How Much Training and Feedback are Needed in MIMO Broadcast Channels?

Mari Kobayashi,
SUPELEC
Gif-sur-Yvette, France

Giuseppe Caire,
University of Southern California
Los Angeles CA, 90089 USA

Nihar Jindal
University of Minnesota
Minneapolis MN, 55455 USA

***Abstract*— We consider a MIMO fading broadcast channel where channel state information is acquired at user terminals via downlink training and explicit analog feedback is used to provide transmitter channel state information (CSIT) to the base station. The feedback channel (the corresponding uplink) is modeled as a MIMO multiple access channel. Under the assumption that data transmission, downlink training, and feedback are performed within the same channel coherence interval of length T symbols, the optimization of a lower bound on the achievable ergodic rate sum yields a non-trivial resource allocation tradeoff. We solve this tradeoff and provide the optimal training and feedback resource allocation for the case of zero-forcing beamforming. When the same power level is used during all stages, it is found that the optimal length of the training + feedback phases increases as $O(\sqrt{T})$ for large T. On the other hand, when different power levels can be used for different stages, for sufficiently large T it is optimal to use the minimum number of symbols for training + feedback but to use power of order $O(\sqrt{T})$.**

I. Introduction

We consider a MIMO Gaussian broadcast channel modeling the downlink of a system where a Base Station (BS) has M antennas and serves K single-antenna User Terminals (UTs). A channel use of such system is described by

$$y_k = \mathbf{h}_k^H \mathbf{x} + z_k, \quad k = 1, \dots, K \tag{1}$$

where y_k is the channel output at UT k, $z_k \sim \mathcal{CN}(0, N_0)$ is the corresponding Additive White Gaussian Noise (AWGN), $\mathbf{h}_k \in \mathbb{C}^M$ is the vector of channel coefficients from the BS antenna array to the k-th UT antenna and $\mathbf{x}$ is the vector of channel input symbols transmitted by the BS, subject to the average power constraint $\mathbb{E}[|\mathbf{x}|^2] \leq P$. We assume a block fading model, i.e., the channel remains constant over a coherence interval of T channel uses. Albeit suboptimal, we focus on zero-forcing (ZF) beamforming with $K = M$ users for its analytical tractability.

As commonly done in existing wireless communications systems, each UT learns its own downlink channel from a training phase. Furthermore, we assume frequency-division duplexing (FDD), requiring explicit (closed-loop) CSIT feedback from the UTs to the BS. CSIT feedback is performed by transmitting the estimated unquantized channel coefficients on the uplink by suitable quadrature-amplitude modulation, as in [1], [2]. The feedback link (uplink) is a faded MIMO multiple access channel with M users and M receiving antennas, since we assume that the BS can use its antenna array for both downlink and uplink. Moreover, we assume that data transmission, downlink training, and feedback are performed within the same coherence interval of duration T. We shall refer to T as the "frame length" and let T_t denote the total number of training and feedback symbols.

In [2] it is shown that the achievable ergodic sum rate of ZF beamforming with Gaussian inputs and CSI training and feedback (to be specified in Section II) is lower bounded by

$$R_k \geq \left(1 - \frac{T_t}{T}\right)\left(R^{\mathrm{ZF}} - \overline{\Delta R}\right) \tag{2}$$

where the factor T_t/T is the multiplexing gain loss due to the fact that only $T - T_t$ symbols are used for actual data transmission, R^{ZF} denotes the achievable per-user rate of ZF beamforming with ideal CSI, and the upper bound of the rate gap $\overline{\Delta R}$ represents the degradation in rate due to the imperfect CSI. The optimization of the sum-rate (2) stems from the tension between two issues: on one hand, the upper bound of the rate gap $\overline{\Delta R}$ decreases with T_t [2]; on the other hand, the multiplexing gain loss T_t/T becomes non-negligible as T_t increases. This yields a non-trivial tradeoff, the characterization of which is the subject of this paper.

