On Resource Allocation in Two-Way Limited Feedback Beamforming Systems

David J. Love and Chun Kin Au-Yeung
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47906, USA
Email: cauyeung@purdue.edu, djlove@ecn.purdue.edu

Abstract—The benefits employing channel state information (CSI) at the transmitter in a multiple antenna wireless link are well documented in the literature. One of the most popular techniques to provide the transmitter with CSI in frequency division duplexing wireless links is by sending a finite number of feedback bits. However, the effect of the overhead created by these feedback bits on the link performance is still not well understood. In this paper, we study a specific scenario of limited feedback known as limited feedback beamforming. We look at the effect of allocating resources to feedback and the scaling of these resources. Monte Carlo simulations also demonstrate the inherent tradeoff between the forward and reverse links in a wireless system.

I. INTRODUCTION

Over the last ten years, much work has been done to exploit channel state information (CSI) at the transmitter in multiple antenna wireless systems. The challenge with implementing signaling techniques that adapt to CSI is developing approaches to supply the transmitter with CSI. While the receiver can obtain CSI by training, the transmitter can only obtain CSI if it is a transceiver (i.e., it acts as both a transmitter and receiver). This means that if the forward link in a wireless system uses CSI adaptive signaling the transmitter will have to use the reverse link to obtain this CSI. Thus, understanding the two-way nature of wireless systems is critical.

In frequency division duplexing (FDD) systems, one of the main applications of the two-way structure is the use of limited feedback (see for example [1]–[3] and the references therein). The basic idea with this work has been that the forward link leverages CSI at the transmitter by having the receiver send some form of quantized CSI information on the reverse link. Most of this research, however, has made the often questionable assumption that the feedback channel is ‘free’ (i.e., available without a loss in time, power, frequency, etc.). This kind of assumption generally has several problems. When a user allocates resources for feedback that same user encounters a corresponding loss in resources available for transmitting data. Even though the feedback increases the maximum achievable rate on the forward link, feedback decreases the maximum achievable rate on the reverse link. Therefore, there exists a tradeoff between the maximum achievable rate in the forward and reverse links.

There has only been limited work that takes into account the two-way nature. A two-way channel estimation scheme is proposed in [4]. Ref. [5] discusses using a quasi-symmetric channel mode. A model for deducting a feedback rate penalty in a symmetric wireless system is proposed in [6]. Feedback design that takes into account available resources is available in [7]. Ref. [8] studies a two-stage bidirectional training scheme suitable for time division duplexing (TDD) and derives the diversity-multiplexing tradeoff is derived. An analytical study of allocating power to training, feedback, and data is given in [9].

In this paper, we look at the at the asymptotic scaling of time and power resources required to optimize a limited feedback beamforming multiple antenna system. We propose models for allocating feedback that takes into account parameters such as coherence time and number of antennas. Numerical results demonstrate the tradeoff. Interestingly, a related set-up was independently developed in [10] that looks at scaling with a large number of transmit antennas.

In Section 2, we set-up the limited feedback system that is analyzed. Section 3 addresses the asymptotic scaling of the number of feedback channel uses. The amount of power that should be allocated is asymptotically characterized in Section 4. Section 5 presents simulations, and we conclude in Section 6.

II. LIMITED FEEDBACK BEAMFORMING

The two-way system under consideration is described herein.

A. System Set-up

The system under consideration is shown in Fig. 1. Consider two mobiles, labeled Mobile $A$ and Mobile $B$, signaling to each other wirelessly using a frequency division duplexing $M$ transmit antenna by $M$ receive antenna wireless link with beamforming and combining. We refer to Link 1 as the link where Mobile $A$ transmits to Mobile $B$ and Link 2 as the link where Mobile $B$ transmits to Mobile $A$.

