# BLAST Training: Estimating Channel Characteristics for High Capacity Space-Time Wireless *

Thomas L. Marzetta

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

**Abstract**

BLAST (Bell Labs Layered Space-Time) is a multiple-antenna communication scheme whose outage capacity in a Rayleigh flat fading environment grows linearly with the minimum of the number of transmit and receive antennas, with no increase in bandwidth or transmitted power. Based on its knowledge of the matrix of propagation coefficients, the receiver performs two critical operations: nulling and cancellation, that in effect create independent virtual subchannels.

Assume that the receiver estimates the propagation matrix from a known set of transmitted training signals, and then uses the estimate as though it were correct for nulling and cancellation. How much training is needed for satisfactory operation?

The optimal training signals are orthogonal with respect to time among the transmit antennas, and each transmit antenna is fed equal energy. Errors in estimating the propagation matrix manifest themselves as crosstalk among the virtual subchannels. If its magnitude is too large, the crosstalk constitutes an outage event that is independent of the primary outage event (e.g., that the value of the propagation matrix cannot support the transmission rate). We show that the training interval required to control the probability of the estimation-error induced outage is approximately proportional to the number of transmit antennas, and is independent of the number of receive antennas. Contrary to what is implied by their names, the operations of nulling and cancellation are, in fact, robust with respect to estimation errors.

Our results have an interesting implication for BLAST, either in a mobile environment, or for TDMA, where both training and data transmission have to occur within a fixed interval: if one wishes to maximize the overall transmission rate, then the number of transmit antennas should be chosen such that half of the interval is used for training, and half of the interval for data transmission.

*Proceedings 37th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Sept. 22-24, 1999

# 1 Introduction

Rayleigh flat fading can be beneficial for a multiple-antenna communication link. Specifically, Telatar [1] and Foschini [2] showed that, in a Rayleigh flat-fading environment, a link has a theoretical capacity that increases linearly with the smaller of the number of transmitter and receiver antennas, provided that the complex-valued propagation coefficients between all pairs of transmitter and receiver antennas are statistically independent and known to the receiver. Moreover Foschini invented an architecture for realizing a significant fraction of this capacity, with a small probability of outage, using only ordinary modulation and coding techniques. This theory was confirmed strikingly in experimental demonstrations at Bell Labs of a particular embodiment called BLAST (Bell Labs Layered Space Time), where data was sent at a rate of 600 kbits/s, in a 30 kHz bandwidth, over a link comprising 8 transmit antennas and 12 receive antennas.

Assume that the receiver estimates the matrix of propagation coefficients from a set of known training signals that are sent by the transmitter. This paper provides answers to two related questions: 1) How much training is required for satisfactory operation of BLAST? 2) What effects do estimation errors have on the performance of the scheme?

# 2 Multiple-Antenna Link: Signal Model

Consider a communication link comprising $M$ transmitter antennas and $N$ receiver antennas that operates in a Rayleigh flat-fading environment as shown in Fig. 1. Each receiver antenna responds to each transmitter antenna through a statistically independent fading coefficient. The received signals are corrupted by additive noise that is statistically independent among the $N$ receivers. In complex baseband representation, at time $t$ we transmit a $M$-dimensional row vector $S_t$ and we receive an $N$-dimensional row vector $X_t$, where

$$X_t = \sqrt{\frac{\rho}{M}} S_t H + W_t, \tag{1}$$

$H$ is a $M \times N$ matrix of propagation coefficients, which are statistically independent, zero-mean, unit-variance, circularly symmetric complex Gaussian ($\mathrm{CN}(0,1)$), and $W_t$ is a $N$-dimensional row vector of additive receiver noise whose elements are independent $\mathrm{CN}(0,1)$. The expected power that is fed to each transmit antenna is equal to one,

$$\mathrm{E}\left\{|s_{tm}|^2\right\} = 1. \tag{2}$$

The quantities in the signal model are normalized so that $\rho$ represents the expected signal-to-noise ratio (SNR) at each receiver antenna, independently of the number of transmitter antennas.

In what follows, we assume that there are at least as many receive antennas as transmit antennas, $N \geq M$, and that the propagation matrix remains constant over some interval of duration $T = T_\tau + T_d$, where training occupies $T_\tau$ symbols, and data transmission occupies $T_d$ symbols. In all cases the propagation matrix remains unknown to the transmitter.

