Noncoherent Modulation for MIMO Training and Capacity Analysis

Yingqun Yu
Department of Electrical and Computer Engineering
University of Minnesota, Minneapolis, MN 55455; USA
Tel: 612-626-7174; Email: yingqun@umn.edu

Abstract

In this paper, we deal with noncoherent communications over multi-input/multi-output (MIMO) wireless links where the fading coefficients are not available to either the transmitter or the receiver. For a Rayleigh flat block-fading channel with M transmit and N receive antennas and a channel coherence interval of length T, it is well-known that for $T \gg M$, or, at high signal-to-noise-ratio (SNR) ($\rho \gg 1, M \leq \min\{N, \lfloor T/2 \rfloor\}$), the random unitary space-time modulation (USTM) is capacity-achieving, but incurs demodulation complexity that is exponential in T. On the other hand, conventional training-based schemes that rely on known pilot symbols for channel estimation simplify the receiver design, but they induce certain SNR loss due to channel estimation errors, which is asymptotically as high as 2.17dB. To achieve desirable tradeoffs between performance and complexity, we propose a novel training approach where USTM symbols with a short length T_τ ($< T$) are used as pilots which can carry information to the receiver. This new approach can reduce considerably the receiver complexity since T_τ is usually a fraction of T, and it can also recover some of the SNR loss experienced by the conventional training-based schemes. When $\rho \to \infty$ and $T \geq T_\tau \geq 2M = 2N \to \infty$, but the ratios $\alpha = M/T, \alpha_1 = T_\tau/T$ are fixed, we obtain analytical expressions of the asymptotic SNR loss for both the conventional and new training-based approaches, which serve as a guideline for practical designs. Numerical results are also given, to verify the usefulness of our asymptotic results.

Index Terms

Capacity, multiple antennas, coherent detection, noncoherent detection, unitary space-time modulation, channel estimation

I. Introduction

Due to fading of the channel strength caused by constructive and destructive interference of the multiple signal paths between the transmitter and the receiver, a major challenge in wireless communications is coping with channel uncertainties. Pilot symbol assisted modulation (PSAM) is a standard training-based approach for communications over time-varying channels [1], [2], [3]. In PSAM, pilot symbols *known* to both the transmitter and the receiver are multiplexed with data symbols and used as training for channel acquisition. Since known pilot symbols carry no data information, they reduce power and bandwidth resources during data transmission. Clearly, there is a tradeoff in allocating these resources between pilot symbols and data symbols. Sending more pilots with increased power improves the quality of channel estimation as well as the reliability of communication. However, over-increasing the overhead for training reduces the amount of channel uses and power for information-carrying data symbols, resulting in decreased data throughput.

An basic information-theoretic question for PSAM is how much training is necessary when using Shannon's capacity as the performance metric. For a given channel estimation accuracy, lower bounds on channel capacity were found for a general setting [4], for Rayleigh flat block-fading multiple-input/multiple-output (MIMO) wireless channels [5], [6], and for perfectly interleaved MIMO channels [7]. The optimal power allocation between the pilot and data symbols as well as the number (equal to M) of training symbols optimizing a lower bound on capacity were obtained in [6]. Similar lower bounds are also available for single-antenna and multiple-antenna frequency-selective fading channels, based on which the optimal training design has been derived [8], [9], [10].

Although training-based schemes like PSAM simplify transceiver design for noncoherent multiple-antenna systems, information-theoretic studies have revealed that in general they are not capacity-achieving. Marzetta and Hochwald [11] investigated the capacity of a Rayleigh flat block-fading channel with M transmit, N receive antennas and a channel coherence interval of length T, and found that the *noncoherent* channel capacity is achieved when the $T \times M$ transmitted signal matrix is expressible as a product of two statistically independent matrices: a $T \times M$ isotropically distributed (*i.d.*) unitary matrix times a real, diagonal and nonnegative $M \times M$ random matrix. The asymptotic capacity for such a channel at high SNR can be achieved using only M^* antennas, and increases linearly with $M^*(1 - M^*/T)$, where $M^* = \min\{M, N, \lfloor T/2 \rfloor\}$ [12]. In comparison, for the same channel model except that the receiver knows the channel coefficients perfectly, it is well-known that the ergodic *coherent* channel capacity increases linearly with $\min\{M, N\}$ [13], [14].

Motivated by results in [11], a class of isotropic unitary space-time modulation (USTM) signals was proposed in [15], [16] to encode the transmitted signals using $T \times M$ isotropic unitary matrices. For $T \gg M$ [11], and for high SNR $\rho \gg 1$ with $M \le \min\{N, \lfloor T/2 \rfloor\}$ [12], the optimal input has indeed a USTM form. The main drawback of USTM signals is that they are typically designed based on numerical optimizations [16], [17], [18], and because they possess no particular algebraic structure, they require relatively high complexity. More problematically, their demodulations incurs exponential complexity since the constellation size grows exponentially with the block length T (the number of signal points is 2^{RT} for a given rate of R bits per symbol). For this reason, USTM is practically applicable only for small block lengths or low rates. Differential USTMs [19], [20] and alternative training-like constellations [21], [22] that can be explained as training codes [23] enjoy polynomial complexity in T; however, they are generally not capacity-achieving for the block-fading channel.

Compared with the case when USTM is optimal, a training-based scheme suffers SNR degradation due to imperfect channel state information (CSI), but gains the benefit of simplified receiver design. If T_τ out of T symbols in a fading block are used to send known training symbols for channel estimation, it has been shown that at high SNR only M^* antennas should used for transmission, and the achievable rate also increases linearly with $M^*(1 - M^*/T)$ similar to the noncoherent case; however, due to channel estimation errors there is a SNR loss compared with the optimal noncoherent scheme [12], [6]. A training-based scheme can be capacity-achieving only when T is sufficiently large, but the rate at which it attains this optimality as T grows has not been quantified yet.

