

On the Overhead of Interference Alignment: Training, Feedback, and Cooperation

Omar El Ayach, Angel Lozano, and Robert W. Heath, Jr.

Abstract

Interference alignment (IA) is a cooperative transmission strategy that, under some conditions, achieves the interference channel's maximum number of degrees of freedom. Realizing IA gains, however, is contingent upon providing transmitters with sufficiently accurate channel knowledge. In this paper, we study the performance of IA in multiple-input multiple-output systems where channel knowledge is acquired through training and analog feedback. We design the training and feedback system to maximize IA's effective sum-rate: a non-asymptotic performance metric that accounts for estimation error, training and feedback overhead, and channel selectivity. We characterize effective sum-rate with overhead in relation to various parameters such as signal-to-noise ratio, Doppler spread, and feedback channel quality. A main insight from our analysis is that, by properly designing the CSI acquisition process, IA can provide good sum-rate performance in a very wide range of fading scenarios. Another observation from our work is that such overhead-aware analysis can help solve a number of practical network design problems. To demonstrate the concept of overhead-aware network design, we consider the example problem of finding the optimal number of cooperative IA users based on signal power and mobility.

I. INTRODUCTION

Interference alignment (IA) for the multiple-input multiple-output (MIMO) interference channel is a cooperative transmission strategy that attempts to structure interfering signals such that they occupy a reduced dimensional space when observed at the receivers [1], [2]. Alignment often enables achieving the maximum number of degrees of freedom (DoF) [1], [2]. Precoding

O. El Ayach and R.W. Heath, Jr are with The University of Texas at Austin, Austin, TX 78712 USA (e-mail: {omarayach, rheath}@mail.utexas.edu). A. Lozano is with Universitat Pompeu Fabra, Barcelona, Spain (e-mail: angel.lozano@upf.edu)

The authors at The University of Texas were supported by the Office of Naval Research (ONR) under grant N000141010337. The work of A. Lozano is supported by the FET FP7 Project 265578 "HIATUS".

transmitted signals to carefully align them at the receivers, however, requires knowledge of the interfering channels in the system, collectively known as channel state information (CSI). Perfect CSI is assumed to be available when designing most IA algorithms [1], [3]–[5] or reporting genie-aided IA gains. Practical systems, however, acquire receiver CSI with the help of training sequences or pilots [6]. Such CSI can then be shared with the transmitters via feedback. As a result, practical CSI is imperfect and comes with an overhead signaling cost, both of which penalize the effective data rates achieved. Realizing the gains of IA is therefore contingent upon providing systems with sufficiently accurate CSI at a manageable overhead cost.

Several approaches have been proposed to fulfill IA's transmit CSI requirement [7]–[9], typically assuming perfect CSI at the receiver. The feedback strategy in [7] proposes to use Grassmannian codebooks to compress and improve CSI feedback in single-antenna frequency extended IA systems. The feedback strategy was then extended to multiantenna frequency extended systems in [8]. Both [7] and [8] guarantee that limited feedback preserves the number of DoF by scaling the number of feedback bits with SNR, thus making codebooks prohibitively large [10]. To overcome the problem of scaling codebook size, and relax the reliance on frequency selectivity for quantization, [9] proposed an analog feedback strategy for constant MIMO interference channels. Using analog feedback, a constant data rate gap from perfect CSI performance was shown, as long as the SNRs on the forward and feedback links are order-wise equal. A limitation of the analysis in [7]–[9], however, is that the number of DoF remains the primary performance metric considered. IA's sum-rate performance at finite SNR, especially when accounting for the time spent on overhead signaling, has yet to be considered.

Attempts to more directly analyze or reduce overhead are limited to [11]–[14]. To analyze the effect of overhead, [11] considers the effective number of spatial DoF of an IA system with training and feedback. By considering DoF, however, [11] implicitly characterizes performance at infinitely high SNR. Alternatively, [12] reduces codebook size to limit overhead in limited feedback IA systems by leveraging temporal correlation without providing any overhead-aware analysis. In another line of work, information about the network topology is used to partition users into optimally sized alignment groups [13]. In [14], IA is applied to partially connected interference channels. User grouping and partial connectivity, however, only reduces the number of channels that must be shared without suggesting an efficient training and feedback strategy.

In this paper, we characterize the performance of a MIMO IA system that is designed

for perfect CSI operation yet only has access to imperfect CSI through training and analog feedback [9], [15], [16]. Thus, the performance demonstrated in this paper constitutes a lower bound for systems that are designed to be more robust to imperfect CSI through improved precoding strategies such as [31] for example. We adopt a block-fading model wherein the channel remains constant over the block length, and varies independently across blocks. In contrast with earlier work on IA with feedback, we precisely model channel selectivity by leveraging the relationship between block-fading and continuous-fading channels shown in [17]. This relationship allows us to define the concept of Doppler spread in a block fading channel and explicitly relate the size of the coherence block to that Doppler spread. Since both CSI acquisition and data transmission must now occur within the limits of a single coherence block, the IA system is faced with a non-trivial tradeoff: too much overhead leaves little time for payload data transmission, whereas too little overhead results in large sum-rate losses due to poor CSI quality [17]–[21]. In this paper, we design the training and analog feedback system to maximize IA's *effective sum-rate*, a non-asymptotic performance metric that accounts for both CSI quality and CSI acquisition overhead. CSI acquisition overhead is a fundamental concept that was largely neglected in earlier work on IA with imperfect CSI.

We begin by giving a tractable expression for the IA sum-rate in genie-aided systems with perfect CSI, and extend the analysis under a general model for imperfect CSI. We then specialize our results to a system with training and analog feedback by characterizing CSI quality as a function of system parameters such as training overhead, feedback overhead and transmit power on both forward and reverse links. This results in a tractable expression for IA's effective sum-rate, which we proceed to optimize. To give a closed-form solution for the optimal effective sum-rate, we build on the method in [17] and optimize a series expansion of the objective function. Initial results were reported in our previous work [22]. In this paper, we complete IA's performance analysis by analytically characterizing its maximum achievable effective sum-rate and the corresponding optimum overhead budget. The main insights and conclusions that can be drawn from the effective sum-rate analysis can be summarized as follows:

- Practical IA performance is not only a function of basic system parameters such as network size and SNR, but is tightly related to quantities such as Doppler spread, and feedback channel quality. Moreover, the dependence of both the maximum effective sum-rate, and the corresponding optimal overhead budget, on the various system parameters can be char-

acterized accurately.

- By properly designing the training and feedback stages, IA can be made both feasible and beneficial in a wide range of fading scenarios, even when its relatively high overhead is considered.
- Overhead-aware analysis is essential to the design of IA networks. As an example of this observation, we use the overhead analysis to give simple results on the optimal number of cooperative IA users for channels with varying levels of selectivity.

Throughout this paper, we use the following notation: \mathbf{A} is a matrix; \mathbf{a} is a vector; a is a scalar; $(\cdot)^*$ denotes the conjugate transpose; $\|\mathbf{a}\|$ denotes the 2-norm of \mathbf{a} ; $|a|$ is the absolute value of a ; \mathbf{I}_N is the $N \times N$ identity matrix; $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ is a complex Gaussian random vector with mean \mathbf{a} and covariance matrix \mathbf{A} ; (a_1, \dots, a_k) is an ordered set; $\mathbb{E}[\cdot]$ denotes expectation.

II. SYSTEM MODEL

Consider the K -user narrowband MIMO interference channel shown in Fig. 1 in which transmitter i communicates with its paired receiver i and interferes with all other receivers, $\ell \neq i$. For simplicity of exposition, consider a homogeneous network where all transmitters are equipped with N_T antennas and all receivers with N_R antennas, and each node pair communicates via $d \leq \min(N_T, N_R)$ independent spatial streams. The results can be generalized to a different number of streams or antennas at each node, provided that IA remains feasible [23].

Assuming perfect time and frequency synchronization, the sampled baseband signal at receiver i can be written as

$$\mathbf{y}_i = \sqrt{\frac{P}{d}} \mathbf{H}_{i,i} \mathbf{F}_i \mathbf{s}_i + \sum_{\ell \neq i} \sqrt{\frac{P}{d}} \mathbf{H}_{i,\ell} \mathbf{F}_\ell \mathbf{s}_\ell + \mathbf{v}_i, \quad (1)$$

where \mathbf{y}_i is the $N_R \times 1$ received signal vector, P is the transmit power, $\mathbf{H}_{i,\ell}$ is the $N_R \times N_T$ discrete-time effective baseband channel matrix from transmitter ℓ to receiver i , $\mathbf{F}_i = [\mathbf{f}_i^1, \dots, \mathbf{f}_i^d]$ is transmitter i 's $N_T \times d$ precoding matrix, \mathbf{s}_i is the $d \times 1$ transmitted symbol vector at node i such that $\mathbb{E}[\mathbf{s}_i \mathbf{s}_i^*] = \mathbf{I}_d$, and \mathbf{v}_i is a vector of i.i.d complex Gaussian noise samples with covariance matrix $\sigma^2 \mathbf{I}_{N_R}$. The channels $\mathbf{H}_{i,\ell}$ are assumed to be independent across users and each with i.i.d $\mathcal{CN}(0, 1)$ entries. Large-scale fading can be included in the system model at the expense of a more involved exposition in Section IV.

The received signal at *transmitter i* on the feedback channel is

$$\overleftarrow{\mathbf{y}}_i = \sqrt{\frac{P_F}{N_R}} \mathbf{G}_{i,i} \overleftarrow{\mathbf{x}}_i + \sum_{\ell \neq i} \sqrt{\frac{P_F}{N_R}} \mathbf{G}_{\ell,i} \overleftarrow{\mathbf{x}}_\ell + \overleftarrow{\mathbf{v}}_i, \quad (2)$$

where P_F is the feedback power available such that $P_F/P = \gamma$, $\mathbf{G}_{\ell,i}$ is the $N_T \times N_R$ discrete time feedback channel between receiver ℓ and transmitter i with i.i.d $\mathcal{CN}(0,1)$ entries, $\overleftarrow{\mathbf{x}}_i$ is the symbol vector with unit variance entries sent by receiver i , and $\overleftarrow{\mathbf{v}}_i$ is a complex vector of i.i.d circularly symmetric white Gaussian noise with covariance matrix $\sigma^2 \mathbf{I}_{N_T}$. The forward and feedback channels are assumed to be independent in the error analysis of Section IV, i.e., a frequency division duplexed system or a general non-reciprocal system is assumed.