We assume a block fading model for both links. The channels for Links 1 and 2 at the $i$th block are denoted by the $M \times M$ matrices $H_{1i}$ and $H_{2i}$, respectively. Both matrices have independent and identically distributed $CN(0,1)$ entries. By the FDD and block fading assumptions, we assume that $H_{1i}[1]$ and $H_{2i}[2]$ are independent when $l_1 \neq l_2$ or $i_1 \neq i_2$. Both channels are constant for $T$ channel uses, and with a
slight abuse of terminology, we refer to \( T \) as the coherence time of the channel. Mobile \( A \) (\( B \)) is assumed to have perfect knowledge of \( H_2[i] \) (\( H_1[i] \)) throughout the entire \( T \) channel uses of the \( i \)th block. Therefore, this analysis will ignore the role of channel estimation in feedback optimization.

In addition to symmetry in the numbers of antenna and channel model, we will assume that both mobiles have the same transmit power constraints. These symmetries motivate both links to use identical framing and power allocation. The below discussion will apply to both links.

The \( T \) channel uses will be divided into two different phases. The first phase for channel uses \( 1, 2, \ldots, T_f \) (with \( 0 \leq T_f \leq T \)) is the feedback phase. During this phase, the transmitter for Link \( l \) has no knowledge of \( H_l[i] \) but needs to convey information about the beamformer that should be used to transmit over \( H_{1+\text{mod}(l,2)}[i] \). For this reason, beamforming can not be optimally designed during the feedback phase. The analysis will assume that each mobile transmits only from its first antenna giving an input-output relationship for link \( l \) at channel use \( k \) (with \( k \in \{1, \ldots, T_f\} \)) of block \( i \) as

\[
y_{f,l}[k, i] = \sqrt{\rho_f} H_l[i] e_1 s_{f,l}[k, i] + n_l[k, i]
\] (1)

where \( e_1 = [1 \ 0 \ \cdots \ 0]^T \), \( n_l[k, i] \) is an \( M \)-dimensional vector with independent \( \mathcal{CN}(0,1) \) entries that are also independent in both \( k \) and \( i \), \( s_{f,l}[k, i] \) is the feedback information symbol containing information about the other link satisfying \( E[|s_{f,l}[k, i]|^2] = 1 \), and \( \rho_f \) is the SNR during the feedback phase.

The second phase is the data phase. During this, the signal received on Link \( l \) at channel use \( k \) (with \( k \in \{T_f+1, \ldots, T\} \)) of block \( i \) is given by

\[
y_{d,l}[k, i] = \sqrt{\rho_d} H_l[i] f_i[k] s_{d,l}[k, i] + n_l[k, i]
\] (2)

where \( s_{d,l}[k, i] \) is the transmitted data symbol satisfying \( E[|s_{d,l}[k, i]|^2] = 1 \), \( f_i[k] \) is the beamforming vector for block \( i \) satisfying \( \|f_i[k]\| = 1 \), and \( \rho_d \) is the SNR during the data phase.

With this framework, the feedback bits on Link \( l \) are used to convey \( f_{1+\text{mod}(l,2)}[i] \) for the opposite link. Therefore, letting \( B \) denote the number of bits sent per feedback channel use during the feedback phase, the \( T_f B \) feedback bits describing \( f_{1+\text{mod}(l,2)}[i] \) are encoded as \([s_{f,l}[1, i] \ \cdots \ s_{f,l}[T_f, i]]\). This means that no time diversity is available during feedback.

The value of \( B \) is fixed for all blocks. This value must be chosen to accommodate some sort of average probability of error constraint on the feedback codeword. Thus, the number of feedback bits is a function of the feedback SNR \( \rho_f \). To model this, we will assume that \( B \) is chosen as

\[
B(\rho_f) = (B_0 + \log_2(\rho_f))^+
\] (3)

where \( B_0 \) is a constant dependent on \( M \) and \((\cdot)^+ = \max(\cdot, 0)\). For the purpose of analysis, we assume that the feedback codeword is received without error when feedback bits are scaled according to (3).

In contrast, we assume that the data blocklength \( T_d \) is much larger than \( T \). For convenience, let \( T_d = K(T - T_f) \) for some integer \( K \). If \( R \) is the number of bits per data channel use, \( T_d R \) bits are encoded as \([s_{d,l}[T_f + 1, 1] \ \cdots \ s_{d,l}[T, 1] \ \cdots \ s_{d,l}[T_f + 1, K], s_{d,l}[T, K]\]. Therefore, the encoded data will be able to leverage time diversity.

