15)

Since (11), (12), (13) are satisfied, however, the total interference term \( I_{i,m} \) can be simplified to include only residual interference due to the channel estimation errors \( \tilde{H}_{i,\ell} \). Equation (15) can be further upper bounded by noticing that
\[
\left\lVert (\tilde{w}_i^m)^* (H_i,k + \tilde{H}_i,k) \tilde{f}_k \rightVert^2 = \left\lVert (\tilde{w}_i^m)^* \tilde{H}_i,k \tilde{f}_k \right\rVert^2 \quad \forall k, \forall \ell \neq m
\] (16)
which gives
\[
\Delta R_{\text{sum}} \leq \sum_{i,m} \log_2 \left( 1 + \frac{\sum_{\ell=1}^{K} \frac{P_d}{d_i} (d_\ell - \delta_{i,\ell}) \mathbb{E}_{H,\tilde{H}} \lVert \tilde{H}_{i,\ell} \lVert_F^2}{\sigma^2} \right). \] (17)

From (8), however, we have \( \mathbb{E}_{H,\tilde{H}} \lVert \tilde{H}_{i,\ell} \lVert_F^2 = N_t N_r \sigma_f^2 = \frac{c}{P_f} \), where \( c \) is a constant, independent of \( P_f \) at high enough SNR, given by
\[
c = N_t N_r \frac{\sigma_f^2 \left( \frac{N_r^2}{\tau_p} + \frac{KN_t N_r (1+\epsilon)}{\tau_c} \right)}{(KN_t - N_r)}. \] (18)

Combining (17) and (18) gives the final upper bound on throughput loss
\[
\Delta R_{\text{sum}} \leq \sum_{i,m} \log_2 \left( 1 + \frac{P}{\sigma^2 d_i} \left( (d_i - 1) \frac{c}{P_f} + \sum_{\ell \neq i} d_\ell \frac{c}{P_f} \right) \right).
\]

Therefore, if we have \( P_f = \alpha^{-1} P \) we get
\[
\Delta R_{\text{sum}} \leq \sum_i d_i \log_2 \left( 1 + \frac{(\lVert d \rVert_1 - 1)\alpha c}{d_i \sigma^2} \right). \] (19)

The bound has been presented at high SNR for simplicity of exposition only; (18) can be adapted for any SNR > 0 by using (8) instead of (9).

In summary, Theorem 1 states that if feedback power is equal to any constant fraction of transmit power, the cost of imperfect CSI is a constant number of bits, independent of SNR. Since transmit and feedback power are likely to be comparable in practice, this result is promising. Using analog feedback allows the system to overcome the problem of exploding complexity and
codebook sizes in quantization based schemes, which remains even if the actual feedback stage operates error-free and at capacity [27], [31]. In case such CSI quality scaling is not possible, however, systems become interference limited and the multiplexing gain is reduced. We present this case in the following theorem.

**Theorem 2:** Interference alignment on the $K$-user $N_r \times N_t$ interference channel using analog feedback when only $P_f = \alpha P^\beta$ with $0 \leq \beta \leq 1$ is available for feedback achieves at least a $\beta$-fraction of its original multiplexing gain.

**Proof:** The multiplexing gain achieved by a system using naive interference alignment and analog feedback with feedback power $P_f = \alpha P^\beta$ can be written as

$$M = \lim_{P \to \infty} \frac{\mathbb{E}_{\mathbf{H},\mathbf{H}}[R_{\text{analog},\beta}]}{\log_2(P)} = \lim_{P \to \infty} \frac{\mathbb{E}_{\mathbf{H},\mathbf{H}} \sum_{i,m} \log_2 \left( 1 + \frac{P}{2 \mathbb{E}|(\hat{w}_i^m)^\ast \mathbf{H}_{i,i} f_i^m|^2}{I_{i,m} + \sigma^2} \right)}{\log_2(P)}$$

$$= \lim_{P \to \infty} \frac{\mathbb{E}_{\mathbf{H},\mathbf{H}} \sum_{i,m} \log_2 \left( \frac{P}{2 \mathbb{E}|(\hat{w}_i^m)^\ast \mathbf{H}_{i,i} f_i^m|^2 + I_{i,m} + \sigma^2} \right) - \mathbb{E}_{\mathbf{H},\mathbf{H}} \sum_{i,m} \log_2 (I_{i,m} + \sigma^2)}{\log_2(P)}$$

$$\geq \sum_{k=1}^K d_k \cdot \lim_{P \to \infty} \frac{\mathbb{E}_{\mathbf{H},\mathbf{H}} \mathbb{E}_{\mathbf{H}_{i,i},\mathbf{H}_{i,k}} [\| \mathbf{H}_{i,i} \|_F^2 + P \mathbb{E}_{\mathbf{H},\mathbf{H}} [\| \mathbf{H}_{i,k} \|_F^2 + \sigma^2]]}{\log_2(P)}$$

$$\geq \sum_{k=1}^K d_k \cdot \lim_{P \to \infty} \frac{\sum_{i,m} \log_2 \left( \frac{(d_i-1)P^{1-\beta} - c d_i + \sigma^2}{d_i} \right)}{\log_2(P)} = \beta \left( \sum_{k=1}^K d_k \right),$$