We adopt a block-fading channel model in which channels remain fixed for a period, T_{frame} , but vary independently from block to block. To model the effect of channel selectivity on IA performance, we set the block length to $T_{\text{frame}} = \frac{1}{2f_D}$, where f_D plays the role of the block fading channel's *effective Doppler spread*. The definition of f_D is motivated by the results in [17] showing a relationship between continuous fading and block fading systems. To enable IA over such a channel, both CSI acquisition and payload data transmission must occur within the coherence time T_{frame} , or else the CSI acquired becomes obsolete. The IA system then encounters a well-known tension between CSI acquisition and data transmission [17]–[21], and must allocate resources to each of the processes to optimize overall performance.

To account for CSI acquisition overhead, and to accurately characterize the *effective data rate* achieved by IA, we adopt the overhead model shown in Fig. 2. In this model, overhead signaling consumes time resources that could otherwise be used for data transmission, i.e., CSI acquisition penalizes effective sum-rate. For such an overhead model, the effective sum-rate (in bits/s/Hz) can be written as [17]–[19]

$$\bar{R}_{\text{eff}}(P, T_{\text{OHD}}) = \left(\frac{T_{\text{frame}} - T_{\text{OHD}}}{T_{\text{frame}}} \right) \bar{R}_{\text{sum}}(P, T_{\text{OHD}}) \quad (3)$$

where T_{OHD} is the total time spent on training and feeding back channels, and $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ is the average sum-rate in bits/s/Hz achieved by IA on the channel uses allocated for payload transmission. Using (3), and previous insights into IA performance, we highlight the tradeoff between overhead signaling and data transmission. Increasing overhead improves CSI quality and in turn improves $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$, but the relative period over which $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ can

be achieved shrinks. A similar tension exists when lowering overhead; less overhead allows more channel uses for data transmission but the sum-rate per channel use suffers due to poor CSI quality. The objective then becomes maximizing the effective sum rate given in (3) by optimally trading off overhead with data transmission [17]–[21]. Throughout this paper, we treat $\bar{R}_{\text{sum}}(P, T_{\text{OHD}})$ as an information-theoretic quantity, and thus derive mutual information-based sum-rates achievable without errors. IA performance can also be analyzed from the perspective of fixed-rate transmission where metrics such as bit error rate may be of interest [24].

III. INTERFERENCE ALIGNMENT: AN AVERAGE SUM-RATE ANALYSIS

This section derives the average sum-rate achieved by IA in both genie-aided networks where channels are known perfectly, as well as practical systems where CSI is imperfect.

A. Interference Alignment with Perfect CSI

IA often achieves the full number of DoF supported by MIMO interference channels. In cases where the full DoF cannot be guaranteed, IA has been shown to provide significant gains in high-SNR sum-rate [3], [4], [25]. While this paper focuses on IA, even better performance could be achieved with other precoding algorithms that seek a balance between interference minimization and signal power maximization [3], [26], [27]. The algorithms in [3], [26], [27], however, do not readily lend themselves to average sum-rate analysis.

To analyze IA sum-rates, we begin by examining the effective channels created after precoding and combining. For tractability, we focus on IA with a simple per-stream zero-forcing (ZF) receiver. Recall that in the high (but finite) SNR regime, where IA is most useful, gains from more involved receiver designs are limited. In such a system, receiver i projects its signal onto the columns of the zero-forcing combiner $\mathbf{W}_i = [\mathbf{w}_i^1, \dots, \mathbf{w}_i^m, \dots, \mathbf{w}_i^d]$ which gives

$$(\mathbf{w}_i^m)^* \mathbf{y}_i = \sqrt{\frac{P}{d}} (\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m s_i^m + \sqrt{\frac{P}{d}} \sum_{(k,\ell) \neq (i,m)} (\mathbf{w}_i^m)^* \mathbf{H}_{i,k} \mathbf{f}_k^\ell s_k^\ell + (\mathbf{w}_i^m)^* \mathbf{v}_i. \quad (4)$$

At the output of these linear receivers \mathbf{w}_i^m , the conditions for perfect IA can be stated as [4]

$$(\mathbf{w}_i^m)^* \mathbf{H}_{i,k} \mathbf{f}_k^\ell = 0, \quad \forall (k, \ell) \neq (i, m) \quad (5)$$

$$|(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m| \geq c > 0, \quad \forall i, m, \quad (6)$$

where alignment is guaranteed by (5), and (6) is satisfied almost surely [1], [4].

As a result of conditions (5) and (6), the combination of IA and ZF effectively creates Kd non-interfering scalar channels. The maximum mutual information across these channels is achieved via Gaussian signaling which yields an instantaneous sum-rate given by

$$R_{\text{sum}} = \sum_{i=1}^K \sum_{m=1}^d \log_2 \left(1 + \frac{\frac{P}{d} |(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2}{\sigma^2} \right). \quad (7)$$

To derive an expression for the average sum-rate, i.e., $\bar{R}_{\text{sum}} = \mathbb{E}[R_{\text{sum}}]$, we first give the following lemma.

Lemma 1 ([9, Appendix A]): The effective direct channels $(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m$ are independent and Gaussian distributed with unit variance if: (i) the precoders \mathbf{F}_i are unitary and are generated by an IA solution that does not consider the direct channels $\mathbf{H}_{i,i}$, and (ii) the combiners \mathbf{W}_i are calculated to simply zero-force inter-user and inter-stream interference.

The conditions Lemma 1 places on precoder and combiner calculation are satisfied by most IA solutions such as [1], [3]–[5]. Hence, as a result of Lemma 1, the scalar point-to-point channels created by the combination of IA and ZF experience Rayleigh fading. As a result, the average sum-rate can be written in exponential integral form as [28], [29]

$$\bar{R}_{\text{sum}}(\rho) = \sum_{i=1}^K \sum_{m=1}^d \mathbb{E} \left[\log_2 \left(1 + \frac{\frac{P}{d} |(\mathbf{w}_i^m)^* \mathbf{H}_{i,i} \mathbf{f}_i^m|^2}{\sigma^2} \right) \right] = Kd \log_2(e) e^{1/\rho} E_1 \left(\frac{1}{\rho} \right), \quad (8)$$

which is written as a function of the per-stream SNR, $\rho = \frac{P}{d\sigma^2}$, and $E_1(\eta) = \int_1^{\infty} t^{-1} e^{-\eta t} dt$ is an exponential integral.

B. Interference Alignment with CSI from Training and Feedback

When the channels are not known perfectly, interference cannot be aligned perfectly. Misalignment leads to “leakage interference”, which reduces the signal-to-interference-plus-noise ratio (SINR) in the desired signal space. Moreover, imperfect knowledge of the direct channel implies that receivers will perform mismatched decoding [30], again reducing effective SINR. In this section, we examine the effect of imperfect CSI on the performance of an IA system that is optimized for perfect-CSI operation, i.e., a system that does not consider CSI imperfection in its design. Thus, the performance results demonstrated in this paper can be improved upon by adopting precoding algorithms that are more robust to CSI errors such as [31].

Consider an IA system in which transmitters use a common set of channel estimates as input to an IA solution such as [1], [3]–[5], i.e., they calculate imperfect IA precoders $\widehat{\mathbf{F}}_i$ and combiners $\widehat{\mathbf{W}}_i$. Denote the channel estimates as $\widehat{\mathbf{H}}_{i,\ell}$ and the corresponding error as $\widetilde{\mathbf{H}}_{i,\ell} = \mathbf{H}_{i,\ell} - \widehat{\mathbf{H}}_{i,\ell}$. In this system, the IA solution satisfies

$$(\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell = 0, \quad \forall (k, \ell) \neq (i, m) \quad (9)$$

$$\left| (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m \right| \geq c > 0, \quad \forall i, m. \quad (10)$$

We assume receivers obtain perfect knowledge of the combiners $\widehat{\mathbf{W}}_i$ and the imperfect effective direct channels $\widehat{\mathbf{w}}_i^m \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m$ for detection, an assumption similar to [7]–[9], [19], [32]¹ whose relaxation is a topic of future work. In general, receiver side information about the effective channels can be acquired blindly [33] or via additional training or silent phases [16]. For such an IA system, the received signal after projection is

$$(\widehat{\mathbf{w}}_i^m)^* \mathbf{y}_i = \sqrt{\frac{P}{d}} (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m s_i^m + \sqrt{\frac{P}{d}} \sum_{k,\ell} (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell s_k^\ell + (\widehat{\mathbf{w}}_i^m)^* \mathbf{v}_i, \quad (11)$$

where we have used the fact that conditions (9) and (10) are satisfied, thus $(\widehat{\mathbf{w}}_i^m)^* \mathbf{H}_{i,k} \widehat{\mathbf{f}}_k^\ell = (\widehat{\mathbf{w}}_i^m)^* (\widehat{\mathbf{H}}_{i,k} + \widetilde{\mathbf{H}}_{i,k}) \widehat{\mathbf{f}}_k^\ell = (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell$.

Analyzing the maximum sum-rates achievable on the channel in (11) is in general difficult, as it requires optimizing the distribution of the input symbols s_i for the interference channel in (11). Recall, however, that our objective is to analyze a system optimized for *perfect CSI operation*, i.e. one that does not account for CSI imperfection. This enables making the following assumptions that would be expected from a system optimized for perfect CSI operation.

Assumption 1: Transmitters use a typical Gaussian codebook made up of i.i.d. symbols to form the symbol vectors s_i . Such a signaling codebook, which was optimal for the interference free channels created by IA with perfect CSI, may no longer be optimal now that CSI is imperfect.

Assumption 2: Receivers perform nearest neighbor decoding using the estimates $\widehat{\mathbf{H}}_{i,i}$. Nearest neighbor decoding would again be optimal with perfect CSI. The nearest neighbor decoder, the channel estimates and the signaling codebook together satisfy the conditions outlined in [30] for

¹In fact [7], [8], [32] place a stronger assumption summarized by the receivers' knowledge of the exact imperfect CSI known to the transmitters. The two assumptions are functionally equivalent from the perspective of the sum-rate analysis, i.e., all that is needed is the receivers' knowledge of $\widehat{\mathbf{w}}_i^m$ and of the scalars $\widehat{\mathbf{w}}_i^m \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m$.

Corollary 3.0.1 of [30] to hold with equality, meaning that the estimation error plays the role of an additional source of Gaussian noise irrespective of its actual distribution.