To take into account power limitations, we assume that each mobile can support an average power of \( \rho \). Therefore,

\[
T_f \rho_f + (T - T_f) \rho_d = T \rho
\] (4)

With this set-up, the power available for data is

\[
\rho_d = \frac{1}{1 - \alpha} \rho - \frac{\alpha}{1 - \alpha} \rho_f
\] (5)

where \( \alpha = \frac{T_f}{T} \). As the power allocated for feedback increases, this causes a loss in the power available for data.

### B. Codebook Approach and Maximum Achievable Rate

The \( T_f B(\rho_f) \) bits of feedback per block sent on Link \((1 + \text{mod}(l, 2)) \) convey the choice of the beamformer \( f_l[i] \) for use on Link \( l \). We will model this as being done using a codebook \( \mathcal{F}[i] \). The codebook will consist of \(2^{T_f B(\rho_f)}\) unit vectors that are known at both Mobile \( A \) and \( B \).

Often, this codebook is designed offline and fixed (meaning that \( \mathcal{F}[i] = \mathcal{F} \)). In this paper, however, we will assume that the codebook is designed using random vector quantization \([11], [12]\) In this model, the codebook \( \mathcal{F} \) will be generated randomly at each block using a uniform distribution on the \( M \)-dimensional unit sphere in \( \mathbb{C}^M \).

The maximum achievable rate (in bits per data channel use) as \( T_d \to \infty \) for Link \( l \) is

\[
C_1(\rho_d, T_f B(\rho_f)) = E\left[ \log_2 \left( 1 + \max_{f_i \in \mathcal{F}} \rho_d \|H_l f_i\|^2 \right) \right]
\] (6)

where the expectation is over \( H_l \) and \( \mathcal{F} \). Note that because of the symmetry in the channel distributions,

\[
C(\rho_d, T_f B(\rho_f)) \triangleq C_1(\rho_d, T_f B(\rho_f)) = C_2(\rho_d, T_f B(\rho_f)).
\] (7)

The maximum achievable rate can be written as

\[
C(\rho_d, T_f B(\rho_f)) = C(\rho_d, \infty) - L(\rho_d, T_f B(\rho_f))
\] (8)

where

\[
L(\rho_d, T_f B(\rho_f)) = C(\rho_d, \infty) - C(\rho_d, T_f B(\rho_f))
\] (9)
is the rate loss incurred from using finite rate feedback. This is given by
\[
L(\rho_d, T_f B(\rho_f)) = E \left[ \log_2 \left( 1 + \rho_d \|H\|^2 \right) - \log_2 \left( 1 + \max_{f \in F} \rho_d \|Hf\|^2 \right) \right]
\]
where the dependence on the link has been omitted for clarity.

The loss can actually be bounded as [12]–[14]
\[
L(\rho_d, T_f B(\rho_f)) \leq \log_2(e) E \left[ \rho_d \left( \|H\|^2 - \max_{f \in F} \|Hf\|^2 \right) \right] \leq \log_2(e) \rho_d \mu E \left[ \max_{f \in F} \|v^*f\|^2 \right]
\]
where \(v\) is a vector that is uniformly distributed on the \(M\)-dimensional complex unit sphere and \(\mu = E \|\|H\|^2\|_2\).

The bound in (11) becomes very loose as \(\rho_d\) grows large (but very tight as \(T_f B(\rho_f)\) grows large for fixed \(\rho_d\)). To accommodate the high SNR case, note that
\[
L(\rho_d, T_f B(\rho_f)) \rightarrow E \left[ \log_2 \left( \|H\|^2 \right) - \log_2 \left( \max_{f \in F} \|Hf\|^2 \right) \right] \leq - \log_2 \left( \max_{f \in F} \|v^*f\|^2 \right) \leq \log_2(e)(M-1)2^{M-2} \left( 2^{T_f B(\rho_f)} - 1 \right)^{-1/(M-1)}.
\]

This lower bound can be shown to be very tight when \(T_f\) grows large.

**Lemma 2:** To optimize (14) with fixed \(\rho_f\) (with \(B(\rho_f) > 0\)) and \(\rho\), the temporal feedback resources should scale as \(T_f = O(\log_2(T))\).

