# Fast Transfer of Channel State Information in Wireless Systems

THOMAS L. MARZETTA

Bell Laboratories
Lucent Technologies
600 Mountain Ave., Rm. 2C-373
Murray Hill, NJ 07974
marzetta@lucent.com

BERTRAND M. HOCHWALD

Bell Laboratories
Lucent Technologies
600 Mountain Ave., Rm. 2C-363
Murray Hill, NJ 07974
hochwald@lucent.com

June 14, 2004

**Abstract**

Knowledge of accurate and timely channel state information (CSI) at the transmitter is becoming increasingly important in wireless communication systems. While it is often assumed that the receiver (whether basestation or mobile) needs to know the channel for accurate power control, scheduling and data-demodulation, it is now known that the transmitter (especially the basestation) can also benefit greatly from this information. For example, recent results in multi-antenna multi-user systems show that large throughput gains are possible when the basestation uses multiple antennas and a known channel to transmit distinct messages simultaneously and selectively to many single-antenna users.

In time-division duplex systems, where the basestation and mobiles share the same frequency band for transmission, the basestation can exploit reciprocity to obtain the forward channel from pilots received over the reverse channel. Frequency-division duplex systems are more difficult because the basestation transmits and receives on different frequencies and therefore cannot use the received pilot to infer anything about the multi-antenna transmit channel. Nevertheless, we show that the time occupied in frequency-duplex CSI transfer is generally less than one might expect, and falls as the number of antennas increases. Thus, although the total amount of channel information increases with the number of antennas at the basestation, the burden of learning this information at the basestation paradoxically decreases.

Thus, the advantages of having more antennas at the basestation extend from having network gains to learning the channel information. We quantify our gains using linear analog modulation which avoids digitizing and coding the channel state information and therefore can convey information very rapidly and can be readily analyzed. The old paradigm that it is not worth the effort to learn channel information at the transmitter should be revisited since the effort decreases and the gain increases with the number of antennas.

*Index Terms*—Wireless communications, broadcast channels, channel state information (CSI), analog modulation

# 1 Introduction

We consider the problem of learning the forward-link channel state information (CSI) at a basestation transmit antenna array and terminal receivers (users) of a cellular system. Recent theoretical and algorithmic advances suggest that channel knowledge at the basestation and terminals can lead to significant network and multi-user gains in system performance and capacity, especially if the basestation is equipped with multiple antennas. For example, results in [1, 2, 3, 4] show that in a flat-fading Rayleigh scattering environment, sum-rates (total of rates to all users) that grow linearly with the number of basestation antennas are possible if the forward link is known, even if the mobile terminals have only one antenna. Conversely, without forward channel knowledge at the transmitter, the sum rate for the users is no greater than the rate for a single user irrespective of the number of transmit antennas. A practical approach to multiple messaging [5] yields a sum-throughput within 4 dB of the theoretical achievable rate for a ten-antenna basestation and ten single-antenna terminals, Thus, the linear throughput improvements that are enjoyed by point-to-point MIMO [6, 7] can be achieved collectively by the multiple-antenna basestation and the autonomous single-antenna terminals.

While the network-level implications of linear growth are very compelling, a stumbling block to achieving this growth is often thought to be the excessive burden of learning the channel at the transmitter. In an $M$-antenna, $K$-user system, the forward channel CSI between the basestation and the users in a flat-fading environment is a $K \times M$ matrix $H$ containing $K \cdot M$ distinct complex channel coefficients. Often $K$ is significantly larger than $M$. Even if we decide to service only a subset of, say, $M$ users at a time [5], the choice of $M$ users is best made when the channel to all $K$ users is known. We therefore seek an efficient and fast way to learn the entire $K \times M$ matrix.

We explicitly compare time-division duplex (TDD) and frequency-division duplex (FDD) systems. In a TDD system, the forward and reverse link share the same frequency band, and therefore the two links can be modeled as reciprocals and consequently simple transposes of each other. Hence, the TDD system has the advantage of using knowledge gained about the channel on the reverse link to infer the channel on the forward link. An FDD system uses different bands for the forward and reverse links and therefore generally cannot use reverse-link information for the forward link. An FDD system requires an explicit transfer of the forward CSI on the reverse link. However, we show that the advantage of a TDD system over an FDD decreases as $M$ grows.

Since we require the terminals to learn the CSI in both TDD and FDD systems, the basestation sends

prearranged training or pilot signals on the forward link that are known to the terminals. This forward-link training is similar in both systems. TDD and FDD systems differ most markedly on the reverse link where the TDD system may again use pilots whereas the FDD system must first send pilots to learn the reverse channel at the basestation and then transfer the CSI that the terminals obtained during the forward-link training.

The transfer of CSI is subject to noise, interference, and fluctuations in the reverse channel, and is often viewed as a major obstacle towards learning the CSI in an FDD system. While we confirm the conventional view for traditional systems that use $M = 1$ basestation antenna, we show that our ability to transfer the CSI improves dramatically with $M > 1$. This holds, paradoxically, despite the fact that the total amount of CSI transferred increases with $M$. On the other hand, TDD system training requirements on the reverse link do not change with $M$.

In our model, the FDD system uses an analog linear transmission of CSI in which the channel coefficients are sent over the reverse link channel directly, and not digitally quantized and coded. This allows us to compute the mean-square error of the received channel estimates without having explicitly to encode/decode quantized data and compute the probability of receiving it in error. While most systems currently use some form of quantization and coding (to reduce the wide dynamic range of the channel information and for error detection and correction), we believe that our analog method offers three major advantages: i) It is readily analyzed and gives a good indication of the important issues in learning the channel as a function of $K$ and $M$ in both TDD and FDD systems, ii) It allows very fast CSI transfer (no delay is needed to encode/decode a channel code), and iii) it is surprisingly effective—under certain conditions it requires only a little more reverse-link resources than TDD. The second advantage is especially significant in a fast-fading environment where the channel can change appreciably in as little as 300 $\mu$s, and there is no time to encode/decode long sequences of channel information.

We mention that there is some precedent for using uncoded analog linear modulation to transfer the CSI. In particular, if an uncorrelated discrete-time Gaussian random process has to be transmitted over a discrete-time additive Gaussian white noise channel having the same sampling-rate, and if the criterion for reproducing the message at the receiver is minimum mean-square error, then direct uncoded analog linear modulation is better than any digital coding and modulation scheme [14, 16].

## 2 Problem Formulation and Notation

At one end of the link there is an $M$-element antenna array (*basestation*), and at the other end of the link are $K$ single-antenna terminals (*users*). The forward link is from the antenna array to the terminals, and the reverse link from the terminals to the antenna array. The biggest issue in CSI transfer is for the array to learn the forward propagation matrix in a timely manner. Typically there are many more terminals than array elements $K >> M$.

We require the basestation to learn the propagation matrix (which we denote $H$) from its $M$ antennas to the $K$ users. We also require each user to learn the row of $H$ that corresponds to the propagation channel from the $M$ basestation antennas to its single antenna. The users generally need to know this channel to demodulate forward-link data signals. Furthermore, in an FDD system, the user transmits this channel information back to the basestation on the reverse link. We refer to the forward link matrix $H$ as "the CSI", but when we refer to the users we implicitly assume that each user needs only to know its own row of $H$. We are interested in computing the time it takes for all parties to learn their CSI as a function of $M$ and $K$ and the desired mean-square error.

We concentrate our efforts on the reverse link from the terminals to the basestation because it is here that the FDD and TDD systems differ most. As the next section briefly shows, the forward link of either system requires the same effort to teach the users their CSI.

We do not make any assumptions on the time evolution of the channel $H$. We are interested in obtaining accurate CSI as quickly as possible and assume that the channel is approximately constant during the necessary training intervals.

### 2.1 Forward link

In the forward link, the basestation array transmits and the terminals receive. At integer time $t$ an $M \times 1$ complex vector $\mathbf{s}_t$ is transmitted and the $K$ terminals collectively receive a $K \times 1$ complex vector $\mathbf{x}_t$,

$$\mathbf{x}_t = H\mathbf{s}_t + \mathbf{w}_t\,, \tag{1}$$

where $H$ is the $K \times M$ forward propagation matrix and $\mathbf{w}_t$ is a $K \times 1$ vector comprising both receiver noise and interference. The forward propagation matrix is partitioned into $K$ row vectors,

$$H = \begin{bmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_K^T \end{bmatrix}, \tag{2}$$

where $\mathbf{h}_k^T$ is the $1 \times M$ propagation vector from the array to the $k$th terminal. We model the elements of $H$ as iid $\mathcal{CN}(0, 1)$ (complex normal with zero mean and unit variance).