Under Assumptions 1-2, and combining the results of [30] and [34], the average sum-rate achieved can be written as

$$\bar{R}_{\text{sum}}(\rho) = \sum_{i=1}^K \sum_{m=1}^d \mathbb{E} \left[\log_2 \left(1 + \frac{\frac{P}{d} \left| (\widehat{\mathbf{w}}_i^m)^* \widehat{\mathbf{H}}_{i,i} \widehat{\mathbf{f}}_i^m \right|^2}{\sum_{k,\ell} \mathbb{E} \left[\frac{P}{d} \left| (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell \right|^2 \right] + \sigma^2} \right)} \right) \right] \quad (12)$$

where we note that the outer expectation is now only over the fading on the direct channel and not the interference. Therefore, the leakage interference terms $(\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell$ indeed play the role of independent sources of additive Gaussian noise, regardless of their distribution.

When the entries of $\widetilde{\mathbf{H}}_{i,k}$ are zero-mean and uncorrelated with a variance of $\sigma_{\widetilde{\mathbf{H}}}^2$, it follows that $\mathbb{E} \left[\frac{P}{d} \left| (\widehat{\mathbf{w}}_i^m)^* \widetilde{\mathbf{H}}_{i,k} \widehat{\mathbf{f}}_k^\ell \right|^2 \right] = \frac{P}{d} \sigma_{\widetilde{\mathbf{H}}}^2$, thus the denominator in (12) is simply $KP\sigma_{\widetilde{\mathbf{H}}}^2 + \sigma^2$. Moreover, if the estimates $\widehat{\mathbf{H}}_{i,k}$ are MMSE estimates of $\mathbf{H}_{i,k}$, the entries of $\widehat{\mathbf{H}}_{i,k}$ have a variance of $1 - \sigma_{\widetilde{\mathbf{H}}}^2$. This results in an effective average SINR that can be written as

$$\rho_{\text{eff}} = \frac{\rho(1 - \sigma_{\widetilde{\mathbf{H}}}^2)}{\rho K d \sigma_{\widetilde{\mathbf{H}}}^2 + 1}, \quad (13)$$

where ρ is the per-stream SNR defined in (8). If the estimated direct channels $\widehat{\mathbf{H}}_{i,i}$ is Gaussian, the average sum-rate achieved by IA with imperfect CSI is again given in exponential integral form as

$$\bar{R}_{\text{sum}}(\rho_{\text{eff}}) = K d \log_2(e) e^{1/\rho_{\text{eff}}} E_1 \left(\frac{1}{\rho_{\text{eff}}} \right). \quad (14)$$

To evaluate sum-rate achieved by IA, one must now characterize ρ_{eff} or equivalently $\sigma_{\widetilde{\mathbf{H}}}^2$. In Section IV we specialize our result for a system with training and analog CSI feedback and later optimize IA's *effective data rate with overhead* in Section V.

IV. TRAINING AND ANALOG FEEDBACK

We propose to split the acquisition of CSI at the transmitter into three main phases. First, the transmitters train the forward channels via pilots. Second, the receivers train the feedback channels via pilots, setting the stage for the forward transmitters to estimate the feedback information in the next stage. Finally, the receivers feedback information about the forward

channels in an analog fashion, i.e., as unquantized complex symbols. We can characterize the CSI error introduced in the CSI acquisition phase by examining the three stages.

A. Forward and Feedback Channel Training

In the first training phase, each transmitter k sends an orthogonal pilot sequence matrix Φ_k , i.e., $\Phi_i \Phi_k^* = \delta_{ik} \mathbf{I}_{N_T}$, over a training period τ_t [35]. Pilot orthogonality imposes the constraint $\tau_t \geq KN_T$. Each receiver i then observes the $N_R \times \tau_t$ matrix

$$\mathbf{Y}_i = \sqrt{\frac{\tau_t P}{N_T}} \sum_{k=1}^K \mathbf{H}_{i,k} \Phi_k + \mathbf{V}_i, \quad \forall i, \quad (15)$$

where \mathbf{V}_i is an $N_R \times \tau_t$ matrix of noise terms. Using \mathbf{Y}_i , receiver i calculates an MMSE estimate of its incoming channels $\mathbf{H}_{i,k} \forall k$ given by

$$\hat{\mathbf{H}}_{i,k}^r = \frac{\sqrt{\frac{\tau_t P}{N_T}}}{\sigma^2 + \frac{\tau_t P}{N_T}} \mathbf{Y}_i \Phi_k^*, \quad \forall k. \quad (16)$$

where the superscript $(\cdot)^r$ emphasizes that $\hat{\mathbf{H}}_{i,k}^r$ are the channel estimates gathered at the receiver before they are relayed back to the transmitters and further corrupted. At the output of this first training stage, the channel estimates $\hat{\mathbf{H}}_{i,k}^r$ have i.i.d. $\mathcal{CN}(0, \frac{\tau_t P / N_T}{\sigma^2 + \tau_t P / N_T})$ entries with corresponding errors $\tilde{\mathbf{H}}_{i,k}^r \sim \mathcal{CN}(0, \frac{\sigma^2}{\sigma^2 + \tau_t P / N_T})$.

The feedback channel training phase proceeds similarly. Namely, the receivers transmit orthogonal pilot sequences over a training period $\tau_p \geq KN_R$. The transmitters independently compute MMSE estimates of their incoming channels, resulting in estimates $\hat{\mathbf{G}}_{k,i} \sim \mathcal{CN}\left(0, \frac{\frac{\tau_p P_F}{N_R}}{\sigma^2 + \frac{\tau_p P_F}{N_R}}\right)$ with corresponding error terms $\tilde{\mathbf{G}}_{k,i} \sim \mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_p P_F}{N_R}}\right)$.

B. Analog Feedback

After forward and feedback channel training, the receivers feedback their channel estimates $\hat{\mathbf{H}}_{i,k}^r$ in an analog fashion during a feedback period τ_f . This is accomplished by first post-multiplying each $N_R \times KN_T$ feedback matrix $[\hat{\mathbf{H}}_{i,1}^r \dots \hat{\mathbf{H}}_{i,K}^r]$ with a $KN_T \times \tau_f$ matrix Ψ_i such that $\Psi_i \Psi_i^* = \delta_{i,k} \mathbf{I}_{KN_T}$ [9], [15]. The spreading matrices Ψ_i orthogonalize the feedback from different users and facilitate estimation. This orthogonality constraint requires that $\tau_f \geq K^2 N_T$.

The transmitted $N_R \times \tau_f$ feedback matrix $\overleftarrow{\mathbf{X}}_i$ from receiver i can be written as [9], [15]

$$\overleftarrow{\mathbf{X}}_i = \sqrt{\frac{\tau_f P_F}{KN_T N_R} \left(\frac{\tau_t P / N_T}{\sigma^2 + \tau_t P / N_T} \right)^{-1}} \left[\hat{\mathbf{H}}_{i,1}^r \cdots \hat{\mathbf{H}}_{i,K}^r \right] \boldsymbol{\Psi}_i, \quad (17)$$

where the leading scalar is to ensure that the average transmit power constraints are satisfied with equality, i.e., one can verify that $\mathbb{E} \left[\text{trace} \left(\overleftarrow{\mathbf{X}}_i \overleftarrow{\mathbf{X}}_i^* \right) \right] = \tau_f P_F$. We write the concatenated $KN_T \times \tau_f$ matrix of feedback symbols observed by all transmitters as

$$\overleftarrow{\mathbf{Y}}_f = \sqrt{\frac{\tau_f P_F}{KN_T N_R} \left(\frac{\tau_t P / N_T}{\sigma^2 + \tau_t P / N_T} \right)^{-1}} \sum_{i=1}^K \begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix} \left[\hat{\mathbf{H}}_{i,1}^r \cdots \hat{\mathbf{H}}_{i,K}^r \right] \boldsymbol{\Psi}_i + \mathbf{V}, \quad (18)$$

where \mathbf{V} is the $KN_T \times \tau_f$ matrix of i.i.d Gaussian noise.

To simplify the performance analysis, we make the same assumption as in [9]: at the end of the feedback phase, the transmitters cooperate by sharing their rows of the received feedback matrix $\overleftarrow{\mathbf{Y}}_f$ which enables them to form a unified estimate of the forward channels $\mathbf{H}_{i,k}$. We refer the reader to [9] for a discussion of this cooperative assumption and for alternative non-cooperative approaches that are shown to perform close to this special case.

Under this cooperative assumption, the transmitters estimate $\mathbf{H}_{i,k} \forall k$ by first isolating the feedback sent by receiver i . They post-multiply their received symbols by $\boldsymbol{\Psi}_i^*$ to compute

$$\overleftarrow{\mathbf{Y}}_f \boldsymbol{\Psi}_i^* = \sqrt{\frac{\tau_f P_F}{KN_T N_R} \left(\frac{\tau_t P / N_T}{\sigma^2 + \tau_t P / N_T} \right)^{-1}} \underbrace{\begin{bmatrix} \mathbf{G}_{i,1} \\ \vdots \\ \mathbf{G}_{i,K} \end{bmatrix}}_{\mathbf{G}_i} \left[\hat{\mathbf{H}}_{i,1}^r \cdots \hat{\mathbf{H}}_{i,K}^r \right] + \mathbf{V} \boldsymbol{\Psi}_i^*. \quad (19)$$

The transmitters then compute a common linear MMSE estimate of the forward channels $\mathbf{H}_{i,k} \forall i, k$ using their feedback channel estimates $\hat{\mathbf{G}}_{i,k} \forall i, k$, and assuming that $KN_T \geq N_R$ so that the estimation problem is well posed. After a lengthy yet standard application of the orthogonality principle and the matrix inversion lemma, the MMSE estimate is given by

$$\hat{\mathbf{H}}_i = \sqrt{\frac{KN_T N_R}{\tau_f P_f} \left(\frac{\tau_t P / N_T}{\sigma^2 + \tau_t P / N_T} \right)^{-1}} \left(\hat{\mathbf{G}}_i^* \hat{\mathbf{G}}_i + \gamma_1 \hat{\mathbf{G}}_i^* \hat{\mathbf{G}}_i + \gamma_2 \mathbf{I}_{N_R} \right)^{-1} \hat{\mathbf{G}}_i^* \overleftarrow{\mathbf{Y}}_f \boldsymbol{\Psi}_i^*, \quad (20)$$

where we have written (20) in terms of $\hat{\mathbf{H}}_i = \left[\hat{\mathbf{H}}_{i,1}, \dots, \hat{\mathbf{H}}_{i,K} \right] \forall i$, the concatenated estimate

of the channels $\mathbf{H}_i = [\mathbf{H}_{i,1}, \dots, \mathbf{H}_{i,K}] \forall i$, for the sake of notational brevity. The constants γ_1 and γ_2 are the MMSE regularization factors. For completeness, γ_1 and γ_2 are given by

$$\gamma_1 = \frac{N_T \sigma^2}{P \tau_t}, \quad \gamma_2 = \left(1 + \frac{N_T \sigma^2}{\tau_t P}\right) \left(\frac{\sigma^2 K N_T N_R}{\tau_f P_F} + \frac{N_R \sigma^2}{\sigma^2 + \tau_p P_F / N_R}\right). \quad (21)$$

In essence, γ_1 captures the effect of the noise in the transmitted estimates $\widehat{\mathbf{H}}_{i,k}^r$, while γ_2 captures the effect of the noise in the estimates $\widehat{\mathbf{G}}_{i,k}$ as well as the noise observed during feedback.