We start by considering the setting where the same (per-symbol) power is used during the training/feedback and data transmission stages. In this case, we show that the optimal number $T_t^\star$ of training and feedback symbols that maximizes (2) increases as $O(\sqrt{T})$ for $T \to \infty$. Next an average power constraint across the training/feedback and data stages is considered, in which case we must optimize over the number of symbols as well as the power allocated to the training, feedback and data phases. It is shown that for sufficiently large frame length, it is optimal to allocate the minimum number of symbols to training/feedback but to increase the training/feedback power as $O(\sqrt{T})$ as $T \to \infty$; these results agree to some extent with [3], [4].

In order to put this work in the context of relevant literature, we note here that similar questions have been addressed in [3], [4], [5], [6], [7]. In [3], point-to-point MIMO communication is considered and only downlink training is addressed (imperfect CSIR, no CSIT). On the other hand, in [5] perfect CSIR is assumed and the resources to be used for channel feedback are investigated. In [6], [7] the model of [3] is extended to also incorporate quantized channel feedback and transmitter beamforming. Although the setup is quite similar to ours, the emphasis of [6], [7] on the asymptotic regime where the

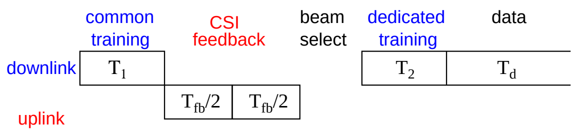

Fig. 1. Channel state estimation and feedback model.

number of antennas and T are simultaneously taken to infinity leads to rather different conclusions as compared to the present work. In [4] a MIMO broadcast with downlink training and perfect channel feedback (i.e., the BS is also able to view the received training symbols) is considered. It is shown that the sum rate achievable with a dirty paper coding-based strategy has a form very similar to the achievable rate expressions in [3], and thus many of the conclusions from [3] directly carry over. On the other hand, we consider the more practical case where there is imperfect feedback from each UT to the BS and also study achievable rates with ZF beamforming, which has lower complexity than dirty-paper coding.

II. Channel State Estimation and Feedback

The training and feedback scheme that allows all UTs to estimate their channel, feedback CSIT to the BS and eventually perform coherent detection is illustrated in Fig. 1 and it is formed by the following phases (see [2] for further details):

1) Common training: T_1 shared pilot symbols (essentially $\frac{T_1}{M}$ pilots are sent per BS antenna, and thus we require $T_1 \geq M$) are sent to all UTs to estimate their downlink channel vectors $\{\mathbf{h}_k\}$ based on the observation

