

MASTER COPY: PLEASE KEEP THIS "MEMORANDUM OF TRANSMITTAL" BLANK FOR REPRODUCTION PURPOSES. WHEN REPORTS ARE GENERATED UNDER THE ARO SPONSORSHIP, FORWARD A COMPLETED COPY OF THIS FORM WITH EACH REPORT SHIPMENT TO THE ARO. THIS WILL ASSURE PROPER IDENTIFICATION. NOT TO BE USED FOR INTERIM PROGRESS REPORTS; SEE PAGE 2 FOR INTERIM PROGRESS REPORT INSTRUCTIONS.

MEMORANDUM OF TRANSMITTAL

U.S. Army Research Office
ATTN: AMSRL-RO-BI (TR)
P.O. Box 12211
Research Triangle Park, NC 27709-2211

Reprint (Orig + 2 copies)

Technical Report (Orig + 2 copies)

Manuscript (1 copy)

Final Progress Report (Orig + 2 copies)

Related Materials, Abstracts, Theses (1 copy)

CONTRACT/GRANT NUMBER:

REPORT TITLE:

is forwarded for your information.

SUBMITTED FOR PUBLICATION TO (applicable only if report is manuscript):

Sincerely,

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE			5. FUNDING NUMBERS	
6. AUTHOR(S)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12 b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)				
14. SUBJECT TERMS			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

Enclosure 1

Throughput-Optimal Training for a Time-Varying Multi-Antenna Channel

Christian B. Peel and A. Lee Swindlehurst

Abstract—A lower bound on the capacity of trained space-time modulation is presented for the case of a time-varying channel. We consider the case where the channel consists of two components of variable strength: a specular component from line-of-sight or strong coherent multipath, and a time-varying diffuse component due to a mobile receiver or scatterers. The training period and the data period are assumed quasi-static, with the time-autocorrelation function of the channel used to model the variation between each sub-block. The training signal, training signal length, power allocation and training frequency that optimize the capacity bound are derived. We compare the effective SNR of this bound to that of differential modulation at high SNR, and find it to be lower. We find the best number of antennas to use at high SNR with differential modulation, and compare it with the optimal value for trained modulation. For our time-varying channel model, trained modulation often has a higher achievable rate than differential modulation. Our results are supported by several numerical examples.

Index Terms—MIMO systems, time-varying channels, fading channels, Rician channels, Rayleigh channels.

I. INTRODUCTION

Time-varying multiple-antenna channels have received increased interest recently, motivated by demand for high bandwidth mobile wireless data links. This has prompted investigation of modulation techniques that can handle temporally selective fading. Though trained modulation is not ideal for a time-varying channel, it is often used because of its simplicity. The parameters that optimize a lower bound on capacity for a training-based scheme have been found by Hassibi and Hochwald [1]. They assume quasi-static fading, where the channel coefficients are constant for a given coherence interval before changing to new independent values. Though a reasonable approximation in some slow-fading scenarios, this first-order model does not enable determination of the best training frequency; a model utilizing second-order statistics would more closely approximate practical channel conditions.

Another approach for time-varying fading is to assume that the receiver has no knowledge of the channel through training or other channel tracking techniques; this is often referred to as the *noncoherent* case. In this situation, unitary signals have been found to be optimal [2]. Several signaling schemes have been developed using this assumption, including differential space-time modulation [3]–[5]. The capacity of

differential modulation for a quasi-static channel has recently been found [5] and provides insight on the capacity of the noncoherent channel [6]. A capacity bound for training over a quasi-static noncoherent Rician channel is used in [7], [8].

In [9] the time-autocorrelation function of the channel coefficients and multiple first-order AR systems are used to model the channel. An effective SNR is derived that includes the penalty due to channel estimation error and the effects of the time-varying channel. In this paper, we use a similar model for analysis; the channel is modeled as quasi-static during both training and data sub-blocks, with the variation between these sub-blocks given by the time-autocorrelation of the channel. Using this model, a lower bound on capacity is derived for training-based multi-antenna systems which is an extension of results in [1] to the case of a time-varying diffuse channel with a known autocorrelation function. We find the properties of the training signal (duration, power, *etc.* as in [1]) as well as the training frequency and number of antennas that maximizes the bound. We also use a high-SNR approximation to the capacity of differential modulation [5] to derive the best number of antennas to use with differential modulation, and find it to be similar to that found for trained modulation with a coding interval of $T = 2M$ samples.

The next section presents the channel model and the basic assumptions underlying our analysis. In Section III we derive the lower capacity bound and maximize it over the training parameters. In Section IV the capacity of differential modulation is investigated, and is compared with the lower bound on capacity for trained modulation. We present numerical examples of the capacity and capacity bounds as functions of the optimized parameters in Section V, and conclude with a discussion of possible extensions and future work.

II. CHANNEL MODEL

A. Fading Channel Model

Assume a flat-fading communications environment with M transmit and N receive antennas. A complex channel coefficient that is assumed constant for a symbol of length T_c describes the effect of the propagation between each pair of transmit and receive antennas. At each receive antenna, interference and other disturbances add temporally and spatially independent noise to the signal. In this case, the channel equation can be written in matrix form:

$$X = \sqrt{\frac{\rho}{M}} SH + V, \quad (1)$$

where X is a $T_c \times N$ matrix of received data, S is a $T_c \times M$ signal matrix, H is a $M \times N$ matrix of channel coefficients,

The authors are with ArrayComm LLC, 2480 N. First Street Suite 200, San Jose, CA 95131 (email: cpeel@arraycomm.com, swindle@arraycomm.com). We thank the anonymous reviewers for their many helpful recommendations.

This work was supported by the National Science Foundation under Information Technology Grant CCR-0081476, and by the U. S. Army Research Office under the Multi-University Research Initiative (MURI) grant W911NF-04-1-0224.

and V is a $T_c \times N$ matrix of noise and interference. We say an $M \times N$ matrix A has unit mean-square if

$$E[\|A\|_F^2] = E[\text{tr}(AA^H)] = MN,$$

where the Frobenius norm is denoted by $\|\cdot\|_F$. The matrices S , H , and V all are assumed to have unit mean-square; the SNR at each receive antenna is thus ρ .

B. Specular and Diffuse Components

We decompose the channel into specular and diffuse components as follows:

$$H = \sqrt{1-\beta}H^s + \sqrt{\beta}H^r, \quad (2)$$

where the specular part H^s is assumed to be fixed and known at the receiver, while the diffuse component H^r is random and unknown. This occurs, for example, in a mobile environment where the specular component is the result of a relatively easy-to-acquire line-of-sight ray and the diffuse component depends on local scatterers and is changing more rapidly. Preliminary analysis of the capacity of the channel for this model is found in [8], [10]. The specular and diffuse power are defined separately, as follows:

$$\rho^s = (1-\beta)\rho \quad (3)$$

$$\rho^r = \beta\rho, \quad (4)$$

where $\rho = \rho^s + \rho^r$. The parameter β indicates the relative contribution of the diffuse vs. specular components, and can be chosen to correspond to a fully specular ($\beta = 0$) or a fully diffuse ($\beta = 1$) channel, or something in between. The Rician K -factor can also be used for the same purpose, but it is convenient for our analysis to have a quantity between zero and one. One can translate between the two models in a simple way: $K = (1-\beta)/\beta$ and $\beta = 1/(1+K)$. We assume no specific distribution for the channel, except that H^s and H^r are independent, and that both are left and right rotationally invariant, i.e., their probability density functions satisfy $p(\Theta H \Phi) = p(H)$ for all unitary matrices Θ and Φ .

