[15] M. Pohst and H. Zassenhaus, *Algorithmic algebraic number theory, Encyclopedia of Mathematics and its Applications*. Cambridge, U.K.: Cambridge University Press, 1989.
[16] J. Proakis, *Digital Communications*, 4 ed. New York: McGrawHill, 2000.
[17] P. Sebastian, H. Sampath, and A. Paulraj, "Adaptive modulation for multiple antenna systems," in *Proc. Asilomar Conf. on Signals, Syst. Comput.*, Oct. 2000, vol. 1, pp. 506–510.
[18] G. Taricco and E. Biglieri, "Exact pairwise error probability of space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 510–513, Feb. 2002.
[19] ——, "Correction to ëxact pairwise error probability of space-time codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 766–766, Mar. 2003.
[20] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[21] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the performance of lattice codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 670–681, Mar. 1999.
[22] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.
[23] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.
[24] E. Viterbo and E. Biglieri, "Computing the Voronoi cell of a lattice: The diamond-cutting algorithm," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 161–171, Jan. 1996.

Information-Bearing Noncoherently Modulated Pilots for MIMO Training

Yingqun Yu, *Student Member, IEEE*,
Georgios B. Giannakis, *Fellow, IEEE*, and
Nihar Jindal, *Member, IEEE*

***Abstract*—In this correspondence, we deal with noncoherent communications over multiple-input–multiple-output (MIMO) wireless links. 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) $\rho \gg 1$ and $M \leq \min\{N, \lfloor T/2 \rfloor\}$, unitary space–time modulation (USTM) is capacity-achieving, but incurs exponential demodulation complexity 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. To achieve desirable tradeoffs between performance and complexity, we propose a novel training approach where USTM symbols over a short length $T_\tau (< T)$ are used as pilots which carry information to the receiver, and coherent communication is used for the remainder of the block. This new approach considerably reduces the receiver complexity when T_τ is a small fraction of T, and recovers part 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, serving as a guideline for practical designs.**

***Index Terms*—Capacity, channel estimation, coherent detection, multiple antennas, noncoherent detection, unitary space–time modulation.**

Manuscript received May 11, 2005; revised September 13, 2006. This work was supported in part by the ARL/CTA under Grant DAAD19-01-2-0011. This work was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: yingqun@umn.edu; georgios@umn.edu; nihar@umn.edu).

Color versions of Figs 1–7 are available online at http://ieeexplore.ieee.org.
Communicated by B. Hassibi, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.890693

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.[1] Pilot symbol-assisted modulation (PSAM) is a standard training-based approach when communicating over time-varying channels [4], [23], [26]. 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, overincreasing the overhead for training reduces the amount of channel uses and power for information-carrying data symbols, which decreases data throughput.

A 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 are available for a general setting [19], for Rayleigh flat block-fading multiple-input–multiple-output (MIMO) wireless channels [29], [22], [9], and for perfectly interleaved MIMO channels [2]. The optimal power allocation between pilot and data symbols as well as the number (equal to M) of training symbols optimizing a lower bound on capacity were obtained in [9]. 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 [16], [21], [27], [1].

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 [17] 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 at high SNR can be achieved using only M^* antennas, and increases linearly with $M^*(1 - M^*/T)$, where $M^* = \min\{M, N, \lfloor T/2 \rfloor\}$ [29]. In comparison, if the receiver knows the channel coefficients perfectly, it is well-known that the ergodic *coherent* channel capacity increases linearly with $\min\{M, N\}$ under the same channel model [6], [25].

Motivated by results in [17], a class of isotropic unitary space-time modulation (USTM) signals was proposed in [11], [12], [15] *et. al* to encode the transmitted signals using $T \times M$ isotropic unitary matrices. For $T \gg M$ [17], and for high SNR $\rho \gg 1$ with $M \leq \min\{N, \lfloor T/2 \rfloor\}$ [29], the optimal input has indeed a USTM form.