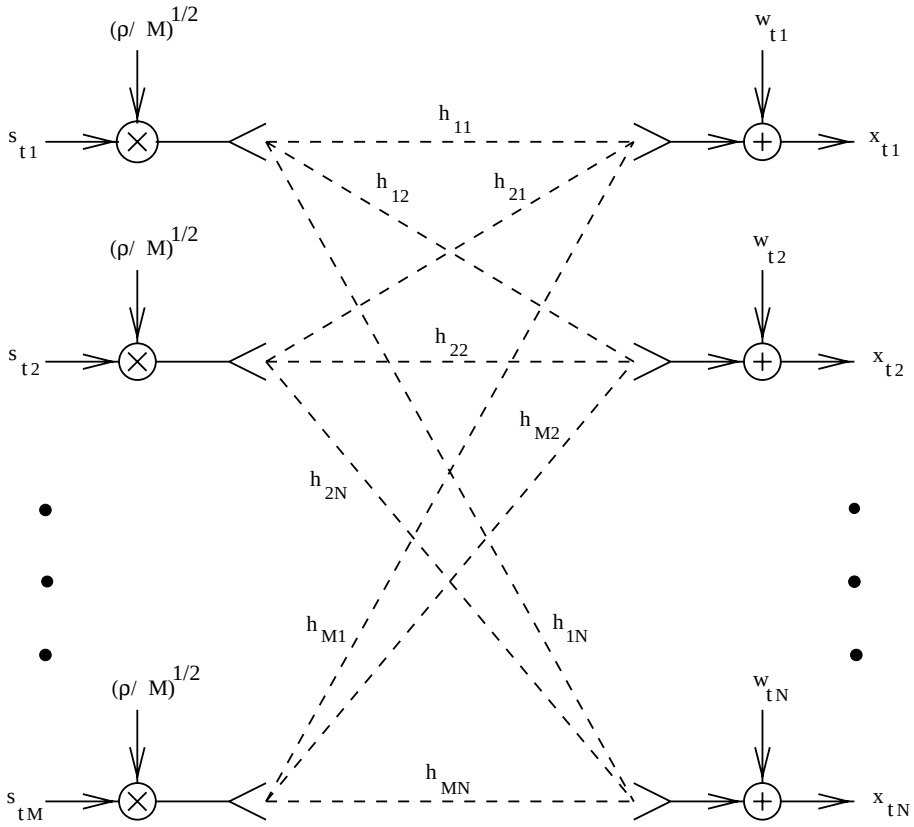



Figure 1: Wireless link comprising $M$ transmitter and $N$ receiver antennas. Every receiver antenna is connected to every transmitter antenna through an independent, random, propagation coefficient having Rayleigh distributed magnitude and uniformly distributed phase. Normalization ensures that the total expected transmitted power is independent of $M$ for a fixed $\rho$.

# 3 Review of BLAST: Nulling and Cancellation

The message bits to be transmitted are divided equally among the $M$ transmit antennas, and the modulation and coding for each transmit antenna occurs independently of the modulation and coding for the other transmit antennas. The message is sent at some pre-arranged rate. Thus the scheme is characterized by an outage probability, since the actual value of the propagation matrix may not support the transmission rate [5]. Assume in this section that the receiver has perfect knowledge of the propagation matrix $H$. The receiver performs a QR factorization of the propagation matrix, and then it implements two operations: nulling and cancellation.

The QR factorization of the propagation matrix has the form (recall that $N \geq M$):

$$H = RQ, \tag{3}$$

where $R$ is an $M \times M$ upper triangular matrix, $r_{mn} = 0,\ m > n$, and the diagonal elements of $R$ are real-valued and nonnegative, and $Q$ is a $M \times N$ matrix whose row vectors are orthonormal.

The nulling operation is a coordinate rotation that multiplies the received vector by $Q^\dagger$ to produce a $M$-dimensional row vector $Y_t$ that constitutes a sufficient statistic,

$$\begin{aligned} Y_t &= X_t Q^\dagger \\ &= \sqrt{\frac{\rho}{M}} S_t H Q^\dagger + W_t Q^\dagger \\ &= \sqrt{\frac{\rho}{M}} S_t R + \nu_t \end{aligned} \tag{4}$$

where the components of $\nu_t$ are independent, $\mathrm{CN}(0,1)$. The effect of nulling is to render the propagation matrix in upper triangular form, with no amplification of receiver noise. One could now multiply $Y_t$ by $R^{-1}$, which would constitute a zero-forcing operation. However Foschini [2] as shown that significant improvements over zero-forcing can be obtained through an operation called cancellation.