As an example of a channel that satisfies this model, consider the rank-one channel that will be assumed for our numerical examples. The diffuse component of this model is a matrix composed of $\mathcal{CN}(0,1)$ random variables, and the specular component is rank-one and is written as the outer product of two unit vectors [2]:

$$H^s = \sqrt{MN} \mathbf{u} \mathbf{v}^H, \quad (5)$$

where $\mathbf{u}^H \mathbf{u} = 1$ and $\mathbf{v}^H \mathbf{v} = 1$. A rotationally invariant realization of the specular component can be obtained using vectors whose elements are independently drawn from a $\mathcal{CN}(0,1)$ distribution, then normalized to have unit length. Rank-one models of the specular component similar to this have been used for analysis of capacity in [8], and analysis of time-varying channels in [9].

C. Channel Estimation

Following [1], we apply the fading model (1) to both training and data transmission phases, resulting in the following joint channel model

$$\begin{bmatrix} X_\tau \\ X_d \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{\rho_\tau}{M}} S_\tau H_\tau \\ \sqrt{\frac{\rho_d}{M}} S_d H_d \end{bmatrix} + \begin{bmatrix} V_\tau \\ V_d \end{bmatrix}, \quad (6)$$

where the training signal S_τ has dimension $T_\tau \times M$ and the data signal S_d has dimension $T_d \times M$. Conservation of time and energy yields the following:

$$T = T_\tau + T_d \quad (7a)$$

$$\rho T = \rho_\tau T_\tau + \rho_d T_d, \quad (7b)$$

where ρ_τ and ρ_d are the relative signal powers employed during training and data transmission, respectively. An important issue is the allocation of energy between training and data transmission. We use the parameter κ to make the tradeoff:

$$\rho_d T_d = \kappa \rho T \quad (8a)$$

$$\rho_\tau T_\tau = (1-\kappa) \rho T. \quad (8b)$$

Another aspect of our analysis is the penalty due to channel estimation error. In particular, we consider the performance of the linear minimum mean-squared error (LMMSE) estimator. Since the columns of X_τ all have the same distribution, the noise is temporally white, and the specular part of the channel and β are assumed to be known, the LMMSE estimator of the diffuse channel has the following form:

$$\hat{H}^r = R_{HX_\tau} R_{X_\tau X_\tau}^{-1} \left(X_\tau - \sqrt{(1-\beta) \frac{\rho_\tau}{M}} S_\tau H^s \right)$$

where

$$R_{HX_\tau} = N \sqrt{\frac{\rho_\tau \beta}{M}} S_\tau^H,$$

and

$$R_{X_\tau X_\tau} = N \left(\frac{\rho_\tau \beta}{M} S_\tau S_\tau^H + I_{T_\tau} \right)$$

are standardly defined covariance matrices. Combining the above equations, and applying a standard linear algebra result, we obtain

$$\hat{H}^r = \sqrt{\frac{\rho_\tau \beta}{M}} \left(I_M + \frac{\rho_\tau \beta}{M} S_\tau^H S_\tau \right)^{-1} S_\tau^H \left(X_\tau - \sqrt{(1-\beta) \frac{\rho_\tau}{M}} S_\tau H^s \right). \quad (9)$$

D. Time-Varying Diffuse Element

The quasi-static model of a time-varying channel assumes that T is the coherence length of the channel, or that the channel is approximately constant for T samples. In our model we go one step further, by allowing the channel to change between the data and training phases of transmission. This means that T is no longer a coherence interval, but is still the coding interval or training period. We assume that the diffuse channel component varies according to the following first-order auto-regressive (AR) or *Gauss-innovations* model:

$$H_d^t = \sqrt{\alpha_T} H_\tau^t + \sqrt{1-\alpha_T} W_d, \quad (10)$$

where H_τ^r and W_d have the same distribution, are zero-mean, and have unit mean-square; we also constrain $0 \leq \alpha_T \leq 1$. Under this model, H_d^r has the same distribution as H_τ^r . Note that $\alpha_T = 1$ produces a time-invariant channel, and $\alpha_T = 0$ indicates a channel with no temporal correlation.

The parameter α_T can be chosen to match the second-order statistics of models based on the mechanisms of physical propagation. Let $r_{hh}(t)$ denote the time autocorrelation function of an element of H_d^r . Solving the Yule-Walker equations for α_T in the first-order AR process (10) we obtain

$$\alpha_T = \left| \frac{r_{hh}(T/2)}{r_{hh}(0)} \right|^2, \quad (11)$$

which provides a reasonable choice for α_T . For example, assuming the Jakes/Clarke model of the land mobile fading channel [11], $r_{hh}(t) = J_0(2\pi ft)$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, $f = f_d T_S$, f_d is the maximum Doppler frequency in the fading environment, and T_S is the sampling period. Under this model, (11) leads to

$$\alpha_T = [J_0(\pi T f)]^2. \quad (12)$$

This model is an appropriate approximation when using decoders that depend only on a single reference channel (i.e., decoders that assume the channel is known). This fact is borne out by the simulation results of [12] (see Figures 1, 2, and 3), where excellent agreement is obtained with data generated according to the Jakes/Clarke model, but analyzed with the Gauss-innovations model.

III. CAPACITY

An analysis of the capacity achievable with trained modulation over the time-varying channel described above is now presented. A lower bound on capacity is derived, extending the work of [1] to that of a time-varying diffuse channel with known time-autocorrelation function and a fixed specular component. We find that the lower bound is significantly different than that in [1].