### 2.1.1 Forward link training

The problem of channel estimation at the users on the forward link has received considerable attention in the literature [8]–[13]. The terminals learn the forward propagation matrix through known pilots that are transmitted by the basestation array. It is most efficient for the basestation antennas to transmit power-scaled signals of duration $\tau \geq M$ that are mutually orthonormal [8]. The $K$ users listen and use their knowledge of the orthonormal signals to form their individual estimates of their respective rows of the propagation matrix $H$. The required number of training symbols is proportional to $M$ and independent of $K$. The quality of the estimates can be improved as much as desired by increasing the training interval $\tau$. We assume that this training time cost is absorbed into the overall system and omit further consideration of the forward link. We are interested primarily in the reverse link where the bulk of the training and CSI transfer effort lies since it is strongly dependent on the number of users $K$ in the system. Generally $K \gg M$ so the cost of forward-link training per user is relatively small. We therefore assume that the users learn their row of $H$ perfectly.

## 2.2 Reverse link

In the reverse link the terminals transmit scalars and the basestation array receives their transmissions collectively. We assume the $K$ terminals are synchronous and collectively transmit a $K \times 1$ complex vector $\mathbf{s}_t$ and the array receives a $M \times 1$ complex vector $\mathbf{x}_t$,

$$\mathbf{x}_t = G\mathbf{s}_t + \mathbf{w}_t\,, \tag{3}$$

where $G$ is the $M \times K$ reverse propagation matrix, and $\mathbf{w}_t$ is the $M \times 1$ vector comprising both receiver noise and interference. (We reuse the notation of $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{w}$ from Section 2.1 since we do not need to refer

to the quantities in Section 2.1 again in the rest of the paper. However, we distinguish the channel matrix on the reverse link $G$ from the forward link $H$ since we need to refer to both later.)

The reverse propagation matrix is partitioned into $K$ column vectors,

$$G = \left[ \begin{array}{ccc} \mathbf{g}_1 & \cdots & \mathbf{g}_K \end{array} \right] , \tag{4}$$

where $\mathbf{g}_k$ is the $M \times 1$ propagation vector from the $k$th terminal to the array. We model the elements of $G$ as iid $\mathcal{CN}(0,1)$. We are interested in learning $G$ only in-so-far that it helps us learn $H$. The components of $\mathbf{w}_t$ are iid $\mathcal{CN}(0,\beta)$, where $\beta$ represents the variance of the interference-plus-noise at the basestation. We assume the basestation antennas are far enough apart so that the noise and interference is uncorrelated between antennas. Each terminal observes a power constraint of the form

$$\mathrm{E}\left\{|s_{tk}|^2\right\} \leq P\,. \tag{5}$$

The variance of the interference-plus-noise $\beta$ can vary depending on whether communication is noise-limited or interference-limited. If noise-limited, $\beta$ does not depend on the number of users $K$, while if interference-limited, $\beta$ is proportional to the number of interferers, which is often proportional to the number of users $K$. We leave this $K$ dependence of $\beta$ unspecified.

# 3 Learning $H$ at the Basestation

## 3.1 TDD

For a TDD link, we assume a reciprocal channel where the forward and reverse channels are unconjugated transposes of each other

$$G = H^T\,. \tag{6}$$

In this case, the terminals do not need to transmit their known $\mathbf{h}_k$ (which they learn on the forward link), but may instead transmit pre-arranged orthonormal pilot signals to the basestation. From the received pilots the basestation estimates the reverse propagation matrix $G$ and uses (6) to form an estimate of $H$.

It is assumed that the users may all transmit simultaneously during some training interval $\tau_{\mathsf{T}}$ (the subscript "$\mathsf{T}$" denotes "TDD"). In order to be able to form an estimate of the $M \times K$ matrix $G$, the basestation must have at least as many training symbols as users, or $\tau_{\mathsf{T}} \geq K$. Let $S = [\mathbf{s}_1 \ \cdots \ \mathbf{s}_{\tau_{\mathsf{T}}}]$ denote the $K \times \tau_{\mathsf{T}}$

matrix of collective training signals. Then $S$ may be written

$$S = \sqrt{\tau_{\mathrm{T}} P}\, \Phi \tag{7}$$

where $\Phi$ is a $K \times \tau_{\mathrm{T}}$ unitary matrix,

$$\Phi \Phi^* = I_K. \tag{8}$$

The $M \times \tau_{\mathrm{T}}$ signal matrix received at the basestation is

$$X = \sqrt{\tau_{\mathrm{T}} P} G \Phi + W\,, \tag{9}$$

where $X = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_{\tau_{\mathrm{T}}}]$ and $W = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_{\tau_{\mathrm{T}}}]$, which yields the minimum mean-square error (MMSE) estimate

$$\hat{G} = \hat{H}^T = \left( \frac{\sqrt{\tau_{\mathrm{T}} P}}{\beta + \tau_{\mathrm{T}} P} \right) X \Phi^*\,. \tag{10}$$

The resulting mean-square error for the $(k, m)$th entry of $\hat{H}$ is

$$\begin{aligned} \sigma_{\mathrm{T}}^2 &= \mathrm{E}\,|\hat{h}_{km} - h_{km}|^2 \\ &= \frac{\beta}{\beta + \tau_{\mathrm{T}} P} \qquad k = 1, \ldots, K, \quad m = 1, \ldots, M \qquad (11) \\ &= \frac{1}{1 + \rho_{\mathrm{r}} \tau_{\mathrm{T}}}\,, \qquad \tau_{\mathrm{T}} = K, K+1, \ldots, \qquad (12) \end{aligned}$$

where $\rho_{\mathrm{r}}$ is the reverse link SINR per terminal

$$\rho_{\mathrm{r}} = \frac{P}{\beta}. \tag{13}$$

The training matrix $S$ has the power constraint $\mathrm{tr}\,\{SS^*\} = K \tau_{\mathrm{T}} P$. We may choose $\tau_{\mathrm{T}}$ as large as needed to give the desired quality of the channel estimate while still obeying the power constraint (5). Unlike the forward link, the required number of reverse link training symbols is proportional to $K$ and is independent of $M$. Although we use the MMSE estimate for $H$, we note that for sufficiently large $\rho_{\mathrm{r}} \tau_{\mathrm{T}}$ the MMSE and the maximum likelihood (ML) estimates coincide. We prefer the MMSE estimate because it has some desirable properties that we exploit in our analysis.

## 3.2 FDD

Because the forward and reverse channels in an FDD system are generally not reciprocally related, the terminals need to transmit $H$ (which they learn on the forward link) back to the basestation on the reverse link. Therefore, CSI transfer on the reverse link entails two separate transmissions: First, sending training signals so that the basestation may learn the reverse channel $G$; second, transmitting $H$.

Let $\tau_{\mathrm{F}}$ denote the interval allocated for teaching the basestation $H$ in an FDD system. Then $\tau_{\mathrm{F}} = \tau_{\mathrm{p}} + \tau_{\mathrm{c}}$, where $\tau_{\mathrm{p}}$ is the time needed to learn $G$ and $\tau_{\mathrm{c}}$ is the time needed to transmit the entire $H$ on the reverse link.

### 3.2.1 Reverse link pilot transmission

The basestation learns $G$ in a method identical to the TDD training in Section 3.1. The terminals collectively transmit a pre-arranged $K \times \tau_{\mathrm{p}}$ unitary matrix with $\tau_{\mathrm{p}} \geq K$,

$$S_{\mathrm{p}} = \sqrt{\tau_{\mathrm{p}} P}\,\Phi\,, \tag{14}$$

where $\Phi\Phi^* = I_K$. The received pilot is then

$$X_{\mathrm{p}} = \sqrt{\tau_{\mathrm{p}} P}\,G\Phi + W_{\mathrm{p}}\,. \tag{15}$$

The MMSE estimate for the reverse propagation matrix is

$$\hat{G} = \left(\frac{\sqrt{\tau_{\mathrm{p}} P}}{\beta + \tau_{\mathrm{p}} P}\right) X_{\mathrm{p}}\,\Phi^*\,. \tag{16}$$

The entries of the estimation error matrix $\tilde{G} = \hat{G} - G$ are iid $\mathcal{CN}(0, \sigma_{\mathrm{p}}^2)$ where

$$\sigma_{\mathrm{p}}^2 = \frac{\beta}{\beta + \tau_{\mathrm{p}} P}. \tag{17}$$

We use the MMSE estimate because the estimation error $\tilde{G}$ is independent of the estimate $\hat{G}$. We note that the entries of the matrix $\hat{G}$ are also iid zero-mean complex Gaussian but with variance

$$\mathrm{E}\,|\hat{g}_{mk}|^2 = \frac{\tau_{\mathrm{p}} P}{\beta + \tau_{\mathrm{p}} P} \qquad m = 1, \ldots, M, \quad k = 1, \ldots, K. \tag{18}$$

### 3.2.2 Reverse link CSI transmission

Transmission of CSI in a communication link is often done digitally: the CSI is quantized and then channel-coded and modulated onto a standard PSK or QAM constellation [17, 18]. This form of transmission is resilient to noise and interference and is convenient since the transmitter and radio-frequency (RF) chain is usually optimized for sending signals with small dynamic ranges.