Because of the upper-triangular structure of $R$, the first component of $Y_t$ depends only on the signal that was fed to the first transmit antenna,

$$y_{t1} = (\rho/M)^{1/2}\, s_{t1} r_{11} + \nu_{t1}, \tag{5}$$

and it constitutes a virtual subchannel that has no interference from the other subchannels, with an SNR of $\rho r_{11}^2/M$. The sequence $\{y_{t1}\}$ is now decoded to yield the associated message bits. From the decoded message bits, the transmitted signal $\{\hat{s}_{1t}\}$ is reconstructed. By assumption the decoding error is small compared to the outage probability. Consequently, in the absence of an outage event, $\hat{s}_{t1} = s_{t1}$ with a high probability.

Inspection of (4) discloses that $y_{t2}$ is subject to interference from the first subchannel through the off-diagonal term $r_{12}$, which can be removed with high probability through the following cancellation operation

$$y_{t2} - (\rho/M)^{1/2} \hat{s}_{t1} r_{12} = (\rho/M)^{1/2} s_{t2} r_{22} + \nu_{t2}. \tag{6}$$

The decoding is performed on the successive subchannels, preceeded by a cancellation of the off-diagonal terms in (4).

We represent the upper-triangular matrix $R$ as the sum of a diagonal matrix $D$, and a strictly above-triangular matrix $U$,

$$R = D + U. \tag{7}$$

Then the cancellation operation is summarized as

$$\begin{aligned} \left(\frac{\rho}{M}\right)^{-1/2} \left[Y_t - (\rho/M)^{1/2} \hat{S}_t U\right] D^{-1} &= S_t + (\rho/M)^{-1/2} \nu_t D^{-1} + \left(S_t - \hat{S}_t\right) U D^{-1} \\ &= S_t + (\rho/M)^{-1/2} \nu_t D^{-1}, \end{aligned} \tag{8}$$

where the second equality holds if $\hat{S}_t = S_t$, i.e., if no bit-errors have occurred. Thus nulling and cancellation together produce $M$ independent virtual subchannels, and the SNR of the $m$-th subchannel is equal to $\rho r_{mm}^2/M$.[1]

# 4 Training

We assume that $T_\tau \geq M$ symbols are used for sending known training signals, represented by a $T_\tau \times M$ matrix $S$, and that the receiver obtains the maximum likelihood (ML) estimate for the propagation matrix from the $T_\tau \times N$ received signal $X$. The ML estimate is

$$\begin{aligned} \hat{H} &= \left(\frac{\rho}{M}\right)^{-1/2} (S^\dagger S)^{-1} S^\dagger X \\ &= H + \left(\frac{\rho}{M}\right)^{-1/2} (S^\dagger S)^{-1} S^\dagger W, \end{aligned} \tag{9}$$

where $W$ is the matrix of receiver noise. The estimate is unbiased, the estimation error is uncorrelated among the $N$ columns of $\hat{H}$, and the covariance of each column is

$$\begin{aligned} \mathrm{E}\left\{\tilde{h}_n \tilde{h}_n^\dagger\right\} &= \mathrm{E}\left\{\left(\hat{h}_n - h_n\right)\left(\hat{h}_n - h_n\right)^\dagger\right\} \\ &= \left(\frac{M}{\rho}\right)\left(S^\dagger S\right)^{-1}. \end{aligned} \tag{10}$$

The training signals that minimize the trace of the covariance (10) subject to the total energy constraint $\mathrm{tr}\left\{S^\dagger S\right\} = T_\tau \cdot M$, have been shown to be proportional to signals that are orthonormal among the $M$ transmit antennas [4]. To see this we represent $S$ according to its singular value decomposition, $S = \Phi V \Psi^\dagger$, where $\Phi$ is $T_\tau \times M$, with orthonormal columns, $\Psi$ is $M \times M$ unitary, and $V$ is $M \times M$ diagonal, with real nonnegative elements. The error covariance is

$$\left(\frac{M}{\rho}\right)\left(S^\dagger S\right)^{-1} = \left(\frac{M}{\rho}\right) \Psi V^{-2} \Psi^\dagger, \tag{11}$$

whose trace is proportional to $\sum_{m=1}^{M} v_{mm}^{-2}$. Subject to the equivalent energy constraint

[1] Foschini has shown that it is advantageous to order the nulling and cancellation, so that at each step, the subchannel that is decoded has the greatest SNR of the remaining subchannels. We do not explicitly consider this more complicated scheme, but our subsequent analysis applies equally well to it.