A. Capacity and Effective SNR

Section II describes in detail the relationship between the channel seen during the information-bearing and training phases:

$$\begin{aligned} X_d &= \sqrt{\frac{\rho_d}{M}} S_d H_d + V_d \\ &= \sqrt{(1-\beta) \frac{\rho_d}{M}} S_d H^s + \sqrt{\beta \alpha_T \frac{\rho_d}{M}} S_d \hat{H}_\tau^r + \\ &\quad \sqrt{\beta(1-\alpha_T) \frac{\rho_d}{M}} S_d W_d + V_d. \end{aligned}$$

The difference between the estimated diffuse channel and that seen during the training phase is defined to be $\tilde{H}_\tau^r = H_\tau^r - \hat{H}_\tau^r$. We separate the data into two terms, one due to the desired signal, and one that comprises an effective “noise” term due to

channel estimation error, channel time-variation, and additive noise:

$$\begin{aligned} X_d &= \underbrace{\sqrt{(1-\beta) \frac{\rho_d}{M}} S_d H^s + \sqrt{\beta \alpha_T \frac{\rho_d}{M}} S_d \hat{H}_\tau^r}_{X'_d} + \\ &\quad \underbrace{\sqrt{\beta \alpha_T \frac{\rho_d}{M}} S_d \tilde{H}_\tau^r + \sqrt{\beta(1-\alpha_T) \frac{\rho_d}{M}} S_d W_d + V_d}_{V'_d}. \end{aligned} \quad (13)$$

Let $R_S = \frac{1}{T_d} E[S_d^H S_d]$, and define

$$\sigma_{\tilde{H}_\tau^r, R_S}^2 = \frac{1}{MN} E \text{tr} \left[\left(\tilde{H}_\tau^r \right)^H R_S \tilde{H}_\tau^r \right].$$

The power of the effective noise term is given by:

$$\begin{aligned} \sigma_{V'_d}^2 &= \frac{1}{T_d N} E \text{tr} \left[V_d^H V_d \right] \\ &= \rho_d \alpha_T \beta \sigma_{\tilde{H}_\tau^r, R_S}^2 + \rho_d \beta (1 - \alpha_T) + 1. \end{aligned} \quad (14)$$

The power of the signal term X'_d is

$$\begin{aligned} \sigma_{X'_d}^2 &= \frac{1}{T_d N} E \text{tr} \left[X_d^H X_d \right] \\ &= \rho_d (1 - \beta) \sigma_{H^s, R_s}^2 + \rho_d \alpha_T \beta \sigma_{\tilde{H}_\tau^r, R_S}^2, \end{aligned} \quad (15)$$

where

$$\sigma_{\hat{H}_\tau^r, R_S}^2 = \frac{1}{MN} E \text{tr} \left[\left(\hat{H}_\tau^r \right)^H R_S \hat{H}_\tau^r \right],$$

and

$$\sigma_{H^s, R_s}^2 = \frac{1}{MN} E \text{tr} \left[\left(H^s \right)^H R_S H^s \right].$$

It is useful to compare the equations for X_d in (6) and (13); in the former, the channel contains an unknown diffuse channel component, while in (13) the channel components H^s and \hat{H}_τ^r are completely known. The analysis of a system with known channel coefficients is much easier than that with an unknown channel. Another significant difference between X_d in (6) and (13) is that the noise in (13) is not guaranteed to be white. Fortunately, Theorem 1 from [1] provides a means to bound the capacity for a training-based scheme with a known channel and unknown noise distribution as in (13). The theorem indicates that the capacity (C_τ) of a training-based system with a channel that is rotationally invariant is lower bounded by a system with i.i.d. white Gaussian noise of the same power. In particular,

$$C_\tau \geq \frac{T - T_\tau}{T} E \left[\log \left| I_N + \frac{\rho_d^{\text{eff}}}{M} \bar{H}^H \bar{H} \right| \right] \triangleq C_t, \quad (16)$$

where

$$\bar{H} = \sqrt{\frac{1-\beta}{\sigma_{X'_d}^2}} \rho_d H^s + \sqrt{\frac{\beta \alpha_T}{\sigma_{X'_d}^2}} \rho_d \hat{H}_\tau^r, \quad (17)$$

and the effective SNR is

$$\rho_d^{\text{eff}} = \frac{\sigma_{X'_d}^2}{\sigma_{V'_d}^2} = \frac{\rho_d (1 - \beta) \sigma_{H^s, R_s}^2 + \rho_d \alpha_T \beta \sigma_{\tilde{H}_\tau^r, R_S}^2}{1 + \rho_d \beta (1 - \alpha_T) + \rho_d \alpha_T \beta \sigma_{\tilde{H}_\tau^r, R_S}^2}. \quad (18)$$

The value of R_S that maximizes (16) depends on \bar{H} and therefore on S_τ . Instead of maximizing R_S and S_τ jointly,

we follow [1] and set $R_S = I_M$, assuming that since the transmitter does not know the channel, the only reasonable strategy is to use all transmit antennas equally. In fact we will see in the next section that the S_τ that maximizes ρ_{eff} results in $R_S = I_M$. Thus these choices of R_S and S_τ are consistent, though not chosen jointly. Let \bar{H}_τ^r be the diffuse channel estimate normalized to be unit mean square: $\bar{H}_\tau^r = \frac{1}{\sigma_{\bar{H}_\tau^r}} \hat{H}_\tau^r$. With this definition we can write

$$\bar{H} = \sqrt{1 - \beta_{\text{eff}}} H^s + \sqrt{\beta_{\text{eff}}} \bar{H}_\tau^r, \quad (19)$$

where

$$\beta_{\text{eff}} = \frac{\alpha_T \beta \sigma_{\bar{H}_\tau^r}^2}{1 - \beta + \alpha_T \beta \sigma_{\bar{H}_\tau^r}^2} = \frac{1}{1 + (1 - \beta) \frac{1}{\alpha_T \beta \sigma_{\bar{H}_\tau^r}^2}}, \quad (20)$$

and the effective SNR is now

$$\rho_{\text{eff}} = \frac{\rho_d (1 - \beta) + \rho_d \alpha_T \beta \sigma_{\bar{H}_\tau^r}^2}{1 + \rho_d \beta (1 - \alpha_T) + \rho_d \alpha_T \beta \sigma_{\bar{H}_\tau^r}^2} \quad (21)$$

$$= \rho_d \left(\frac{1 - \beta (1 - \alpha_T \sigma_{\bar{H}_\tau^r}^2)}{1 + \rho_d \beta (1 - \alpha_T \sigma_{\bar{H}_\tau^r}^2)} \right), \quad (22)$$

where we have dropped the R_S subscript from the symbols for the channel power, and used the facts that $R_S = I_M$ implies $\sigma_{H^s}^2 = 1$, and that for the given LMMSE channel estimator, $\sigma_{\bar{H}_\tau^r}^2 = 1 - \sigma_{\hat{H}_\tau^r}^2$. For the special case of $\beta = 1$ and $\alpha_T = 1$, we obtain the lower bound in [1] which does not model the time variation of the channel nor a specular versus diffuse separation of the channel. Since $0 \leq \alpha_T, \beta, \sigma_{\bar{H}_\tau^r}^2 \leq 1$, it is clear from the above expressions that the effect of the channel estimation error and time variation is to reduce both the effective SNR and the effective specular parameter β_{eff} . The channel appears more diffuse due to the presence of the channel estimation error \hat{H}_τ^r and the error due to mobility W_d , both of which have the same distribution as H_d^r . The remainder of this section will focus on maximizing this lower bound over the parameters S_τ , κ , T_τ , M , and T .