In this paper, we first analytically compute the asymptotic SNR loss for the conventional training-based methods when $\rho \to \infty$ and $T \ge 2M = 2N \to \infty$, but the ratio $\alpha = M/T$ is fixed. We show that as α decreases, the asymptotic SNR loss drops monotonically but also slowly from 2.17dB ($\alpha = 0.5$) to zero ($\alpha \to 0$). Further, we introduce a novel scheme that combines noncoherent and coherent detection for the block fading channel, and thus is useful in trading off performance for complexity. A channel coherence interval T is divided into two parts, the noncoherent part with T_τ symbols and the coherent part with $T_d (= T - T_\tau)$ symbols. The noncoherent symbols carry information unknown to the receiver and are encoded over multiple fading blocks. A key observation is that after those T_τ noncoherent symbols are correctly decoded without CSI, they can be further used to estimate the channel coefficients in their own block, thus enabling the coherent detection of the remaining T_d coherent symbols. There are three advantages of the proposed scheme. First, unlike conventional training where the pilots are known sequences only for the purpose of channel estimation and are incapable of carrying data information, here those noncoherent symbols do carry information. Second, since T_τ is only a small fraction of T, the cardinality of the noncoherent constellation is reduced considerably, leading to low decoding complexity. Finally, one is flexible to control the tradeoff between complexity and SNR loss by selecting a suitable T_τ.

The rest of the paper is organized as follows. In Section II, we introduce the system model and provide some preliminary results. In Section III, we dwell on the training-based scheme and compare it with USTM. In Section IV, we introduce and analyze the novel noncoherent/coherent scheme. Numerical examples are given in Section V, and conclusions are drawn in Section VI.

II. System Model and Preliminaries

A. System Model

We consider a single-user transmission with M transmit and N receive antennas over a frequency-nonselective (flat) Rayleigh block-fading channel, as in [11]. The channel coefficients, which are unknown to both the transmitter and the receiver, are assumed to remain constant over a block of T symbols, but are allowed to change independently from block to block. Within a block of T symbols, given that a signal matrix $\mathbf{\Phi} \in \mathcal{C}^{T \times M}$ is transmitted[1], the received

[1] Here $\mathcal{C}$ denotes the complex field.

signal matrix $\boldsymbol{X} \in \mathcal{C}^{T\times N}$ can be written as

$$\boldsymbol{X} = \sqrt{\frac{\rho T}{M}}\boldsymbol{\Phi}\boldsymbol{H} + \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{H} \in \mathcal{C}^{M\times N}$ is the channel matrix, and $\boldsymbol{W} \in \mathcal{C}^{T\times N}$ is the additive noise matrix. Both $\boldsymbol{H}$ and $\boldsymbol{W}$ are complex Guassian matrices with independent and identically distributed (*i.i.d.*) $\mathcal{CN}(0,1)$ entries. The power constraint on the transmitted signal is assumed to be $\mathbb{E}\{\text{Tr }(\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi})\} = M$, and thus ρ is the average received SNR at each receive antenna since $\boldsymbol{\Phi}, \boldsymbol{H}$ and $\boldsymbol{W}$ are independent. Because the receiver does not know the channel matrix $\boldsymbol{H}$, the model in (1) is often referred to as a *noncoherent* channel; otherwise, it is called a *coherent* channel.

B. Known Results on Coherent Capacity, Noncoherent Capacity and Mutual Information of USTM

When perfect knowledge of the channel coefficients is available at the receiver (but not at the transmitter), the channel capacity, often called *coherent capacity*, is computed in [24], [14] and is summarized in the following lemma.

Lemma 1: If $\boldsymbol{H}$ is known to the receiver, the coherent capacity in bits per symbol is given by

$$C_{coherent}(\rho) = \mathbb{E}\{\log_2\det(\boldsymbol{I}_M + \frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^\dagger)\} = \mathbb{E}\{\log_2\det(\boldsymbol{I}_N + \frac{\rho}{M}\boldsymbol{H}^\dagger\boldsymbol{H})\}. \tag{2}$$

When $M = N$, the normalized asymptotic capacity for high SNR and large M satisfies

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{C_{coherent}(\rho)}{M} - \log_2(\frac{\rho}{e})\right] = 0. \tag{3}$$

For the noncoherent channel model described by (1), it has been shown that at high SNR the degrees of freedom per symbol for each noncoherent block is $M^*(1 - M^*/T)$, where $M^* = \min\{M, N, \lfloor T/2\rfloor\}$ [12]. This result indicates that at high SNR, the optimal strategy is to use only M^* out of M available antennas. The capacity-achieving input matrix can be written as $\boldsymbol{\Phi} = \boldsymbol{\Theta}\boldsymbol{D}$, where $\boldsymbol{\Theta}$ is a $T \times M$ isotropically distributed (i.d.) unitary matrix, i.e., $\boldsymbol{\Theta}^\dagger\boldsymbol{\Theta} = \boldsymbol{I}_M$, and $\boldsymbol{D}$ is an $M \times M$ random real nonnegative diagonal matrix with $\mathbb{E}\{\text{Tr }(\boldsymbol{D})\} = 1$ [11]. The distribution of $\boldsymbol{D}$ is generally unknown, except for the asymptotic case $T \gg M$ [11] and for high SNR with $M \le \min\{N, T/2\}$ [12], where $\boldsymbol{D}$ becomes a deterministic identity matrix $\boldsymbol{D} = \sqrt{1/T}\boldsymbol{I}_M$, suggesting the so-called USTM inputs for noncoherent channels [15]. The result is summarized in the following lemma for the case $T \ge 2M = 2N$ [12, Theorem 9, Corollary 11].

Lemma 2: Assume $T \ge 2M = 2N$. If $T \gg M$ and/or $\rho \gg 1$, the unitary space-time modulation with $\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi} = \boldsymbol{I}_M$ achieves the noncoherent channel capacity of (1). In particular, for $\rho \gg 1$ the capacity is given by

$$C_{M,M}(\rho) = (1 - \frac{M}{T})C_{coherent}(\rho) + c(T, M) + o(1), \tag{4}$$

where $c(T, M)$ is a constant that depends only on M and T, and goes to zero as $T \to \infty$; and $o(1)$ is a term that goes to zero as $\rho \to \infty$. If we let both T and M go to infinity but keep the ratio $\alpha = M/T$ fixed, then we have [2]

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{C_{M,M}(\rho)}{M} - \left(\frac{k(\alpha)}{\ln 2} + (1-\alpha)\log_2(\frac{\rho}{e})\right)\right] = 0, \tag{5}$$

where

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{c(T,M)}{M} - \frac{k(\alpha)}{\ln 2}\right] = 0, \text{ and } k(\alpha) = \frac{(1-\alpha)^2}{2\alpha}\ln(1-\alpha) + \frac{\alpha}{2}\ln\alpha + \frac{1-\alpha}{2} < 0 \tag{6}$$

for all $0 < \alpha \le 1/2$.