However, the digital approach carries two disadvantages that we would like to avoid: i) An analysis of the accuracy of the received CSI becomes very dependent on the exact quantization, channel-coding and modulation techniques used; ii) The delay caused by quantization and coding of the CSI may cause the CSI to become too dated for use at the transmitter. The CSI is essentially an analog quantity and the usual measure of fidelity is mean-square error rather than probability-of-error [14, 16, 11].

To make our analysis generic and the CSI transfer very fast, we assume the transfer from the terminals to the array is a form of "analog transmission", where the terminals directly modulate their carrier with their estimate of $\mathbf{h}_k$. We ignore the possible practical difficulties of this approach such as the large dynamic range needed by the RF transmitter. We note that rate distortion theory suggests that it is optimum (in mean-square) to use this approach when a sequence of iid Gaussian channel variables is sent over a Gaussian channel [14, 15, 16]. While our tranmission is not necessarily a sequence of iid random variables, any trends and conclusions that we glean from the analysis of our analog approach will have implications for digital approaches as well.

We assume that a total of $\tau_{\mathrm{c}}$ symbols are used on the reverse link for an estimate of $H$ to be received at the basestation. In order to minimize the time needed for CSI transfer, it is advantageous to have all of the terminals transmitting continuously during the $\tau_{\mathrm{c}}$ interval. (The terminals are not allowed to conserve their power at one interval and transmit more power later to make up for the lost interval.) We therefore "spread" the terminal's $M \times 1$ forward channel information by premultiplying $\mathbf{h}_k$ by a pre-arranged $\tau_{\mathrm{c}} \times M$ unitary spreading matrix (this is analogous to the spreading performed in a CDMA system). While it would be desirable to make the $K$ unitary spreading matrices mutually orthogonal, this is only possible if $\tau_{\mathrm{c}} \geq K \cdot M$.

In order to shorten $\tau_{\mathrm{c}}$ the users can "share" spreading matrices. This sharing causes enormous interference in a standard CDMA system, but we are able to benefit from sharing because the multiple antennas at the basestation can separate the users with the same spreading matrix. For $\tau_{\mathrm{c}} < KM$ we divide the $K$ terminals into $Q = K/L$ distinct groups with $L$ terminals in each group. The $L$ terminals in each group share the same unitary spreading matrix. We need $Q$ (assumed integer-valued) mutually orthogonal $M \times \tau_{\mathrm{c}}$

unitary spreading matrices,

$$\{\Psi_1,\ \Psi_2,\ \cdots,\ \Psi_Q\}\,, \quad \Psi_{q_1}\Psi_{q_2}^* = I_M\delta_{q_1q_2}, \qquad q_1, q_2 = 1,\ldots,Q. \tag{19}$$

The construction of these matrices requires $QM$ mutually orthogonal vectors of length $\tau_{\rm c}$, so $\tau_{\rm c} \geq QM$ or

$$\tau_{\rm c} \geq \frac{KM}{L}\,, \qquad L = 1,\ldots,K. \tag{20}$$

The $q$th set comprises the terminals numbered $\{(q-1)L+1, (q-1)L+2, \ldots, qL\}$, and the $1\times\tau_{\rm c}$ signal transmitted by the $k$th terminal in this set is

$$\sqrt{\frac{\tau_{\rm c}P}{M}}\mathbf{h}_k^T\Psi_q, \qquad k = (q-1)L+1,\ldots,qL, \qquad q = 1,\ldots,Q\,, \tag{21}$$

The power constraint (5) is observed since

$$\mathrm{E}\left\|\sqrt{\frac{\tau_{\rm c}P}{M}}\mathbf{h}_k^T\Psi_q\right\|^2 = \frac{\tau_{\rm c}P}{M}\mathrm{E}\,\|\mathbf{h}_k\|^2 = \tau_{\rm c}P.$$

Collectively the $K$ terminals transmit the $K\times\tau_{\rm c}$ CSI-bearing signal $S_{\rm c}$ whose $k$th row is given by (21), and the resulting $M\times\tau_{\rm c}$ signal that is received by the array is

$$\begin{aligned} X_{\rm c} &= GS_{\rm c} + W_{\rm c} \\ &= \sqrt{\frac{\tau_{\rm c}P}{M}}\sum_{q=1}^{Q}\sum_{k=(q-1)L+1}^{qL}\mathbf{g}_k\mathbf{h}_k^T\Psi_q + W_{\rm c} \end{aligned}$$

Despreading with the matrix $\Psi_q^*$ at the basestation yields

$$\begin{aligned} X_{\rm c}\Psi_q^* &= \sqrt{\frac{\tau_{\rm c}P}{M}}\sum_{k=(q-1)L+1}^{qL}\mathbf{g}_k\mathbf{h}_k^T + {W_{\rm c}}^{(q)} \\ &= \sqrt{\frac{\tau_{\rm c}P}{M}}G^{(q)}H^{(q)} + {W_{\rm c}}^{(q)}\,, \end{aligned} \tag{22}$$

where ${W_{\rm c}}^{(q)} = W_{\rm c}\Psi_q^*$, and $G^{(q)}$ is an $M\times L$ matrix whose columns are the reverse channels $\mathbf{g}_k$ for the terminals in the $q$th group, and $H^{(q)}$ is an $L\times M$ matrix whose rows are the forward channels $\mathbf{h}_k^T$ for the

terminals in the $q$th group,

$$G^{(q)} = [\mathbf{g}_{(q-1)L+1} \ \cdots \ \mathbf{g}_{qL}] \tag{23}$$

$$H^{(q)} = \begin{bmatrix} \mathbf{h}_{(q-1)L+1}^T \\ \vdots \\ \mathbf{h}_{qL}^T \end{bmatrix}.$$

The expression (22) comprises $M^2$ measurements (the elements of $X_{\rm c}\Psi_q^*$) and $ML$ unknowns (the elements of $H^{(q)}$). To avoid an underdetermined system of equations, we therefore have the constraint $L \leq M$. Since $L \leq K$, we have

$$L \leq \min(M, K). \tag{24}$$

We use a "pseudo" maximum-likelihood estimate for $H^{(q)}$, which uses the estimate $\hat{G}$ in (16) for the reverse propagation matrix as though it were correct in (23) to obtain

$$\hat{H}^{(q)} = \sqrt{\frac{M}{\tau_{\rm c}P}}(\hat{G}^{(q)*}\hat{G}^{(q)})^{-1}\hat{G}^{(q)*}X_{\rm c}\Psi_q^*. \tag{25}$$

The despreading operation (22) is performed for each $\Psi_q$, and (25) is formed for each $q = 1, \ldots, Q$, thus forming our entire estimate $\hat{H}$. This is the estimate of $H$ that we wish to analyze as a function of $\tau_{\rm p}$ and $\tau_{\rm c}$.

### 3.2.3 Performance analysis

To obtain the mean-square error of the entries of the estimate (25) as a function of $\tau_{\rm p}$ and $\tau_{\rm c}$, we substitute (22) into (25) for $X_{\rm c}\Psi_q^*$ and let $\tilde{G}^{(q)} = \hat{G}^{(q)} - G^{(q)}$ to obtain

$$\begin{aligned}
\hat{H}^{(q)} &= \sqrt{\frac{M}{\tau_{\rm c}P}}(\hat{G}^{(q)*}\hat{G}^{(q)})^{-1}\hat{G}^{(q)*}\left[\sqrt{\frac{\tau_{\rm c}P}{M}}G^{(q)}H^{(q)} + W_{\rm c}{}^{(q)}\right] \\
&= \sqrt{\frac{M}{\tau_{\rm c}P}}(\hat{G}^{(q)*}\hat{G}^{(q)})^{-1}\hat{G}^{(q)*}\left[\sqrt{\frac{\tau_{\rm c}P}{M}}(\hat{G}^{(q)} - \tilde{G}^{(q)})H^{(q)} + W_{\rm c}{}^{(q)}\right] \\
&= H^{(q)} + (\hat{G}^{(q)*}\hat{G}^{(q)})^{-1}\hat{G}^{(q)*}\left[\sqrt{\frac{M}{\tau_{\rm c}P}}W_{\rm c}{}^{(q)} - \tilde{G}^{(q)}H^{(q)}\right] \qquad (26) \\
&= H^{(q)} + \tilde{H}^{(q)}. \qquad (27)
\end{aligned}$$