$\sum_{m=1}^{M} v_{mm}^2 = T_\tau \cdot M$, the trace is minimized when the $M$ singular values are equal, $v_m = \sqrt{T_\tau}$. This gives $S = \sqrt{T_\tau}\Phi\Psi^\dagger$, and the factor $\Psi^\dagger$ can be absorbed into $\Phi$ without loss of generality. The resulting error covariance is

$$\mathrm{E}\left\{\tilde{h}_n \tilde{h}_n^\dagger\right\} = \left(\frac{M}{\rho T_\tau}\right) I_M. \tag{12}$$

To summarize, the duration of the training interval must be at least as great as the number of transmit antennas $T_\tau \geq M$, and the optimum training signals are of the form $S = \sqrt{T_\tau}\Phi$, where $\Phi$ is $T_\tau \times M$ unitary, i.e., $\Phi^\dagger\Phi = I_M$. The resulting ML estimate for the propagation matrix is unbiased, the performance is independent of the number of receive antennas, and the $M \cdot N$ errors are independent, and distributed as $\mathrm{CN}\left(0, \frac{M}{\rho T_\tau}\right)$.

# 5 Effects of Channel Estimation Error

Starting with its estimate for the propagation matrix, the receiver performs the QR factorization,

$$\hat{H} = \hat{R}\hat{Q} = \left(\hat{D} + \hat{U}\right)\hat{Q}. \tag{13}$$

## 5.1 Channel estimation errors, nulling, and cancellation

The nulling operation gives

$$Y_t = X_t \hat{Q}^\dagger. \tag{14}$$

Together, the nulling and cancellation operations are summarized by the left-hand side of (8), with $Y_t$ according to (14), and with $U$ and $D$ replaced by $\hat{U}$ and $\hat{D}$ respectively. In turn this expression, by the successive substitution of $X_t$ according to (1), the replacement of $H$ by $\hat{H} - \tilde{H}$, and the use of the identity $\hat{H}\hat{Q}^\dagger = \hat{R} = \hat{D} + \hat{U}$, becomes

$$\begin{aligned}
& \left(\frac{\rho}{M}\right)^{-1/2}\left[X_t\hat{Q}^\dagger - (\rho/M)^{1/2}\hat{S}_t\hat{U}\right]\hat{D}^{-1} \\
= \; & \left(\frac{\rho}{M}\right)^{-1/2}\left[(\rho/M)^{1/2}S_tH\hat{Q}^\dagger + W_t\hat{Q}^\dagger - (\rho/M)^{1/2}\hat{S}_t\hat{U}\right]\hat{D}^{-1} \\
= \; & \left(\frac{\rho}{M}\right)^{-1/2}\left[(\rho/M)^{1/2}S_t\hat{H}\hat{Q}^\dagger - (\rho/M)^{1/2}S_t\tilde{H}\hat{Q}^\dagger + W_t\hat{Q}^\dagger - (\rho/M)^{1/2}\hat{S}_t\hat{U}\right]\hat{D}^{-1} \\
= \; & \left(\frac{\rho}{M}\right)^{-1/2}\left[(\rho/M)^{1/2}S_t\hat{D} + W_t\hat{Q}^\dagger - (\rho/M)^{1/2}S_t\tilde{H}\hat{Q}^\dagger + (\rho/M)^{1/2}\left(S_t - \hat{S}_t\right)\hat{U}\right]\hat{D}^{-1} \\
= \; & S_t + \left[(\rho/M)^{-1/2}W_t - S_t\tilde{H}\right]\hat{Q}^\dagger\hat{D}^{-1} + \left(S_t - \hat{S}_t\right)\hat{U}\hat{D}^{-1}.
\end{aligned} \tag{15}$$