[1]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

The main drawback of USTM transmissions is that very often their design requires numerical optimization [12], [18], [3], and because they possess no particular algebraic structure, they incur relatively high complexity. Furthermore, their demodulation is exponentially complex 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 USTM's [14], [13], [8] *et. al* and alternative training-like constellations [28], [24], [5] 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 be 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 an SNR loss compared with the optimal noncoherent scheme [29], [9]. A training-based scheme can be capacity-achieving at high SNR 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 correspondence, we first analytically compute the asymptotic SNR loss for the conventional training-based methods when $\rho \to \infty$ and $T \geq 2M = 2N \to \infty$, but the ratio $\alpha = M/T$ is fixed. We show that as α decreases, the asymptotic SNR loss drops monotonically but also slowly from 2.17 dB $(\alpha = 0.5)$ to zero $(\alpha \to 0)$. Further, we introduce a novel scheme that combines noncoherent and coherent detection for the block fading channel, and offers flexibility 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 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 used only for channel estimation, 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 correspondence 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 hybrid 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 [17]. 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 $\boldsymbol{\Phi} \in \mathcal{C}^{T\times M}$ is transmitted,[2] the received signal matrix $\boldsymbol{X} \in \mathcal{C}^{T\times N}$ can be written as

$$\boldsymbol{X} = \sqrt{\frac{\rho T}{M}}\boldsymbol{\Phi}\boldsymbol{H} + \boldsymbol{W} \quad (1)$$

where $\boldsymbol{H} \in \mathcal{C}^{M\times N}$ is the channel matrix, and $\boldsymbol{W} \in \mathcal{C}^{T\times N}$ denotes the additive white noise matrix. Both $\boldsymbol{H}$ and $\boldsymbol{W}$ are complex Gaussian 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 $\mathbb{E}\{\mathrm{Tr}(\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi})\} = M$, 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 [7], [25] and is summarized in the following.

Lemma 1: If $\boldsymbol{H}$ is known to the receiver but not the transmitter, the coherent capacity in bits per symbol is given by

$$\begin{aligned} C_{\mathrm{coherent}}(\rho) &= \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I}_M + \frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^\dagger\right)\right] \\ &= \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I}_N + \frac{\rho}{M}\boldsymbol{H}^\dagger\boldsymbol{H}\right)\right]. \end{aligned} \quad (2)$$

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

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

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 $M^* = \min\{M, N, \lfloor T/2 \rfloor\}$ [29]. 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 $\mathbb{E}\{\mathrm{Tr}(\boldsymbol{D})\} = 1$ [17]. The distribution of $\boldsymbol{D}$ is generally unknown, except for the asymptotic case $T \gg M$ [17] and for high SNR with $M \leq \min\{N, T/2\}$ [29], where $\boldsymbol{D}$ becomes a scaled deterministic identity matrix $\boldsymbol{D} = \sqrt{1/T}\boldsymbol{I}_M$, suggesting the so-called USTM inputs for noncoherent channels [11]. The result is summarized in the following lemma for the case $T \geq 2M = 2N$ [29, Th. 9, Corollary 11].

Lemma 2: Assume $T \geq 2M = 2N$. 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

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

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

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

we let both T and M go to infinity but keep the ratio $\alpha = M/T$ fixed, then we have[3]

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

where

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{c(T,M)}{M} - \frac{k(\alpha)}{\ln 2}\right] = 0, \quad (6)$$

and

$$k(\alpha) = \frac{(1-\alpha)^2}{2\alpha}\ln(1-\alpha) + \frac{\alpha}{2}\ln\alpha + \frac{1-\alpha}{2} < 0 \quad (7)$$

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 at both high and low SNR. It is shown that in terms of achievable rates USTM outperforms training under both situations.

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 [9]

$$\mathbf{\Phi} = \begin{pmatrix}\sqrt{\frac{\rho_\tau T_\tau}{M}}\mathbf{\Phi}_\tau \\ \sqrt{\frac{\rho_d T_d}{M}}\mathbf{\Phi}_d\end{pmatrix} \quad (8)$$

where $\mathbf{\Phi}_\tau \in \mathcal{C}^{T_\tau\times M}$ with $T_\tau \ge M, \mathrm{Tr}(\mathbf{\Phi}_\tau^\dagger\mathbf{\Phi}_\tau) = M$ is the training matrix known to both the transmitter and the receiver, and $\mathbf{\Phi}_d \in \mathcal{C}^{T_d\times M}$ with $\mathbb{E}\{\mathrm{Tr}(\mathbf{\Phi}_d^\dagger\mathbf{\Phi}_d)\} = M$ 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

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_\tau \\ \boldsymbol{X}_d\end{pmatrix}, \text{ and } \boldsymbol{W} = \begin{pmatrix}\boldsymbol{W}_\tau \\ \boldsymbol{W}_d\end{pmatrix} \quad (9)$$

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

$$\boldsymbol{X}_\tau = \sqrt{\frac{\rho_\tau T_\tau}{M}}\mathbf{\Phi}_\tau\boldsymbol{H} + \boldsymbol{W}_\tau \quad (10)$$

and for the data phase as

$$\boldsymbol{X}_d = \sqrt{\frac{\rho_d T_d}{M}}\mathbf{\Phi}_d\boldsymbol{H} + \boldsymbol{W}_d. \quad (11)$$

[3]We believe that in [29, eq. (22)] the term $\log_2 e$ should be $\ln 2$.