All of the distinct matrices in (26) are mutually independent. The elements of $W_{\rm c}{}^{(q)}$ are iid $\mathcal{CN}(0, \beta)$, the elements of $H^{(q)}$ are iid $\mathcal{CN}(0, 1)$, while the elements of $\tilde{G}^{(q)}$ are iid $\mathcal{CN}(0, \frac{\beta}{\beta+\tau_{\rm p}P})$ (compare equation

(11)). The product $\tilde{G}^{(q)}H^{(q)}$ is an $M \times M$ matrix whose elements are uncorrelated, each with variance $\frac{L\beta}{\beta+\tau_{\rm p}P}$. Hence, the column vectors that comprise the estimation error $\tilde{H}^{(q)}$ are uncorrelated, and the $L \times L$ covariance matrix for the $m$th column $\tilde{\mathbf{h}}_m^{(q)}$, conditioned on $\hat{G}^{(q)}$, is

$$\mathrm{Cov}\left\{\tilde{\mathbf{h}}_m^{(q)} \mid \hat{G}^{(q)}\right\} = \left[\frac{M\beta}{\tau_{\rm c}P} + \frac{L\beta}{\beta+\tau_{\rm p}P}\right]\left(\hat{G}^{(q)*}\hat{G}^{(q)}\right)^{-1}.$$

From (18) the elements of the $M \times L$ matrix $\hat{G}^{(q)}$ are iid $\mathcal{CN}\left(0, \frac{\tau_{\rm p}P}{\beta+\tau_{\rm p}P}\right)$, and any diagonal element of $(\hat{G}^{(q)*}\hat{G}^{(q)})^{-1}$ is proportional to the reciprocal of a chi-square random variable with $2(M+1-L)$ degrees of freedom, so

$$\mathrm{E}\left\{|\tilde{h}_{mk}|^2\right\} = \left[\frac{M\beta}{\tau_{\rm c}P} + \frac{L\beta}{\beta+\tau_{\rm p}P}\right]\mathrm{E}\left\{\left(\frac{\beta+\tau_{\rm p}P}{\tau_{\rm p}P}\right)\frac{2}{\chi^2_{2(M+1-L)}}\right\}. \tag{28}$$

But $\mathrm{E}\left\{2/\chi^2_{2(M+1-L)}\right\} = 1/(M-L)$, so the mean-square estimation error of any element of $\hat{H}$ is

$$\sigma_{\rm F}^2 = \frac{1}{(M-L)\rho_{\rm r}}\left[\frac{L}{\tau_{\rm p}} + \frac{M}{\tau_{\rm c}}\left(1+\frac{1}{\tau_{\rm p}\rho_{\rm r}}\right)\right], L = 1, \ldots, \min(M,K), \quad \tau_{\rm p} \geq K, \quad \tau_{\rm c} \geq \frac{KM}{L}. \tag{29}$$

### 3.2.4 Comments on the mean-square error

Note that the mean-square error (29) is infinite when $L = M$. In particular, CSI transfer in FDD systems is disastrous when the array comprises only $M = 1$ antenna. (In this case we must have $L = 1$.) When $M = 1$ the basestation has only one antenna communicating with each user; the reverse link constitutes only one fading coefficient per user and because this coefficient is Rayleigh distributed and may have arbitrarily small square-magnitude with high probability, the transfer of the CSI may be arbitrarily poor. Thus, when $M = 1$ some form of diversity to improve the reverse link (such as interleaving combined with channel coding to harden the channel) becomes important when transferring channel information. As we see in Section 3.3, the antennas at the basestation provide the necessary diversity, and conditions improve dramatically when $M > 1$.

We also observe that (29) depends on the reverse-link SINR in two places, but the dominant contribution at high SINR is in the first multiplicative factor. As expected, the mean-square error is inversely proportional to the SINR at high values.

### 3.3 TDD versus FDD

To reduce the number of parameters and ease interpretation of the mean-square error expressions, we rewrite (12) and (29) in terms of the normalized quantities

$$\bar{\tau}_{\mathrm{T}} = \frac{\tau_{\mathrm{T}}}{K}, \;\; \bar{\tau}_{\mathrm{F}} = \frac{\tau_{\mathrm{F}}}{K}, \;\; \bar{\tau}_{\mathrm{p}} = \frac{\tau_{\mathrm{p}}}{K}, \;\; \bar{\tau}_{\mathrm{c}} = \frac{\tau_{\mathrm{c}}}{K}, \;\; \bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}}. \tag{30}$$

The TDD mean-square error (12) becomes

$$\sigma^2_{\mathrm{T}} = \frac{1}{1 + \bar{\rho}_{\mathrm{r}}\bar{\tau}_{\mathrm{T}}} \quad \bar{\tau}_{\mathrm{T}} = 1, 2, \ldots \tag{31}$$

and the FDD mean-square error (29) becomes

$$\sigma^2_{\mathrm{F}} = \frac{1}{(M-L)\bar{\rho}_{\mathrm{r}}} \left[ \frac{L}{\bar{\tau}_{\mathrm{p}}} + \frac{M}{\bar{\tau}_{\mathrm{c}}} \left( 1 + \frac{1}{\bar{\tau}_{\mathrm{p}}\bar{\rho}_{\mathrm{r}}} \right) \right], L = 1, \ldots, \min(M, K), \;\; \bar{\tau}_{\mathrm{p}} \geq 1, \;\; \bar{\tau}_{\mathrm{c}} \geq \frac{M}{L}. \tag{32}$$

The $\tau$-quantities normalized by $K$ in (30) can be thought of as "per-user" quantities and they are no longer necessarily integer. We later assume that they may be arbitrary fractions, subject to the constraints in (31) and (32). This is equivalent to assuming that $K$ is large and allows us to treat the quantities in (30) as continuous when creating performance curves in Section 3.3.2.

In addition to looking at the performance of TDD and FDD systems for CSI transfer on an absolute basis, we also find it instructive to compare the relative performance of TDD and FDD. For a given total FDD interval $\bar{\tau}_{\mathrm{F}} = \bar{\tau}_{\mathrm{p}} + \bar{\tau}_{\mathrm{c}}$ we choose the combination of parameters $\{L, \bar{\tau}_{\mathrm{p}}, \bar{\tau}_{\mathrm{c}}\}$ that minimizes the mean-square error $\sigma^2_{\mathrm{F}}$ (32). (Equivalently, we can choose a $\sigma^2_{\mathrm{F}}$ and find the combination of parameters that makes $\bar{\tau}_{\mathrm{F}}$ smallest.) Then we find the TDD interval $\bar{\tau}_{\mathrm{T}}$ whose mean-square error $\sigma^2_{\mathrm{T}}$ (12) is equal to $\sigma^2_{\mathrm{F}}$. Because an FDD system requires both pilot signals and channel transfer signals and a TDD system requires only pilot signals, generally $\bar{\tau}_{\mathrm{F}} > \bar{\tau}_{\mathrm{T}}$. Of interest is the relative effort associated with the two schemes $\bar{\tau}_{\mathrm{F}}/\bar{\tau}_{\mathrm{T}} = \tau_{\mathrm{F}}/\tau_{\mathrm{T}}$. One interesting conclusion is that the FDD pilot $\bar{\tau}_{\mathrm{p}}$ can be small leading to a poor $\hat{G}$, but we can still receive an accurate $\hat{H}$.

#### 3.3.1 Analysis of $\sigma^2_{\mathrm{F}}$ and comparison with $\sigma^2_{\mathrm{T}}$

We are interested in (32) primarily when there are at least as many users as antennas $K \geq M$. In this case, $K$ disappears from (32) and the optimized mean-square errors of the TDD and FDD schemes depend only

on the normalized quantities (30).