(20)

where (a) follows from disregarding interference in the first term and applying Jensen’s to the second; (b) follows from the observation in (16) and from the fact that each term in the first summation has a multiplexing gain of 1. Finally, (c) is due to the resulting scaling of residual interference with $P^{1-\beta}$.

In the case where $\beta = 0$, feedback power is constant, and the interference limited system achieves zero multiplexing gain. In fact, the system simulations in Section V indicate that the sum-rate achieved is upper bounded by a constant. As for the case of $\beta = 1$, Theorem 2 confirms the preservation of full multiplexing gain proven in Theorem 1, but does not establish the $O(1)$
loss in sum rate. As for any $0 < \beta < 1$, Theorem 2 shows that analog feedback and interference alignment can still achieve linear sum rate scaling even when feedback power is much smaller than transmit power. This may be the case in certain non-homogeneous networks such as cellular networks in which mobile and base station powers do not match.

IV. DEGREES OF FREEDOM WITH OVERHEAD

While the analysis done in Section III-B is a good indicator of the cost of imperfect CSI obtained through training and analog feedback, it does not directly predict the expected throughput achieved by this strategy. Namely, the analysis done thus far neglects the cost of training and feedback overhead. In this section we define the expected throughput with overhead and use it to optimize training and feedback.

A. Definition of Overhead

As shown in Section III-B, the performance of IA is tightly related to the mean square error in the channel estimates at the transmitter. When operating in a time varying channel, training and feedback must be done periodically to ensure the validity of the channel estimate at the transmitter. Depending on the channel’s coherence time, the overhead due to training and feedback may consume an arbitrarily large fraction of resources such as time or frequency slots, resulting in low net throughput. In this section we consider the case in which training, feedback, and data transmission are all orthogonal in time, in the same coherence time or frame $T$ [13], [32]. Using this model for training and feedback overhead we compute the expected effective sum rate as

$$R_{\text{eff}} = \left( \frac{T - (\tau_c + \tau_p)}{T} \right) (\mathbb{E}_H R_{\text{sum}} - \Delta R_{\text{sum}}).$$

(21)

Using the upper bound on the mean loss in sum rate at high SNR in (19), we lower bound the expected sum rate with overhead as

$$R_{\text{eff}}(\tau_p, \tau_c) \geq \left( \frac{T - (\tau_c + \tau_p)}{T} \right) \left( \mathbb{E}_H R_{\text{sum}} - \sum_i d_i \log_2 \left( 1 + \frac{(\|d_i\|_1 - 1)\alpha c(\tau_p, \tau_c)}{d_i \sigma^2} \right) \right),$$

(22)

which we explicitly write as a function of the training and feedback lengths. As expected, and as can be seen from (21), insufficient training and feedback may result in poor channel estimates at
the receiver, and thus a large loss in sum rate, whereas excessive training and feedback becomes too costly as a large portion of the frame is spent on overhead.

B. Training and Feedback Optimization

Given the expression for system throughput with overhead, we propose solving the following optimization problem

\[
\max_{\tau_p, \tau_c} \left( \frac{T - (\tau_c + \tau_p)}{T} \right) \left( \mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \sum_i d_i \log_2 \left( 1 + \frac{\alpha(\|d\|_1 - 1)N_tN_r \left( \frac{N_r^2}{\tau_p} + \frac{KN_tN_r}{\tau_c} \right)}{d_i(KN_t - N_r)\left( N_r + \sqrt{KN_tN_r} \right)} \right) \right),
\]

(23)

over the set of feasible training and feedback lengths. The sum rate expression defined in (21) is not convex as it is defined on a non-convex non-continuous closed set of bounded integers, as would be the case if one were optimizing over the number of feedback bits. Nevertheless, we seek to maximize a continuous relaxation of the defined cost function by typical methods of convex optimization.