The capacity in bits per symbol for the training-based scheme is given by [9]

$$\begin{aligned}C_{\text{known}} &= \sup_{\mathbf{\Phi}_\tau, p(\mathbf{\Phi}_d)}\frac{1}{T}I(\mathbf{\Phi}_\tau,\mathbf{\Phi}_d;\boldsymbol{X}_\tau,\mathbf{\Phi}_d)\\ &= \sup_{\mathbf{\Phi}_\tau, p(\mathbf{\Phi}_d)}\frac{1}{T}I(\mathbf{\Phi}_d;\boldsymbol{X}_d|\mathbf{\Phi}_\tau,\boldsymbol{X}_\tau)\end{aligned} \quad (12)$$

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 (12) 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 $\mathrm{Tr}(\mathbf{\Phi}_\tau^\dagger\mathbf{\Phi}_\tau) = M$ and $\mathbb{E}\{\mathrm{Tr}(\mathbf{\Phi}_d^\dagger\mathbf{\Phi}_d)\} = M$. 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 minimum mean square error (MMSE) estimate of the channel matrix, and then absorb the estimation error in 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 [9] that the training matrix with orthonormal columns $\mathbf{\Phi}_\tau^\dagger\mathbf{\Phi}_\tau = \boldsymbol{I}_M$ simultaneously maximizes this lower bound and minimizes MSE. If optimal power allocation between training and information-bearing symbols can be afforded, the optimal training interval should be $T_\tau^{\text{opt}} = M$, and the corresponding lower bound is

$$C^L_{\text{known}}(\rho) = \left(1 - \frac{M}{T}\right)C_{\text{coherent}}(\rho_{\text{eff}}) \quad (13)$$

where

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

$$\gamma = \frac{(M+\rho T)(T-M)}{\rho T(T-2M)}. \quad (14)$$

Compared with the noncoherent capacity in (4) at high SNR for the case $T \ge 2M = 2N$, the training-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-B1, an asymptotic expression for SNR loss will be found.

If equal training and data power $\rho = \rho_\tau = \rho_d$ is mandatory, e.g., in order to ensure constant-modulus transmissions, then

$$C^L_{\text{known}}(\rho) = \left(1 - \frac{T_\tau}{T}\right)C_{\text{coherent}}(\rho_{\text{eff}}) \quad (15)$$

where

$$\rho_{\text{eff}} = \frac{\rho^2 T_\tau/M}{1 + (1 + T_\tau/M)\rho}. \quad (16)$$

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.

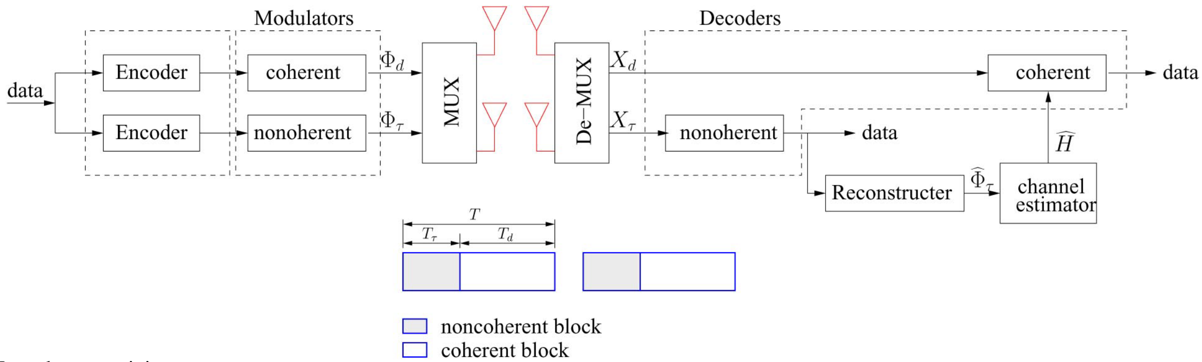

Fig. 1. Noncoherent training.