We first express the multiplexing factor $L$ in (32) as a fraction of the number of basestation antennas

$$L = \alpha M, \qquad \alpha = 1/M, 2/M, \ldots, 1. \tag{33}$$

Substituting (33) into (32), we obtain

$$\sigma_{\mathtt{F}}^2 = \frac{1}{(1-\alpha)\bar{\rho}_{\mathrm{r}}} \left[ \frac{\alpha}{\bar{\tau}_{\mathrm{p}}} + \frac{1}{\bar{\tau}_{\mathtt{F}} - \bar{\tau}_{\mathrm{p}}} \left( 1 + \frac{1}{\bar{\tau}_{\mathrm{p}} \bar{\rho}_{\mathrm{r}}} \right) \right], \tag{34}$$

where we also make the substitution $\bar{\tau}_{\mathrm{c}} = \bar{\tau}_{\mathtt{F}} - \bar{\tau}_{\mathrm{p}}$. It is straightforward to minimize (34) over $\bar{\tau}_{\mathrm{p}}$ to obtain

$$\bar{\tau}_{\mathrm{p}} = \left[ \sqrt{1 + \frac{(1-\alpha)\bar{\tau}_{\mathtt{F}}}{\alpha \bar{\tau}_{\mathtt{F}} + 1/\bar{\rho}_{\mathrm{r}}}} - 1 \right] \frac{\alpha \bar{\tau}_{\mathtt{F}} + 1/\bar{\rho}_{\mathrm{r}}}{1-\alpha}. \tag{35}$$

(We ignore the constraint $\bar{\tau}_{\mathrm{p}} \geq 1$ in our discussion here.) This equation is much more readily understood when $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}}$ is large, which is equivalent to saying either the number of users $K$ is large or the reverse-link SINR per user $\rho_{\mathrm{r}}$ is large. For large $\bar{\rho}_{\mathrm{r}}$ we may ignore the $1/\bar{\rho}_{\mathrm{r}}$ terms in (35) to obtain

$$\bar{\tau}_{\mathrm{p}} = \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}} \bar{\tau}_{\mathtt{F}}, \qquad \bar{\tau}_{\mathrm{c}} = \frac{1}{1+\sqrt{\alpha}} \bar{\tau}_{\mathtt{F}}, \tag{36}$$

whence (34) becomes

$$\sigma_{\mathtt{F}}^2 = \frac{1+\sqrt{\alpha}}{(1-\sqrt{\alpha})\bar{\rho}_{\mathrm{r}} \bar{\tau}_{\mathtt{F}}}. \tag{37}$$

To get a feeling of how TDD and FDD systems compare, we equate $\sigma_{\mathtt{F}}^2$ given in (37) with $\sigma_{\mathtt{T}}^2$ given in (31) to see the relative effort TDD and FDD need to obtain the same mean-square error. For large $\bar{\rho}_{\mathrm{r}}$, we obtain

$$\sigma_{\mathtt{T}}^2 = \frac{1}{\bar{\rho}_{\mathrm{r}} \bar{\tau}_{\mathtt{T}}} = \sigma_{\mathtt{F}}^2 = \frac{1+\sqrt{\alpha}}{(1-\sqrt{\alpha})\bar{\rho}_{\mathrm{r}} \bar{\tau}_{\mathtt{F}}}$$

Solving for $\bar{\tau}_{\mathtt{F}}$ in terms of $\bar{\tau}_{\mathtt{T}}$, we obtain

$$\bar{\tau}_{\mathtt{F}} = \frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}} \bar{\tau}_{\mathtt{T}}. \tag{38}$$

Rather surprisingly, it appears that we can make learning the channel in FDD cost only a little more time than in TDD when $\alpha$ is small. From (33), the smallest $\alpha$ may be is $1/M$ when $L = 1$, and as we let $M$ grow, $\bar{\tau}_{\mathtt{F}}/\bar{\tau}_{\mathtt{T}} \to 1$. It is easy to see from (37) that, for a given $\bar{\tau}_{\mathtt{F}}$, FDD also improves in an absolute sense as $M$ increases.

At the extreme of $\alpha = 1/M$ (36) becomes

$$\bar{\tau}_{\mathrm{p}} = \frac{1}{1+\sqrt{M}}\bar{\tau}_{\mathtt{F}}, \qquad \bar{\tau}_{\mathrm{c}} = \frac{\sqrt{M}}{1+\sqrt{M}}\bar{\tau}_{\mathtt{F}}.$$

We see that in FDD much less time is spent on pilot training to learn the reverse channel $G$ when $M$ is large (small $\bar{\tau}_{\mathrm{p}}$) than is spent on transmitting the channel coefficients $H$ (large $\bar{\tau}_{\mathrm{c}}$). This is dramatically different from TDD, where the channel is learned at the basestation only through the reverse pilot. We may infer that when $M$ is large, we do not need to know $G$ very well to get a good estimate of $H$. On the other hand if $\alpha \approx 1$ then $L \approx M$ (we do not want $\alpha = 1$ for then $\sigma_{\mathtt{F}}^2 = \infty$) and the estimate of $H$ at the basestation becomes much more sensitive to imperfections in the knowledge of $G$. At this extreme, $\bar{\tau}_{\mathrm{p}} \approx \bar{\tau}_{\mathrm{c}} \approx \bar{\tau}_{\mathtt{F}}/2$ for any $M$.

### 3.3.2 Performance computations and curves

For given values of $\bar{\tau}_{\mathrm{p}}$ and $\bar{\tau}_{\mathrm{c}}$, (32) shows that the number of users assigned to each spreading matrix $L$ should be as small as possible, subject to $L \geq M/\bar{\tau}_{\mathrm{c}}$. Since $L$ must be an integer,

$$L = \max\left(1, \left\lceil \frac{M}{\bar{\tau}_{\mathrm{c}}} \right\rceil\right). \tag{39}$$

To minimize $\bar{\tau}_{\mathtt{F}}$ it sometimes pays for a given mean-square error to have users share the same spreading codes ($L > 1$), thus allowing $\bar{\tau}_{\mathrm{c}}$ to be shortened. Conversely, for a given $\bar{\tau}_{\mathtt{F}}$ we may choose $L > 1$ to minimize $\sigma_{\mathtt{F}}^2$.

Having chosen $L$ optimally as a function of $\bar{\tau}_{\mathrm{c}}$, we numerically minimize $\sigma_{\mathtt{F}}^2$ with respect to $\bar{\tau}_{\mathrm{p}}$ and $\bar{\tau}_{\mathrm{c}}$ for a given value of $\bar{\tau}_{\mathtt{F}} = \bar{\tau}_{\mathrm{p}} + \bar{\tau}_{\mathrm{c}}$ subject to $\bar{\tau}_{\mathrm{p}} \geq 1$ and $\bar{\tau}_{\mathrm{c}} \geq M/L$. Because all normalized $\bar{\tau}$ quantities are discrete-valued with resolution $1/K$ but we treat them as continuous, our minimization results will be most accurate for large $K$.

Figures 1, 2 and 3 display the required TDD training interval $\bar{\tau}_{\mathtt{T}} = \tau_{\mathtt{T}}/K$ and FDD training interval $\bar{\tau}_{\mathtt{F}} = \tau_{\mathtt{F}}/K$, measured in symbols per user, as functions of the root-mean-square (rms) estimation error $\sigma_{\mathtt{F}}$ and $\sigma_{\mathtt{T}}$ for $M = 2, 4, 8, 16$ basestation antennas, and for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}}$ equal to 0, 10 and 20 dB. There are several features in the figures we comment upon. The rms error on the abscissa can be thought of as either absolute or relative error since the elements of the forward channel $H$ obey $\mathrm{E}\,|h_{km}|^2 = 1$ for all $k = 1, \ldots, K$ and $m = 1, \ldots, M$. Typically the integer-valued $L$ increases as $\sigma$ increases (see Figure 5).

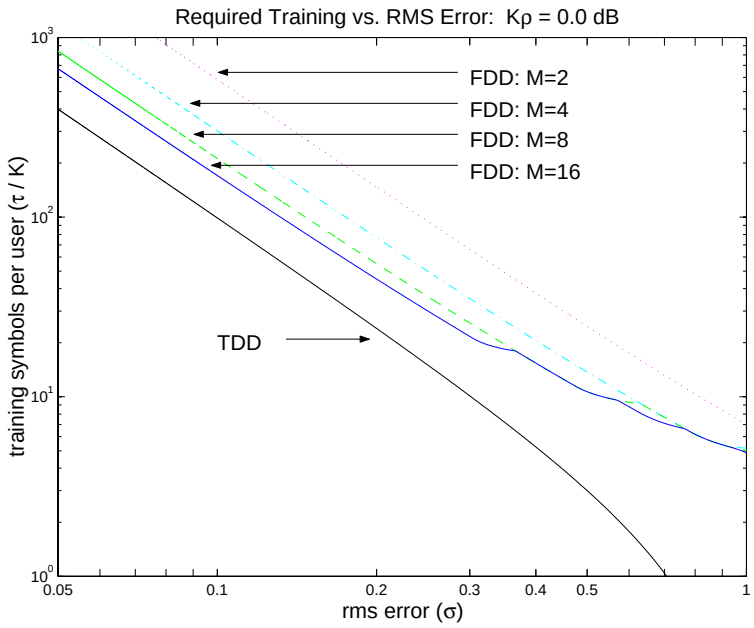



Figure 1: Required FDD and TDD training symbols per user $\bar{\tau}_{\mathrm{F}} = \tau_{\mathrm{F}}/K$ and $\bar{\tau}_{\mathrm{T}} = \tau_{\mathrm{T}}/K$ versus rms error $\sigma$ for $M = 2, 4, 8, 16$ basestation antennas, where the normalized SINR is $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 0$ dB. The TDD training interval is independent of $M$ and the FDD training interval is nearest to the TDD at large $M$ and small $\sigma$. The abrupt changes in slope of the curves in this and other figures is due to changes in the integer-valued $L$. Note that the SINR in this figure is very low.