$$\mathbf{s}_k = \sqrt{\frac{T_1 P}{M}}\,\mathbf{h}_k + \mathbf{z}_k \tag{3}$$

where $\mathbf{z}_k \sim \mathcal{CN}(0, N_0\mathbf{I})$. Each UT performs linear MMSE of $\mathbf{h}_k$ from the observation $\mathbf{s}_k$.
2) CSI feedback: in this paper we treat only the case of "analog" channel state feedback [1], [2] where each UT sends a scaled version of its channel MMSE estimate using unquantized quadrature-amplitude modulation over the uplink. The uplink is modeled as a MIMO MAC with a common SNR $\frac{P_{\text{fb}}}{N_0}$ for all M users. Uplink transmission is organized as follows. The M UTs are partitioned into groups of size L (such that M/L is an integer). The UTs in the same group simultaneously transmit their feedback signal over LT_{fb}/M channel uses (each UT transmits each of its M channel coefficient estimates LT_{fb}/M^2 times), and the BS performs MMSE estimation on the received vectors to acquire imperfect CSIT. Different groups access the feedback channel in time-division, such that the feedback spans a total of T_{fb} uplink symbols. In [2] we have shown that the choice $L = M/2$ minimizes the rate gap for this type of feedback; thus we consider only this choice of L and furthermore have the requirement $T_{\text{fb}} \geq 2M$ because each of the two groups requires at least M uplink symbols. The uplink and downlink channels are assumed to be identically distributed (i.i.d. coefficients $\sim \mathcal{CN}(0,1)$) and independent, due to the assumption that in FDD the uplink and downlink are separated in frequency by much more than the channel coherence bandwidth.
3) Beamformer selection: the BS selects the beamforming vectors by treating its available CSIT $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_M]$ as if it was the true channel and computes the transmit vector $\mathbf{x} = \widehat{\mathbf{V}}\mathbf{u}$, where $\widehat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_M]$ is the beamforming matrix, such that $\hat{\mathbf{v}}_k$ is a unit-norm vector orthogonal to the subspace $\mathcal{S}_k = \text{span}\{\widehat{\mathbf{h}}_j : j \neq k\}$ and $\mathbf{u} \in \mathcal{CN}(\mathbf{0}, \frac{P}{M}\mathbf{I})$ contains the symbols from M independently generated Gaussian codewords.
4) Dedicated training: because $\widehat{\mathbf{V}}$ is a function of a noisy version of all UT channels, it is not known by any UT. In order to enable coherent detection, a dedicated downlink training sequence is inserted in each beam so that each UT k can estimate its useful signal coefficient $a_k = \mathbf{h}_k^H \widehat{\mathbf{v}}_k$ using linear MMSE estimation. The total dedicated training phase length is T_2 (essentially $\frac{T_2}{M}$ pilots are sent per beam, and thus we require $T_2 \geq M$), so that each UT k observes

$$r_k = \sqrt{\frac{T_2 P}{M}} a_k + z_k$$

5) Data transmission: after the dedicated training, the BS sends the coded data symbols for the rest of the frame. Inspired by the setting of [3], we assume that the overall procedure of training and feedback occurs in the same downlink frame of duration T, over which the channels are constant. Hence, we let $T_t = T_1 + T_2 + T_{\text{fb}}$ to denote the total length of common training, dedicated training, and feedback and $T_d = T - T_t$ to denote the length of downlink data symbols[1].

III. Optimizing Training and Feedback

A. Equal Training/Feedback and Data Power

In [2, Section VIII] it is shown that the rate gap with analog feedback over MIMO-MAC with the optimal group size choice $L = M/2$ is upper bounded by

$$\overline{\Delta R} = \log\left(1 + \frac{M-1}{T_1} + \frac{1}{T_2} + \frac{4(M-1)}{MT_{\text{fb}}}\right). \tag{4}$$

Our objective is to find T_1, T_2, T_{fb} such that the achievable rate of ZF (RHS of (2)) is maximized. This leads to the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \left(1 - \frac{T_t}{T}\right)\left[R^{\text{ZF}} - \log(1 + g(T_1, T_2, T_{\text{fb}}))\right] \\ \text{subject to} \quad & T_1 + T_2 + T_{\text{fb}} \leq T_t \\ & T_i \geq T_i^{\min}, \quad i = 1, 2, \text{fb} \end{aligned} \tag{5}$$

where $T_1^{\min} = T_2^{\min} = M, T_{\text{fb}}^{\min} = 2M$, and where we define

$$g(T_1, T_2, T_{\text{fb}}) = \frac{w_1}{T_1} + \frac{w_2}{T_2} + \frac{w_3}{T_{\text{fb}}} \tag{6}$$

with $w_1 = M-1, w_2 = 1, w_3 = \frac{4(M-1)}{M}$. Although the objective is neither concave nor convex, it is possible to solve

[1] In settings where uplink and downlink symbols are valued differently, it is straightforward to extend our analysis to the case where $T_t = T_1 + T_2 + \Gamma T_{\text{fb}}$ for some constant Γ.

through a two-step optimization: we first optimize T_1, T_2, T_{fb} for a fixed T_t, and then optimize with respect to T_t.