B. Unitary Training Signals

We begin by optimizing the bound over the training signal S_τ . When $\beta \simeq 0$, it is best to disregard the diffuse part of the channel, do no training, and to communicate over the (known) specular part of the channel. For now, we assume that β is large enough that training is necessary; we will discuss the $\beta \simeq 0$ case in Section III-D. Equation (16) indicates that the choice of training signal affects the bound primarily through the effective SNR ρ_{eff} . In the discussion below, we find the training signal that maximizes ρ_{eff} , and show that it also maximizes β_{eff} .

It is clear from (21) that maximizing $\sigma_{\bar{H}_\tau^r}^2$ (i.e., minimizing $\sigma_{\hat{H}_\tau^r}^2$) will maximize the effective SNR; it is also clear from (20) that maximizing $\sigma_{\bar{H}_\tau^r}^2$ also maximizes β_{eff} . As shown in [1], this problem can be restated in terms of the eigenvalues of $S_\tau^H S_\tau$ (denoted by $\lambda_1, \dots, \lambda_M$), as follows:

$$\min_{\lambda_1, \dots, \lambda_M} \sum_{m=1}^M \frac{1}{1 + \frac{\rho \beta}{M} \lambda_m} \quad \text{such that} \quad \sum_{m=1}^M \lambda_m = T_\tau M. \quad (23)$$

When $T_\tau \geq M$, it is shown in [1] that the solution is achieved by setting $\lambda_1 = \dots = \lambda_M = T_\tau$, resulting in an optimal training signal with orthogonal columns: $S_\tau^H S_\tau = T_\tau I_M$. Unlike [1], we are also interested in the case where $T_\tau \leq M$, for which the minimum is achieved by $\lambda_1 = \dots = \lambda_{T_\tau} = M$ and $\lambda_{T_\tau+1} = \dots = \lambda_M = 0$, resulting in a training signal with orthogonal rows: $S_\tau S_\tau^H = M I_{T_\tau}$. We will assume unitary training signals such as these throughout the remainder of this work, which results in a channel estimate with variance

$$\sigma_{\bar{H}_\tau^r}^2 = \frac{\rho_\tau \beta T_\tau}{M + \rho_\tau \beta K}, \quad (24)$$

where $K = \max(M, T_\tau)$. This result provides a more explicit expression for the effective SNR, as presented in the following theorem.

Theorem 1 (Unitary Training Signals): Given the channel model of Section II, the training signal that maximizes the effective SNR (21) satisfies $S_\tau^H S_\tau = T_\tau I_M$ for $T_\tau \geq M$ or $S_\tau S_\tau^H = M I_{T_\tau}$ for $T_\tau \leq M$. In this case the effective SNR is

$$\rho_{\text{eff}} = \frac{\rho_d (1 - \beta) (M + \rho_\tau \beta K) + \alpha_T \rho_d \rho_\tau \beta^2 T_\tau}{(1 + \rho_d \beta) (M + \rho_\tau \beta K) - \alpha_T \rho_d \rho_\tau \beta^2 T_\tau}, \quad (25)$$

and the effective specular/diffuse parameter is

$$\beta_{\text{eff}} = \frac{\alpha_T \rho_\tau \beta^2 T_\tau}{(1 - \beta) (M + \rho_\tau \beta K) + \alpha_T \rho_\tau \beta^2 T_\tau}. \quad (26)$$

We recover the results of [1] when $\beta = 1$ and $\alpha_T = 1$. However, unlike [1], we discover a high-SNR error floor due to the time-variation of the channel:

$$\lim_{\rho \rightarrow \infty} \rho_{\text{eff}} = \frac{K(1 - \beta) + \alpha_T \beta T_\tau}{(K - \alpha_T T_\tau) \beta}. \quad (27)$$

In the general case, the effective SNR ρ_{eff} is a complicated function of various quantities, and maximizing it over all of the system design parameters is a difficult task. In the sections that follow, we examine the behavior of ρ_{eff} with respect to the power allocation (κ), training length (T_τ) and training frequency (T), and optimize these parameters where possible.

C. Optimizing the Power Allocation

Assigning more power to training leads to a better channel estimate, but less power available for data transmission. This tradeoff enters into the effective SNR (25) through the dependence of ρ_d and ρ_τ on κ , the power allocation between on training and data periods. Maximizing (25) with respect to κ , the following result is obtained.

Theorem 2 (Optimal Power Allocation): Given the channel model of Section II and unitary training signals, the power allocation which maximizes the effective SNR (25) is $\kappa = \min(1, \hat{\kappa})$, where for $\gamma \neq 0$

$$\hat{\kappa} = \frac{1 - \sqrt{1 - \delta \gamma}}{\gamma}, \quad (28)$$

with

$$\gamma = \frac{K}{M} - \frac{\alpha_T T_\tau^2}{T_d (K - K\beta + \alpha_T \beta T_\tau)} \frac{K}{M} + \frac{T_\tau}{\rho T \beta} \quad (29)$$

$$\delta = 1 + \frac{M T_\tau (1 - \beta)}{\rho T \beta (K - K\beta + \alpha_T \beta T_\tau)}. \quad (30)$$

The resulting effective SNR and β are obtained by substituting the above power allocation into (25) and (26), respectively. For $\gamma = 0$

$$\hat{\kappa} = \frac{1}{2} + \frac{1}{2} \frac{KT_d(1-\beta)}{\alpha_T \rho T \beta T_\tau}. \quad (31)$$

Proof: The result is readily found by setting the derivative of the effective SNR with respect to κ equal to zero. ■

As $\beta \rightarrow 0$ the optimal power allocation goes to one ($\hat{\kappa} \rightarrow 1$), more power is allocated to data transmission, and less power is allocated to training. The transition from doing some training to not training at all occurs at specular parameter

$$\hat{\beta} = \frac{\sqrt{MT_d(MT_d + 4\alpha_T \rho T(T_d + \rho T))} - MT_d}{2\alpha_T T \rho (T_d + \rho T)}.$$

As $\rho \rightarrow \infty$ the power allocation becomes

$$\lim_{\rho \rightarrow \infty} \hat{\kappa} = \frac{\sqrt{KT_d(K(1-\beta) + \alpha_T \beta T_\tau)}}{\sqrt{KT_d(K(1-\beta) + \alpha_T \beta T_\tau)} + T_\tau M}, \quad (32)$$

and the resulting effective SNR is

$$\lim_{\rho \rightarrow \infty} \rho_{\text{eff}}(\hat{\kappa}) = \frac{K(1-\beta) + \alpha_T \beta T_\tau}{\sqrt{\beta(K - \alpha_T T_\tau)} + T_\tau \sqrt{M \alpha_T}}. \quad (33)$$

The asymptotic value (32) is valid in typical scenarios where M is not large, $\rho > 1$, and β is not too small ($\beta > 0.5$).