III. Training with Known Pilot Symbols versus USTM

In this section, we introduce the conventional training-based scheme with known pilot symbols and compare it with USTM. Based on a lower bound C^L_{known} of the training-based schemes, we compare the asymptotic behavior of the two options when both M and T go to infinity, but their ratio $\alpha = M/T$ remains fixed.

[2]We believe that in eq. (22) of [12] the term $\log_2 e$ should be $\ln 2$.

A. Training-Based Schemes

In a typical training-based system, the transmitted signal matrix $\mathbf{\Phi}$ is partitioned into a training submatrix $\mathbf{\Phi}_\tau$ and a data submatrix $\mathbf{\Phi}_d$ as follows

$$\mathbf{\Phi} = \begin{pmatrix} \sqrt{\frac{\rho_\tau T_\tau}{M}}\mathbf{\Phi}_\tau \\ \sqrt{\frac{\rho_d T_d}{M}}\mathbf{\Phi}_d \end{pmatrix}, \tag{7}$$

where $\mathbf{\Phi}_\tau \in \mathcal{C}^{T_\tau \times M}$ with $T_\tau \geq M, \text{Tr }(\mathbf{\Phi}_\tau^\dagger \mathbf{\Phi}_\tau) = M$ is the training matrix known to both the transmitter and the receiver, and $\mathbf{\Phi}_d \in \mathcal{C}^{T_d \times M}$ with $\mathbb{E}\{\text{Tr }(\mathbf{\Phi}_d^\dagger \mathbf{\Phi}_d)\} = M$ is the data matrix carrying information from the transmitter to the receiver. The parameters ρ_τ and ρ_d are the SNR values during the training phase and the data transmission phase, respectively. In addition, we have the equations of time and energy conservation: $T_\tau + T_d = T$, and $\rho T = \rho_\tau T_\tau + \rho_d T_d$.

Similarly, the received signal matrix $\boldsymbol{X}$ and the noise matrix $\boldsymbol{W}$ are also partitioned into two submatrices

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_\tau \\ \boldsymbol{X}_d \end{pmatrix}, \text{ and } \boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_\tau \\ \boldsymbol{W}_d \end{pmatrix}, \tag{8}$$

where $\boldsymbol{X}_\tau \in \mathcal{C}^{T_\tau \times N}, \boldsymbol{X}_d \in \mathcal{C}^{T_d \times N} (\boldsymbol{W}_\tau \in \mathcal{C}^{T_\tau \times N}, \boldsymbol{W}_d \in \mathcal{C}^{T_d \times N})$ are the received signal (noise) matrices during the training phase and the data transmission phase, respectively. We can thus write the signal model for the training phase as

$$\boldsymbol{X}_\tau = \sqrt{\frac{\rho_\tau T_\tau}{M}}\mathbf{\Phi}_\tau \boldsymbol{H} + \boldsymbol{W}_\tau, \tag{9}$$

and for the data phase as

$$\boldsymbol{X}_d = \sqrt{\frac{\rho_d T_d}{M}}\mathbf{\Phi}_d \boldsymbol{H} + \boldsymbol{W}_d. \tag{10}$$

The capacity in bits per symbol for the training-based scheme is given by [6]

$$C_{known} = \sup_{\mathbf{\Phi}_\tau, p(\mathbf{\Phi}_d)} \frac{1}{T} I(\mathbf{\Phi}_\tau, \mathbf{\Phi}_d; \boldsymbol{X}_\tau, \mathbf{\Phi}_d) = \sup_{\mathbf{\Phi}_\tau, p(\mathbf{\Phi}_d)} \frac{1}{T} I(\mathbf{\Phi}_d; \boldsymbol{X}_d | \mathbf{\Phi}_\tau, \boldsymbol{X}_\tau), \tag{11}$$

since $\mathbf{\Phi}_\tau$ is known to both the transmitter and the receiver and $\mathbf{\Phi}_d$ is independent of $\mathbf{\Phi}_\tau$ and $\boldsymbol{X}_\tau$. The optimization in (11) is performed over all choices of the deterministic training matrix $\mathbf{\Phi}_\tau$ and the input distributions $p(\mathbf{\Phi}_d)$ of the data matrix $\mathbf{\Phi}_\tau$ under the constraints that $\text{Tr }(\mathbf{\Phi}_\tau^\dagger \mathbf{\Phi}_\tau) = M$ and $\mathbb{E}\{\text{Tr }(\mathbf{\Phi}_d^\dagger \mathbf{\Phi}_d)\} = M$. However, such an optimization problem is very difficult to solve.

One option is to form an explicit channel estimate $\hat{\boldsymbol{H}}$ first and use it as if it were correct. In this process, information may be thrown away, which results in a suboptimal scheme. Nevertheless, this method enables us to compute a tight lower bound on channel capacity C_{known}. We first compute a minimum mean square error (MMSE) estimate of the channel matrix, and then absorb the estimation error to the additive noise to obtain an equivalent noise term. Further, this new noise term is replaced by a worst case noise, yielding a lower bound on mutual information. It has been shown in [6] that the training matrix with orthonormal columns $\mathbf{\Phi}_\tau^\dagger \mathbf{\Phi}_\tau = \boldsymbol{I}_M$ simultaneously maximizes this lower bound and minimizes MMSE. If optimal power allocation between the training phase and the data phase can be afforded, the optimal training interval length should be $T_\tau^{opt} = M$, and the corresponding lower bound is given by

$$C_{known}^L(\rho) = (1 - \frac{M}{T}) C_{coherent}(\rho_{eff}) \tag{12}$$

where

$$\rho_{eff} = \begin{cases} \frac{\rho T}{T-2M}(\sqrt{\gamma} - \sqrt{\gamma - 1})^2, & \text{for } T > 2M; \\ \frac{\rho^2}{1+2\rho}, & \text{for } T = 2M; \\ \frac{\rho T}{2M-T}(\sqrt{-\gamma} - \sqrt{-\gamma + 1})^2, & \text{for } T < 2M. \end{cases}$$
$$\gamma = \frac{(M + \rho T)(T - M)}{\rho T(T - 2M)}. \tag{13}$$