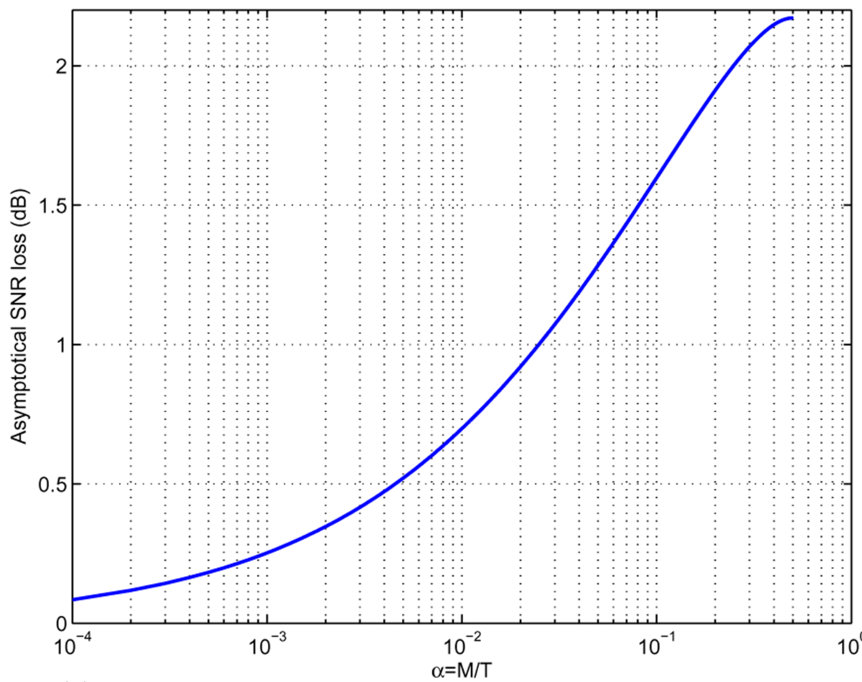

Fig. 2. Asymptotic SNR loss $\rho_{\text{loss}}(\alpha)$.

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^L_{known} can be expressed as

$$\beta := \frac{\rho}{\rho_{eff}} = \frac{T-2M}{T}(\sqrt{\gamma}-\sqrt{\gamma-1})^{-2} = \frac{T-2M}{T}(\sqrt{\gamma}+\sqrt{\gamma-1})^{2}, \text{ for } T>2M. \quad (17)$$

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

$$\beta_\infty := \lim_{\rho\to\infty}\beta = 1+2\sqrt{\frac{M}{T}\left(1-\frac{M}{T}\right)} \quad (18)$$

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 $\alpha = M/T$ is fixed. From Lemma 2, incorporating the term $k(\alpha)$ into $\log_2(\cdot)$ yields

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{C_{M,M}(\rho)}{M}-(1-\alpha)\log_2\left(\left(\frac{\rho}{e}\right)\cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}}\right)\right]=0. \quad (19)$$

Also from Lemma 1, we have a similar asymptotic result for $C^L_{\text{known}}(\rho)$

$$\lim_{M\to\infty}\lim_{\rho\to\infty}\left[\frac{C^L_{\text{known}}(\rho)}{M}-(1-\alpha)\log_2\left(\left(\frac{\rho}{e}\right)\cdot\beta_\infty^{-1}\right)\right]=0. \quad (20)$$

Upon comparing (19) with(20), we obtain the following result.

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

$$\rho_{\text{loss}}(\alpha) := \beta_\infty \cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}} = \left[1+2\sqrt{\alpha(1-\alpha)}\right]\cdot 2^{\frac{k(\alpha)}{(1-\alpha)\ln 2}} \quad (21)$$

where $k(\cdot)$ is given in (7).

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

We plot $\rho_{\text{loss}}(\alpha)$ in Fig. 2 with α in logarithmic 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 1.5980 dB at $\alpha = 10^{-1}(T = 10\ \text{M})$, and drops to 0.6981 dB at $\alpha = 10^{-2}(T = 100\ \text{M})$. Even when $\alpha = 10^{-3}(T = 1000\ \text{M})$, there is still 0.2523 dB SNR loss. For the region $T = 2 \sim 20$ M, there is always an SNR loss less than 2.17 dB, but more than 1.5 dB. Fig. 3 compares the asymptotic SNR loss with the case $\rho = 25$ dB,

