Reciprocal Training and Scheduling Protocol for MIMO systems

Raman Venkataramani* and Thomas L. Marzetta†

∗ Division of Engineering and Applied Sciences, Harvard University
33 Oxford St, Cambridge, MA 02138, raman@deas.harvard.edu

† Math. Sciences Research Center, Bell Laboratories, Lucent Technologies
600 Mountain Ave, Murray Hill, NJ 07974, tlm@research.bell-labs.com

Abstract

We propose a new training based scheme for multiple antenna broadcast channels consisting of a base station (transmitter) and many users (receivers) in a flat fading environment. Both the transmitter and the users transmit at a common carrier frequency so that, by virtue of reciprocity, the downlink and uplink propagation matrices are equal. We show that by using appropriate training the base station and users to learn the channel state information efficiently. With the knowledge of the channel at both ends of the link, they diagonalize the channel optimally and schedule power settings in order to maximize a performance criterion. This permits independent coding over resulting sub-channels of the diagonalized channel. The proposed training scheme also automatically allows both ends of the link to agree on the rate scheduling. The practical advantages of this scheme are (a) the simplicity of coding and (b) its robustness to the nature of the flat fading environment. In particular, the scheme provides a seamless transition between Rayleigh propagation to specular (beamforming) propagation.

1 Introduction

It is well known that the capacity of a multiple antenna (MIMO) communication link grows linearly with the smaller of the number of transmitter and receiver antennas in a Rayleigh flat fading environment [1, 2]. The amount of channel state information (CSI) that is available plays a critical role in MIMO. Unitary space-time modulation, which dispenses with CSI and therefore all training, nearly achieves the capacity under certain conditions [3, 4, 5, 6]. The alternative to space-time modulation is to use a training-based scheme where the transmitter sends a training signal from which the receiver estimates the channel propagation coefficients [7]. If the coherence intervals are short (fast fading environment) and there are many antennas, training based-schemes are not very feasible. However, training-based schemes permit greatly simplified coding techniques compared to space-time coding for unknown fading. BLAST is an example of a practical training-based scheme that achieves high data rates [2, 7].

In this paper we consider a scenario where the transmitter and receivers uses a common carrier frequency, so that the uplink and downlink channels are equal by virtue of reciprocity. We assume that the coherence interval is long enough that a training-based scheme is affordable. We propose a new training-based scheme where both ends of the link learn the channel state information by sending each other training signals. The transmitter's possession of the CSI permits simple and entirely different coding strategies.

The following is the outline of the paper. We first present uplink and downlink models for the reciprocal channel. For a simple superposition coding strategy, we derive expressions for the capacity and the optimal input distribution. Finally, we demonstrate that the reciprocity of the channel allows very efficient training for both ends of the link to learn the channel matrix, as well as agree the rate and power scheduling.

2 Signal and Channel Models

Consider a multiple antenna communication link consisting of a base station and K users operating in a flat fading environment, where the base station has M antennas and the k-th user has N_k antennas. In the complex baseband representation, the downlink received signals are modeled as

$$X_t^k = S_t H_k + W_t^k, \quad k = 1, \dots, K \tag{1}$$

where S_t is the $1 \times M$ vector transmitted at time t, X_t^k is the $1 \times N_k$ vector received by user k, W_t^k is the additive receiver noise, and H_k is the $M \times N_k$ matrix of propagation coefficients between the base station and the k-th user. The matrices H_k and W_t^k are isotropically random and statistically independent of each other with entries being $\mathcal{CN}(0, \rho_k)$ and $\mathcal{CN}(0, \sigma_k^2)$ respectively. The power constraint on the transmitted signal is

$$\mathrm{E} \sum_{m=1}^{M} |S_{t,m}|^2 = 1. \tag{2}$$

where $\mathrm{E}(\cdot)$ denotes the expected value.