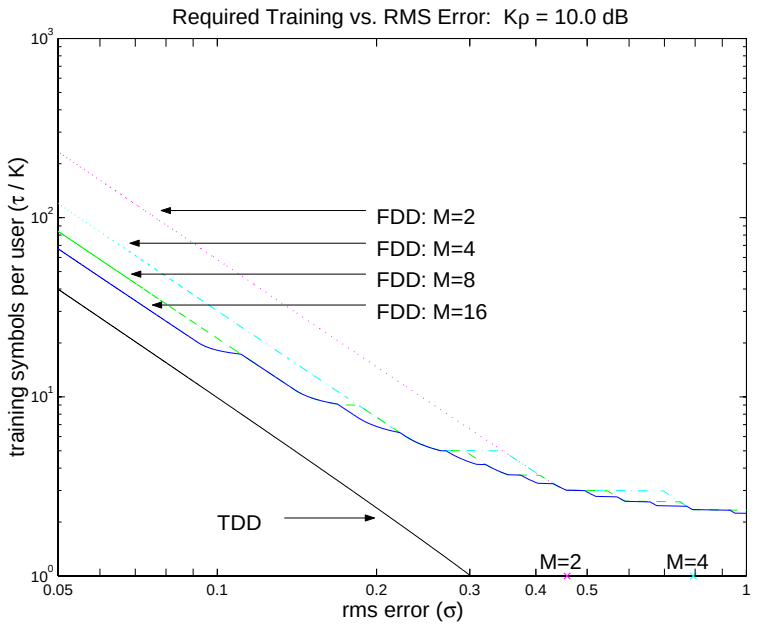



Figure 2: Required FDD and TDD training symbols per user $\tau_{\mathtt{F}}/K$ and $\tau_{\mathtt{T}}/K$ versus rms error $\sigma$ for $M = 2, 4, 8, 16$ basestation antennas, where the normalized SINR is $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 10$ dB. This curve is similar to Figure 1 except for the higher SINR. The "x" marks on the $\sigma$ axis represent points where the FDD $M = 2$ and $M = 4$ curves terminate (see text).

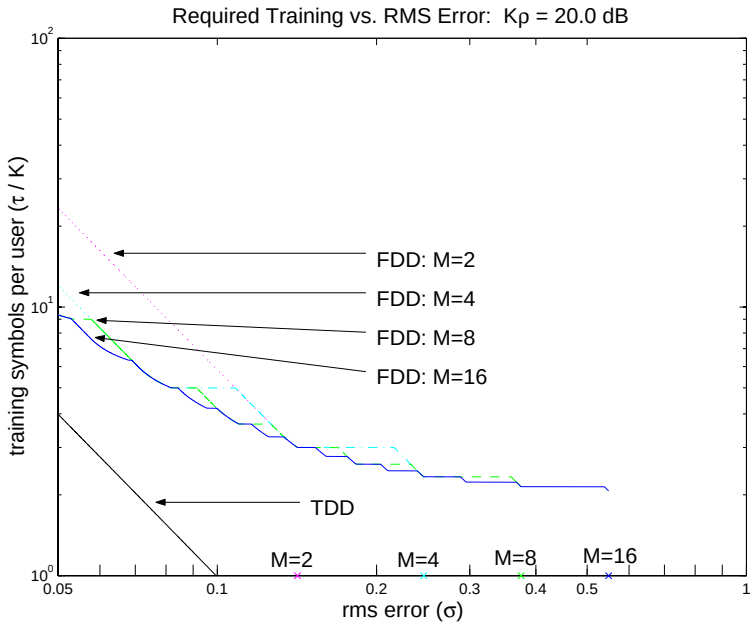



Figure 3: Required FDD and TDD training symbols per terminal $\tau_{\mathrm{F}}/K$ and $\tau_{\mathrm{T}}/K$ versus rms error $\sigma$ for $M = 2, 4, 8, 16$ basestation antennas, where the normalized SINR is $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 20$ dB. The "x" marks on the $\sigma$ axis represent points where the FDD $M = 2, 4, 8, 16$ curves terminate.

The curves have abrupt changes in their slopes when $L$ changes. The curves representing $\tau_{\mathrm{F}}/K$ and $\tau_{\mathrm{T}}/K$ terminate when $\bar{\tau}_{\mathrm{T}}$, $\bar{\tau}_{\mathrm{p}}$ or $\bar{\tau}_{\mathrm{c}}$ reach their minimum possible values. This is easily seen for TDD and is marked with an "x" on the abscissa for FDD. (This use of the "x" appears in later figures as well.) The FDD curves for different values of $M$ merge when their respective values of $\alpha$ given in (33) and (34) are equal. The figures all show that the required number of training symbols per user in FDD CSI transfer decreases as $M$ increases. This improved performance comes despite the fact that when $M$ is larger, the amount of CSI to be transferred on the reverse link is also larger. This increased amount of CSI is apparently easily overcome by the increased throughput of the reverse link as a function of $M$.

At still larger values of $\bar{\rho}_{\mathrm{r}}$, new training-versus-error plots can be generated by using the curves of Figure 3 with little qualitative modification; simply shift the curves to the left as given by the high-SINR approximations of (31) and (32)

$$\sigma_{\mathrm{T}}^2 \approx \frac{1}{\bar{\rho}_{\mathrm{r}}\bar{\tau}_{\mathrm{T}}} \quad \bar{\tau}_{\mathrm{T}} = 1, 2, \ldots$$

$$\sigma_{\mathrm{F}}^2 \approx \frac{1}{(M-L)\bar{\rho}_{\mathrm{r}}}\left[\frac{L}{\bar{\tau}_{\mathrm{p}}} + \frac{M}{\bar{\tau}_{\mathrm{c}}}\right], L = 1, \ldots, \min(M, K), \;\; \bar{\tau}_{\mathrm{p}} \geq 1, \;\; \bar{\tau}_{\mathrm{c}} \geq \frac{M}{L}.$$

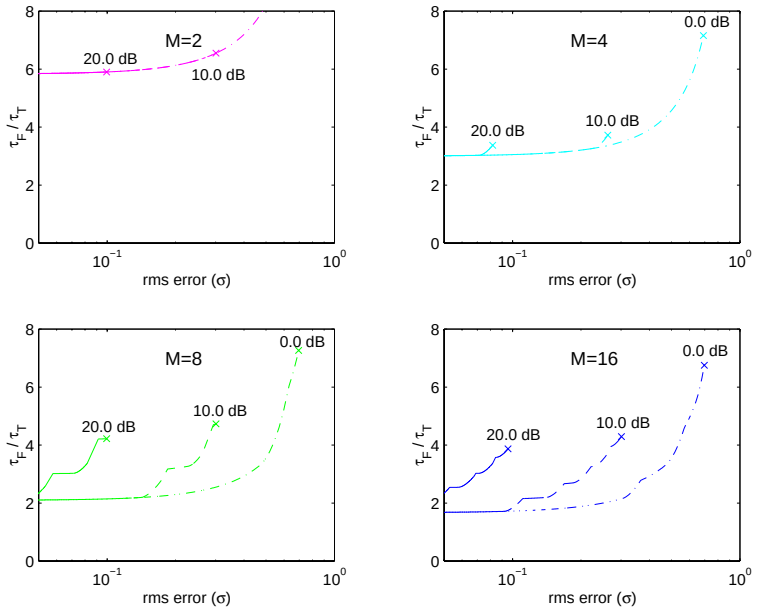



Figure 4: Ratio of the FDD to TDD training $\tau_{\mathrm{F}}/\tau_{\mathrm{T}}$ versus the rms error $\sigma$ for $M = 2, 4, 8, 16$ transmit antennas, and for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 0, 10, 20$ dB. The "x"-marks denote where the curves end as a function of $\sigma$. Observe that the ratio of FDD to TDD training can be as little as two for reasonable values of $\sigma$, $M$ and the SINR.