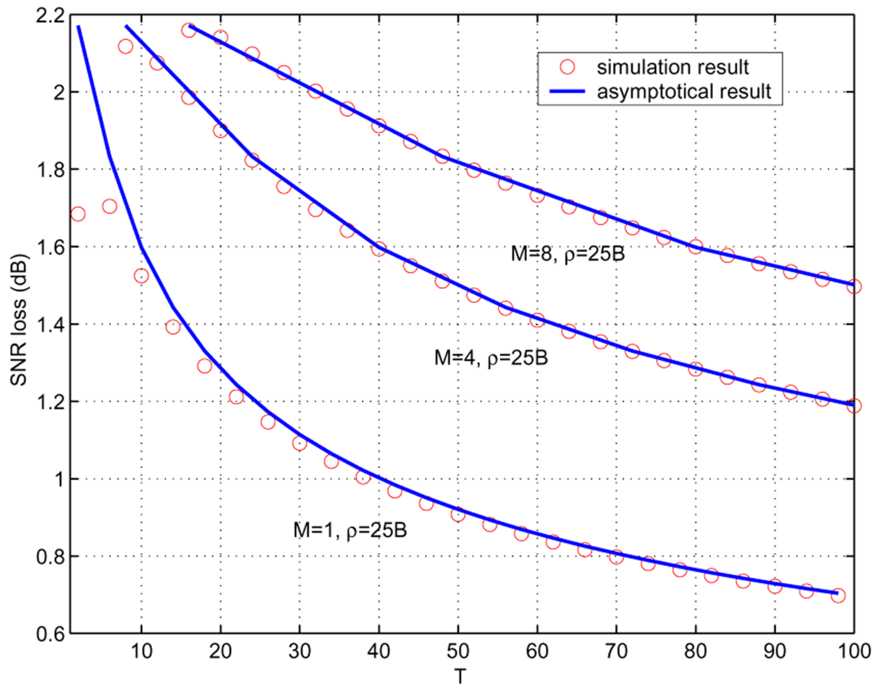

Fig. 3. SNR loss for finite ρ, M and T.

$M = 1, 4, 8$ and $T = 1 \sim 100$. It can be seen that $\rho_{\text{loss}}(\alpha)$ is actually tight except when $M = 1$ and $T < 10$; therefore, $\rho_{\text{loss}}(\alpha)$ 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, which is the scenario we focus on, a training-based scheme is not capacity-optimal, although it can achieve the same degrees of freedom as the optimal noncoherent scheme.

2) Low SNR: We assume that $T > M$. We know that the optimum training length is $T_\tau^{\text{opt}} = M$, but at low SNR, the lower bound of the training-based scheme is not sensitive to the length of training symbols [9], and the lower bound is given by (up to $o(\rho^2)$)

$$C_{\text{known}}^L(\rho) \approx \frac{NT \log e}{4M} \rho^2. \tag{22}$$

At low SNR, the mutual information of USTM inputs can be approximated as (up to $o(\rho^2)$) [20]

$$I_{\text{USTM}}(\rho) \approx \frac{N(T-M)\log e}{2M} \rho^2. \tag{23}$$

From (22) and(23) we obtain

$$\frac{I_{\text{USTM}}(\rho)}{C_{\text{known}}^L(\rho)} \approx \frac{2(T-M)}{T}. \tag{24}$$

Proposition 2: For fixed M and T with $T > M$, the asymptotic ratio of the mutual information of USTM to C_{known}^L at low SNR is

$$\eta(\alpha) := \lim_{\rho \to 0} \frac{I_{\text{USTM}}(\rho)}{C_{\text{known}}^L(\rho)} = 2(1-\alpha), \tag{25}$$

where $\alpha = M/T$.

We can see that at low SNR, USTM is also better than training. The reason that training-based schemes become worse is that at low SNR the channel estimation has very low quality and thus becomes unreliable. Since USTM bypasses channel estimation, it achieves higher rates than training. The complexity of USTM at low SNR can be acceptable since the achievable rate for small ρ is quite low.

IV. Training Via Information-Bearing Noncoherent Space–Time Modulation

In this section, we develop a novel training-based scheme where "pilot" symbols, just like data symbols, can also carry information and thus are unknown to the receiver. It appears impossible to estimate the channel when the receiver does not know the transmitted pilot symbols, unless some kind of blind estimation scheme is used. However, it is certainly possible to do so after the receiver successfully decodes them. The decoding of unknown pilot symbols, though, does not require CSI knowledge, which can be enabled by using any noncoherent communication scheme.