Having formalized the three training and analog feedback stages, we now analyze the variance,² $\sigma_{\widehat{\mathbf{H}}}^2$, of the CSI error $\mathbf{H}_{i,k} - \widehat{\mathbf{H}}_{i,k}$, which automatically yields an estimated CSI variance of $1 - \sigma_{\widehat{\mathbf{H}}}^2$. Unfortunately, writing $\sigma_{\widehat{\mathbf{H}}}^2$ exactly yields rather cumbersome expressions. For this reason, we replace the variance of the MMSE estimation error by that of a zero-forcing estimator in a manner similar to [16], [36]. This ZF simplification intuitively amounts to deriving a high SNR result [16] and mathematically amounts to neglecting the constants γ_1 and γ_2 ; recall that moderately high SNR is after all the main operating region of interest for IA. Numerical results in Section VI will demonstrate that the effect of this simplification is negligible.

By neglecting γ_1 and γ_2 , and after some algebraic manipulation, we find that the error $\widetilde{\mathbf{H}}_i = \mathbf{H}_i - \widehat{\mathbf{H}}_i$ at the end of the three training and feedback phases can be written as

$$\widetilde{\mathbf{H}}_i = \sqrt{1 + \frac{N_T \sigma^2}{\tau_t P}} \left[\widetilde{\mathbf{H}}_i^r + \left(\widehat{\mathbf{G}}_i^* \widehat{\mathbf{G}}_i\right)^{-1} \widehat{\mathbf{G}}_i^* \left(\sqrt{1 + \frac{N_T \sigma^2}{\tau_t P}} \widetilde{\mathbf{G}}_i \widehat{\mathbf{H}}_i^r + \sqrt{\frac{K N_T N_R}{\tau_p P_f}} \mathbf{V} \Psi_i^* \right) \right]. \quad (22)$$

As can be seen from (22), the resulting CSI error is a combination of three terms: the first due to forward channel estimation error $\widetilde{\mathbf{H}}_i^r$, the second due to feedback channel estimation error $\widetilde{\mathbf{G}}_i$, and the third due to feedback noise.

To derive the statistics of $\widetilde{\mathbf{H}}_i$, we note the following three facts about the three terms in (22):

- 1) The entries of $\widetilde{\mathbf{H}}_i^r$ are uncorrelated $\mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_t P}{N_T}}\right)$ variables as shown in Section IV-A.
- 2) Similarly, the entries of $\widetilde{\mathbf{G}}_i$ are $\mathcal{CN}\left(0, \frac{\sigma^2}{\sigma^2 + \frac{\tau_p P_F}{N_R}}\right)$ implying that $\widetilde{\mathbf{G}}_i \widehat{\mathbf{H}}_i^r$ has independent entries with variance equal to $\frac{N_R \sigma^2}{\sigma^2 + \frac{\tau_p P_F}{N_R}} \frac{\tau_t P N_T}{\sigma^2 + \tau_t P / N_T}$.
- 3) The entries of \mathbf{V} are uncorrelated $\mathcal{CN}(0, \sigma^2)$ variables and so are the elements of $\mathbf{V} \Psi_i^*$ since the matrix Ψ_i is unitary.

²We in fact derive the entire covariance matrix for the columns of $\mathbf{H}_{i,k} - \widehat{\mathbf{H}}_{i,k}$. We show that the covariance matrices are scaled identities and thus the second order statistics of the error are entirely described by a scalar variance.

Combining the properties stated, the conditional covariance of each column of $\tilde{\mathbf{H}}_i$ denoted $\tilde{\mathbf{H}}_i^{(\ell)}$, conditioned of $\hat{\mathbf{G}}_i$ is [9], [15]

$$\mathbb{E} \left(\tilde{\mathbf{H}}_i^{(\ell)} \tilde{\mathbf{H}}_i^{(\ell)*} | \hat{\mathbf{G}}_i \right) = \frac{N_T \sigma^2}{\tau_t P} \mathbf{I}_{N_R} + \left(\frac{K N_T N_R \sigma^2}{\tau_f P_F} + \frac{N_R \sigma^2}{\sigma^2 + \tau_p \frac{P_F}{N_R}} \right) \left(\hat{\mathbf{G}}_i^* \hat{\mathbf{G}}_i \right)^{-1}. \quad (23)$$

Since the entries of the MMSE estimate $\hat{\mathbf{G}}_i$ are Gaussian, the matrix $\left(\hat{\mathbf{G}}_i^* \hat{\mathbf{G}}_i \right)^{-1}$ has an inverse-Wishart distribution [37]. Moreover, since $\hat{\mathbf{G}}_i$ has uncorrelated entries with a variance of $\frac{\tau_p P_F}{\sigma^2 + \tau_p \frac{P_F}{N_R}}$, $\left(\hat{\mathbf{G}}_i^* \hat{\mathbf{G}}_i \right)^{-1}$ has a covariance matrix equal to a properly scaled identity [9], [15], [37]. Thus marginalizing (23) over $\hat{\mathbf{G}}_i$, we find that $\tilde{\mathbf{H}}_i$ has independent columns with scaled identity covariance matrices with diagonal entries given by

$$\sigma_{\tilde{\mathbf{H}}}^2 = \frac{N_T \sigma^2}{\tau_t P} + \frac{\sigma^2}{(K N_T - N_R) P_F} \left(\frac{N_R^2}{\tau_p} + \frac{K N_T N_R}{\tau_f} \left(1 + \frac{N_R \sigma^2}{\tau_p P_F} \right) \right). \quad (24)$$

The same high SNR simplification adopted earlier to replace MMSE estimation error by ZF estimation error, however, allows us to further simplify (24) by writing

$$\sigma_{\tilde{\mathbf{H}}}^2 = \frac{N_T \sigma^2}{\tau_t P} + \frac{\sigma^2}{P(K N_T - N_R)} \left(\frac{N_R^2}{\gamma \tau_p} + \frac{K N_T N_R}{\gamma \tau_f} \right), \quad (25)$$

which completes the characterization of the distortion introduced by training and analog feedback.

Note: Finally, a word on applying the results of Section III to the analog feedback system described. First, we note that the analog feedback system satisfies Assumption 2, and the estimates yield $\mathbb{E} \left[\frac{P}{d} \left| (\hat{\mathbf{w}}_i^m)^* \tilde{\mathbf{H}}_{i,k} \hat{\mathbf{f}}_k^\ell \right|^2 \right] = \frac{P}{d} \sigma_{\tilde{\mathbf{H}}}^2$ and $\mathbb{E} \left[\frac{P}{d} \left| (\hat{\mathbf{w}}_i^m)^* \hat{\mathbf{H}}_{i,i} \hat{\mathbf{f}}_i^m \right|^2 \right] = \frac{P}{d} (1 - \sigma_{\tilde{\mathbf{H}}}^2)$ as needed. One subtlety though is that the fading on the feedback channel introduces non-Gaussian terms into the estimates $\hat{\mathbf{H}}_{i,i}$, yet (14) is only exact when the estimates are truly Gaussian. For fairly accurate estimation, however, $\hat{\mathbf{H}}_{i,i}$ can be well approximated by a Gaussian. Moreover, it will be clear from the results of Section VI that the effect of this is negligible.

V. OPTIMIZING OVERHEAD AND EFFECTIVE SUM-RATE

Having formally quantified IA sum-rate as a function of SNR and CSI quality, and characterized CSI quality in terms of training and feedback resources, we redefine both the optimization

problem and objective function as

$$\bar{R}_{\text{eff}}^*(P) = \max_{\tau_t, \tau_p, \tau_f} \left(\frac{T_{\text{frame}} - (\tau_t + \tau_p + \tau_f)}{T_{\text{frame}}} \right) \bar{R}_{\text{sum}}(\rho_{\text{eff}}), \quad (26)$$

where we have used $(\cdot)^*$ to denote optimality. We note from (26) that ρ_{eff} depends on $\sigma_{\mathbf{H}}^2$ and thus on τ_t , τ_p , and τ_f . The problem in (26) can be rewritten in a more tractable form as [19]

$$\bar{R}_{\text{eff}}^*(P) = \max_{\alpha_{\min} \leq \alpha \leq 1} \left[(1 - \alpha) \max_{\substack{\tau_t, \tau_p, \tau_f \\ \tau_t + \tau_p + \tau_f = \alpha T_{\text{frame}}}} \bar{R}_{\text{sum}}(\rho_{\text{eff}}) \right], \quad (27)$$

where $\alpha_{\min} = K(N_T + N_R + KN_T)/T_{\text{frame}}$ and is dictated by the minimum number of training and feedback symbols needed to render the estimation problems in Section IV well defined. The inner maximization in (27) optimizes sum-rate for a fixed overhead length of $T_{\text{OHD}} = \alpha T_{\text{frame}}$ and the outer maximization finds the optimal α thereby completing the solution.