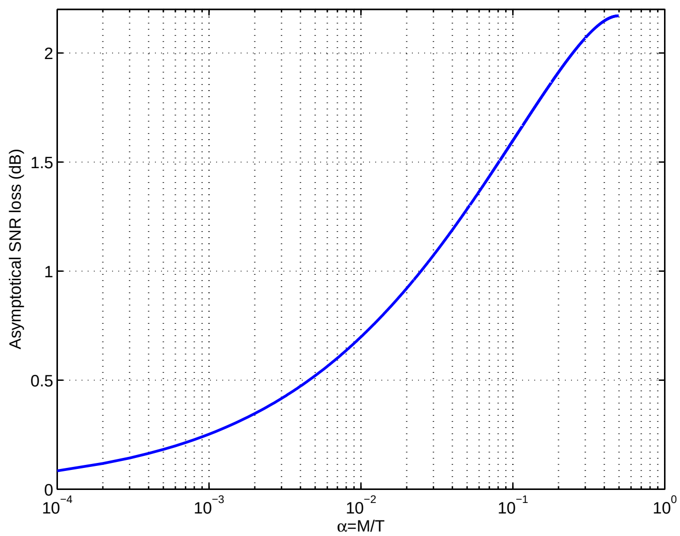

Fig. 1: Asymptotic SNR loss $\rho_{loss}(\alpha)$

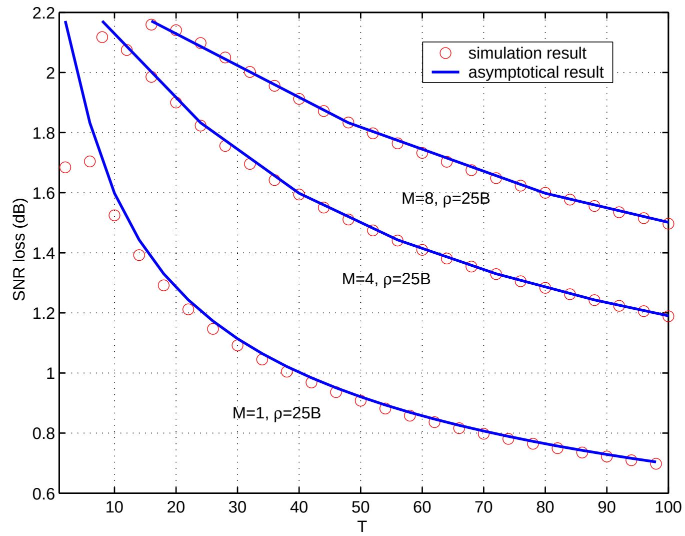

Fig. 2: SNR loss for finite ρ, M and T

Compared with the noncoherent capacity in (4) at high SNR for the case $T \geq 2M = 2N$, the traing-based scheme can achieve the same degrees of freedom $M(1 - M/T)$; however, it incurs an SNR degradation which depends on ρ, M and T. In Section III-B.1, an asymptotic expression for SNR loss will be found.

If equal training and data power $\rho = \rho_\tau = \rho_d$ is mandatory, then

$$C_{known}^L(\rho) = (1 - \frac{T_\tau}{T})C_{coherent}(\rho_{eff}), \tag{14}$$

where

$$\rho_{eff} = \frac{\rho^2 T_\tau / M}{1 + (1 + T_\tau/M)\rho}. \tag{15}$$

In this case, the optimal training interval can be found numerically.

B. Comparison between Conventional Training-Based Schemes and USTM

In this section, we compare the asymptotic behavior of the mutual information of USTM and training-based schemes with optimal power allocation, which motivates our training-based scheme with unknown symbols in Section IV.

1) High SNR: Consider for simplicity that $T \geq 2M = 2N$. From Lemma 2, we know that at high SNR, USTM inputs are capacity-achieving. Compared with the first term in (4), the SNR loss in C_{known}^L can be expressed as

$$\beta := \frac{\rho}{\rho_{eff}} = \frac{T - 2M}{T}(\sqrt{\gamma} - \sqrt{\gamma - 1})^{-2} = \frac{T - 2M}{T}(\sqrt{\gamma} + \sqrt{\gamma - 1})^2, \text{ for } T > 2M. \tag{16}$$

Since $\lim_{\rho\to\infty} \gamma = (T - M)/(T - 2M)$, substitution of the latter into β and simplification leads to

$$\beta_\infty := \lim_{\rho\to\infty} \beta = 1 + 2\sqrt{\frac{M}{T}(1 - \frac{M}{T})} \tag{17}$$

for $T \geq 2M$, since this expression also applies when $T = 2M$.

Note that β is not the real SNR loss at high SNR, since there is another term $c(T, M)$ in (4) that does not depend on ρ. To account for that, let us consider the case when both T and M go to infinity, but the ratio $\gamma = M/T$ is fixed. From Lemma 2, incorporating the term $k(\alpha)$ into $\log_2(\cdot)$ yields

$$\lim_{M\to\infty} \lim_{\rho\to\infty} \left[\frac{C_{M,M}(\rho)}{M} - (1 - \alpha)\log_2\left((\frac{\rho}{e}) \cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}}\right)\right] = 0. \tag{18}$$

Also from Lemma 1, we have a similar asymptotic result for C^L_{known}:

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{C^L_{known}(\rho)}{M}-(1-\alpha)\log_2\left((\frac{\rho}{e})\cdot\beta_\infty^{-1}\right)\right]=0. \tag{19}$$

Upon comparing (18) with (19), we obtain the following result.

Proposition 1: When $\rho\to\infty, M\to\infty$ and $T\to\infty$, but the ratio $\alpha=M/T\le 1/2$ is fixed, compared with the noncoherent capacity, C^L_{known} suffers an asymptotic SNR loss

$$\rho_{loss}(\alpha):=\beta_\infty\cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}}=\left[1+2\sqrt{\alpha(1-\alpha)}\right]\cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}}, \tag{20}$$

where $k(\cdot)$ is given in (6).