Figure 4 displays the ratio of the FDD training to the TDD training $\tau_{\mathrm{F}}/\tau_{\mathrm{T}}$ as a function of the rms error $\sigma$ for $M = 2, 4, 8, 16$ transmit antennas. Within each sub-plot, the FDD/TDD ratio is shown for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 0, 10, 20$ dB, as dash-dot, dashed and solid curves respectively. Each of the curves terminates for a sufficiently large value of $\sigma$ as indicated by an "x" symbol. Over typical ranges of parameters, the FDD scheme requires between two- and four-times as many symbols as the TDD scheme for the same quality of CSI.

Figure 5 shows the multiplexing factor $L$ as a function of the rms error $\sigma$ for $M = 2, 4, 8, 16$ transmit antennas and for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 10$ dB. Of necessity $L$ is integer-valued and less than $M$. The jumps in $L$ coincide with abrupt changes in the rms error. Small values of $L$ are associated with small rms errors and an exploitation of the diversity and beamforming gains of the antenna array. In this regime the CSI is transmitted over a hardened channel. Large values of $L$ are associated with the multiplexing gains that are possible when $M$ is large. In this regime the users share spreading codes and therefore a larger amount of CSI is being transmitted simultaneously to the receive array but with lower diversity than when $L$ is small. We see this effect most clearly in equation (28) where the conditional mean-square error is proportional to

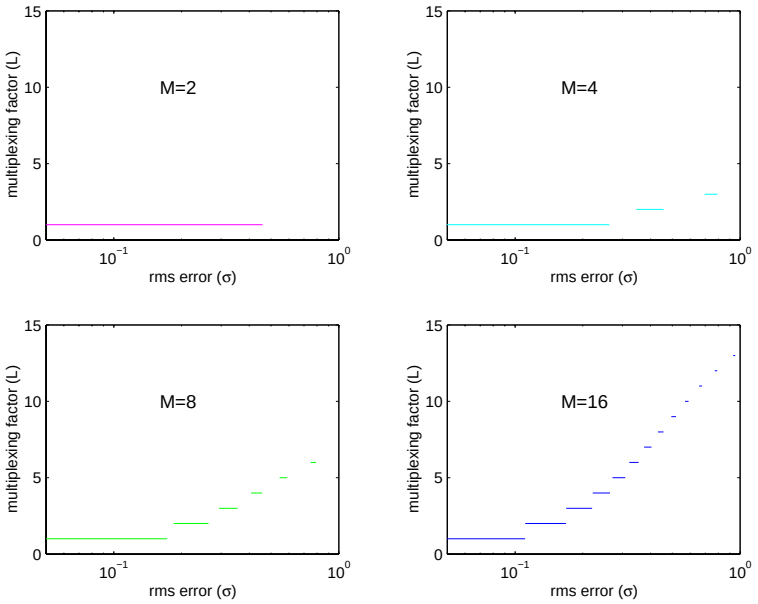



Figure 5: The number of terminals sharing the same spreading matrix $L$ versus the rms error $\sigma_{\mathtt{F}}$ clockwise beginning at the top-left for $M = 2, 4, 8, 16$ transmit antennas, and for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 10$ dB. Because of the discreteness of $L$, $\sigma_{\mathtt{F}}$ may change discontinuously.

the reciprocal of a chi-square random variable with $2(M + 1 - L)$ degrees of freedom. A similar diversity versus multiplexing tradeoff effect has been observed in other contexts with multiple antennas [20].

The solid curves in Figure 6 show the number of reverse-link FDD pilot symbols per user $\bar{\tau}_{\mathrm{p}} = \tau_{\mathrm{p}}/K$ versus the rms error $\sigma$ for $\bar{\rho}_{\mathrm{r}} = K\rho_{\mathrm{r}} = 10$ dB. Included are also plots of the number of TDD training symbols per user $\bar{\tau}_{\mathtt{T}} = \tau_{\mathtt{T}}/K$ as dashed curves. For $M \geq 4$ the FDD system may use fewer reverse link pilot symbols than the TDD scheme at low rms errors. This implies that poor knowledge of the reverse channel $G$ is not necessarily an obstacle to learning $H$ well at the basestation.

Table 1 shows the required FDD CSI transfer parameters for achieving an rms error of $\sigma_{\mathtt{F}} = 0.10$ for $K = 48$ users and $\rho_{\mathrm{r}} = 3$ dB ($\bar{\rho}_{\mathrm{r}} = 19.8$ dB), where the optimization is over the integer-valued $\tau_{\mathtt{F}}$, $\tau_{\mathrm{c}}$ and $\tau_{\mathrm{p}}$, as opposed to their normalized counterparts. We see that the TDD system requires $\tau_{\mathtt{T}} = 50$ whereas the FDD system requires approximately six times as much training ($\tau_{\mathtt{F}} = 293$) when $M = 2$ but only approximately four times as much ($\tau_{\mathtt{F}} = 210$) when $M = 8$ or $M = 16$.

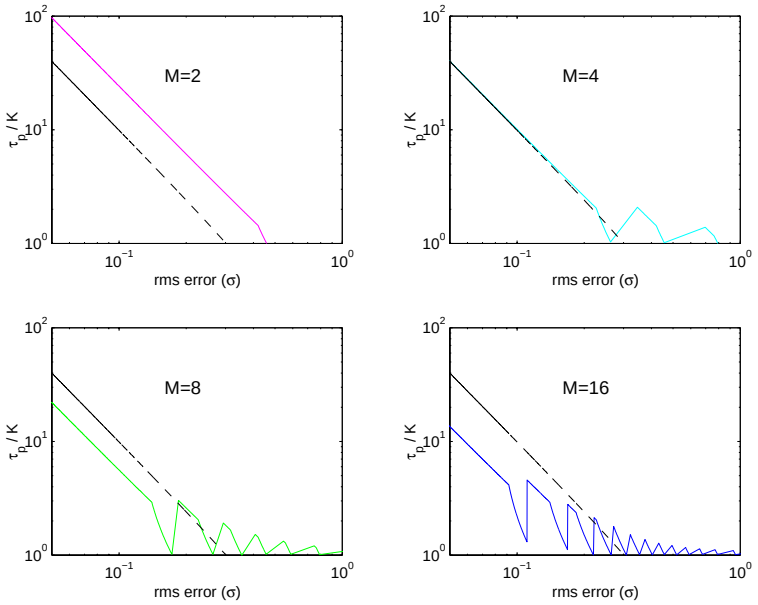



Figure 6: The number of reverse-link FDD pilot symbols per user $\tau_{\mathrm{p}}/K$ (solid curves) and the number of TDD pilot symbols per user $\tau_{\mathtt{T}}/K$ (dashed curves) versus the rms error $\sigma$ clockwise beginning at the top-left for $M = 2, 4, 8, 16$ transmit antennas, and for $K\rho_{\mathrm{r}} = 10$ dB. As before, abrupt changes coincide with changes in $L$.

| $M$ | $\sigma_{\mathtt{F}}$ | $\tau_{\mathtt{F}}$ | $\tau_{\mathrm{c}}$ | $\tau_{\mathrm{p}}$ | $L$ |
|---|---|---|---|---|---|
| 2 | 0.100 | 293 | 172 | 121 | 1 |
| 4 | 0.084 | 240 | 192 | 48 | 1 |
| 8 | 0.100 | 210 | 130 | 80 | 3 |
| 16 | 0.100 | 210 | 130 | 80 | 6 |

Table 1: Required FDD CSI training and transfer times for achieving an rms error of $\sigma_{\mathtt{F}} = 0.10$ for $K = 48$ users and $\rho_{\mathrm{r}} = 3$ dB. The TDD scheme requires $\tau_{\mathtt{T}} = 50$ symbols for all $M$ to achieve the same rms error. We see that the total FDD training interval $\tau_{\mathtt{F}} = \tau_{\mathrm{p}} + \tau_{\mathrm{c}}$ decreases as the number of antenna increases and is approximately four times the TDD interval for $M = 8$ or $M = 16$. We could not achieve $\sigma_{\mathtt{F}} = 0.1$ exactly for $M = 4$ because of the integer constraints on $L$ and $\tau_{\mathtt{F}}$.

### 3.3.3 Near-far effects

Our previous discussion assumes that each terminal enjoys the same expected SINR. Here we briefly extend our performance analysis to shadow or distance-dependent (from the basestation) fading, where each terminal may have a different SINR. We model the forward and reverse propagation matrices as $\sqrt{D}H$ and $G\sqrt{D}$ where, as before, $H$ is a $K \times M$ matrix, $G$ is $M \times K$, both comprising iid $\mathcal{CN}(0,1)$ elements, and $D$ is a $K \times K$ diagonal matrix whose diagonal elements are given by the $K \times 1$ vector $\mathbf{d} > 0$ representing the distance effects. In both TDD and FDD systems it is reasonable for $\mathbf{d}$ to be the same for the forward and reverse links since they represent macroscopic environmental effects. We assume that $d_k \leq 1$, $k = 1, \ldots, K$; this constraint prevents the terminals from violating their peak power constraint. We also assume that the basestation array knows $\mathbf{d}$ in order to obtain the MMSE estimate for $G$. The terminals transmit reverse pilots and CSI as before.