D. Optimizing the Training Length

A longer training length T_τ results in higher effective SNR values due to improved channel estimation, while shorter intervals leave more time for data transmission; here we attempt to optimally balance this trade-off by maximizing the bound (16) over T_τ . The following theorem is obtained for the case where $T_\tau \geq M$.

Theorem 3 (Optimal Training Length for $T_\tau > M$): The capacity-maximizing training length when using pilot signals with unitary columns (assuming $T_\tau \geq M$) with a fixed power allocation κ is $T_\tau = M$.

Proof: Let $Q = \min(M, N)$ and λ be an arbitrary non-zero eigenvalue of the matrix $\frac{1}{M} \bar{H}^H \bar{H}$. Then the capacity bound of (16) can be expressed as

$$C_t = \frac{QT_d}{T} E \log_2(1 + \rho_{\text{eff}} \lambda). \quad (34)$$

We will find the T_d that maximizes C_t , and use that result to determine the best training data length. Because $T_\tau \geq M$, $K = T_\tau$ and every place T_τ appears in (26), ρ_τ appears also. We know $\rho_\tau T_\tau = (1-\kappa)\rho T$ with the result that that β_{eff} and \bar{H} do not depend on T_d . Using this, we write the derivative of C_t with respect to T_d :

$$\frac{dC_t}{dT_d} = \frac{Q}{T} E \log(1 + \rho_{\text{eff}} \lambda) + \frac{QT_d}{T} \frac{d\rho_{\text{eff}}}{dT_d} E \left[\frac{\lambda}{1 + \rho_{\text{eff}} \lambda} \right]. \quad (35)$$

The derivative of the effective SNR (25) for $T_\tau \geq M$ is $\frac{d\rho_{\text{eff}}}{dT_d} = \rho_{\text{eff}} \zeta$, where

$$\zeta = \frac{\rho T \beta (1-\kappa) + M}{\alpha (\rho T \beta)^2 \kappa (1-\kappa) - [\rho T \beta \kappa + T_d][M + \rho T \beta (1-\kappa)]}. \quad (36)$$

Using (36) we note that if the inequality

$$\log(1 + \rho_{\text{eff}} \lambda) + T_d \zeta \frac{\rho_{\text{eff}} \lambda}{1 + \rho_{\text{eff}} \lambda} \geq 0 \quad (37)$$

is satisfied, then $\frac{dC_t}{dT_d} \geq 0$. In [1] it was shown that for $\lambda \geq 0$,

$$\log(1 + \rho_{\text{eff}} \lambda) - \frac{\rho_{\text{eff}} \lambda}{1 + \rho_{\text{eff}} \lambda} \geq 0. \quad (38)$$

Thus, if $T_d \zeta \leq 1$, then (37) is satisfied. Expanding the inequality $T_d \zeta \leq 1$, we obtain

$$\underbrace{\alpha}_{1} \underbrace{(\rho T \beta)(1-\kappa)}_{2} < \underbrace{1}_{1} \underbrace{[\rho T \beta (1-\kappa) + M]}_{2}, \quad (39)$$

where under-braces separate each side into two factors. Comparing the first factor from each side, we observe $\alpha \leq 1$; comparing the second factor from each side we find

$$\rho T \beta (1-\kappa) < \rho T \beta (1-\kappa) + M. \quad (40)$$

Multiplying these two inequalities we obtain (39), and the inequality is satisfied: $\frac{dC_t}{dT_d} \geq 0$. Thus, T_d should be made as large as possible, or T_τ as small as possible, which means $T_\tau = M$. ■

A similar analysis for $T_\tau \leq M$ is difficult. Since T_τ must be an integer and M is typically not large, a direct numerical maximization using

$$\hat{T}_\tau = \arg \max_{T_\tau} C_t(\rho_{\text{eff}}(\hat{\kappa})) \quad (41)$$

is sufficient. An extreme case occurs when the capacity is dominated by the specular part of the channel, and it is best to not train, even though (28) may indicate that training should occur. In this situation all power is allocated to data transmission ($\kappa = 1$), the training length is set to $T_\tau = 0$, and there is no training signal used.

The model presented in Section II allows the determination of the best training frequency which depends on the power, number of transmit antennas, specular parameter, and autocorrelation function of the channel coefficients:

$$\hat{T} = \arg \max_T \frac{T-M}{T} E \left[\log \left| I_N + \rho_{\text{eff}} \frac{\bar{H}^H \bar{H}}{M} \right| \right]. \quad (42)$$

Using a high-SNR approximation for the capacity, analytic values for \hat{T} can be found in the case with no specular channel [13] (see also Section III-E), but no similar approximation is currently available for the general channel.

E. Diffuse Channel

We now specify several results for the case where no specular component of the channel is present ($\beta = 1$). Many results that are difficult to obtain for the general channel are tractable for this case. We begin by optimizing the length of the training signal for the case of $T_\tau \leq M$. Using techniques similar to the general case, we find that making T_d as small as possible, or T_τ as large as possible ($T_\tau = M$), maximizes the lower bound C_t . For $T_\tau \geq M$, the general analysis shows that T_τ should be made as small as possible. Thus, the best amount of training for the model of Section II is $T_\tau = M$, regardless of the speed of the channel, or training frequency.

The optimal power allocation may be obtained as a special case of Theorem 2.

We now turn to determination of the best training frequency, which is enabled by our model, in contrast to that of [1]. The model of the time variation of the channel in Section II shows the dependence of the training frequency on the power, number of transmit antennas, and autocorrelation function of the channel coefficients:

$$\hat{T} = \arg \max_T C_t(T, M, N, T_\tau, \rho, \kappa, \alpha_T). \quad (43)$$

An analytic solution of (43) for an arbitrary autocorrelation function and SNR is difficult to obtain. We find an approximation for the specific case of the Jakes/Clarke model [11] in the high-SNR case; similar approximations can be made for other autocorrelation functions. A high-SNR, large- M approximation to the capacity of a non-coherent system [6],

$$C \approx M \left(1 - \frac{M}{T}\right) \log \left(\frac{\rho}{e}\right), \quad (44)$$

is valid for $M \leq N$ and $M \leq T/2$. The term $M(1 - \frac{M}{T})$ can be considered to be the number of degrees of freedom for the training-based scheme [6]. This enables numerical estimates of \hat{T} using the effective SNR:

$$\hat{T} = \arg \max_T \left(1 - \frac{M}{T}\right) M \log \left(1 + \frac{\rho_{\text{eff}}}{e}\right), \quad (45)$$

where the addition of 1 in the argument of the log function recovers some of the low-SNR behavior. In Section V we compare the result of this equation with our high-SNR value, which we derive next.