Since $\bar{R}_{\text{sum}}(\rho_{\text{eff}})$ is decreasing in $\sigma_{\mathbf{H}}^2$, the inner maximization step simplifies to

$$\begin{aligned} \sigma_{\mathbf{H}}^{2*} = \min_{\tau_t, \tau_p, \tau_f} & \frac{N_T \sigma^2}{\tau_t P} + \frac{\sigma^2}{P(KN_T - N_R)} \left(\frac{KN_T N_R}{\gamma \tau_f} + \frac{N_R^2}{\gamma \tau_p} \right) \\ \text{s.t.} & \quad \tau_t + \tau_p + \tau_f = \alpha T_{\text{frame}}. \end{aligned} \quad (28)$$

Although (28) is an integer problem, its continuous relaxation is convex. Applying standard convex optimization techniques, the Lagrangian for the inner maximization is

$$\Lambda(\tau_t, \tau_p, \tau_f, \lambda) = \frac{N_T \sigma^2}{\tau_t P} + \frac{\sigma^2}{P(KN_T - N_R)} \left(\frac{KN_T N_R}{\gamma \tau_f} + \frac{N_R^2}{\gamma \tau_p} \right) + \lambda(\tau_t + \tau_p + \tau_f - \alpha T_{\text{frame}}). \quad (29)$$

Solving for the first order KKT conditions, we obtain the optimal training and feedback times as a function of the total overhead budget αT_{frame} as

$$\tau_t^* = \frac{\sqrt{\gamma N_T (KN_T - N_R)}}{\mu} \alpha T_{\text{frame}}, \quad \tau_p^* = \frac{N_R}{\mu} \alpha T_{\text{frame}}, \quad \tau_f^* = \frac{\sqrt{KN_T N_R}}{\mu} \alpha T_{\text{frame}},$$

where $\mu = \sqrt{\gamma N_T (KN_T - N_R)} + N_R + \sqrt{KN_T N_R}$. After solving the problem's continuous relaxation, convexity implies that for any given feasible overhead budget αT_{frame} simply examining the few integer neighbors of the points τ_t^* , τ_p^* , τ_f^* yields the integer training and feedback times that minimize CSI distortion, i.e., optimal integer training and feedback times can be found by a simple search over the grid neighbors of the non-integer solution. Proceeding with the continuous

relaxation, the minimum CSI distortion for an overhead budget αT_{frame} is

$$\sigma_{\tilde{\mathbf{H}}}^{2*} = \frac{\sigma^2 \left(\sqrt{KN_{\text{T}}N_{\text{R}}} + N_{\text{R}} + \sqrt{\gamma N_{\text{T}}(KN_{\text{T}} - N_{\text{R}})} \right)^2}{\gamma P (KN_{\text{T}} - N_{\text{R}}) \alpha T_{\text{frame}}}. \quad (30)$$

Having found the optimal allocation of τ_{t} , τ_{p} , and τ_{f} for a fixed overhead budget, what remains is to optimize the budget itself. The outer optimization in (26), however, does not admit a closed form solution. To circumvent this problem, prior work on single user and broadcast channels has specialized their results to the limiting high or low SNR regimes [18], [20], relied on numerical optimization [9], or resorted to characterizing the scaling of overhead with various system parameters based on sum-rate lower bounds [19]. To give accurate results on finite-SNR sum-rate, we propose to optimize a series expansion of (3) with respect to the channel's Doppler spread around the point $f_{\text{D}} = 0$ [17]. Recall that T_{frame} which we have been using thus far is related to f_{D} by the relationship $T_{\text{frame}} = \frac{1}{2f_{\text{D}}}$. To that end, we give the following result on the series expansion of $\bar{R}_{\text{eff}}(P, T_{\text{OHD}})$.

Proposition 1: The effective sum-rate achieved by IA with training and feedback expands as

$$\begin{aligned} \bar{R}_{\text{eff}}(P, T_{\text{OHD}}) = & (1 - \alpha)(1 + \rho Kd) \left[\frac{\bar{R}_{\text{sum}}(\rho)}{1 + \rho Kd} - \frac{2\beta}{d\alpha} \dot{R}_{\text{sum}}(\rho) f_{\text{D}} \right. \\ & \left. + \left(\frac{2\beta}{d\alpha} \right)^2 \left(\ddot{R}_{\text{sum}}(\rho)(1 + \rho Kd) + 2Kd \dot{R}_{\text{sum}}(\rho) \right) \frac{f_{\text{D}}^2}{2} \right] + O(f_{\text{D}}^3), \end{aligned} \quad (31)$$

where

$$\beta = \frac{\left(\sqrt{KN_{\text{T}}N_{\text{R}}} + N_{\text{R}} + \sqrt{\gamma N_{\text{T}}(KN_{\text{T}} - N_{\text{R}})} \right)^2}{\gamma (KN_{\text{T}} - N_{\text{R}})}, \quad (32)$$

whereas $\dot{R}_{\text{sum}}(\rho)$ and $\ddot{R}_{\text{sum}}(\rho)$ are the first and second derivatives of *perfect CSI sum-rate*, $\bar{R}_{\text{sum}}(\rho)$, which can be conveniently expressed as

$$\dot{R}_{\text{sum}}(\rho) = \frac{1}{\rho} \left(Kd \log_2(e) - \frac{\bar{R}_{\text{sum}}(\rho)}{\rho} \right), \quad \ddot{R}_{\text{sum}}(\rho) = -\frac{1}{\rho^2} \left(Kd \log_2(e) + \dot{R}_{\text{sum}}(\rho) - 2 \frac{\bar{R}_{\text{sum}}(\rho)}{\rho} \right). \quad (33)$$

Proof: Given in Appendix A. ■

Thus, by expanding effective sum-rate w.r.t. f_{D} , we have transformed the complicated non-linear dependence of effective sum-rate on system parameters such as P , T_{frame} , f_{D} , and T_{OHD} to a simpler polynomial dependence. The expansion in Proposition 1 can now be used to derive the

expansion of the optimal overhead budget, α^* , along with the performance it achieves. Relaxing the constraint that the overhead fraction α must be rational, simply differentiating the series expansion in Proposition 1 and equating it to zero yields the optimal overhead budget α^* .

Proposition 2: The optimum overhead fraction α^* for an IA system with training and analog feedback expands as

$$\alpha^* = \sqrt{\frac{2\beta(1 + \rho Kd)}{d} \frac{\dot{R}_{\text{sum}}(\rho)}{\bar{R}_{\text{sum}}(\rho)} f_D} - \frac{\beta}{d} \left(\frac{\ddot{R}_{\text{sum}}(\rho)}{\dot{R}_{\text{sum}}(\rho)} (1 + \rho Kd) + 2Kd \right) f_D + O(f_D^{3/2}), \quad (34)$$

which results in the optimal effective sum-rate

$$\bar{R}_{\text{eff}}^*(P) = \bar{R}_{\text{sum}}(\rho) - 2\sqrt{\frac{2\beta}{d} (1 + \rho Kd) \dot{R}_{\text{sum}}(\rho) \bar{R}_{\text{sum}}(\rho) f_D} + O(f_D). \quad (35)$$

Note that if f_D is large enough that $\alpha^* < \alpha_{\min}$ the optimal overhead budget must be adjusted to α_{\min} and the expression for $\bar{R}_{\text{eff}}^*(P)$ correspondingly updated.

Proof: The proof follows directly from differentiating the expansion in Proposition 1 w.r.t. α and solving the resulting cubic polynomial for its relevant root. ■

Therefore, Proposition 2 along with the solution to (28) gives the effective sum-rate-maximizing amount of forward training, feedback channel training, and analog feedback as simple functions of fundamental system parameters such as SNR, Doppler spread (equivalently T_{frame}), and perfect CSI sum-rate. Numerical results in Section VI will show that the overhead expansion in Proposition 2 is accurate for a wide range of system parameters and can thus obviate the need for numerical overhead optimization. Furthermore, the derived results allow us to draw several interesting insights into IA system design and performance:

- 1) The optimal overhead budget α scales with $\sqrt{f_D}$. As stated, for high enough Doppler α^* , must be adjusted to α_{\min} meaning that overhead subsequently increases with f_D . This scaling behavior is in line with previous results on other single and multiuser channels.
- 2) The sum-rate penalty due to overhead and imperfect CSI behaves similarly, i.e., increases with $\sqrt{f_D}$ initially and with f_D at high Doppler.
- 3) Examining the leading term in α^* we note that, similarly to [17], the term $(1 + \rho Kd) \frac{\dot{R}_{\text{sum}}(\rho)}{\bar{R}_{\text{sum}}(\rho)}$ behaves like $Kd / \log_e(1 + \rho)$ and thus the optimal overhead budget decreases with SNR roughly as $\sqrt{Kd / \log_e(1 + \rho)}$.
- 4) Since overhead decreases with SNR, a minimum overhead interval of $KN_T + KN_R + K^2N_T$

is always optimal at sufficiently high SNR. Thus, the effective number of spatial DoF achieved by IA with the analog feedback strategy described is $\left(1 - \frac{KN_T + KN_R + K^2 N_T N_R}{T_{\text{frame}}}\right) Kd$, i.e., the DoF penalty increases linearly with f_D .

- 5) Again examining the leading term in α^* we note that it increases with $\sqrt{\beta}$. Recalling the definition of β in (32), we conclude that the optimal overhead budget increases with $\sqrt{P/P_F}$. This formalizes the relationship between overhead and feedback link quality.

In addition to highlighting the dependence of overhead and effective sum-rate on various system parameters, the derived results can provide simple answers to various network design questions. For example, by simply comparing IA's effective sum-rate expression to those achieved by other transmission strategies, one can choose the optimal transmission strategy for a given fading environment. Moreover, since overhead and channel selectivity have been shown to place fundamental limits on the gains of cooperation in wireless network [38], the overhead-aware analysis presented in this paper can help in determining the optimal number of cooperative IA users at a given level of selectivity.

Consider, as a simple example, a K -user single-stream cooperation cluster with a variable number of antennas in which extra users are allowed to cooperate via IA if they do not incur a loss in effective sum-rate, else the extra users are not allowed access to the propagation medium and presumably left to transmit on a separate channel. In this model, additional cooperating users can be incorporated into the cluster as long as $\mathcal{I}^*_{K+1}(P) - \mathcal{I}^*_K(P) > 0$ where we have made cluster size explicit in the effective sum-rate subscript. Consequently, the effective sum-rate-maximizing cluster size becomes the smallest K such that $\mathcal{I}^*_{K+1}(P) - \mathcal{I}^*_K(P) < 0$. Moreover, note that minimizing overhead and maintaining IA feasibility imposes the constraint $N_T + N_R = K + 1$ [23]. Thus, writing N_T and N_R in terms of K , e.g. $N_T = \lceil (K + 1)/2 \rceil$, the user admission rule can be simplified to a function of only K , f_D , SNR, and γ . To simplify the user admission rule even further, we make the following approximations: (i) we consider the leading term in $\bar{R}_{\text{eff}}(P, T_{\text{OHD}})$ thus focusing on IA's effective DoF given in the fourth observation after Proposition 2, (ii) we assume that $N_T = (K + 1)/2$ and thus relax its integer constraint. Using these simplifications, the user admission rule $\mathcal{I}^*_{K+1}(P) - \mathcal{I}^*_K(P) > 0$ simplifies to

$$4K^3 + 15K^2 + 17K + 6 < \frac{1}{f_D}, \quad (36)$$

i.e., a K -user cluster can be extended to $K + 1$ as long as (36) is satisfied. Interestingly, this implies that in such single-stream IA scenarios the effective sum-rate-maximizing cluster size grows with $f_D^{-1/3}$. While the approximate admission rule is a rather simplified version of $\mathcal{I}^*_{K+1}(P) - \mathcal{I}^*_K(P) > 0$, we show in Section VI that it is very accurate at predicting optimal cluster size. Finally, we note that while we provide this example to illustrate problems that can be solved using our analysis, the rule in (36) is by no means universal. When parameters such as large-scale fading or uncoordinated interference are considered, both the analysis and the admission rule must be adjusted.