A. Training Via Noncoherent Communication

The proposed system architecture is shown in Fig. 1. 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 $\boldsymbol{X}_d$. The received matrix $\boldsymbol{X}_\tau$ carries data and is decoded first. Since $\mathbf{\Phi}_\tau$ is noncoherently modulated, the receiver can decode it without knowing the fading matrix $\boldsymbol{H}$. Once the transmitted signal matrix is recovered as $\widehat{\boldsymbol{\Phi}}_\tau$ after decoding, the receiver can estimate the channel using $\widehat{\boldsymbol{\Phi}}_\tau$. The estimated channel $\hat{\boldsymbol{H}}$ is subsequently sent to the coherent detector to decode the information carried by $\boldsymbol{X}_d$. In practice, the data bits carried by $\mathbf{\Phi}_\tau$ should be encoded across multiple blocks with relatively strong codes so that the channel estimation error caused by the incorrectly decoded $\widehat{\boldsymbol{\Phi}}_\tau$ is negligible. However, here we only focus on the information-theoretic analysis and the design of practical coding schemes is beyond the scope of this correspondence.

The receiver structure depicted in Fig. 1 is suboptimal in general, since information may be lost, because: 1) $\mathbf{\Phi}_\tau$ and $\mathbf{\Phi}_d$ are decoded not jointly but separately; and 2) 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 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 $\mathbf{X}_\tau$, $\mathbf{X}_d$, i.e.

$$C_{\text{unknown}} = \sup_{p(\mathbf{\Phi}_\tau, \mathbf{\Phi}_d)} \frac{1}{T} I(\mathbf{\Phi}_\tau, \mathbf{\Phi}_d; \mathbf{X}_\tau, \mathbf{X}_d). \tag{26}$$

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

$$\begin{aligned} I(\mathbf{\Phi}_\tau, \mathbf{\Phi}_d; \mathbf{X}_\tau, \mathbf{X}_d) &= I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau, \mathbf{X}_d) + I(\mathbf{\Phi}_d; \mathbf{X}_\tau, \mathbf{X}_d | \mathbf{\Phi}_\tau) \\ &= I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau) + I(\mathbf{\Phi}_\tau; \mathbf{X}_d | \mathbf{X}_\tau) \\ &\quad + I(\mathbf{\Phi}_d; \mathbf{X}_\tau | \mathbf{\Phi}_\tau) + I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau) \\ &\geq I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau) + I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau) \end{aligned} \tag{27}$$

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

$$C_{\text{unknown}} \geq \sup_{p(\mathbf{\Phi}_\tau), p(\mathbf{\Phi}_d)} \frac{1}{T} [(I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau) + I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau)]. \tag{28}$$

The optimization in (28) is taken over all input distributions $p(\mathbf{\Phi}_\tau)$, $p(\mathbf{\Phi}_d)$ adhering to the power constraints that $\mathbb{E}\{\text{Tr}(\mathbf{\Phi}_\tau^\dagger \mathbf{\Phi}_\tau)\} = M$ and $\mathbb{E}\{\text{Tr}(\mathbf{\Phi}_d^\dagger \mathbf{\Phi}_d)\} = M$. Note that the right-hand side (RHS) of (28) is consistent with the receiver structure shown in Fig. 1, in which the data stream $\mathbf{\Phi}_\tau$ is decoded first and $\mathbf{\Phi}_d$ is decoded later based on $\mathbf{X}_\tau$ and the reconstructed $\mathbf{\Phi}_\tau$ (with or without using an explicit estimate of the channel).

B. Training Via Unitary Space–Time Modulation

We do not know what inputs maximize $I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau) + I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau)$ in the RHS of (28). Instead, we choose $\mathbf{\Phi}_\tau$ to be USTM for which $I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau)$ can be calculated at least by Monte Carlo simulations [10], and then compute an analytical lower bound on $I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau)$ as in [29] and [9]. The reason for choosing unitary $\mathbf{\Phi}_\tau$ is twofold. First, the USTM inputs maximize $I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau)$ for large $T_\tau (\gg M)$, and for large ρ_τ with $M \leq \min\{N, \lfloor T_\tau/2 \rfloor\}$. Second, the unitary $\mathbf{\Phi}_\tau$, which is used as training symbols after being successfully decoded, minimizes channel minimum square error (mse) and maximizes a lower bound on capacity of training-based schemes simultaneously.