**Proof:** The proof uses Lemma 1 and the fact that the lower bound in (15) is tight and concave.

\[
\log_2(e) \rho_d \mu \left( \frac{1}{T} \right)^{\frac{B(\rho_f)}{M-1} - \frac{T_f B(\rho_f)}{M-1}} - C(\rho_d, \infty)
\]

As \(T\) grows large, the derivative behaves as
\[
\log_2(e) \rho_d \mu \left( \frac{B(\rho_f) \ln(2)}{M-1} - \frac{T_f B(\rho_f) \ln(2)}{T(M-1)} \right) 2^{-\frac{T_f B(\rho_f)}{M-1}} - C(\rho_d, \infty)
\]

Giving an asymptotically optimal \(T_f\),
\[
T_f = \frac{M-1}{B(\rho_f)} \left( \log_2(T) + \log_2 \left( \frac{\rho_d \mu B(\rho_f)}{(M-1)C(\rho_d, \infty)} \right) \right)
\]

(16)

The result in Lemma 2 makes sense but is a little surprising.

The amount of time resources dedicated to feedback will scale quite slowly with the coherence time.

**IV. FEEDBACK SCALING WITH POWER**

The next step in understanding feedback resources is to understand how to allocate power to the feedback link given a fixed \(T_f\) (with \(T_f > 0\)) and \(T\). First, let us state the following lemma.

**Lemma 3:** When \(\rho_f\) is optimally chosen subject to (4) with fixed \(T_f, T, \rho_d \rightarrow \infty\) as \(\rho \rightarrow \infty\),

Again, this lemma is obvious.

As \(\rho_d \rightarrow \infty\),
\[
R(\rho_d, T_f) \rightarrow \left( 1 - \frac{T_f}{T} \right) E \left[ \log_2 \left( \rho_d \max_{f \in F} \|Hf\|^2 \right) \right]
\]

(17)

Using (3) and the bound (13),
\[
\left( 1 - \frac{T_f}{T} \right) \left( \eta + \log_2(\rho_d) - \log_2(e)(M-1)2^{M-2} \left( 2^{T_f B(\rho_f)} - 1 \right)^{-1/(M-1)} \right)
\]

\[
= \left( 1 - \frac{T_f}{T} \right) \left( \eta + \log_2(\rho_d) - \log_2(e)(M-1)2^{M-2} \left( 2^{T_f B(\rho_f)} - 1 \right)^{-1/(M-1)} \right)
\]

(18)

where \(\eta = E \left[ \log_2 \left( \|H\|^2 \right) \right]\).

The following lemma summarizes the high SNR behavior.

**Lemma 4:** As \(\rho \rightarrow \infty\) with fixed \(T_f > 0\) and \(T\), \(\rho_f\) behaves as \(O(\rho^k)\) where \(k = O \left( \frac{1}{T_f} \right)\).
Proof: To maximize this lower bound, we only need to deal with maximizing,
\begin{align*}
g(\rho_f) &= \ln \left( \frac{1}{1-\alpha} \rho - \frac{\alpha}{1-\alpha} \rho_f \right) \\
&\quad - (M-1)^2 \left( 2^{T_f B_0 \rho_f T_f - 1} - 1 \right)^{1/(M-1)},
\end{align*}
which uses (5). This function can be shown to be asymptotically concave over the region of interest with derivative
\begin{equation}
\frac{dg}{d\rho_f} = -\frac{\alpha}{\rho - \alpha \rho_f} + \left( 2^{M-2-T_f B_0 T_f} \rho_f \right)^{T_f-1} \left( 2^{T_f B_0 \rho_f T_f - 1} - 1 \right)^{M-1}.
\end{equation}
By dealing with only the higher order terms, we find that we should set
\[\rho_f \sim \left( 2^{M-2-T_f B_0 T_f} \rho \right)^{M-1}.
\]
Lemma 4 tells us that should have polynomial growth in the feedback power as a function of total power. However, this growth rate is couple with the amount of channel uses allocated to feedback.

V. SIMULATIONS

In this section, we verify our feedback resource scaling with coherence time and power expressions numerically.