We assume that the base station and the users operate at a common frequency. Then, by reciprocity, the uplink can be modeled as

$$Y_t = \sum_{k=1}^{K} R_t^k H_k^T + V_t \tag{3}$$

where the $1 \times M$ vectors Y_t and $V_t \sim \mathcal{CN}(0, \sigma^2 I)$ are the received signal and receiver noise respectively and R_t^k is the $1 \times N_k$ vector transmitted by the k-th user at time t satisfying the power constraint

$$\mathrm{E} \sum_{n=1}^{N_k} |R_{t,n}^k|^2 = 1. \tag{4}$$

We assume that the users know their receiver noise variances σ_k^2, and the base station knows all the variances. We also assume that the coherence interval is T symbols, i.e., the propagation matrices remain essentially constant over intervals of duration T.

3 Partial Channel Inversion

We shall focus primarily on the broadcast (downlink) problem where the base station sends messages to the K users. We refer to the base station as the *transmitter* and the users as *receivers*. For the moment assume that the base station and the users have perfect knowledge of the propagation matrices H_k and receiver noise variances σ_k^2. We assume that the columns of $(H_1\ H_2\ \ldots\ H_K)$ are linearly independent. In particular, we require that

$$M \geq \sum_{k=1}^{K} N_k. \tag{5}$$

For convenience, we rewrite the downlink model (1) in its normalized form

$$\frac{1}{\sigma_k} X_t^k = S_t H_k^\circ + \frac{1}{\sigma_k} W_t^k \tag{6}$$

where $H_k^\circ = H_k / \sigma_k$ is the normalized propagation matrix and the noise term (W_t^k / σ_k) has covariance I.

Consider the following superposition signaling scheme where the transmitter sends a linear combination of the messages intended for the users on the downlink, i.e.,

$$S_t = \sum_{j=1}^{K} A_t^j G_j \tag{7}$$

where A_t^k is the $1 \times N_k$ vector that constitutes the message to user k and G_k is an $N_k \times M$ matrix. From (6) and (7), we obtain the k-th user's received signal:

$$\frac{1}{\sigma_k} X_t^k = \sum_{j=1}^{K} A_t^j G_j H_k^\circ + \frac{1}{\sigma_k} W_t^k. \tag{8}$$

Since the users are non-cooperating, we enforce the following no-crosstalk condition on G_k:

$$G_j H_k^\circ = \delta_{jk} F_k \tag{9}$$

where F_k is an $N_k \times N_k$ matrix and $\delta_{jk} = 1$ if $j = k$ and zero otherwise. Therefore, (8) reduces to

$$\frac{1}{\sigma_k} X_t^k = \sum_{j=1}^{K} A_t^j G_j H_k^\circ + \frac{1}{\sigma_k} W_t^k = A_t^k F_k + \frac{1}{\sigma_k} W_t^k. \tag{10}$$

Let the messages $(A_t^1, \ldots, A_t^K)$ be zero-mean and statistically independent of each other with covariance Q_k:

$$\mathrm{E}[A_t^{l\dagger} A_t^k] = \delta_{kl} Q_k.$$

Let P_k denote the power allocated to send messages to user k:

$$P_k = \mathrm{E}\|A_t^k G_k\|^2 = \mathrm{E}\,\mathrm{tr}[G_k^\dagger Q_k G_k].$$

Then, the power constraint at the transmitter yields

$$\sum_{k=1}^{K} P_k = 1 \tag{11}$$

Thus, the capacity of the Gaussian channel is (10)

$$C_k = \max_{Q_k, F_k} I(X_t^k; A_t^k) = \max_{Q_k, F_k} \log\det(I + F_k^\dagger Q_k F_k) \tag{12}$$

subject to the power constraint

$$\mathrm{tr}[G_k^\dagger Q_k G_k] \leq P_k. \tag{13}$$

Let us define $H^\circ = (H_1^\circ \; H_2^\circ \; \ldots \; H_K^\circ)$ and its least-squares left inverse $E = (H^{\circ\dagger} H^\circ)^{-1} H^{\circ\dagger}$. Let E be decomposed as