When T_t is fixed, the original optimization problem (5) reduces to the minimization of $g(T_1, T_2, T_{\text{fb}})$. This is readily seen to be a convex optimization, and can be solved by forming the Lagrangian with respect to the constraint (5)

$$\mathcal{L}(T_1, T_2, T_{\text{fb}}, \mu) = g(T_1, T_2, T_{\text{fb}}) + \frac{1}{\mu^2}(T_1 + T_2 + T_{\text{fb}})$$

where $\mu > 0$ is the Lagrangian multiplier. The KKT condition yields the following solution

$$T_i^{\star} = \max\left\{T_i^{\min}, \sqrt{w_i}\mu\right\} \tag{7}$$

which is simply a linear function of μ with a slope $\sqrt{w_i}$ for $\mu \geq \frac{T_i^{\min}}{\sqrt{w_i}}$. Assuming $\mu \geq \max_i \frac{T_i^{\min}}{\sqrt{w_i}}$ such that $T_i \geq T_i^{\min}$ for all i (or equivalently, $T_t \geq \sqrt{\frac{\mathcal{K}}{w_i}} T_i^{\min}$ for all i), we have

$$T_i^{\star} = \sqrt{\frac{w_i}{\mathcal{K}}} T_t \tag{8}$$

where we let $\mathcal{K} = (\sum_{i=1}^{3} \sqrt{w_i})^2$ and the objective value is given by

$$g(T_t) = \frac{\mathcal{K}}{T_t} \tag{9}$$

It is clear that T_t is shared between common/dedicated training and feedback proportional to the square root of the weight w_i.

Using (9) in (5), the original problem can be characterized in terms of a single variable T_t. Namely the second step of the proposed optimization corresponds to maximizing

$$f(T_t) = \left(1 - \frac{T_t}{T}\right)\left[R^{\text{ZF}} - \log\left(1 + \frac{\mathcal{K}}{T_t}\right)\right] \tag{10}$$

Although this f is concave in T_t, the optimal $T_t^{\star}$ cannot be given in closed form. However, it can be easily obtained by numerically solving for $\frac{\partial f}{\partial T_t} = 0$. This amounts to a simple line search. For later use, the derivative is calculated as

$$\frac{\partial f}{\partial T_t} = \frac{\mathcal{K}\left(1 - \frac{T_t}{T}\right)}{T_t^2\left(1 + \frac{\mathcal{K}}{T_t}\right)} - \frac{1}{T}\left[R^{\text{ZF}} - \log\left(1 + \frac{\mathcal{K}}{T_t}\right)\right] \tag{11}$$

It should be noted that this approach can be applied to other scenarios by appropriately choosing the weights w_1, w_2, w_3. In fact, several cases of training and feedback analyzed in [2], including the unfaded AWGN feedback channel and TDD systems that exploit channel reciprocity, differ from (4) only in the value of the weights. Furthermore, the same basic approach can also be used to analyze quantized channel feedback.

B. Asymptotic Analysis

We examine how the optimal length of the feedback and training phase $T_t^{\star}$ scales with the coherence interval T and with the SNR. From (11), the optimal $T_t^{\star}$ satisfies the following equality

$$\frac{\mathcal{K}(T - T_t)}{T_t^2\left(1 + \frac{\mathcal{K}}{T_t}\right)} = R_k^{\text{ZF}} - \log\left(1 + \frac{\mathcal{K}}{T_t}\right) \tag{12}$$

It is easy to see that the derivative (11) is upperbounded by $\frac{1}{T}\widetilde{f}(T_t)$, where

$$\widetilde{f}(T_t) = \frac{\mathcal{K}\,(T - T_t)}{T_t^2} - \left[R^{\text{ZF}} - \frac{\mathcal{K}}{T_t}\right] \tag{13}$$

Since f is concave, it follows that the solution $\widetilde{T}_t$ of the equation $\widetilde{f}(T_t) = 0$ is an upper bound to the optimal value $T_t^{\star}$, solution of $\frac{\partial f}{\partial T_t} = 0$. Explicitly, we find

$$T_t^{\star} \leq \widetilde{T}_t = \sqrt{\frac{\mathcal{K}T}{R^{\text{ZF}}}} \tag{14}$$

Furthermore, when the rate gap is small such that $\log\left(1 + \frac{\mathcal{K}}{T_t}\right) \approx \frac{\mathcal{K}}{T_t}$ (which becomes accurate for large T), the upperbound becomes also a very good approximation.