Corollary 1: It holds that $\rho_{loss}(0.5)=2.1715\text{dB}$, and $\rho_{loss}(0)=\lim_{\alpha\to 0}\rho(\alpha)=0\text{dB}$.

We plot $\rho_{loss}(\alpha)$ in Fig. 1 with α in logarithm scale. We observe that the SNR loss decreases monotonically to zero as $\alpha\to 0$, which is consistent with the intuition that for large T training can be optimal; however, the slope of decrease is very small. For example, the SNR loss is 1.5980dB at $\alpha=10^{-1}(T=10M)$, and drops to 0.6981dB at $\alpha=10^{-2}(T=100M)$. Even when $\alpha=10^{-3}(T=1000M)$, there is still 0.2523dB SNR loss. For the region $T=2\sim 20M$, there is always an SNR loss less than 2.17dB, but more than 1.5dB. Fig. 2 compares the asymptotic SNR loss with the case $\rho=25\text{dB}, M=1,4,8$ and $T=1\sim 100$. It can be seen that $\rho_{loss}(\alpha)$ is actually tight except when $M=1$ and $T<10$; therefore, $\rho_{loss}(\alpha)$ is a good approximation for practical scenarios.

Note that the interest in noncoherent channel models is often limited to the case when the channel varies quickly; that is, when T is small. If $T\gg M$, then we can allocate minimal overhead for channel estimation and feed channel estimates back to the transmitter, which leads to adaptive signaling designs with improved performance. When T is not very large or comparable to M, it is not efficient to sacrifice a significant portion (M out of T) of limited resources to training, and it is not suitable to feed the channel estimates back to the transmitter since the channel changes so quickly over time that it requires very fast feedback which induces considerable overhead. Under this scenario, a training-based scheme is not capacity-optimal, although it can achieve the same degrees of freedom as the optimal noncoherent scheme.

IV. TRAINING VIA INFORMATION-BEARING NONCOHERENT SPACE-TIME MODULATION

In this section, we present a new training-based scheme where "pilot" symbols, just like data symbols, can also carry information and thus are unknown to the receiver. Indeed, it does not make sense to estimate the channel when the receiver does not know the transmitted pilot symbols, unless some kind of blind channel estimation scheme is used. However, it is definitely legitimate to do that after the receiver successfully recovers them. The decoding of unknown pilot symbols, for sure, should not require CSI knowledge, which can be effected by using any noncoherent communication scheme.

A. Training via Noncoherent Communication

The proposed system architecture is shown in Fig. 3. Information data are first encoded and then sent to the coherent and noncoherent modulators, respectively. The modulator outputs $\mathbf{\Phi}_\tau$ and $\mathbf{\Phi}_d$ are multiplexed for transmission. The receiver first demultiplexes the received signal to obtain $\boldsymbol{X}_\tau$ and $\boldsymbol{X}_d$. $\boldsymbol{X}_\tau$ carries data and is decoded first. Since $\mathbf{\Phi}_\tau$ is noncoherently modulated, the receiver can decode it without knowing the fading coefficient matrix $\boldsymbol{H}$. Once the transmitted signal matrix is recovered as $\widehat{\mathbf{\Phi}}_\tau$ after decoding, the receiver can estimate the channel using $\widehat{\mathbf{\Phi}}_\tau$. The estimated channel $\hat{\boldsymbol{H}}$ is subsequently sent to the coherent detector to decode the information carried by $\boldsymbol{X}_d$.

The receiver structure depicted in Fig. 3 is suboptimal in general, since information may be lost because: i) $\mathbf{\Phi}_\tau$ and $\mathbf{\Phi}_d$ are decoded not jointly but separately; and ii) an explicit $\hat{\boldsymbol{H}}$ is formed and used as if it were correct. In the following, we will develop a lower bound on the channel capacity that favors this suboptimal receiver.

The capacity in bits per symbol for the new scheme is the maximum over the distribution of the transmit signals of the mutual information between the transmitted signals $\mathbf{\Phi}_\tau, \mathbf{\Phi}_d$ and the received signals $\boldsymbol{X}_\tau, \boldsymbol{X}_d$; i.e.,

$$C_{unknown}=\sup_{p(\mathbf{\Phi}_\tau,\mathbf{\Phi}_d)}\frac{1}{T}I(\mathbf{\Phi}_\tau,\mathbf{\Phi}_d;\boldsymbol{X}_\tau,\mathbf{\Phi}_d), \tag{21}$$

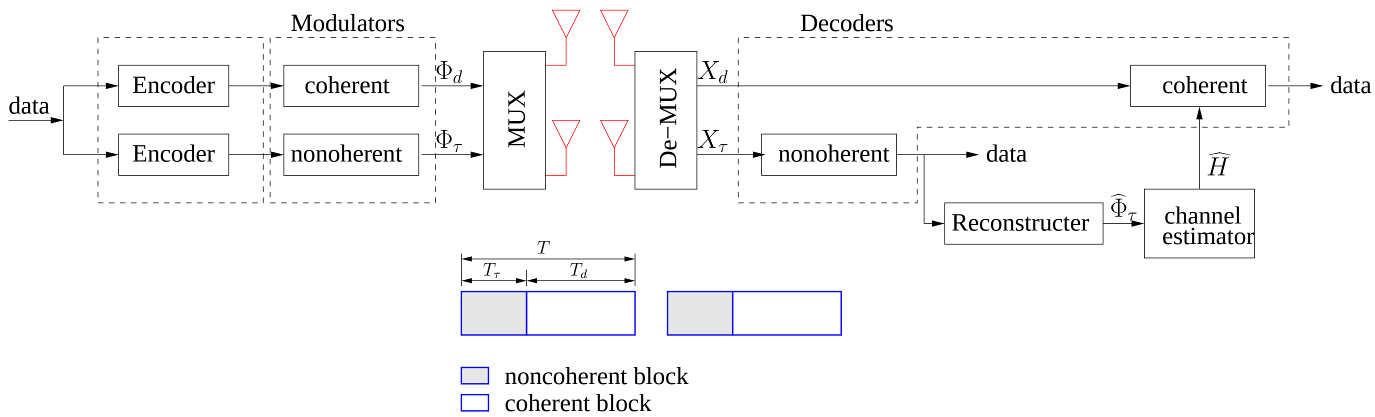

Fig. 3: Noncoherent Training

Compared with the conventional training-based system, it is even harder to compute the capacity for this new one. Similarly, we are only be able to calculate a lower bound on capacity. Using the chain rule of mutual information, we have