The high- ρ effective SNR is:

$$\hat{\rho}_{\text{eff}} \triangleq \lim_{\rho \rightarrow \infty} \rho_{\text{eff}} = \frac{\alpha_T}{1 - \alpha_T}, \quad (46)$$

which is valid for $\rho \gg \alpha_T/(1 - \alpha_T)$. Using (12) for the Jakes/Clarke model, this high-SNR value may be approximated further for small values of fT by

$$\hat{\rho}_{\text{eff}} \approx \frac{2}{(fT\pi)^2} \triangleq \tilde{\rho}_{\text{eff}}. \quad (47)$$

Typical Doppler values satisfy $fT \leq .1$, in which case

$$\frac{\tilde{\rho}_{\text{eff}} - \hat{\rho}_{\text{eff}}}{\hat{\rho}_{\text{eff}}} \leq 0.032, \quad (48)$$

indicating the accuracy of the approximation in (47).

Using the high-SNR approximation (47) in place of ρ in (44) we obtain

$$C_t \approx \left(1 - \frac{M}{T}\right) M \log_2 \left(\frac{2}{e(fT\pi)^2}\right) \triangleq C_a. \quad (49)$$

Differentiating (49) leads to

$$\frac{dC_a}{dT} = \frac{M^2}{T^2} \log_2 \left(\frac{2}{e(fT\pi)^2}\right) - 2 \left(1 - \frac{M}{T}\right) \frac{M}{T} \log_2(e). \quad (50)$$

We maximize C_a by setting this expression to zero and solving for T . This yields:

$$\frac{\sqrt{2e}}{f\pi M} = \frac{T}{M} e^{\frac{T}{M}}. \quad (51)$$

The principal branch of Lambert's W-function [14] for any $x \in \mathcal{C}$ satisfies $w = W_0(x)$ when $we^w = x$. Identifying $x = \sqrt{2e}/(f\pi M)$ and $w = T/M$, we use this function to obtain an expression for the best training period \hat{T} to use at high SNR:

$$\hat{T} = \text{round} \left[MW_0 \left(\frac{\sqrt{2e}}{\pi M f} \right) \right], \quad (52)$$

This approximation is restricted to the case of $\rho \gg \alpha_T/(1 - \alpha_T)$, large M , and $Tf \leq 1/10$.

In the discussion above, we have assumed that the number of transmit antennas M is not a design parameter, but rather fixed by cost or other physical requirements. When a system designer is free to choose M , setting $M = T/2$ maximizes the high-SNR capacity [1], [6], and results in $M/2$ degrees of freedom. Using this optimal value $M = T/2$, the same number of receive as transmit antennas, a training signal length $T_\tau = M$, and the optimal power allocation, we can apply similar approximations as above to determine the best training period T . This high-SNR value depends only on the fading parameter f :

$$\hat{T} = \text{round} \left[\sqrt{\frac{2}{e^3}} \frac{1}{\pi f} \right]. \quad (53)$$

This expression results in $f\hat{T} \approx .1$, satisfying the limitations on fT resulting from (48).

IV. DIFFERENTIAL MODULATION

Differential modulation [3], [4] assumes a channel that is constant over each pair of consecutive square symbols; data at the current and previous time instants are used for decoding rather than explicit information about the channel (as in most work on differential techniques, we focus on a diffuse channel where $\beta = 1$). In this "pairwise-constant" channel, differential modulation suffers from an effective noise term that is twice that seen in a coherent system [4]. For the time-varying model described in Section II with no specular component, the equivalent effective SNR is found in [12] to be:

$$\rho_D = \frac{\rho \alpha_{2M}}{1 + \alpha_{2M} + (1 - \alpha_{2M})\rho} \leq \frac{\rho}{2}. \quad (54)$$

This equation describes the behavior of a time-varying system with a fixed autocorrelation function. Use of this formula in place of ρ in equations derived for the quasi-static model of the channel gives the effective behavior of a time-varying system, as described in [12].

Though a relationship to the effective SNR for training-based modulation is not immediately apparent when comparing (54) with (21), we note that at high SNR with $T = 2M$, the effective SNR for differential modulation in (54) is equal to that for trained modulation given in (46). In general, however, the effective SNR performance ceiling for differential modulation at high SNR will be higher than that for trained modulation [12], since usually $T > 2M$. Though one is tempted to conclude from this that differential is a better option than trained modulation, this is not the case when the capacity of the two schemes are compared.

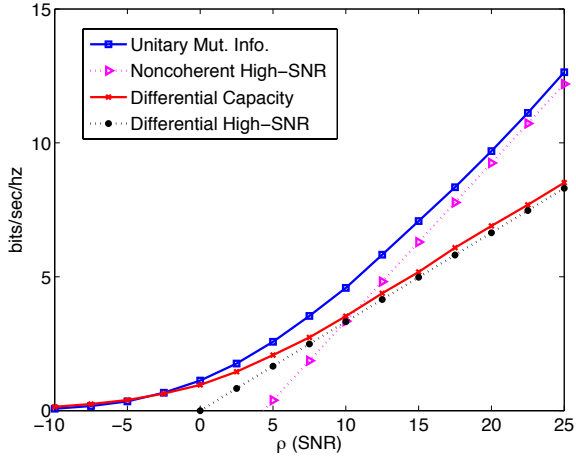


Fig. 1. The lower bound to the capacity of the noncoherent channel from [15] (indicated with squares), the high-SNR approximation (44) (shown with triangles), the capacity for differential modulation [5] (shown with 'x'), and the high-SNR approximation (56) (shown with 'o') are plotted as a function of the power constraint ρ . In these curves $T = 18$, $M = 2$. For ρ above 5dB, the approximation (56) is very close to the true capacity.

The achievable rate $C_D(M, N, \rho)$ for differential modulation has been found recently [5] for a quasi-static channel. Based on the analysis of [12], we assume that the rate for differential modulation with the time-varying model of Section II is given by the expression found in [5] with ρ replaced by the effective SNR from (54):

$$C_{Dt} = C_D(M, N, \rho_D). \quad (55)$$

The analytic expressions from [5] for C_D depend on a Monte-Carlo evaluation of a given expectation. For our analysis, we use a simple high-SNR approximation to the capacity which is valid for $M = N$:

$$\lim_{\rho \rightarrow \infty} C_D(M, M, \rho) \approx \frac{M}{2} \log_2(\rho). \quad (56)$$

In Figure 1 we compare this equation with the capacity for differential modulation from [5]. For SNR values above 5dB, the approximation (56) is close to the true differential capacity from [5]. For comparison, the mutual information for noncoherent signaling with unitary signals from [15], which is tight lower bound to the noncoherent capacity at high SNR, and the approximation (44) are also shown.