The optimization can be simplified by realizing that for a fixed overhead length \(T_{\text{total}}\), optimizing training and feedback lengths simplifies to minimizing \(c(\tau_p, \tau_c)\) and then finding the optimal overhead length. So given a fixed amount of overhead the solution of the first optimization step can be shown to be

\[
\tau_p = \frac{N_r}{N_r + \sqrt{KN_tN_r} T_{\text{total}}},
\]

(24)

\[
\tau_c = \frac{\sqrt{KN_tN_r}}{N_r + \sqrt{KN_tN_r} T_{\text{total}}},
\]

(25)

which gives the optimal value \(c_{\text{opt}} = \frac{(N_r + \sqrt{KN_tN_r})^2}{T_{\text{total}}}\). Replacing \(c_{\text{opt}}\) into (23), the optimal training length \(T_{\text{total}}\) can now be found by solving

\[
\frac{\delta}{\delta T_{\text{total}}} \left( \frac{T - T_{\text{total}}}{T} \right) \left( \mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \sum_i d_i \log_2 \left( 1 + \frac{\alpha(\|d\|_1 - 1)N_tN_r \left( N_r + \sqrt{KN_tN_r} \right)}{d_i(KN_t - N_r)T_{\text{total}}} \right) \right) = 0
\]

\[
-\frac{1}{T} \left( \mathbb{E}_{\mathbf{H}} R_{\text{sum}} - \sum_i d_i \log \left( 1 + \frac{\gamma_i}{T_{\text{total}}} \right) \right) + \frac{T - T_{\text{total}}}{T} \left( \sum_i \frac{d_i}{T_{\text{total}}(T_{\text{total}} + \gamma_i)} \right) = 0
\]

(26)

where \(\gamma_i = \frac{\alpha(\|d\|_1 - 1)N_tN_r \left( N_r + \sqrt{KN_tN_r} \right)^2}{d_i(KN_t - N_r)}\). Though the exact solution to this optimization cannot be found in closed form, it can be shown that \(T_{\text{total}}\) increases with the \(\sqrt{T}, \sqrt{\|d\|_1}\), and initially
decreases with the ratio transmit to feedback power $\frac{P}{P_f}$. Fig. 2 plots the optimal feedback vs. frame length, where we have solved the optimization in (26) numerically for a three $2 \times 2$ user network. Fig. 2 verifies the claimed scaling and shows that for a wide range of frame lengths ($< 10^4$), the solution to the optimization problem is less than the minimum length required to satisfy the dimensionality constraints on the training and feedback matrices, i.e. the optimal overhead is minimal in most cases since frames rarely exceed $10^4$ symbols.

While it may not be surprising that longer frames can support more training [32], and that more interference from more data streams requires more training, it’s interesting to note that small mismatches in forward and feedback SNR initially decrease training and feedback lengths. Hence, a slightly noisy feedback channel does not require extra training to compensate. This, however, is not true of significantly poor feedback channels, as the optimal overhead length does in fact increase to improve the quality of CSI. It can also be shown that, all else fixed, the optimal training length decreases with the achieved sum rate or effectively SNR, making analog feedback especially efficient in the high SNR. This is shown if Fig. 3.

V. SIMULATION RESULTS

In this section we present simulation results that validate the claims and proofs given in Sections III-B and IV. We verify the results shown in Theorems 1 and 2 which state that as long as the transmit power on the feedback channel scales sufficiently with the power on the forward channel, the multiplexing gain achieved by perfect interference alignment is preserved. To show the full potential of both IA and IA with analog feedback we remove the restriction of using linear receivers and thus calculate the sum rate of an optimal receiver with ideal decoding as

$$R_{\text{sum}} = \sum_{k=1}^{K} \log_2 \left| I + \left( \sigma^2 I + \sum_{m \neq k} H_{k,m} F_m F_m^* H_{k,m}^* \right)^{-1} \left( H_{k,k} F_k F_k^* H_{k,k}^* \right) \right|,$$

and the precoders are calculated given ideal or estimated CSI [33].

Fig. 4 shows the sum rate achieved by a 3 user $2 \times 2$ channel for ideal CSI, scaling quality CSI where $P_f = P$, slower scaling CSI with $P_f = P^\beta$ and $\beta = 3/4$, and fixed quality CSI where the SNR on the feedback channel is fixed at 20dB. Fig. 4 confirms that both perfect and scaling feedback exhibit the same sum rate scaling or degrees of freedom. This establishes the multiplexing gain optimality of using analog feedback. Fig. 4 also confirms the fact that
the mean loss in sum rate at high enough SNR is indeed a constant independent of the forward channel transmit power or SNR. Moreover, Fig. 4 shows that our zero-forcing based lower bound is tight at high SNR where the optimal receiver is in fact an interference suppression filter. As for the low SNR case, its possible that another receiver design significantly outperforms zero-forcing, which results in a looser lower bound on sum rate. When perfect scaling feedback is not possible, Fig. 4 shows the slower yet linear scaling in the case of $P_f = P^\beta$ which verifies the result shown in Theorem 2 on the preservation of fractional degrees of freedom. For the case of fixed feedback quality, multiplexing gain is zero and the sum rate saturates at high SNR. Though, not analyzed theoretically, Fig. 4 shows that improvements to the IA algorithm, such as MAX SINR [6] exhibits less gain and are thus affected more by imperfect channel information.