$$\begin{aligned} I(\boldsymbol{\Phi}_\tau, \boldsymbol{\Phi}_d; \boldsymbol{X}_\tau, \boldsymbol{X}_d) &= I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau, \boldsymbol{X}_d) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_\tau, \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau) \\ &= I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau) + I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_d|\boldsymbol{X}_\tau) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_\tau|\boldsymbol{\Phi}_\tau) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau) \\ &\geq I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau), \end{aligned} \tag{22}$$

where the inequality cannot be reduced to equality, since $\boldsymbol{\Phi}_\tau$ is random and may depend on $\boldsymbol{\Phi}_d$. Even when $\boldsymbol{\Phi}_\tau$ and $\boldsymbol{\Phi}_d$ are independent, the term $I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_d|\boldsymbol{X}_\tau)$ can still be nonzero. Nevertheless, supposing that $\boldsymbol{\Phi}_\tau$ and $\boldsymbol{\Phi}_d$ are independent, we find that $C_{unknown}$ satisfies

$$C_{unknown} \geq \sup_{p(\boldsymbol{\Phi}_\tau), p(\boldsymbol{\Phi}_d)} \frac{1}{T}[(I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau)]. \tag{23}$$

The optimization in (23) is taken over all input distributions $p(\boldsymbol{\Phi}_\tau), p(\boldsymbol{\Phi}_d)$ adhering to the power constraints that $\mathbb{E}\{\text{Tr } (\boldsymbol{\Phi}_\tau^\dagger \boldsymbol{\Phi}_\tau)\} = M$ and $\mathbb{E}\{\text{Tr } (\boldsymbol{\Phi}_d^\dagger \boldsymbol{\Phi}_d)\} = M$. Note that the righthand side of (23) is consistent with the receiver structure shown in Fig. 3, in which the data stream $\boldsymbol{\Phi}_\tau$ is decoded first and $\boldsymbol{\Phi}_d$ is decoded later based on $\boldsymbol{X}_\tau$ and the reconstructed $\boldsymbol{\Phi}_\tau$ (with or without using an explicit estimate of the channel).

B. Training via Unitary Space-Time Modulation

We do not know what inputs maximize $I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau) + I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau)$ in (23). Instead, we choose $\boldsymbol{\Phi}_\tau$ to be USTM for which $I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau)$ can be calculated at least by Monte Carlo simulations [25], and then compute an analytical lower bound on $I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau)$ as in [12] and [6]. The reason for choosing unitary $\boldsymbol{\Phi}_\tau$ is two-fold: first, the USTM inputs maximize $I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau)$ for large $T_\tau (>> M)$, and for large ρ_τ with $M \leq \min\{N, \lfloor T_\tau/2 \rfloor\}$. Second, used as training symbols after being successfully decoded, the unitary $\boldsymbol{\Phi}_\tau$ minimizes MMSE and maximizes a lower bound on capacity of training-based schemes simultaneously.

Since optimizing power between the training and non-training parts is difficult, we assume that equal power is used: $\rho = \rho_\tau = \rho_d$, which can also ensure constant modulus transmissions. For the USTM part, we have

$$I(\boldsymbol{\Phi}_\tau; \boldsymbol{X}_\tau) = T_\tau \cdot I_{USTM}(\rho), \tag{24}$$

where $I_{USTM}(\rho)$ is the mutual information in bits per symbol of USTM inputs with block length T_τ.

For the part with channel estimation, due to equal transmission power, we obtain from (14) that

$$I(\boldsymbol{\Phi}_d; \boldsymbol{X}_d|\boldsymbol{\Phi}_\tau, \boldsymbol{X}_\tau) \geq (1 - \frac{T_\tau}{T}) C_{coherent}(\rho_{eff}) \cdot T = (T - T_\tau) C_{coherent}(\rho_{eff}), \tag{25}$$

where $\rho_{eff} = \frac{\rho^2 T_\tau/M}{1+(1+T_\tau/M)\rho}$. Combining (24) and (25) leads to a lower bound in bits per symbol on channel capacity

$$C^L_{unknown}(\rho) := \frac{1}{T}[T_\tau I_{USTM(\rho)} + (T - T_\tau) C_{coherent}(\rho_{eff})] = \alpha_1 I_{USTM(\rho)} + (1 - \alpha_1) C_{coherent}(\rho_{eff}), \tag{26}$$

if we define $\alpha = M/T$ as before, and $\alpha_1 = T_\tau/T \leq 1$.

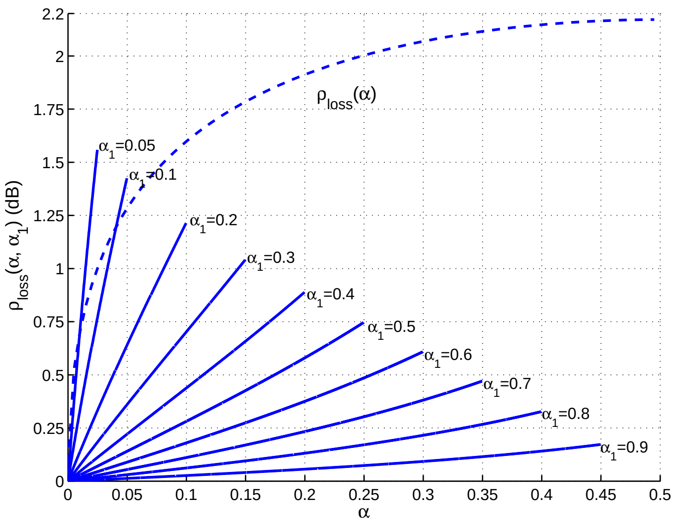

Fig. 4: Asymptotic SNR loss $\rho'(\alpha, \alpha_1)$ for $2\alpha \leq \alpha_1 \leq 1$