In contrast to trained modulation, for differential modulation there is only one design parameter that we are concerned with: the number of transmit antennas to use; in many cases even this will not be under the designer's control. For differential modulation with a large number of transmit antennas M , a symbol (which has length M) is long enough that the channel may have changed significantly from the previous symbol. Thus, there is a trade-off: does a system designer use many antennas in order to (hopefully) obtain a high data rate, or fewer antennas so that the channel more accurately obeys the quasi-static model? A similar analysis to that given in Section III-E results in the following approximation for the number of antennas to employ at high SNR when using differ-

ential modulation over the time-varying channel of Section II:

$$\hat{M} = \text{round} \left[\frac{1}{\sqrt{2e\pi f}} \right]. \quad (57)$$

This expression is very similar to that obtained in (53) for trained modulation when the number of transmit antennas is assumed to be $M = T/2$, and $N = T_\tau = M$. This correspondence follows from the close relationship between the two methods. When training with $T_\tau = M = T/2$, half of the interval T is used in training, with the other half used for information transmission. Though differential transmission with the same number of antennas sends twice as many information-bearing symbols in the same time, each has half the degrees of freedom of the arbitrary space-time symbols possible in the trained modulation case. In this case, the high-SNR error floor for differential and trained modulation is also the same, and the high-SNR approximations (44) and (56) are the same to within a factor of e . In many practical systems, however, cost and other physical factors will result in $T \gg 2M$, and trained modulation will have a higher achievable rate than differential modulation.

V. NUMERICAL EXAMPLES

Numerical results are now presented along with comparisons to the analysis above. Each simulation point in the following plots is the result of a Monte-Carlo evaluation based on 1000 channel realizations. We will make comparisons to the capacity for a system with M transmit antennas, N receive antennas, and a channel H known at the receiver but not the transmitter:

$$C_{\text{coh}}(\rho) = E \left[\log_2 \left| I_N + \frac{\rho}{M} H^H H \right| \right]. \quad (58)$$

Figure 2 illustrates the lower bound (16) as a function of the power allocation factor κ for $\rho = 0$ dB SNR, $M = N = 2$ antennas, training length $T_\tau = 2$, normalized Doppler frequency $f = 0.003$, training period $T = 5$, and specular parameter $\beta = 1$. Theorem 2 indicates that $\kappa = .53$ maximizes the capacity bound, which agrees well with the simulation. For purposes of comparison, capacity is shown for a system with perfect channel state information at the receiver (58), and for differential modulation with a quasi-static model [5] and with the model of (55).

Figure 3 illustrates the bound (16) as a function of the training length T_τ for several values of the specular parameter β . A training period of $T = 8$ was used with $M = N = 3$ antennas, an SNR of 3 dB, and a fading parameter $f = .003$. As our analysis indicates, it is best to set $T_\tau = M$ for a Rayleigh channel ($\beta = 1$), but higher capacity may be achieved with less training for specular channels. Figure 3 shows that a training length of $T_\tau = 1$ is optimal for $\beta = 0.45$, while using no training at all is best for $\beta \leq 0$.

Figure 4 shows the capacity bound as a function of β for three different choices of T_τ . Equation (41) was used to find the best joint values for T_τ and κ . The performance with these optimal values is shown with the solid line and is compared to the result when using $T_\tau = M = 3$, $\kappa = 0.5$, and $T_\tau = 0$. For this example, the following parameters were assumed:

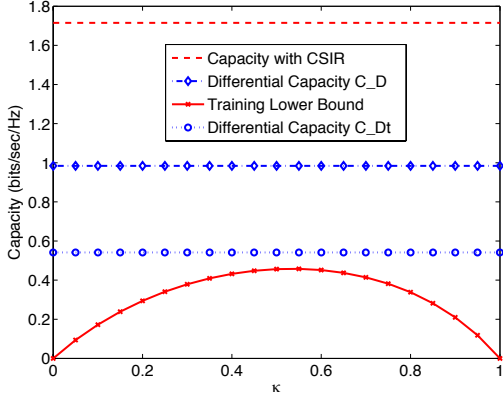


Fig. 2. Plot of the capacity lower bound (16) as a function of the allocation of power between training and data communication is shown with 'x'. The coherent capacity (58) (dashed line), differential capacity from [5] (diamonds), and the approximation (55) (shown with 'o') are also shown for comparison. Theorem 2 indicates that the optimal power allocation is $\kappa = 0.53$, which agrees with the simulation.

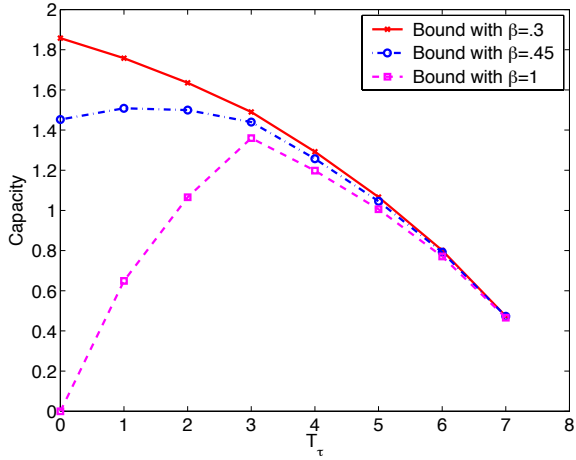


Fig. 3. The lower bound (16) is shown as a function of T_τ for several values of β . For a fully Rayleigh channel ($\beta = 1$), it is best to use $T_\tau = M$. For the intermediate value of $\beta = .45$ the best training length is 1, while for $\beta = .3$ it is best to do no training at all.

$M = N = 3$ antennas, training period $T = 6$, fading speed $f_d = 0.01$ and SNR $\rho = 3$ dB. At low β values it is best to do no training at all, while for high β , the optimal solution is $T_\tau = M$, which agrees with results from (41).

Figure 5 shows the capacity bound (16) as a function of T for $\rho = \{20, 25, 40\}$ dB when $\beta = 1$. The channel is time-varying with normalized Doppler frequency $f = 0.003$, and $M = N = T_\tau = 2$. Numerical evaluation of (45) predicts that the best values of T are $\hat{T} = \{13, 11, 7\}$ for $\rho = \{20, 25, 40\}$ dB respectively, which agree well with simulation as indicated by the vertical lines in the figure. The predicted high-SNR optimal training interval given by (52) is 7, which agrees well with the 40 dB curve.

In Figure 6 we compare differential modulation with trained modulation where training occurs every $T = 2M$ samples. The quasi-static capacity (58), the lower bound (16), the quasi-

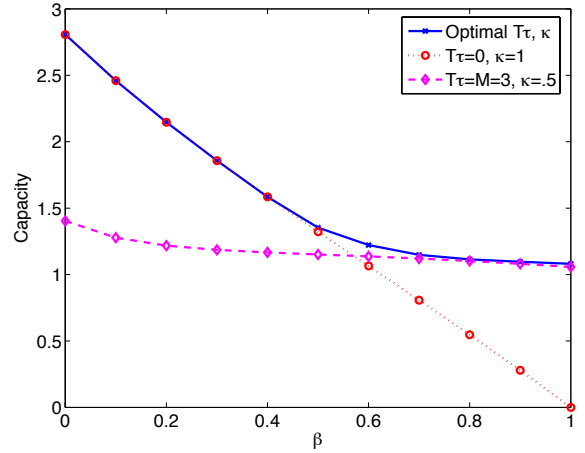


Fig. 4. The bound (16) as a function of β for optimal T_τ (shown with solid lines), for $T_\tau = M = 3$ (shown with diamonds), and for no training (shown with 'o's). At low β values the optimal $T_\tau = 0$, while for high β , the best training length is $T_\tau = M$. The optimal T_τ from (41) gives the highest throughput.