This expression is exact, and it contains four terms. The first term is the desired signal $S_t$. As before we assume that $\hat{S}_t = S_t$, so the fourth term vanishes. The second term involves the receiver noise. The third term, containing $S_t\tilde{H}$, can be interpreted as crosstalk among the nominally decoupled virtual subchannels, due to the estimation error. The estimation errors that are associated with $\hat{D}$ and $\hat{Q}$ have only second-order effect in (15). Hence to first order, nulling and cancellation produces the following

$$\left(\frac{\rho}{M}\right)^{-1/2}\left[X_t\hat{Q}^\dagger - (\rho/M)^{1/2}\hat{S}_t\hat{U}\right]\hat{D}^{-1}$$

$$\begin{aligned} &\approx S_t + \left[(\rho/M)^{-1/2} W_t - S_t \tilde{H}\right] Q^\dagger D^{-1} \\ &= S_t + \left[(\rho/M)^{-1/2} \nu_t - S_t \tilde{H} Q^\dagger\right] D^{-1}. \end{aligned} \tag{16}$$

## 5.2 Outage due to channel estimation error

An inspection of (16) discloses the possibility of two independent outage events, that arise from the independent terms $D$ and $\tilde{H}$. The primary outage event is that the elements of $D$ are too small to support the transmission rate. The secondary outage event is that the elements of the estimation error $\tilde{H}$ are big enough to produce a significant crosstalk term. The $n$-th component of the crosstalk is given by

$$\left[S_t \tilde{H} Q^\dagger\right]_n = \sum_{m=1}^{M} s_{tm} \left[\tilde{H} Q^\dagger\right]_{mn}, \tag{17}$$

and its variance, conditioned on $\tilde{H}$, is

$$\mathrm{E}\left\{ \left| \left[S_t \tilde{H} Q^\dagger\right]_n \right|^2 \right\} = \sum_{m=1}^{M} \left| \left[\tilde{H} Q^\dagger\right]_{mn} \right|^2, \tag{18}$$

where we used the facts that the $s_{tm}$ are zero-mean, with unit variance, and that they are independent over $m$. The $M \times M$ matrix $\tilde{H} Q^\dagger$ has independent components that are zero-mean, complex Gaussian, with variance $M/(\rho T_\tau)$. Consequently the conditional variances of the $M$ crosstalk terms are independent, identically distributed, and proportional to chi-square random variables with $2M$ degrees of freedom,

$$\left(\frac{2\rho T_\tau}{M}\right) \cdot \mathrm{E}\left\{ \left| \left[S_t \tilde{H} Q^\dagger\right]_n \right|^2 \right\} : \quad \chi^2_{2M}. \tag{19}$$

We say that a secondary outage event has occurred if any of the $M$ conditional crosstalk variances is greater than some fraction $\beta$ of the scaled receiver noise variance, where typically $\beta \leq 1$, or

$$\sup_n \mathrm{E}\left\{ \left| \left[S_t \tilde{H} Q^\dagger\right]_n \right|^2 \right\} > \frac{\beta M}{\rho}. \tag{20}$$

Utilizing (19) the outage probability is given in terms of the cumulative distribution for a chi-square random variable with $2M$ degrees of freedom by

$$P_{o2} = 1 - \left[P\left\{\chi^2_{2M} < 2\beta T_\tau\right\}\right]^M. \tag{21}$$

Note that the outage probability is independent of the expected SNR, and that it only depends on $M$ and on the product $\beta T_\tau$. Figure 2 displays the outage probability according to this formula, as a function of $\beta T_\tau / M$ for five values of $M$. The outage probability decays exponentially with the duration of the training interval, and the required training interval increases at a rate that is approximately linear with the number of transmit antennas.