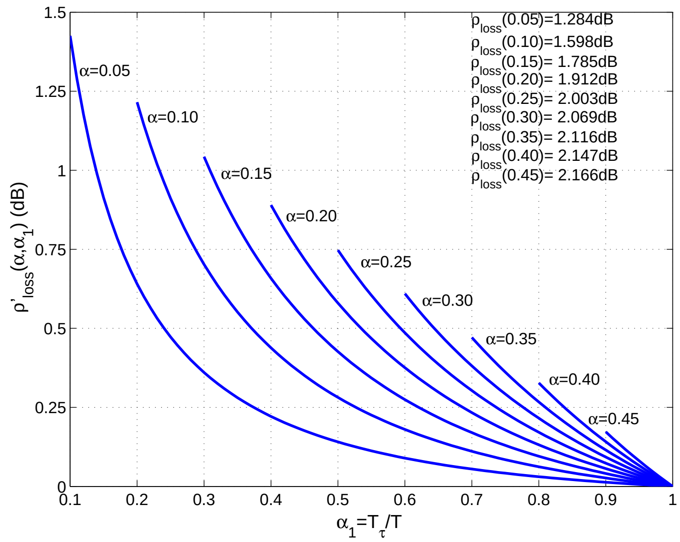

Fig. 5: Asymptotic SNR loss $\rho'(\alpha, \alpha_1)$ for $2\alpha \leq \alpha_1 \leq 1$

C. High SNR

With $T \geq T_\tau \geq 2M = 2N$, we have $2\alpha \leq \alpha_1$. We are interested in the asymptotic behavior when ρ, T, T_τ and M go to infinity, but the ratios α, α_1 are fixed. Note that $M/T_\tau = \alpha/\alpha_1$.

Since $T_\tau \geq 2M$, at high SNR, the first term in $C^L_{unknown}(\rho)$ satisfies [c.f. Lemma 2]

$$\frac{I_{USTM}(\rho)}{M} \to \frac{k(\frac{\alpha}{\alpha_1})}{\ln 2} + (1 - \frac{\alpha}{\alpha_1}) \log_2(\frac{\rho}{e}), \text{ as } \rho, M \to \infty. \tag{27}$$

Note that for large ρ, we have $\rho_{eff} = \frac{\rho}{1+\alpha/\alpha_1}$. Similar to (3), we obtain for large ρ, M

$$\frac{C_{coherent}(\rho_{eff})}{M} \to \log_2 \left[(\frac{\rho}{e})(1 + \frac{\alpha}{\alpha_1})^{-1} \right], \text{ as } \rho, M \to \infty. \tag{28}$$

Therefore,

$$\begin{aligned} \frac{C^L_{unknown}(\rho)}{M} &\to \frac{\alpha_1 k(\frac{\alpha}{\alpha_1})}{\ln 2} + (\alpha_1 - \alpha) \log_2(\frac{\rho}{e}) + (1 - \alpha_1) \log_2 \left[(\frac{\rho}{e})(1 + \frac{\alpha}{\alpha_1})^{-1} \right] \\ &= (1 - \alpha) \log_2 \left[(\frac{\rho}{e}) \cdot (1 + \frac{\alpha}{\alpha_1})^{(\frac{-1+\alpha_1}{1-\alpha})} \cdot 2^{\frac{\alpha_1 k(\frac{\alpha}{\alpha_1})}{(1-\alpha)\ln 2}} \right]. \end{aligned} \tag{29}$$

Compared with (18), we can identify the asymptotic SNR loss, as summarized in the following Theorem.

Theorem 1: When $\rho \to \infty, M \to \infty, T_\tau \to \infty$ and $T \to \infty$, but the ratios $2\alpha = 2M/T \leq \alpha_1 = T_\tau/T \leq 1$ are fixed, $C^L_{unknown}(\rho)$ suffers an asymptotic SNR loss relative to the noncoherent capacity

$$\rho'_{loss}(\alpha, \alpha_1) = (1 + \frac{\alpha}{\alpha_1})^{\frac{1-\alpha_1}{1-\alpha}} \cdot 2^{\frac{k(\alpha) - \alpha_1 k(\frac{\alpha}{\alpha_1})}{(1-\alpha)\ln 2}}, \tag{30}$$

where $k(\cdot)$ is given in (6).

Fig. 4 depicts rate vs. α, for fixed $\alpha_1 (\geq 2\alpha)$, while Fig. 5 depicts rate vs. $\alpha_1 (\geq 2\alpha)$, for fixed α. For fixed α_1, a large α leads to a high SNR loss, while for fixed α, a large α_1 yields a small SNR loss. An interesting fact revealed by Fig. 4 and 5 is that for sufficiently small α and α_1, we find that the conventional training-based scheme with power control is better, since $\rho'_{loss}(\alpha, \alpha_1) > \rho_{loss}(\alpha)$. For example, $\rho'_{loss}(0.05, 0.1) = 1.425$dB while $\rho_{loss}(0.05) = 1.284$dB. The reason is that equal transmission power is used for both training and non-training parts with computing $\rho'_{loss}(\alpha, \alpha_1)$. For very small α and α_1, the advantage of power control is greater than the benefit of noncoherent training. If optimal power allocation is used, which is a difficult optimization problem, we conjecture that $\rho'_{loss}(\alpha, \alpha_1)$ will be always smaller than $\rho_{loss}(\alpha)$. Even without power optimization though, for most interesting (α, α_1) combinations $(\alpha > 0.05, \alpha > 0.1)$, our noncoherent training-based approach outperforms the conventional one.