The upperbound (14) yields two interesting behaviors: 1) for a fixed SNR (i.e., constant R^{ZF}) $T_t^{\star}$ increases as $O(\sqrt{T})$ as $T \to \infty$; 2) for a fixed coherence interval T, $T_t^{\star}$ decreases as $O(1/\sqrt{R^{\text{ZF}}})$ for large SNR, or equivalently, it decreases as $O(1/\sqrt{\log(\mathsf{SNR})})$ since $R^{\text{ZF}} = \log(\mathsf{SNR}) + O(1)$ for large SNR.

Although not explicitly stated in [3], it can be confirmed that the optimal number of training symbols also scales as $O(\sqrt{T})$ for a point-to-point MIMO channel. With respect to the scaling of $T_t^{\star}$ with SNR, recall the requirement $T_t \geq 4M$; thus, for sufficiently large SNR the value of $T_t^{\star}$ converges to this minimum.

Next, we examine the impact of $T_t^{\star}$ on the achievable rate. Using the upperbound (14) into (10), the objective value can be approximated as

$$f(\widetilde{T}_t) \;=\; \left(1 - \sqrt{\frac{\mathcal{K}}{R^{\text{ZF}}T}}\right)\left[R^{\text{ZF}} - \log\left(1 + \sqrt{\frac{\mathcal{K}R^{\text{ZF}}}{T}}\right)\right]$$

After some manipulation, it can be shown that the resulting effective rate gap with respect to R^{ZF} is given by

$$R^{\text{ZF}} - f(T_t^{\star}) \;\leq\; R^{\text{ZF}} - f(\widetilde{T}_t) \approx 2\sqrt{\frac{\mathcal{K}R^{\text{ZF}}}{T}} \tag{15}$$

The effective gap decreases roughly as $O(1/\sqrt{T})$ as T increases. We would like to remark that these scaling laws hold even if downlink training is the only source of CSI imperfection (i.e., even when channel feedback and dedicated training are perfect), and thus apply to all of the feedback schemes considered in [2].

C. Optimized Training/Feedback and Data Power

We now consider the case where power can be optimized over the data and training/feedback phases along the lines of [3], [4]. Although it is possible to assign different powers to the two training phases and the feedback phase, it is readily verified that it is optimal to use only two (per-symbol) power levels: power P_t for training (common and dedicated) and feedback, and P_d for data. Again using the results of [2], the achievable rate maximization can be stated as

$$\begin{array}{ll} \text{maximize} & \left(1-\frac{T_t}{T}\right)\left[R^{\mathrm{ZF}}(P_d)-\log\left(1+g(\{T_i\},P_d,P_t)\right)\right] \\ \text{subject to} & T_t+T_d\le T \\ & T_tP_t+T_dP_d\le PT \end{array} \quad (16)$$

where the last inequality corresponds to the average power constraint, $R^{\mathrm{ZF}}(P_d)$ denotes the ergodic rate of ZF beamforming with ideal CSI at SNR P_d/N_0, and where we define

$$g(\{T_i\},P_d,P_t)=\frac{P_d}{P_t}\left(\frac{w_1}{T_1}+\frac{w_2}{T_2}+\frac{w_3}{T_{\mathrm{fb}}}\right)$$