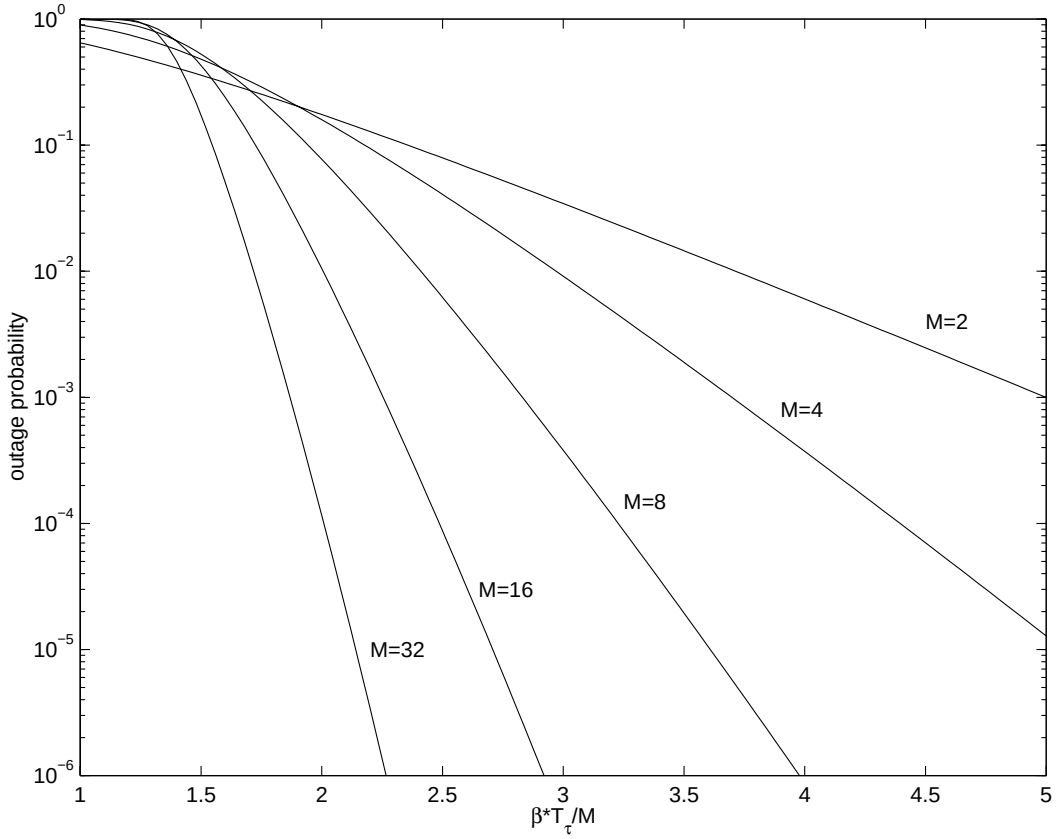



Figure 2: Estimation error-induced outage probability as a function of $\beta T_\tau / M$

# 6 Conclusions

The BLAST scheme does not impose especially stressful training requirements, and the operations of nulling and cancellation are robust with respect to errors in estimating the propagation matrix.

Our findings, notwithstanding, training emerges as a major issue when the total interval for training and for transmitting data is limited, either because of fast fading or because of a TDMA scenario, and where it is desired to maximize the effective transmission rate. We have seen that the required training interval grows approximately linearly with the number of transmit antennas. In turn Foschini [2] has shown that the capacity in units of bits per symbol is approximately proportional to the number of transmit antennas. Thus $T_\tau = \alpha M$, for some constant $\alpha$, and this leave $T_d = T - T_\tau = T - \alpha M$ symbols for sending the message. The capacity is $C = \gamma M$ for some constant $\gamma$, so the total number of message bits that can be sent is equal to $CT_d = \gamma M(T - \alpha M)$. The optimum number of transmit antennas from the standpoint of maximizing the number of message bits is $M = \frac{T}{2\alpha}$, implying that, whatever the values of $\alpha$ and $\gamma$, the training interval is $T_\tau = T/2$. In other words, when trying to achieve the maximum possible throughput, half of the available interval is used for training.

Ideally one would like to achieve BLAST-like transmission rates with multiple antennas, while completely avoiding training and channel estimation. Some steps in this direction are described in [6] and [7].

# References


[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Laboratories internal 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] T. L. Marzetta, "Estimating channel characteristics for a high capacity, layered space-time architecture for wireless communications", Bell Laboratories internal technical memorandum, 1996.

[4] S. D. Silverstein, "Application of orthogonal codes to the calibration of active phased array antennas for communication satellites", *IEEE Trans. on Signal Processing*, vol. 45, no. 1, pp.206–218, 1997.

[5] L. H. Ozarow, S. Shami (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio", *IEEE Trans. Info. Thy.*, vol 43, pp.359–378, 1994.

[6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Info. Thy.*, vol 45, no. 1, pp. 139–157, 1999.

[7] B. M. Hochwald and T. L. Marzetta, "Unitary Space-Time Modulation for multiple-antenna communications in Rayleigh flat fading," to appear in *IEEE Trans. Info. Thy.*