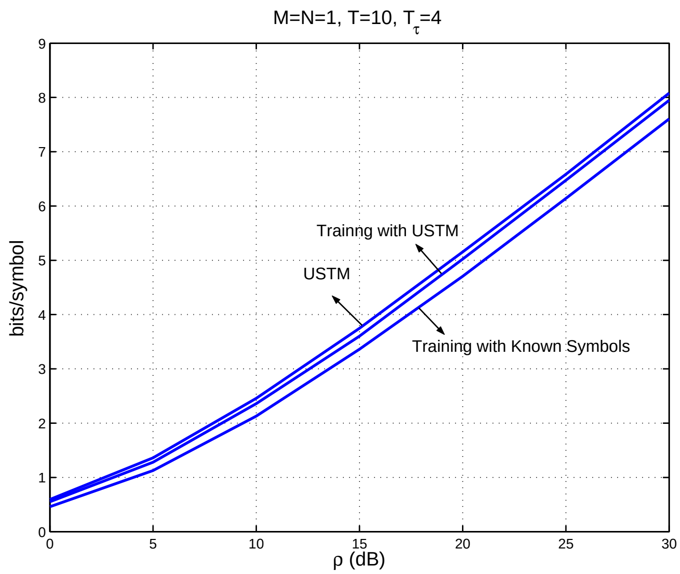

Fig. 6: $M = N = 1, T = 10, T_\tau = 4$

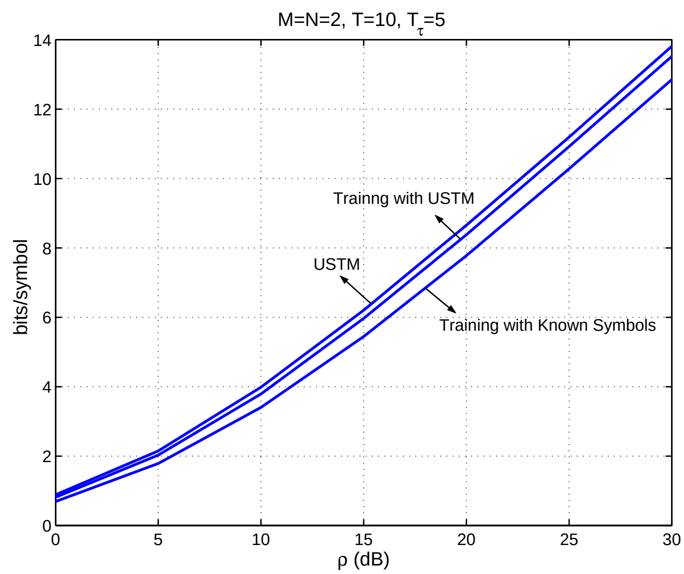

Fig. 7: $M = N = 2, T = 10, T_\tau = 5$

V. Numerical Results

In this section, we present numerical simulation results for the three approaches over noncoherent channels: USTM, training with known pilot symbols, and training with USTM symbols. We obtain the mutual information $I_{USTM}(\rho), C^L_{known}(\rho)$ and $C^L_{unknown}(\rho)$ through Monte Carlo simulations for finite $M = N, T$ and ρ. The results validate the asymptotic SNR loss $\rho_{loss}(\alpha)$ and $\rho'_{loss}(\alpha, \alpha_1)$. For the method of numerical evaluation of the mutual information of USTM inputs, we refer the reader to [25].

Fig. 6 shows the result for $M = N = 1, T = 10$ and $T_\tau = 4$. We observe that compared with USTM, training with known pilot symbols suffers about 1.5dB SNR loss at $\rho = 25$dB, while training with USTM incurs only about 0.4dB penalty. These results are very close to the asymptotic SNR loss given by Proposition 1 and Theorem 1: $\rho_{loss}(0.1) = 1.598\text{dB},\ \rho'_{loss}(0.1, 0.4) = 0.438\text{dB}$.

Fig. 6 depicts the result for $M = N = 2, T = 10$ and $T_\tau = 5$. Compared with USTM, training with known pilot symbols suffers about 1.8dB SNR loss at $\rho = 25$dB, while training with USTM incurs only about 0.55dB loss. Similarly, these results are very close to the asymptotic SNR loss given by Proposition 1 and Theorem 1: $\rho_{loss}(0.2) = 1.912\text{dB},\ \rho'_{loss}(0.2, 0.5) = 0.580\text{dB}$.

VI. Conclusions

We developed a new training scheme that uses information-bearing USTM symbols as "pilots" instead of known symbols as in the conventional training-based approaches. The receiver first decodes these USTM pilot symbols without channel state information, and then uses the decoded symbols as training to estimate the channel. While this new method decreases the complexity of the capacity-achieving approach through a short USTM block $T_\tau < T$, it can also recover some SNR loss that is inherent to conventional training-based strategies. When $T \geq T_\tau \geq 2M = 2N \to \infty$ and $\rho \to \infty$, but the ratios $\alpha = M/T, \alpha_1 = T_\tau/T$ are fixed, the asymptotic expressions of SNR loss were obtained analytically for both conventional and the proposed schemes, and are useful as a guideline for practical MIMO designs. While the current work is only focused on the information-theoretic analysis, in our future work we will pursue practical coding schemes for the proposed approach.

REFERENCES

[1] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 6, pp. 686–693, Nov. 1991.

[2] S. Sampei and T. Sunaga, "Rayleigh fading compensation for QAM in land mobile radio communications," *IEEE Trans. Veh. Technol.*, vol. 42, no. 2, pp. 137–147, May 1993.

[3] L. Tong, B. Sadler, and M. Dong, "Pilot assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[4] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[5] A. Sabharwal, E. Erkip, and B. Aazhang, "On channel state information in multiple antenna block fading channels," in *IEEE Symposium on Information Theory*, Honolulu, HI, November 2000, pp. 116–119.

[6] B. Hassibi and B. Hochwald, "How much training is needed in a multiple-antenna wireless link," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–964, Apr. 2003.

[7] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[8] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.

[9] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for mimo frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453–466, Mar. 2005.

[10] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots and precoders for wireless OFDM over rapidly fading channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2138–2145, Sept. 2004.

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

[12] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[13] G. Foschini and M. Gans, "On limits of wireless communications in fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 331–335, Nov./Dec. 1998.

[14] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[15] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[16] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[17] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for space-time communication on the rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1186–1194, May 2002.

[18] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of non-coherent space-time constellations," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.

[19] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[20] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[21] W. Zhao, G. Leus, and G. B. Giannakis, "Orthogonal design of unitary constellations for uncoded and trellis coded non-coherent space-time systems," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1319–1327, June 2004.

[22] V. Tarokh and I. Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3112–3117, Dec. 2002.

[23] P. Dayal, M. Brehler, and M. K. Varanasi, "Leveraging coherent space-time codes for noncoherent communication via training," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2058–2080, Sept. 2004.

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

[25] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1473–1484, June 2002.