VI. SIMULATION RESULTS

Consider a three-user IA cluster with two transmit antennas, two receive antennas, and one spatial stream per user and let $\gamma = \frac{P_F}{P} = 1$. Fig. 3 shows the effective sum-rate achieved by IA in systems with various levels of mobility or normalized Doppler spreads, f_D . To quantify the degradation in effective sum-rate caused by overhead and imperfect CSI, we include the performance of a baseline genie-aided system in which CSI is both perfect and free. Fig. 3 indicates that IA achieves good performance in a system with vehicular-levels of mobility. In fact, if typical wireless parameters are adopted, such as a wavelength of $\lambda = 0.15$ m (corresponding to a carrier frequency of 2 GHz), a coherence bandwidth of $W_C = 300$ kHz, and a normalized Doppler given by $f_D = \frac{v}{\lambda W_C}$ where v is the user's velocity, Fig. 3 indicates that IA could theoretically perform well even at a speed of more than 160 km/hr. The rate of performance degradation over a wider range of Doppler spread can be seen in Fig. 4. Both Figs. 3 and 4 indicate that the analytical results of Section V are very effective in optimizing the effective sum-rate of IA systems as the resulting performance closely matches that of a numerically optimized system. Finally, Fig. 3 indicates that the effect of the simplifying assumptions made in Section IV is negligible since the effective sum-rate predicted by the derived rate expressions closely matches simulated IA performance. A very slight deviation is noticed at very low SNR where the ZF simplification in Section IV is a less accurate approximation of MMSE performance.

Fig. 5 shows the optimal overhead budget for systems with varying frame lengths and again includes both the analytical overhead budget from Section V as well as the result of numerically optimizing the same system. Fig. 5 confirms that T_{OHD} increases with frame size T_{frame} at a rate proportional to $\sqrt{T_{\text{frame}}}$. Thus α^* indeed decreases with $\frac{1}{\sqrt{T_{\text{frame}}}}$, as shown in Fig. 6, or

equivalently increases with $\sqrt{f_D}$ (and with f_D for sufficiently high Doppler). Fig. 5 also shows that the expansion in Proposition 2 provides an accurate characterization of IA's effective sum-rate-maximizing overhead budget over a wide range of SNRs and frame sizes. Fig. 7 in turn verifies the decrease of α^* with SNR, which as stated in Section V follows the relationship $\alpha^* \sim \sqrt{\log_e(1 + \rho)^{-1}}$. To complete the characterization of overhead and effective sum-rate, Fig. 8 quantifies the deleterious effect of a weak feedback channel on overhead and effective sum-rate. Fig. 8 also indicates that the expansion results of Section V could significantly underestimate α^* in very weak feedback channels, though the final effect on throughput remains limited.

Finally we examine the efficiency of our overhead analysis in further network design. We consider the motivating example given in Section V of a K -user system for which we seek to optimize the cooperation cluster size as a function of mobility. Fig. 9 shows the optimal cluster size as a function of T_{frame} for an IA system at 35 dB SNR. Fig 10 shows the corresponding effective sum-rate achieved. We plot the cluster size and effective sum-rate resulting from (i) an exhaustive search over all possible K , and (ii) the simple overhead-based user admission rule in (36). We note that the cluster size predicted by the two methods are in close agreement, and that the asymptotic cube-root relationship predicted in Section V between optimal cluster size and T_{frame} is quite accurate even for small values of K . While the overhead-based rule tends to underestimate cluster size for small intervals of T_{frame} , Fig. 10 indicates that the resulting rate gap from optimal sizing is negligible. The same can be said about the rate loss when applying the same overhead based rule to a system at an SNR of 10 dB and a system with $\gamma = 10^{-2}$.

VII. CONCLUSION

We considered IA's effective sum-rate in practical systems where CSI is imperfect and comes with an associated overhead cost. We showed that training and feedback overhead can be optimized to ensure good IA performance over a wide range of SNR and Doppler spread. We quantified the dependence between overhead and various system parameters such as feedback link quality. More sophisticated precoding algorithms, designed to be robust to imperfect CSI, could further improve the demonstrated performance and thus remain a promising area for future work. The derived results provide a formal method to gauge true IA performance vs. other transmission strategies, and can thus highlight settings under which IA provides tangible gains. The derived analysis can also be used for further network design as demonstrated by the motivating example

given at the end of Section V on overhead-aware user admission and optimal network sizing.

APPENDIX A

PROOF OF PROPOSITION 1

To expand effective sum-rate around $f_D = 0$, we start computing its first order derivative

$$\frac{\partial \bar{R}_{\text{eff}}(P, T_{\text{OHD}})}{\partial f_D} = (1 - \alpha) \dot{R}_{\text{sum}}(\rho) \frac{\partial \rho_{\text{eff}}}{\partial f_D} \Big|_{f_D=0} = -(1 - \alpha) \dot{R}_{\text{sum}}(\rho) (1 + Kd\rho) \frac{2\beta}{d\alpha} \quad (37)$$

where $\frac{\partial \rho_{\text{eff}}}{\partial f_D} \Big|_{f_D=0}$ is evaluated by noticing that after solving the inner maximization in (28) and obtaining $\sigma_{\mathbf{H}}^{2*}$ in (30) we have $\frac{\partial \sigma_{\mathbf{H}}^{2*}}{\partial f_D} \Big|_{f_D=0} = \frac{2\beta}{d\alpha}$. The term $\dot{R}_{\text{sum}}(\rho)$ can be obtained by a standard derivation of the exponential integral rate expression in (8) w.r.t ρ and is given directly in the statement of Proposition 1; $\ddot{R}_{\text{sum}}(\rho)$ is obtained similarly. As for the second order term, we have

$$\begin{aligned} \frac{\partial^2 \bar{R}_{\text{eff}}(P, T_{\text{OHD}})}{\partial f_D^2} &= (1 - \alpha) \left[\ddot{R}_{\text{sum}}(\rho) \left(\frac{\partial \rho_{\text{eff}}}{\partial f_D} \right)^2 + \dot{R}_{\text{sum}}(\rho) \frac{\partial^2 \rho_{\text{eff}}}{\partial f_D^2} \right] \Big|_{f_D=0} \\ &\stackrel{(a)}{=} (1 - \alpha) \left[\ddot{R}_{\text{sum}}(\rho) \left(\frac{\partial \rho_{\text{eff}}}{\partial f_D} \right)^2 + \dot{R}_{\text{sum}}(\rho) \left(\frac{\partial^2 \rho_{\text{eff}}}{\partial \sigma_{\mathbf{H}}^{2*2}} \left(\frac{\partial \sigma_{\mathbf{H}}^{2*}}{\partial f_D} \right)^2 + \frac{\partial \rho_{\text{eff}}}{\partial \sigma_{\mathbf{H}}^{2*}} \frac{\partial^2 \sigma_{\mathbf{H}}^{2*}}{\partial f_D^2} \right) \right] \Big|_{f_D=0} \\ &\stackrel{(b)}{=} (1 - \alpha) \left(\frac{2\beta}{d\alpha} \right)^2 (1 + \rho Kd) \left(\ddot{R}_{\text{sum}}(\rho) (1 + \rho Kd) + 2Kd \dot{R}_{\text{sum}}(\rho) \right) \end{aligned} \quad (38)$$

where (a) expands $\frac{\partial^2 \rho_{\text{eff}}}{\partial f_D^2}$ for clarity and (b) is by noticing that $\frac{\partial^2 \sigma_{\mathbf{H}}^{2*}}{\partial f_D^2} = 0$ since $\sigma_{\mathbf{H}}^{2*}$ is linear in f_D and otherwise replacing the values of the different variables. Combining (37) and (38) we get the resulting second order expansion. Higher order expansions can be found if additional accuracy is needed, however, the second order expansion is in general sufficient.

REFERENCES

- [1] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [2] T. Gou and S. Jafar, "Degrees of freedom of the K user MxN MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.
- [3] S. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 206–218, Jan. 2011.
- [4] K. Gomadam, V. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3309–3322, 2011.
- [5] R. Tresh, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," *Proc. IEEE/SP 15th Workshop on Statistical Signal Processing*, pp. 277–280, Aug. 31–Sept. 3 2009.
- [6] A. Goldsmith, *Wireless communications*. Cambridge Univ Pr, 2005.
- [7] J. Thukral and H. Bolcskei, "Interference alignment with limited feedback," *Proc. IEEE International Symposium on Information Theory*, pp. 1759–1763, Jun. 28–Jul. 3 2009.
- [8] R. Krishnamachari and M. Varanasi, "Interference Alignment Under Limited Feedback for MIMO Interference Channels," *Arxiv preprint arXiv:0911.5509*, 2009. [Online]. Available: <http://arxiv.org/abs/0911.5509>