The analysis is similar to Section 3.2.3 and we omit the details. The mean-square error for both the TDD and FDD systems now depend on the specific terminal $k$. The mean-square error for the FDD system also depends on $q$, the $L$ terminals that user $k$ is grouped with:

$$\sigma^2_{\mathtt{T},k} = \frac{d_k}{1 + \rho_{\mathrm{r},k}\tau_{\mathtt{T}}}, \qquad k = 1, \ldots, K, \tag{40}$$

$$\begin{aligned} \sigma^2_{\mathtt{F},qk} &= \frac{1}{(M-L)\rho_{\mathrm{r},k}} \left[ \frac{L}{\tau_{\mathrm{p}}} \left( \frac{1}{L} \sum_{\ell=(q-1)L}^{qL} \frac{d_\ell \left(1 + \frac{1}{\tau_{\mathrm{p}}\rho_{\mathrm{r},k}}\right)}{1 + \frac{1}{\tau_{\mathrm{p}}\rho_{\mathrm{r},\ell}}} \right) + \frac{M}{\tau_{\mathrm{c}}} \left( 1 + \frac{1}{\tau_{\mathrm{p}}\rho_{\mathrm{r},k}} \right) \right], \\ & q = 1, \ldots, Q, \qquad k = (q-1)L + 1, \ldots, qL, \end{aligned} \tag{41}$$

$$\rho_{\mathrm{r},k} = \frac{P d_k}{\beta}, \qquad k = 1, \ldots, K. \tag{42}$$

The forms of (40) and (41) are similar to (12) and (29). The mean-square error associated with the CSI for a particular terminal is controlled largely by the SINR for that terminal $\rho_{\mathrm{r},k}$.

# 4 Conclusions and Extensions

We have seen that the training burden for CSI in a TDD system is dependent on the number of users but not the number of antennas at the basestation. As expected, we have seen that the CSI burden in an FDD system is larger than TDD. But surprisingly we have seen that as the number of antennas increases the FDD burden

decreases despite the fact that the users have to transfer increasing amounts of CSI. In certain regimes of SINR's and numbers of antennas, the FDD and TDD CSI burden are, in fact, similar. The multi-antenna effect apparently benefits CSI transfer in the same way that it benefits the transmission of data. We hope that practical FDD systems can therefore benefit from the tremendous network gains that are possible when CSI is known at the transmitter.

We have used uncoded analog linear modulation to transmit the raw channel information on the reverse link in the FDD system. Besides being an effective and rapid alternative to digital techniques, this modulation gives a good indication of the system tradeoffs in mean-square error versus training time, number of antennas and number of users. Many of the tradeoffs should carry over to digital methods.

Our FDD CSI transfer scheme has ramifications even for noncentralized wireless networks that share the same carrier frequency. Consider, for example, three nodes A, B and C, where A wishes to transmit a message to C, possibly through B. The three nodes take turns transmitting pilots on the same frequency for CSI acquisition. Node A will learn the A-B channel and the A-C channel but not the B-C channel. To take advantage of a possible good connection between B and C, node A has to learn the B-C channel. Node A can learn this channel only if nodes B and/or C transfer their estimates of the B-C channel to A. Our analog-type of analysis should be a useful tool in learning the tradeoffs involved.

We have described one particular type of analog linear modulation where groups of $L$ terminals share the same unitary spreading matrix. A more general procedure would assign a unique unitary spreading matrix to *each* terminal. When $\tau_{\mathrm{c}}$ is too small to permit the spreading matrices to be exactly orthogonal to each other, one intuitively expects that the spreading matrices should be "as orthogonal to each other as possible," possibly in a Grassmanian sense.

The basestation could certainly employ more sophisticated signal processing techniques. For example it could perform joint ML estimation of $H$ and $G$ by working with the joint likelihood $p(X_{\mathrm{c}}, X_{\mathrm{p}} \mid H, G)$, and alternately maximizing with respect to $H$ and $G$. This could be especially advantageous when $L << M$ and the CSI measurement $X_{\mathrm{c}}$ is highly overdetermined with respect to $H$.

Our terminals had, by assumption, only one antenna. If each terminal had $N > 1$ antennas then each user's individual CSI would consist of an $M \times N$ forward propagation matrix. One way to transmit this matrix to the basestation is to pre-multiply the matrix by a $\tau_{\mathrm{c}} \times M$ unitary spreading matrix, resulting in a $\tau_{\mathrm{c}} \times N$ matrix. This product could be treated as a space-time signal and transmitted over the terminal's $N$-element array. Clearly there are many other transmission possibilities as well.

We casually mention that extensions to forward and the reverse channels with wide bandwidths and

temporal correlation or memory would be areas of interest. Since we are working in an analog domain, perhaps linear filters at the transmitter and receiver could help design an effective transmission scheme.

Some simplifications in the reverse link channel transfer are possible. For example, when there is a single terminal ($K = 1$) we have a point-to-point link. If the terminal has a single antenna the basestation array would use maximum-ratio weighting on the forward link. To implement maximum-ratio weighting, the basestation needs to know only the normalized (unit-energy) $M \times 1$ forward propagation vector. If the single terminal has $N > 1$ antennas then the two arrays could jointly implement the singular value decomposition of the propagation matrix on the forward link $H = UD_vV^\dagger$. Clearly the basestation needs to know only the product $UD_v$ [19] rather than the full $H$. Once again, analog modulation could be used to transmit the product on the reverse link.

As a final comment, we ignored the effort needed on the forward link for the users to learn their row of $H$ because it is dependent on $M$ and not $K$. A more detailed analysis of CSI transfer could include forward-link estimation errors that a FDD system might have in the CSI that it is transferring on the reverse link.

## References


[1] G. Caire and S. Shamai (Shitz), "On the achievable Throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Information Theory*, vol. 49, pp. 1691 – 1706, July, 2003.

[2] W. Yu and J. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. Information Theory*, pp. 2580–2595, Oct., 2003.

[3] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Information Theory*, pp. 1912–1921, Aug. 2003.

[4] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, achievable rates and sum capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Information Theory*, pp. 2658–2668, Oct., 2003.

[5] C. Peel, B. Hochwald and L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication, Parts I & II" submitted to *IEEE Trans. on Comm.* 2003. Also *Proc. 41st Annual Allerton Conf. Communication, Control, Computing*, Monticello, IL, Oct. 2003. Download available at `http://mars.bell-labs.com`.

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, Vol. 6, pp. 311–335, Mar. 1998.

[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 586–595, Nov./Dec. 1999. Download available on `http://mars.bell-labs.com`.

[8] T. L. Marzetta, "BLAST Training: Estimating Channel Characteristics for High Capacity Space-Time Wireless," *Proc. 37th Annual Allerton Conf. Communication, Control, Computing*, Monticello, IL, Sept. 1999.

[9] Y. Li, N. Seshadri and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Sel. Area Communications*, vol. 17, pp. 461–471, Mar. 1999.

[10] H. Bolcskei, R. Heath and A. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 96–109, Jan. 2002.

[11] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Information Theory*, vol. 49, pp. 951–963, Apr. 2003.

[12] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communication*, Cambridge: University Press, 2003.

[13] X. Ma, L. Yang and G. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Comm.* to appear.

[14] T. J. Goblick, "Theoretical Limitations on the Transmission of Data From Analog Sources," *IEEE Trans. Information Theory*, vol. IT-11, pp. 558–567, Oct. 1965.

[15] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons: New York, 1991.

[16] M. Gastpar, B. Rimoldi, and M. Vetterli, "To Code, or Not to Code: Lossy Source-Channel Communication Revisited," *IEEE Trans. Information Theory*, vol. 49, pp. 1147–1158, May 2003.

[17] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmanian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Trans. Information Theory*, vol. 49, pp. 2735–2746, Oct. 2003.

[18] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On Beamforming With Finite Rate Feedback in Multiple-Antenna Systems," *IEEE Trans. Information Theory*, vol. 49, pp. 2562–2579, Oct. 2003.

[19] R. Venkataramani and T. L. Marzetta, "Reciprocal Training and Scheduling Protocol for MIMO Systems," *Proc. 41st Annual Allerton Conf. Communication, Control, Computing*, Monticello, IL, Oct. 2003.

[20] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antennas," *IEEE Trans. Information Theory*, vol. 49, pp. 1073–1096, May 2003.