$$E = (E_1^T \; E_2^T \; \ldots \; E_K^T)^T \tag{14}$$

where the sub-matrix E_k has size $N_k \times M$. Furthermore, let $E_k = \alpha_k \Gamma_k \beta_k^\dagger$ be the singular value decomposition (SVD) where α_k and β_k have orthonormal columns and Γ_k is diagonal. Then, it can be easily shown that the input distribution that maximizes C_k in (12) is Gaussian. We can take arbitrarily take $Q_k = I$, but this fixes the optimal G_k:

$$G_k = \sqrt{\Pi_k} \beta_k^\dagger \tag{15}$$

where Π_k is yet to be determined. The substitution of (15) into (7) yields the optimal superposition signaling scheme for the transmitter:

$$S_t = \sum_{j=1}^{K} A_t^j \sqrt{\Pi_j} \beta_j^\dagger \tag{16}$$

Clearly, the diagonal entries of Π_k represent the powers allocated to the individual message components of A_t^k and the power constraint in (13) reduces to $\mathrm{tr}[\Pi_k] \leq P_k$.

Now, a simple calculation yields $F_k = G_k H_k^\circ = \sqrt{\Lambda_k}\alpha_k^\dagger$ where $\Lambda_k = \Pi_k \Gamma_k^{-2}$. Using this result in (10) we obtain $X_t^k/\sigma_k = A_t^k \sqrt{\Lambda_k}\alpha_k^\dagger + W_t^k/\sigma_k$. Now, the k-th user applies a unitary transformation to this equation to obtain a diagonal channel

$$\tilde{X}_t^k = \frac{1}{\sigma_k} X_t^k \alpha_k = A_t^k \sqrt{\Lambda_k} + \frac{1}{\sigma_k}\tilde{W}_t^k \tag{17}$$

where the transformed noise vector $\tilde{W}_t^k = W_t^k \alpha_k$ has the same statistics as W_t^k. Equivalently, the transformed channel is a set of parallel independent Gaussian channels. Since the covariance matrix Π_k is also diagonal, the coding of messages on each of these "virtual sub-channels" can be performed independently of the others. The obvious advantage of independent coding is the simplicity of implementation. From (17), we directly obtain the capacity to the k-th user:

$$C_k = \max_{\Pi_k} \sum_{n=1}^{N_k} \log(1+\lambda_{kn}) = \max_{\Pi_k} \sum_{n=1}^{N_k} \log\left(1 + \frac{\pi_{kn}}{\gamma_{kn}^2}\right) \quad \text{s.t.} \quad \sum_{n=1}^{N_k} \pi_{kn} \le P_k. \tag{18}$$

where $\Gamma_k = \text{diag}\{\gamma_{kn} : n = 1, \ldots, N_k\}$ and $\Pi_k = \text{diag}\{\pi_{kn} : n = 1, \ldots, N_k\}$. We briefly examine the problem of optimal power allocation to attain maximum throughput. We consider two cases (a) maximize each C_k for given powers P_k and (b) maximize the sum-capacity $C_{\text{sum}} = \sum_k C_k$ over all power policies.

Maximum capacity for fixed P_k: Suppose that the powers P_k are fixed. What are the optimal powers $\Pi_k^\star$? The problem (18) reduces to computing the capacities of K independent sets of parallel Gaussian channels, whose solutions are given by the familiar "water-filling" rule [1]: $\pi_{kn}^\star = (\mu_k - \gamma_{kn}^2)^+$ for $n = 1, \ldots, N_k$, where $(\cdot)^+$ denotes the positive part and μ_k are chosen to satisfy the power constraint with equality

$$\sum_{n=1}^{N_k} \pi_{kn}^\star = \sum_{n=1}^{N_k} (\mu_k - \gamma_{kn}^2)^+ = P_k.$$