- [9] O. El Ayach and R. W. Heath, Jr., "Interference alignment with analog channel state feedback," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 626–636, 2012.
- [10] G. Caire, N. Jindal, and S. Shamai, "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," *Proc. of the Asilomar Conference on Signals, Systems and Computers*, pp. 287–291, Nov. 2007.
- [11] B. Xie, Y. Li, H. Minn, and A. Nosratinia, "Interference alignment under training and feedback constraints," *In Proc. of IEEE Global Telecommunications Conference*, Dec. 5–9 2011.
- [12] O. El Ayach and R. W. Heath, Jr., "Grassmannian differential limited feedback for interference alignment," *submitted to IEEE Transactions on Signal Processing*, Nov. 2011. [Online]. Available: <http://arxiv.org/abs/1111.4596>
- [13] S. Peters and R. W. Heath, Jr., "User partitioning for less overhead in MIMO interference channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 592–603, 2012.
- [14] M. Guillaud and D. Gesbert, "Interference alignment in the partially connected K-user MIMO interference channel," *Proc. of European Signal Processing Conference (EUSIPCO), Barcelona, Spain*, pp. 1–5, Sept. 2011.
- [15] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
- [16] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.
- [17] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, 2010.
- [18] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [19] M. Kobayashi, N. Jindal, and G. Caire, "Training and Feedback Optimization for Multiuser MIMO Downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.
- [20] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5020–5029, 2008.
- [21] W. Santipach and M. Honig, "Optimization of training and feedback overhead for beamforming over block fading channels," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6103–6115, 2010.
- [22] O. El Ayach, A. Lozano, and R. W. Heath, Jr., "Optimizing Training and Feedback for MIMO Interference Alignment," *in Proc. of Asilomar Conference on Signals, Systems, and Computers*, Nov. 6–9 2011.
- [23] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 812–821, Feb. 2012.
- [24] A. Mahmoud, M. El-Khamy, and K. Elsayed, "Interference alignment performance on MIMO X channels with imperfect channel knowledge," *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications, Cesme, Turkey*, Jun. 2012. [Online]. Available: http://4gpp-project.net/administrator/components/com_jresearch/files/publications/Interference Alignment Performance on MIMO X Channels with Imperfect Channel Knowledge.pdf
- [25] O. El Ayach, S. Peters, and R. W. Heath, Jr., "The feasibility of interference alignment over measured MIMO-OFDM channels," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4309–4321, Nov. 2010.
- [26] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Minimum mean squared error interference alignment," *Proc. of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 1106–1110, Nov. 2009.
- [27] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sept. 2011.
- [28] H. Shin and J. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2636–2647, 2003.
- [29] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, 1994.
- [30] A. Lapidath and S. Shamai, "Fading channels: How perfect need perfect side information be?" *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1118–1134, 2002.
- [31] M. Rezaee, R. Couillet, M. Guillaud, and G. Matz, "Sum-Rate Optimization for the MIMO IC under Imperfect CSI: a Deterministic Equivalent Approach," *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications, Cesme, Turkey*, Jun. 2012. [Online]. Available: http://couillet.romain.perso.sfr.fr/docs/conf/SPAWC_Mohsen.pdf
- [32] R. Tresch and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," *in Proc. of IEEE International Conference on Communications*, pp. 1–5, Jun. 2009.
- [33] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 944–960, Jul. 1995.
- [34] A. Lapidath, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1520–1529, 1996.
- [35] T. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," *Proceedings of the Annual Allerton Conference on Communication Control and Computing*, vol. 37, pp. 958–966, 1999.

- [36] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.
- [37] A. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004.
- [38] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental limits of cooperation," *Arxiv preprint arXiv:1204.0011*, Mar. 2012. [Online]. Available: <http://arxiv.org/abs/1204.0011>

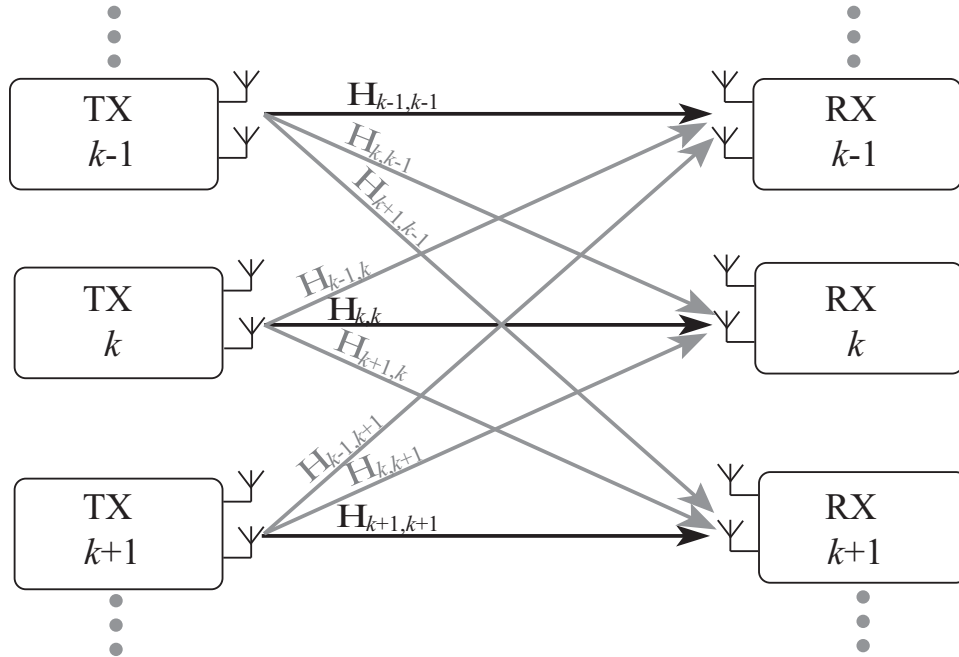


Fig. 1. K -User MIMO interference channel model

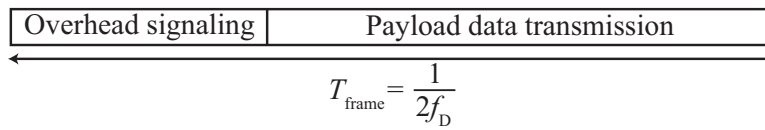


Fig. 2. The overhead model adopted in which training and feedback consume resources that would otherwise be used for data transmission.

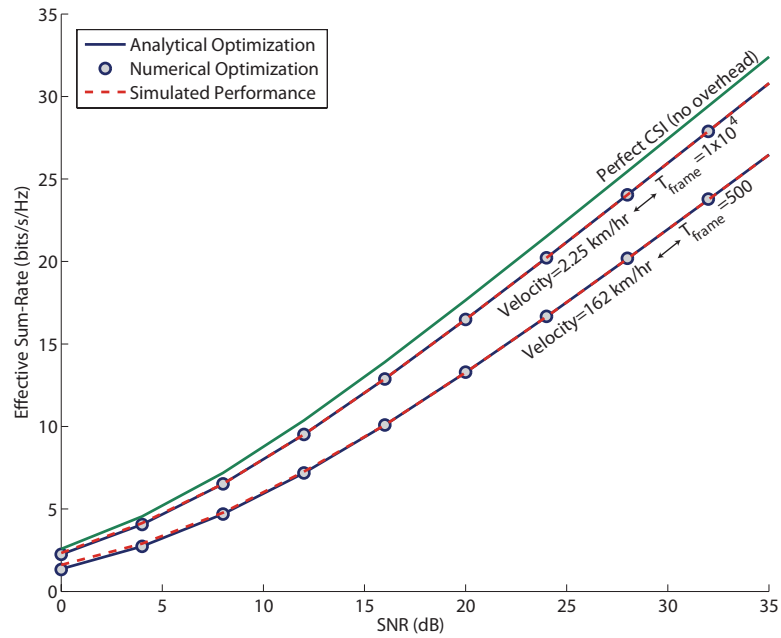


Fig. 3. Effective Sum-Rate vs. SNR for systems with different normalized Doppler spreads. This quantifies the loss in sum-rate due to both imperfect CSI and overhead and shows that the performance predicted by the analytical results presented is an accurate representation of optimal performance.

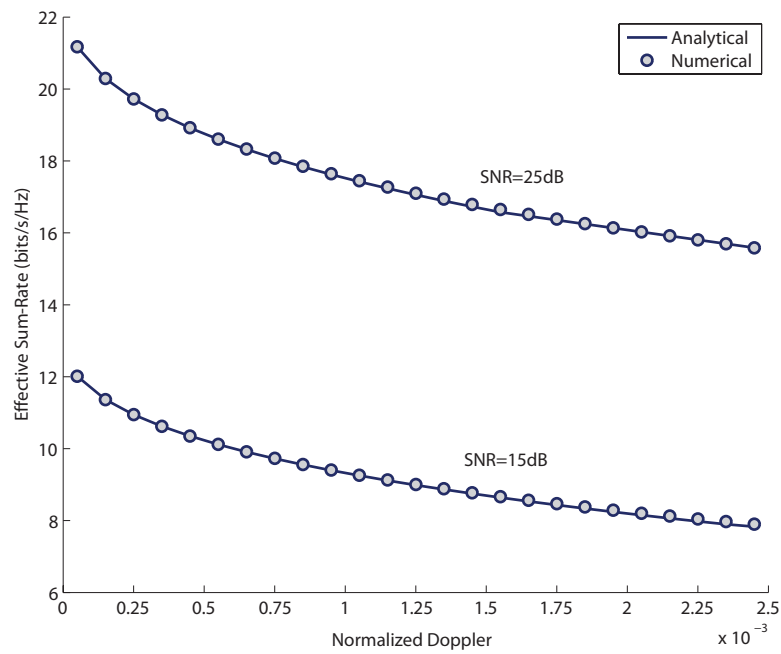


Fig. 4. Effective Sum-Rate vs. Normalized Doppler for IA systems at different SNR levels. This quantifies the degradation in sum-rate as mobility increases resulting in an increased overhead penalty.

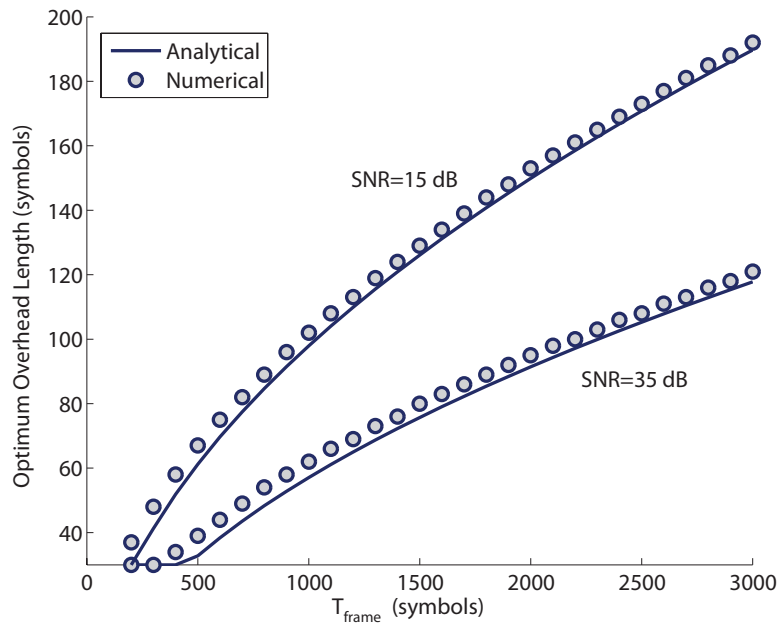


Fig. 5. T_{OHD} vs. T_{frame} . This confirms that the optimal value of T_{OHD} scales with $\sqrt{T_{\text{frame}}}$ as predicted, and shows that optimizing a series expansion of the objective yields remarkably accurate results.