We solve this using a three-step optimization: 1) optimize $T_1, T_2, T_{\mathrm{fb}}$ with T_t, P_t, P_d fixed; 2) optimize P_t, P_d with T_t fixed; 3) optimize T_t. The first step is identical to what was done before. By using the solution (9), the second step reduces to maximizing

$$R^{\mathrm{ZF}}(P_d)-\log\left(1+\frac{P_d}{P_t}\frac{\mathcal{K}}{T_t}\right) \quad (17)$$

subject to the energy constraint (16) for fixed T_t, T_d. If we use the high SNR approximation $R^{\mathrm{ZF}}(P_d)\approx\log(P_d/N_0M)+\gamma$ where $\gamma\approx 0.5772$ is the Euler-Mascheroni constant [8] and rearrange terms, maximizing (17) is equivalent to minimizing $\frac{1}{P_d}+\frac{\mathcal{K}/T_t}{P_t}$, which is clearly convex in P_d, P_t. We form the Lagrangian

$$\mathcal{L}(P_d,P_t,\mu)=\frac{1}{P_d}+\frac{\mathcal{K}/T_t}{P_t}+\frac{1}{\mu^2}(T_tP_t+T_dP_d)$$

and the KKT conditions yield the following solution

$$P_d^\star = \frac{TP}{\sqrt{T_d}(\sqrt{T_d}+\sqrt{\mathcal{K}})} \quad (18)$$

$$P_t^\star = \frac{\sqrt{\mathcal{K}}TP}{T_t(\sqrt{T_d}+\sqrt{\mathcal{K}})} \quad (19)$$

Finally, by using the above solution and replacing $T_d=T-T_t$, the third step (again using the approximation for $R^{\mathrm{ZF}}(P_d)$) reduces to the maximization of

$$f(T_t) = \left(1-\frac{T_t}{T}\right)\log\left(\frac{TP}{N_0M\gamma}\frac{1}{(\sqrt{T-T_t}+\sqrt{\mathcal{K}})^2}\right).$$

Since the above objective function is concave in T_t, the optimal length $T_t^\star$ can be again found by a simple line search (e.g., using the bisection method). Perhaps surprisingly, $T_t^\star$ depends on the operating SNR P/N_0 and on T and M, and is not always equal to the minimum possible length $4M$, contrary to the observations made in [3], [4]. The condition for which the achievable rate with power allocation is maximized by $T_t^\star=T_t^{\min}=4M$, is given by $\frac{\partial f}{\partial T_t}(T_t^{\min})<0$, i.e.

$$\exp\left(\frac{\sqrt{T-T_{\min}}}{\sqrt{T-T_{\min}}+\sqrt{\mathcal{K}}}\right)<\frac{PT}{MN_0\gamma}\left(\sqrt{T-T_{\min}}+\sqrt{\mathcal{K}}\right)^{-2}$$

For $T\gg T_{\min}$, we let $T-T_{\min}\approx T$, approximate the exponential with its second-order Taylor expansion, and perform a further approximation to get the approximated condition

$$1+\sqrt{\frac{\mathcal{K}}{T}}<\frac{2P/(MN_0\gamma)-1}{1+\sqrt{4P/(MN_0)-1}}\approx\frac{1}{2}\sqrt{\frac{P}{MN_0\gamma}}$$

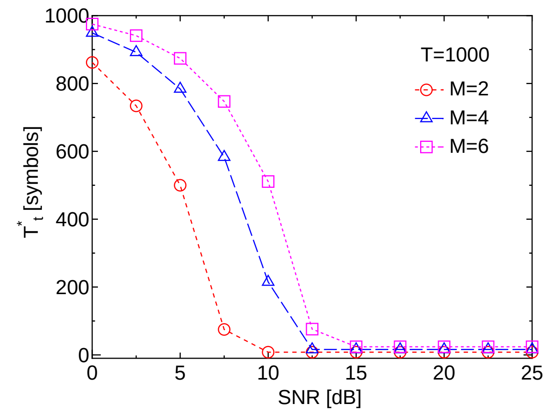

Fig. 2. The optimal $T_t^\star$ with power allocation.