The capacities are given by

$$C_k^\star = \sum_{n=1}^{N_k} \left[\log\left(\frac{\mu_k}{\gamma_{kn}^2}\right)\right]^+.$$

Maximum sum-capacity: Alternatively, we can also vary the total powers P_k allocated to the users and maximize the sum of all capacities. Thus, the set of achievable capacities (18) is convex and the best achievable sum-capacity is the solution to

$$C_{\text{sum}} = \max \sum_{k=1}^{K} \sum_{n=1}^{N_k} \log\left(1 + \frac{\pi_{kn}}{\gamma_{kn}^2}\right) \quad \text{s.t.} \quad \sum_{k=1}^{K} \sum_{n=1}^{N_k} \pi_{kn} \le 1.$$

Again, the solution is given by the water-filling rule: $\pi_{kn}^\star = (\mu - \gamma_{kn}^2)^+$ for $k = 1, \ldots, K$, and $n = 1, \ldots, N_k$ where μ is chosen to satisfy

$$\sum_{k=1}^{K} \sum_{n=1}^{N_k} \pi_{kn}^\star = \sum_{k=1}^{K} \sum_{n=1}^{N_k} (\mu - \gamma_{kn}^2)^+ = 1.$$

The maximum sum-capacity and the individual capacities to the users are

$$C^\star_{\text{sum}} = \sum_{k=1}^{K} C^\star_k, \quad C^\star_k = \sum_{n=1}^{N_k} \left[\log\left(\frac{\mu}{\gamma^2_{kn}}\right)\right]^+.$$

4 Training Scheme

In Section 3 we assumed that the base station and users had complete knowledge of all the propagation matrices. The base station can easily learn the propagation matrices by means of orthonormal training on the uplink. On the downlink, however, it would be a formidable task for each user to learn the channel matrices corresponding to all the users. However, a careful examination of (17) reveals that the k-th user needs to learn the quantities α_k and $\Lambda_k = \Pi_k \Gamma_k^{-2}$ alone in order to diagonalize the channel. In this section, we present a scheme that enables the base station and users to acquire the knowledge to implement (16) and (17) using a small number of training symbols.

4.1 Uplink Training

In this phase of training, the all users send training signals simultaneously on the uplink over a period of T_u symbols. Let us use the subscript τ for signal and noise quantities occurring during training. The training signal of the k-th user is the matrix R^k_τ of size $T_u \times N_k$, where each rows of R^k_τ represents one training symbol. According to (3), the corresponding received signal Y_τ (of size $T_u \times M$) is given by

$$Y_\tau = \sum_{k=1}^{K} R^k_\tau H^T_k + V_\tau = R_\tau H^T + V_\tau,$$

where V_τ is the $T_u \times M$ matrix of receiver noise values (having the same statistics as V_t) and $R_\tau = (R^1_\tau \; R^2_\tau \; \ldots \; R^K_\tau)$ and $H = (H_1 \; H_2 \; \ldots \; H_K)$. Note that we require $T_u \geq \sum_k N_k$ to obtain a meaningful estimate of H. The maximum likelihood (ML) estimate of H is then given by

$$\hat{H} = \left((R^\dagger_\tau R_\tau)^{-1} R^\dagger_\tau Y_\tau\right)^T = H + \left((R^\dagger_\tau R_\tau)^{-1} R^\dagger_\tau V_\tau\right)^T. \tag{19}$$