Finally, Fig. 4 shows the performance of the distributed processing approach introduced in Section III-A. As stated, the cooperation assumed in Section III-A is not practical. In a real system, either a central node can calculate precoders and send them to all others, adding yet more overhead, or nodes calculate precoders independently. The analysis of the centralized processor strategy is straightforward. It only adds noise to the precoders in (16), which decays with power, thus similar bounds on $\Delta R_{sum}$ can be derived. In the case of distributed processing, the extra loss is due to the mismatch of CSI between nodes; nodes calculating vectors based on different perturbations of the same channels, may not reach the same exact precoders, leading to extra interference leakage. Fig. 4, however, shows that the extra loss in sum rate due to distributed processing is small, and no degrees of freedom are lost. The performance of distributed processing is not theoretically surprising and can be seen by inspecting the closed form IA solution in [2] for example. It can be shown using Wedin’s theorem [34] and results in [35], which bound the angles between singular and invariant subspaces of different perturbed matrices, that even if nodes compute precoders using different perturbations of the same matrices, the angle between the precoders they each calculate is small and decays with feedback power. Such error decay is all that is needed to prove the preservation of multiplexing gain and a constant loss in sum rate, though computing the bound on $\Delta R_{sum}$ becomes more involved. The small loss due to distributed processing, and the fact that it does not require extra overhead, make such an approach practical.

We now simulate the system’s total throughput according to the overhead model presented in Section IV in which training, feedback, and data transmission all happen in the same coherence
interval. Fig. 5 shows that when the forward and reverse channel’s SNR scale together, the optimal feedback length is close to the theoretical minimum required for a frame length of 10,000, as predicted by Fig. 2. Although analog feedback may provide less than optimal overhead scaling, the importance of such scaling results is often overstated, since for realistic frame lengths, overhead is minimal [32]. Fig. 6, which plots sum rate for a fixed feedback power, verifies that a large mismatch in forward and reverse SNR induces more overhead. Finally, Figs 5 and 6 show the optimal feedback length assuming fixed training length for ease of exposition. We know, however, from Section IV, that this is not optimal. Fig. 7 shows that while less resources are used for training, both training and feedback must scale together for maximum throughput.

VI. CONCLUSIONS

In this paper we have proposed a low overhead feedback strategy for the interference channel. We showed that when combined with interference alignment, analog feedback can achieve full multiplexing gain when the forward and reverse channel SNR levels are comparable. When such symmetry is not possible, we showed that a system can still retain a fraction of the its degrees of freedom. Thus, using analog feedback, the cost of imperfect channel knowledge at the transmitter is bounded and quickly becomes negligible at high SNR. The mild assumption on feedback power implies that analog feedback performs well with constant overhead, in the high SNR regime where IA is optimal. In addition to quantifying the cost of imperfect CSI, we show the scaling of required overhead with several network variables such as users, streams, SNR, and frame length. Using this overhead analysis, we show that the throughput loss due to the overhead of training and analog feedback is often minimal.

REFERENCES

[31] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Quantized vs. analog feedback for the MIMO broadcast channel: a comparison between zero-forcing based achievable rates.”
Fig. 1. K-User MIMO interference channel model

Fig. 2. Optimal Overhead length vs. Frame Length for a 3 user $2 \times 2$ system with $T=7000$ and $P_f = P$. 
Fig. 3. Optimal Overhead length vs. $R_{sum}$ for a 3 user $2 \times 2$ system with SNR=40dB and $P_f = P$.

Fig. 4. The sum rate achieved by interference alignment and its hybrid algorithms (MAX SINR) with perfect CSI, scaling feedback quality ($SNR_f = SNR$), and fixed quality feedback with $SNR_f = 20dB$. 
Fig. 5. The effective throughput achieved by interference alignment with $SNR_f = SNR$ and a frame length $T = 10000$. This confirms that the amount of feedback needed to achieve optimal throughput is often minimal.

Fig. 6. The effective throughput achieved by interference alignment with $SNR_f = 10dB$ and a frame length $T = 10000$. This confirms that the amount of feedback needed to achieve optimal throughput does not increase as the difference between forward and reverse channel SNR increases.
Fig. 7. This plots the sum rate achieved by interference alignment at 30dB with and feedback SNR of 10dB and a frame length $T = 10000$. This shows that feedback and training lengths increase simultaneously, i.e. increasing feedback or training length, while keeping the other fixed, quickly becomes counterproductive.