which is satisfied for sufficiently large P/M and/or large T. From (19) we see that $P_t^\star$, the per-symbol SNR during training/feedback, increases as $O(\sqrt{T})$ when $T_t^\star=T_t^{\min}$. When power allocation is not allowed, the same order SNR is achieved by allocating $O(\sqrt{T})$ symbols to training/feedback. Interestingly, one would think that for very large T some $T_t>T_t^{\min}$ could be used, in order to improve channel estimation and feedback without any appreciable penalty in multiplexing gain. In contrast the above result contradicts this intuition. The key issue is the possibility of power allocation: for very large T, we can allocate a very large symbol power P_t to the training and feedback phases, and yet impact very little the energy allocated to the data.

Fig. 2 plots the optimal number $T_t^\star$ of feedback and training symbols under power allocation as a function of P/N_0 for $T=1000$. It is observed that for sufficiently large SNR the optimal $T_t^\star$ coincides with the minimum $4M$.

IV. Numerical Examples

This section provides some numerical examples to illustrate the analysis of the previous section. We let $M=4$ and consider equal power over data, training, and feedback phases.

a) Performance vs. SNR: Fig. 3 plots the optimized lengths $T_1, T_2, T_{\mathrm{fb}}$ versus SNR for $T=2000$ along with the corresponding analytical values obtained from (14). The length of common training T_1 and feedback T_{fb} coincide since they have the same weight $w_1=w_2=3$ when $M=4$. As expected, the optimal amount of training/feedback decreases towards the minimum value of $4M=16$, although rather slowly. Fig 4 shows the corresponding achievable sum rate versus SNR for $T=2000$, along with those for $T=500$ and $T=100$, computed by numerical search and the analytical expression (15), along with the sum rate of ZF with CSI.

b) Performance vs. coherence interval T: Fig. 5 shows $T_1, T_2, T_{\mathrm{fb}}$ as a function of the coherence interval T for a fixed SNR of 10 dB. Note that $T_t^\star$ grows as the square root of T, as expected from (14). The corresponding achievable rates are shown at 10 dB and 20 dB in Fig. 6, and the gap to ideal CSI is seen to decrease rather slowly with T. Also shown in this figure are the achievable rates for a system in which the

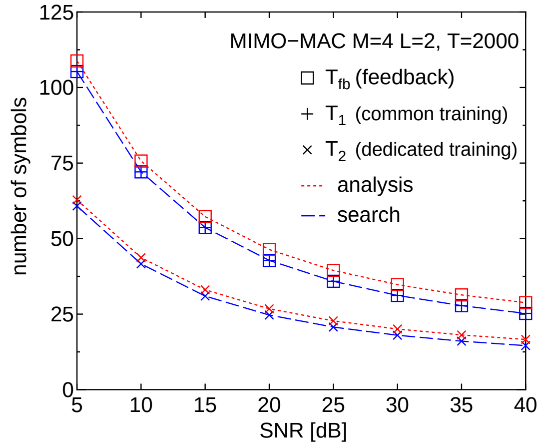

Fig. 3. Number of feedback/training symbols with $T = 2000$.

uplink SNR is 10 dB smaller than the downlink SNR; this very practical consideration reduces w_3 by a factor of 10 and has a significant effect on performance.

V. Conclusions

We addressed the optimization of training and feedback in a MIMO downlink system that makes use of explicit downlink training and CSI feedback. One of the key findings is that it is optimal to scale the resources dedicated to training and feedback per coherence time according to the square root of the coherence time via allocation of either power and/or symbols. For example, if the coherence time is doubled, it is advantageous to increase the training/feedback by a factor of $\sqrt{2}$ rather than using the same resources per coherence time; this corresponds to decreasing the total fraction of resources devoted to training and feedback by only a factor of $\sqrt{2}$ rather than the factor of 2 suggested by conventional wisdom. This result appears to apply quite generally and may have interesting implications on wireless system design.