The covariance of any row of the estimation error $(H - \hat{H})$ is $\sigma^2 (R^\dagger R)^{-1}$. It is easy to verify that this quantity is minimized for $R^k_\tau = \sqrt{T_u/N_k}\Phi_k$ where $\Phi = (\Phi_1 \; \Phi_2 \; \ldots \; \Phi_K)$ has orthonormal columns. Thus, the optimal training signals are orthogonal among all the transmit antennas (the columns of R^k_τ). We impose the above structure on R^k_τ and use (19) to obtain

$$\hat{H}_k = H_k + \sqrt{N_k/T_u}(\Phi^\dagger_k V_\tau)^T.$$

The estimation errors $\tilde{H}_k = \hat{H}_k - H_k = \sqrt{N_k/T_u}(\Phi^\dagger_k V_\tau)^T$ are independent of each other because Φ has orthonormal columns. The entries of $\tilde{H}_k$ are independent $\mathcal{CN}(0, \sigma^2 N_k/T_u)$

random variables. These are the lowest estimation errors attainable individually for any unbiased estimate of each H_k from the training signals meeting the power constraints. Note that if the base station does not know the variances σ_k^2, the users additionally needs to encode and transmit these quantities separately on the uplink.

4.2 Downlink Training

Suppose that the base station (transmitter) had exact knowledge of the normalized propagation matrices. Then, (16) could be implemented exactly. To implement (17), the k-th user would need to learn (a) the unitary matrix α_k and (b) the signal powers $\Lambda_k = \Pi_k \Gamma_k^{-2}$ which determine the transmission rates on the virtual sub-channels. More precisely, the user estimates the matrix $J_k := \sqrt{\Lambda_k}\alpha_k^\dagger$ which can be decomposed uniquely into its factors $\sqrt{\Lambda_k}$ and $\alpha_k^\dagger$.

Let the base station choose Π_j using some power allocation rule such that the power constraint is satisfied, i.e., $\sum_{k=1}^K \operatorname{tr}\Pi_k \leq 1$. Consider the following downlink training signal of length $T_d \geq \max_j N_j$ symbols:

$$S_\tau = \sqrt{T_d}\sum_{j=1}^{K} \Psi_j \sqrt{\Pi_j}\beta_j^\dagger \tag{20}$$

where Ψ_j is a $T_d \times N_j$ matrix of orthonormal columns known to the j-th user and β_j is the right unitary factor of the SVD $E_k = \alpha_k \Gamma_k \beta_k^\dagger$. The matrices Ψ_j are used to extend the training interval. Note that the rows of the matrix $\sqrt{T_d}\Psi_j$ appear like they are drawn from a $\mathcal{CN}(0, I)$ distribution for large T_d. Now, consider the signal power for a general T_d:

$$\begin{aligned}\frac{1}{T_d}\operatorname{tr}[S_\tau S_\tau^\dagger] &= \sum_{j=1}^{K}\sum_{k=1}^{K} \operatorname{tr}[\Psi_j\sqrt{\Pi_j}\beta_j^\dagger\beta_k\sqrt{\Pi_k}\Psi_k^\dagger] \\ &= \sum_{j=1}^{K}\sum_{k=1}^{K} \operatorname{tr}[(\beta_j^\dagger\beta_k)(\sqrt{\Pi_k}\Psi_k^\dagger\Psi_j\sqrt{\Pi_j})]\end{aligned}$$

If $T_d \geq \sum_k N_k$, we can choose the matrices Ψ_j such that $\Psi = (\Psi_1 \;\; \Psi_1 \;\; \ldots \;\; \Psi_K)$ has orthonormal columns. This would imply that the terms corresponding to $j \neq k$ in the double summation vanish and we obtain

$$\frac{1}{T_d}\operatorname{tr}[S_\tau S_\tau^\dagger] = \sum_{j=1}^{K}\operatorname{tr}[\Pi_j] \leq 1.$$

Now, if $T_d < \sum_k N_k$, we cannot guarantee that the training signal meets the power constraint in a particular coherence interval. However, the average power over many coherence intervals can be made to satisfy the constraint:

$$\frac{1}{T_d}\operatorname{E}\operatorname{tr}[S_\tau S_\tau^\dagger] = \sum_{j=1}^{K}\sum_{k=1}^{K}\operatorname{E}\operatorname{tr}[\Psi_j\sqrt{\Pi_j}\beta_j^\dagger\beta_k\sqrt{\Pi_k}\Psi_k^\dagger].$$