Since optimizing power between the training and nontraining parts is difficult, we assume that equal power is used: $\rho = \rho_\tau = \rho_d$. For the USTM part, we have

$$I(\mathbf{\Phi}_\tau; \mathbf{X}_\tau) = T_\tau \cdot I_{\text{USTM}}(\rho) \tag{29}$$

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 (15) that

$$\begin{aligned} I(\mathbf{\Phi}_d; \mathbf{X}_d | \mathbf{\Phi}_\tau, \mathbf{X}_\tau) &\geq \left(1 - \frac{T_\tau}{T}\right) C_{\text{coherent}}(\rho_{\text{eff}}) \cdot T \\ &= (T - T_\tau) C_{\text{coherent}}(\rho_{\text{eff}}) \end{aligned} \tag{30}$$

where $\rho_{eff} = \frac{\rho^2 T_\tau / M}{1 + (1 + T_\tau/M)\rho}$. Combining (29) and (30) leads to a lower bound on channel capacity

$$\begin{aligned} C_{\text{unknown}}^L(\rho) &:= \frac{1}{T} [T_\tau I_{\text{USTM}(\rho)} \\ &\quad + (T - T_\tau) C_{\text{coherent}}(\rho_{\text{eff}})] \\ &= \alpha_1 I_{\text{USTM}(\rho)} + (1 - \alpha_1) C_{\text{coherent}}(\rho_{\text{eff}}) \end{aligned} \tag{31}$$

if we define $\alpha_1 = T_\tau / T \leq 1$.

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 $C_{\text{unknown}}^L(\rho)$ satisfies [cf. Lemma 2]

$$\frac{I_{\text{USTM}}(\rho)}{M} \to \frac{k(\frac{\alpha}{\alpha_1})}{\ln 2} + \left(1 - \frac{\alpha}{\alpha_1}\right) \log_2\left(\frac{\rho}{e}\right), \text{ as } \rho, M \to \infty. \tag{32}$$

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

$$\frac{C_{\text{coherent}}(\rho_{\text{eff}})}{M} \to \log_2\left[\left(\frac{\rho}{e}\right)\left(1 + \frac{\alpha}{\alpha_1}\right)^{-1}\right], \text{ as } \rho, M \to \infty. \tag{33}$$

Therefore

$$\begin{aligned} \frac{C_{\text{unknown}}^L(\rho)}{M} &\to \frac{\alpha_1 k(\frac{\alpha}{\alpha_1})}{\ln 2} + (\alpha_1 - \alpha) \log_2\left(\frac{\rho}{e}\right) \\ &\quad + (1 - \alpha_1) \log_2\left[\left(\frac{\rho}{e}\right)\left(1 + \frac{\alpha}{\alpha_1}\right)^{-1}\right] \\ &= (1 - \alpha) \log_2\left[\left(\frac{\rho}{e}\right) \cdot \left(1 + \frac{\alpha}{\alpha_1}\right)^{(\frac{-1 + \alpha_1}{1 - \alpha})} \cdot 2^{\frac{\alpha_1 k(\frac{\alpha}{\alpha_1})}{(1 - \alpha)\ln 2}}\right]. \end{aligned} \tag{34}$$

Compared with (19), we can identify the asymptotic SNR loss, as summarized in the following theorem.

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

$$\rho'_{\text{loss}}(\alpha, \alpha_1) = \left(1 + \frac{\alpha}{\alpha_1}\right)^{\frac{1 - \alpha_1}{1 - \alpha}} \cdot 2^{\frac{k(\alpha) - \alpha_1 k(\frac{\alpha}{\alpha_1})}{(1 - \alpha)\ln 2}} \tag{35}$$

where $k(\cdot)$ is given in (7).

Fig. 4 depicts rate versus α, for fixed $\alpha_1 (\geq 2\alpha)$, while Fig. 5 depicts rate versus $\alpha_1 (\geq 2\alpha)$, 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 Figs. 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 $\rho'_{\text{loss}}(\alpha, \alpha_1) > \rho_{\text{loss}}(\alpha)$. For example, $\rho'_{\text{loss}}(0.05, 0.1) = 1.425$ [dB] while $\rho_{\text{loss}}(0.05) = 1.284$ dB. The reason is that equal transmission power is used for both training and nontraining parts in computing $\rho'_{\text{loss}}(\alpha, \alpha_1)$. For very small α and α_1, the advantage of power control outweighs the benefits of noncoherent training. If optimal power allocation is used, which is a difficult optimization problem, we conjecture that $\rho'_{\text{loss}}(\alpha, \alpha_1)$ will be always smaller than $\rho_{\text{loss}}(\alpha)$. Even without power optimization though, for