To simulate the scaling of the coherence time, we used the lower bound in (15). In Fig. 2, we plot the optimal number of feedback channel uses \(T_f\) against the coherence time \(T\) while keeping the power used during the feedback and the data phases constant at \(\rho_f = 10\) and fixing \(\rho_f\) such that \(B(\rho_f) = 1\). The logarithmic behavior exhibited matches the prediction in Lemma 2. Note also that increasing the number of antennas \(M\) increases the optimal \(T_f\) for any given \(T\). Intuitively, this is because increasing \(M\) increases the dimension of the space to be represented during the feedback phase. Hence, one requires more feedback bits to represent the space satisfactorily.

In Fig. 3, we plot the optimal feedback power \(\rho_f\) against the available power \(\rho\) while keeping the coherence time constant at \(T = 10\). Both scales show the power in decibels. The asymptotic lower bound resulting from (12) and (17) was used. The straight line behavior in the plot satisfies the prediction from Lemma 4 where the optimal feedback power \(\rho_f\) grows polynomially with available power \(\rho\). Also observe that the optimal feedback power decreases as \(T_f\) increases. This is intuitive as less power for each feedback channel use is needed when feedback spans more channel uses.

VI. CONCLUSION

In this paper, we looked at how feedback resources should scale to optimize the two-way symmetric rate. This scaling is important because too much or too little feedback can cause a significant rate loss. We looked at optimizing both time and power resources. In future analysis, it would be interesting to more thoroughly take into account feedback probability of error scaling with \(T_f\) and \(\rho_f\). We have assumed that the effect of the probability of error scaling is negligible.

VII. APPENDIX

We bound \(E \left[ -\log_2 \left( \max_{f \in \mathcal{F}} |v^* f|^2 \right) \right] \) in this section.

Lemma 5: The expected value of the inner product satisfies
\begin{align*}
\log_2(e) (N-1) B \left( N-1, \frac{M}{M-1} \right) \\
\leq E \left[ -\log_2 \left( \max_{f \in \mathcal{F}} |v^* f|^2 \right) \right] \\
\leq \log_2(e) (M-1) 2^{M-2} (N-1) B \left( N-1, \frac{M}{M-1} \right)
\end{align*}
where \( B(\cdot, \cdot) \) is the beta function and \( \mathcal{F} \) is an RVQ codebook with \( N = 2^{T_f} B(\rho T_f) \) random unit vectors.

Proof: Let \( w_N \) denote a random variable defined as \( \max_{v \in \mathcal{F}} |v^T w| \) given an \( N \) vector random unit vector codebook \( \mathcal{F} \) and random unit vector \( v \). From [12],

\[
E [-\ln (w_N)] = \int_0^1 \frac{1}{x} \left( 1 - (1 - x)^{M-1} \right)^N \, dx \\
= \int_0^1 \frac{1}{x} \left( 1 - (1 - x)^{M-1} \right) \left( 1 - (1 - x)^{M-1} \right)^{N-1} \, dx.
\]

Note that \( 1 \leq \frac{1-(1-x)^{M-1}}{x} \leq (M-1)2^{M-2} \). Therefore,

\[
E [1 - w_{N-1}] \leq E [-\ln (w_N)] \leq (M-1)2^{M-2} E [1 - w_{N-1}].
\]

The proof is completed by noting that

\[
E [1 - w_N] = NB \left( N, \frac{M}{M-1} \right) \leq (N)^{-\frac{1}{M-1}}.
\]

Substituting \( N = 2^{T_f} B(\rho T_f) \) completes the proof. \( \blacksquare \)

The bound in (13) follows from this lemma because [14] \( (N-1)B \left( N-1, \frac{M}{M-1} \right) \leq (N-1)^{-\frac{1}{M-1}} \). In fact, as \( N \to \infty \), \( (N-1)B \left( N-1, \frac{M}{M-1} \right) \to (N-1)^{-\frac{1}{M-1}} \).

**ACKNOWLEDGMENT**

This work was supported in part by the National Science Foundation under CCF0513916 and the AT&T Foundation. Chun Kin Au-Yeung is supported by a Motorola UPR grant.

**REFERENCES**