Now, the matrices β_j can be *chosen* such that the terms corresponding to $j \neq k$ in the above summation vanish. For instance, let $\alpha'_k = \alpha_k e^{i\phi_k}$ and $\beta'_k = \beta_k e^{i\phi_k}$, where $i = \sqrt{-1}$. Then $\alpha'_k \Gamma_k \beta'^{\dagger}_k = E_k$ is also a valid SVD of E_k. If the angles $\phi_k \in [0, 2\pi]$ are uniform and independent of all other random variables in question, then $\mathrm{E}\,\mathrm{tr}[\Psi_j \sqrt{\Pi_j} \beta'^{\dagger}_j \beta'_k \sqrt{\Pi_k} \Psi^{\dagger}_k] = 0$ for $j \neq k$ and the above summation reduces to

$$\frac{1}{T_d} \mathrm{E}\,\mathrm{tr}[S_\tau S_\tau^\dagger] = \mathrm{E} \sum_{j=1}^{K} \mathrm{tr}[\Pi_j] \leq 1$$

If our SVD algorithm does not produce factors that are uniformly random over the possible values, it may be necessary to generate the angles ϕ_k explicitly. This is however not an issue for $K = 1$ because there are no cross-terms in the summation.

Now, the SVD representation $E_j = \alpha_j \Gamma_j \beta_j^\dagger$ and $E_j H_k^\circ = \delta_{jk} I$ imply that

$$\beta_j^\dagger H_k^\circ = \delta_{jk} \Gamma_k^{-1} \alpha_k^\dagger \tag{21}$$

The substitution of (20) and (21) into (6) gives the corresponding signal received by the k-th user as

$$\begin{aligned} X_\tau^k &= \sigma_k S_\tau H_k^\circ + W_\tau^k = \sigma_k \sqrt{T_d} \sum_{j=1}^{K} \Psi_j \sqrt{\Pi_j} \beta_j^\dagger H_k^\circ + W_\tau^k \\ &= \sigma_k \sqrt{T_d} \Psi_k \sqrt{\Pi_k} \Gamma_k^{-1} \alpha_k^\dagger + W_\tau^k = \sigma_k \sqrt{T_d} \Psi_k \sqrt{\Lambda_k} \alpha_k^\dagger + W_\tau^k \end{aligned}$$

where X_τ^k and W_τ^k are $T_d \times N_k$ matrices representing the received signal and noise for the downlink training phase. Therefore,

$$\hat{J}_k = \frac{1}{\sigma_k \sqrt{T_d}} \Psi_k^\dagger X_\tau^k = \sqrt{\Lambda_k} \alpha_k^\dagger + \frac{1}{\sigma_k \sqrt{T_d}} \Psi_k^\dagger W_\tau^k \tag{22}$$

is an estimate for J_k. We can now estimate Λ_k and α_k from $\hat{J}_k$.

In the above analysis, we assumed that the transmitter had exact knowledge of the propagation matrices. In reality, it has only estimates of these quantities. We omit the exact perturbation analysis that incorporates the effects of estimation errors in the uplink training phase, but refer the reader to [10].

4.3 Rate Scheduling and Outage

We briefly examine the problem of rate scheduling for the downlink transmission. The main advantage of diagonalizing the channel is the resulting simplified coding: the messages on the individual sub-channels are encoded independently and the users perform independent decoding to recover the messages. The base station determines the transmission rates for each of the sub-channels from its estimates of the signal to noise ratios, namely $\Lambda_k = \Pi_k \Gamma_k^{-2}$. Likewise, the users can estimate the link capacities from their estimates of Λ_k.