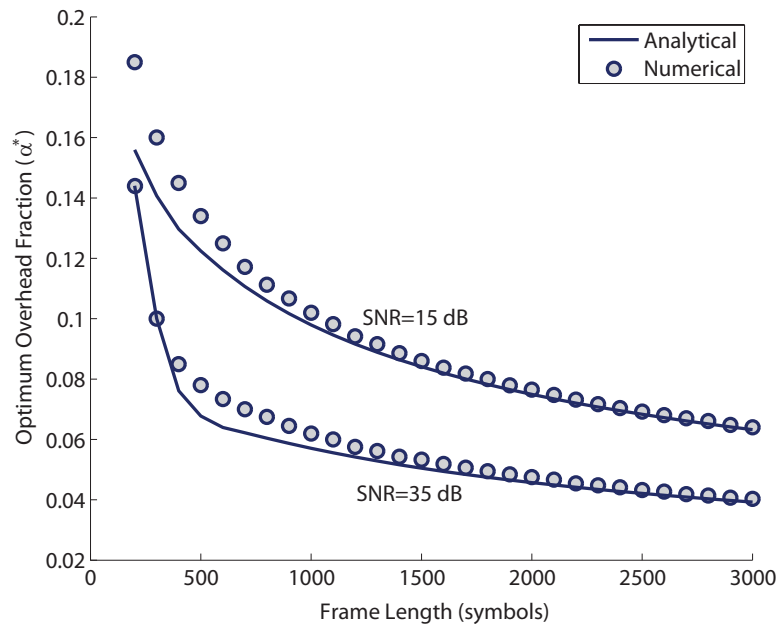


Fig. 6. α^* vs. T_{frame} . This confirms that the optimal value of α scales with $\frac{1}{\sqrt{T_{\text{frame}}}}$ and thus scales with $\sqrt{f_D}$ as predicted.

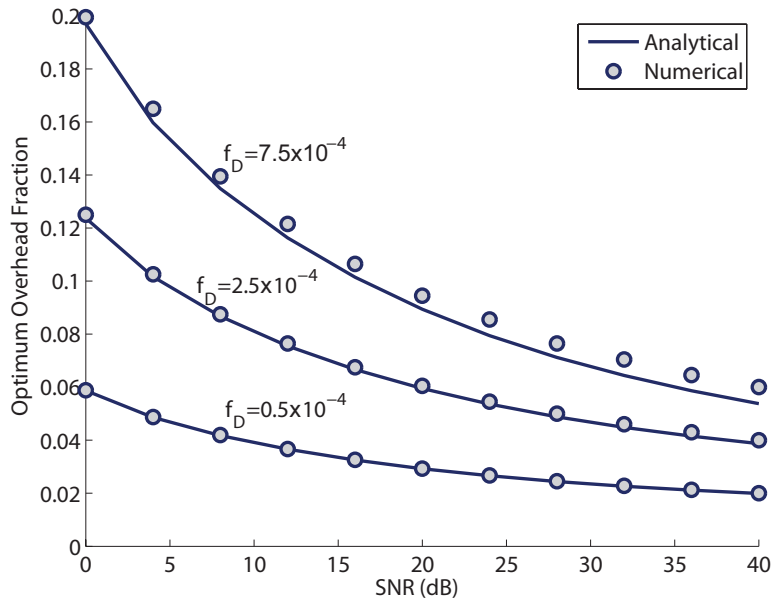


Fig. 7. α^* vs. SNR. This shows the decrease of the optimal overhead budget with SNR. As stated in Section V, it can be shown that the decrease is logarithmic with SNR. The figure also demonstrates that our expansion-base results are very accurate, deviating only slightly in high-SNR high-mobility scenarios.

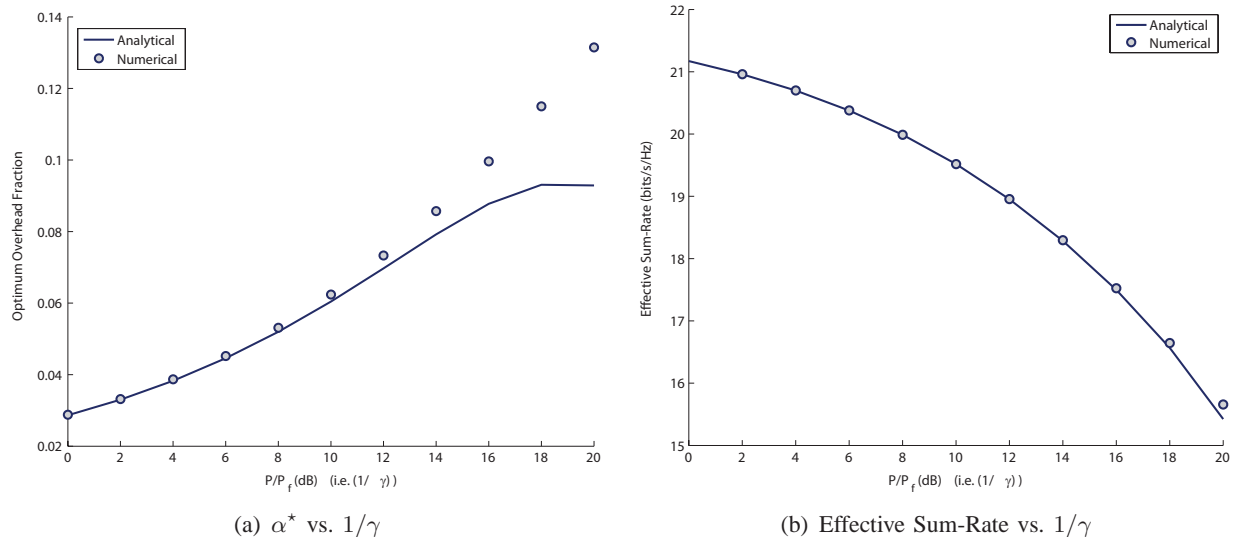


Fig. 8. This figure shows the relationship between α^* and $\bar{R}_{\text{eff}}^*(P)$ with the feedback channel's relative quality for a system with $T_{\text{frame}} = 10^4$. Plot (a) verifies the increase of overhead with $1/\gamma$, when plot in linear scale the square root rate of increase can be verified. Plot (b) verifies the rate of decrease of optimal effective sum-rate with feedback link quality.

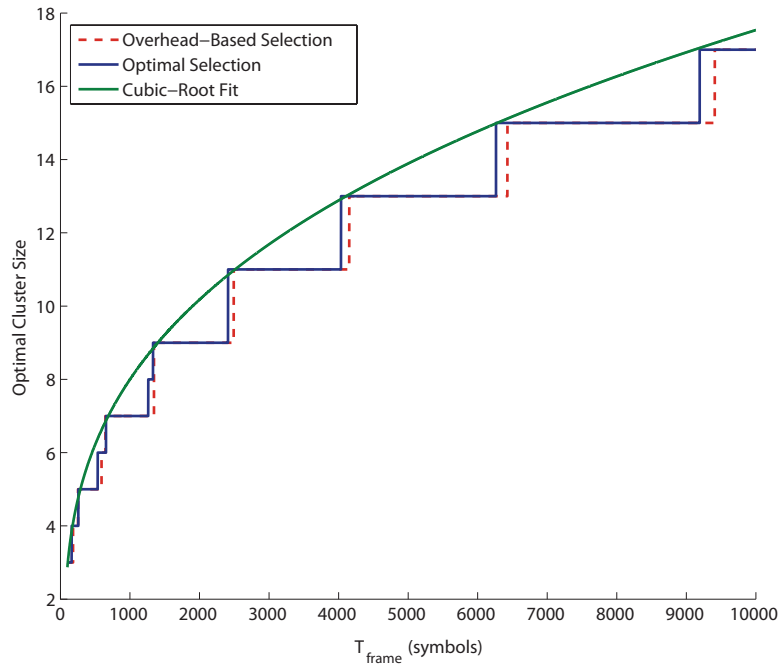


Fig. 9. Optimal Cluster Size vs. T_{frame} . This shows the optimal number of users to coordinate via IA which increases channels coherence time. This also shows that comparing overhead, i.e., overhead based selection, provides accurate decisions on optimal cluster size.

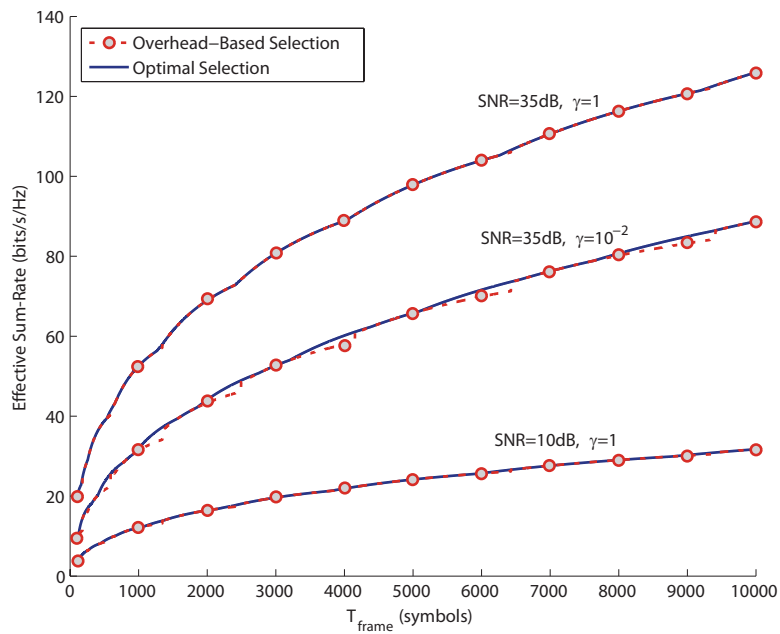


Fig. 10. Effective Sum-Rate with Cluster Size Optimization vs. T_{frame} . This shows the increase in effective sum-rate as a function of T_{frame} when the cluster size is chosen to maximize rate. This also quantifies the minimal sum-rate loss due to sub-optimal overhead-only based cluster sizing.