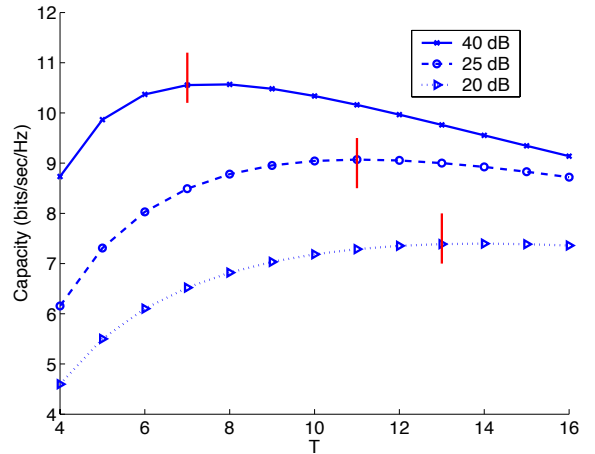


Fig. 5. The lower bound (16) for SNR values of 40, 25, and 20 dB. These curves agree well with (45) which indicates that the best T values are $\hat{T} = \{13, 11, 7\}$. The best training frequency from the high-SNR expression (52) is 7 which agrees well with the curve for $\rho = 40$ dB.

static capacity of differential modulation [5], and the approximation (55) to the capacity of differential modulation with the model of Section II are shown as a function of the number of transmit antennas. Other parameters used include setting the training period T_τ and number of receive antennas to M , SNR of 20dB, a fading speed of $f = .003$, and a fully diffuse channel ($\beta = 1$). Equation (57) predicts that the best number of transmit antennas to use is $M = 8$, while (53) indicates that $M = 10$ is best for trained modulation; the numerical results shown in the figure agree with these theoretical values.

VI. DISCUSSION

We have presented a new lower bound on the capacity of trained modulation, extending the results of [1] to the case of a generalized channel having a fixed specular component and a time-varying diffuse component with known time-autocorrelation function. We optimized this lower bound over

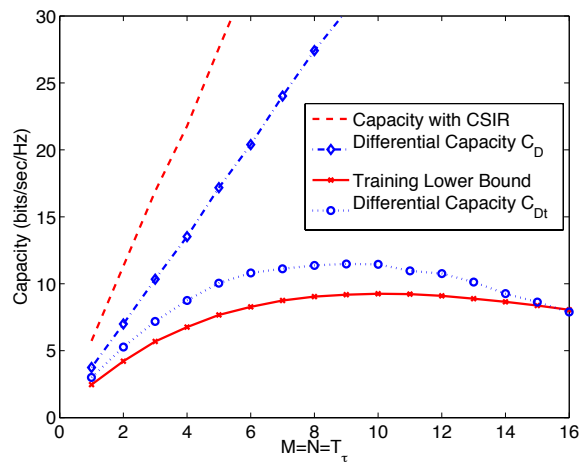


Fig. 6. Plot of the quasi-static capacity (58) (the dashed line), the lower bound (16) (shown with 'x'), the quasi-static capacity of differential modulation capacity [5] (shown with diamonds), and the approximation (55) (shown with circles) to the capacity of differential modulation as a function of M .

the training signal, training length, training frequency, and power allocation between training and data transmission. The bound provides practical insight for the design of multiple-antenna communications systems with time-varying channels.

We compared our bound with the capacity of differential modulation over a quasi-static channel, and noted that for slow channels with $T \gg M$, trained modulation does better, while for large f_d and $T = 2M$, differential coding performs best. We noted that though the effective SNR for differential is greater at high SNR than for trained modulation, its capacity is often lower. We derived a high-SNR expression for the number of antennas that maximizes capacity, and found it to be similar to that for trained modulation when $M = T/2$ antennas are used.

An open question is how the comparison between training and differential techniques would change with channel tracking. Multiple-symbol differential techniques [16] are known to eliminate both the high-SNR performance ceiling and the 3 dB SNR penalty, though they would not overcome the inherent capacity penalty due to the use of unitary signals. While comparisons would be algorithm-dependent, coherent schemes that track the channel such as [17], [18] would be expected to maintain a capacity advantage over differential techniques. In addition, though we have assumed a simple time-autocorrelation function, analysis with more physically realistic channel models [19] which include joint space-time correlation would be useful.

REFERENCES

- [1] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 1, pp. 951–963, April 2003.
- [2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, May 1999.
- [3] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, November 2000.

- [4] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2041–2052, March 2000.
- [5] A. L. Moustakas, S. H. Simon, and T. Marzetta, "Capacity of differential versus nondifferential unitary space-time modulation for mimo channels," *IEEE Trans. Information Theory*, vol. 52, no. 8, pp. 3622–3634, Aug. 2006.
- [6] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multi-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.
- [7] M. Godavarti, T. L. Marzetta, and S. Shamai (Shitz), "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3330–3334, Dec 2003.
- [8] M. Godavarti, "Multiple-antennas in wireless communications: Array signal processing and channel capacity," Ph.D. dissertation, University of Michigan, 2001.
- [9] C. B. Peel and A. L. Swindlehurst, "Performance of space-time modulation for a general time-varying Rician channel model," *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 1003–1012, May 2004.
- [10] M. Godavarti, T. L. Marzetta, and S. Shamai (Shitz), "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," in *Proceedings of the 2001 IEEE International Symposium on Information Theory*, June 2001.
- [11] W. C. Jakes, *Microwave Mobile Communications*. IEEE Press, 1993.
- [12] C. B. Peel and A. L. Swindlehurst, "Effective SNR for Space-Time Modulation over a Time-Varying Rician Channel," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 17–23, Jan 2004.
- [13] —, "Capacity-optimal training for space-time modulation in time varying fading," in *The 2003 International Conference on Communications*, vol. 4, May 2003, pp. 2698–2702.
- [14] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.
- [15] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1473–1484, June 2002.
- [16] R. Schober and L. H.-J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 768–777, May 2002.
- [17] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 183–186, February 2002.
- [18] A. Duel-Hallen, S. Hu, and H. Hallen, "Long range prediction of fading signals," *IEEE Signal Processing Magazine*, pp. 62–75, May 2000.
- [19] T.-A. Chen, M. P. Fitz, W.-Y. Kuo, M. D. Zoltowski, and J. H. Grimm, "A space-time model for frequency nonselective rayleigh fading channels with applications to space-time modems," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1175 – 1190, July 2000.