In a practical system, these capacities would have to be quantized to a set of finite set of transmission rates for implementation purposes. An advantage of quantization also generally ensures that both ends of the link to agree on the transmission rates in the presence of estimation errors. However, near the edges of the quantization cells, the estimation errors can cause the transmitter and receiver to arrive at quantized transmission rates. Such an event constitutes a transmission outage for the entire coherence interval. A second source of outage occurs when the estimation errors cause the transmission rate exceeds the capacity of the link. We provide a detailed analysis of the perturbation analysis and outage errors in [10].

4.4 Special Cases

We briefly examine the following two special cases of interest: (a) many users with single antennas, i.e., $N_k = 1$, and (b) point-to-point MIMO link, i.e., $K = 1$.

In case (a) each E_k is a row-vector because $N_k = 1$. Thus, the right SVD factor β_k of E_k and the message A_t^k are both scalars. The transmitter learns the propagation matrices on the uplink, and each user learns a single coefficient (representing the SNR for the corresponding sub-channel) on the downlink. Since we enforce the no-cross talk among users, this scheme essentially reduces to channel-inversion at the transmitter.

In case (b) there is only one receiver. The transmitter first learns $H = H_1^\circ$ from the uplink training and then computes $E_1 = E = (H^\dagger H)^{-1} H^\dagger$ and its SVD factors $E_1 = \alpha_1 \Gamma_1 \beta_1^\dagger$. Since E_1 is the least-squares left-inverse of H_1°, we clearly also have $H_1^\circ = \beta_1 \Gamma_1^{-1} \alpha_1^\dagger$. In the subsequent downlink training the receiver learns α_1 (the right-SVD factor of H_1°). Thus, both ends of the link have learned their respective unitary factors of the SVD. Thus, they essentially undo the unitary transformations to render the channel diagonal. Note that, even though the SVD, both ends of the link agree on the same factorization that the transmitter chooses.

In the most general case, with many users with multiple antennas, the optimum scheme can be viewed as a partial channel inversion at the transmitter, which renders the channel matrix H° block-diagonal, followed by diagonalization of the blocks using the SVD. The important feature of this training scheme is that the users learn just the essential information rather than all the propagation matrices.

5 Conclusion

We proposed a new training based scheme for the multiple antenna communication link consisting of a base station (transmitter) and K users (receivers). In a reciprocal setting, where the uplink and downlink carrier frequencies are the same, the uplink and downlink channel matrices. We showed that this property can be exploited in our training-scheme to enable the base station and all the users to learn the relevant channel state information using minimal training. Subsequently, the channel is diagonalized and one can perform independent coding over the resulting sub-channels. Our proposed training also automatically allows both ends of the link to agree on the rate scheduling. Apart from the simplicity of coding and scheduling, the other advantage of this scheme is its robustness to the nature of environment. It works well whether the fading is Rayleigh or specular because the transmitter has complete knowledge of the channel and can choose the optimal encoding strategy for each coherence interval.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Laboratories Technical Memorandum,* 1995.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[3] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Info. Theory*, vol. 46, pp. 543–565, March 2000.

[4] B. M. Hochwald and T. L. Marzetta, "Space-time modulation for unknown fading," *Proc. SPIE Aerosense Conf.*, 1999.

[5] L. Zheng and D. Tse, "Communication on the Grassman manifold: a geometric approach to the noncoherent multi-antenna channel," *IEEE Trans. Info. Theory*, vol.48, no. 2, pp. 359–383, 2002.

[6] B. Hassibi and B. Hochwald, "Caley differential unitary space-time codes," *IEEE Trans. Info. Theory*, vol. 48, no. 6, June 2000.

[7] T. L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Allerton Conference on Communication, control, and Computing,*, 1999.

[8] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *submitted to IEEE Trans. Info. Theory.*

[9] B. Hassibi and B. Hochwald, "High rate codes that are linear in space and time," *IEEE Trans. Info. Theory*, vol. 48, no. 7, July 2002.

[10] R. Venkataramani and T. L. Marzetta, "Reciprocal training and scheduling for MIMO," *Lucent Bell Laboratories Technical Memorandum*, 2002.