References

[1] T.L.Marzetta and B.M.Hochwald, "Fast Transfer of Channel State Information in Wireless Systems," *IEEE Trans. on Sig. Proc.*, pp. 1268–1278, April 2006.

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Downlink Made Practical : Achievable Rates with Simple Channel State Estimation and Feedback Schemes," *Arxiv preprint cs.IT/0710.2642*.

[3] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[4] A. Dana, M. Sharif, and B. Hassibi, "On the capacity region of multi-antenna gaussian broadcast channels with estimation error," in *IEEE Int. Symp. on Inform. Theory*, July 2006.

[5] D. Love, "Duplex distortion models for limited feedback MIMO communication," *IEEE Trans. on Sig. Proc.*, vol. 54, pp. 766–774, Feb 2006.

[6] W. Santipach and M. Honig, "Capacity of beamforming with limited training and feedback," in *IEEE Int. Symp. on Inform. theory*, Seattle, Washington, 2006.

[7] ——, "Optimization of training and feedback for beamforming over a MIMO channel," in *IEEE Wireless Comm. and Networking Conf.*, 2007.

[8] A. Lozano, A. M. Tulino, and S. Verdú, "High SNR power offset in multiantenna communication," *IEEE Trans. on Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

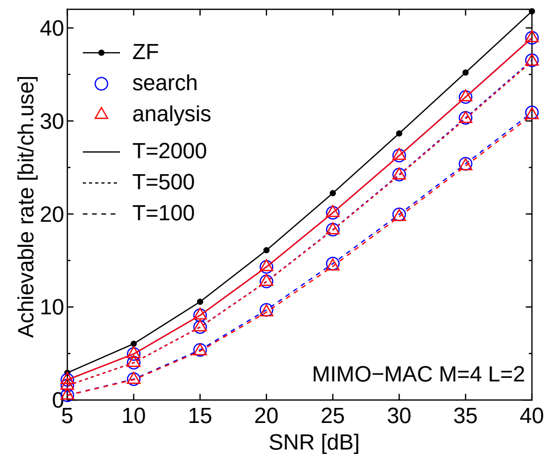

Fig. 4. Achievable sum rate vs. SNR.

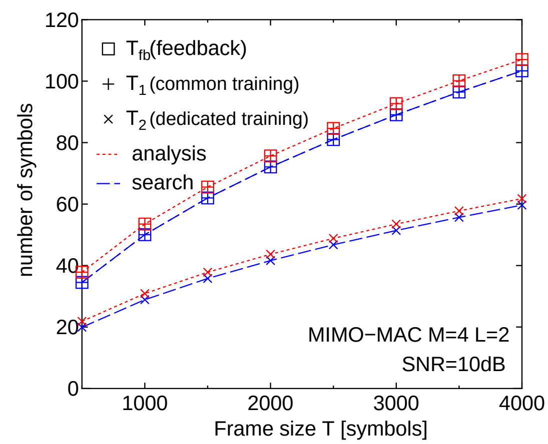

Fig. 5. Number of feedback/training symbols vs. T ($M = 4$).

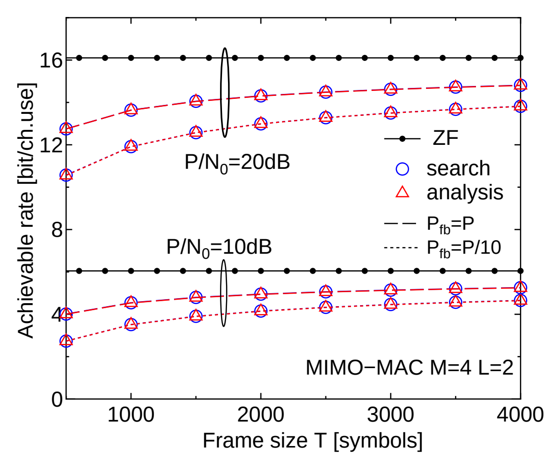

Fig. 6. Achievable sum rate vs. T under asymmetric SNR.