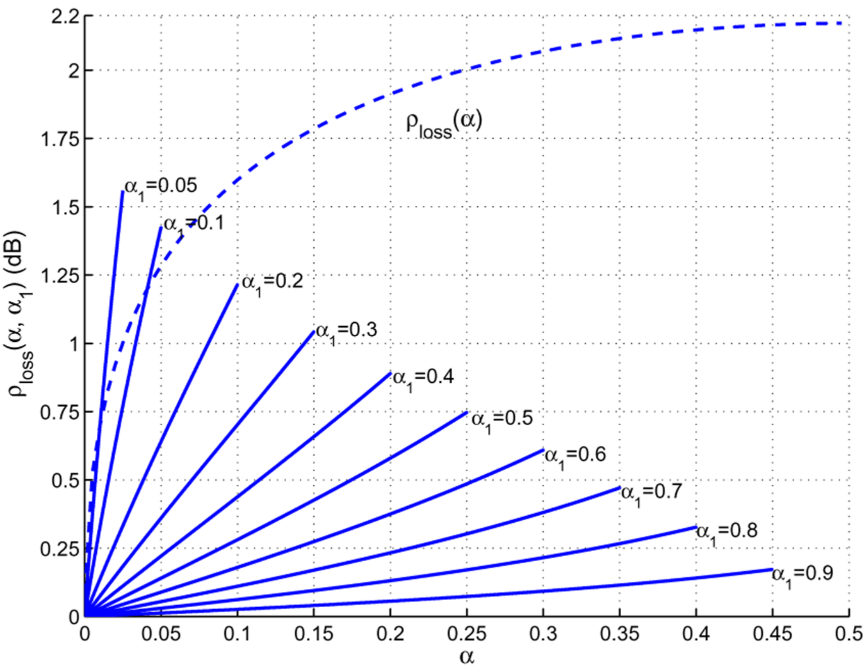

Fig. 4. Asymptotic SNR loss $\rho'(\alpha, \alpha_1)$ for $2\alpha \le \alpha_1 \le 1$

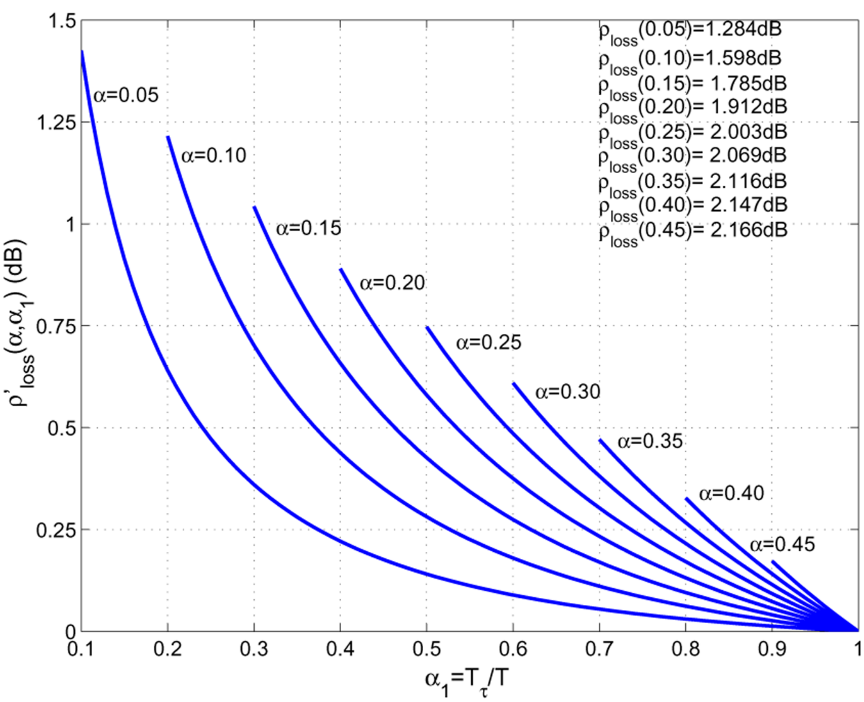

Fig. 5. Asymptotic SNR loss $\rho'(\alpha, \alpha_1)$ for $2\alpha \le \alpha_1 \le 1$

most interesting (α, α_1) combinations $(\alpha > 0.05, \alpha > 0.1)$, our noncoherent training-based approach outperforms the conventional one.

V. Numerical Results

In this section, we present simulations for the three approaches over noncoherent channels: USTM, training with known pilot symbols, and training via USTM symbols. We obtain the mutual information $I_{\mathrm{USTM}}(\rho)$, $C^L_{\mathrm{known}}(\rho)$ and $C^L_{\mathrm{unknown}}(\rho)$ through Monte Carlo simulations for finite $M = N$, T and ρ. The results validate the asymptotic SNR loss $\rho_{\mathrm{loss}}(\alpha)$ and $\rho'_{\mathrm{loss}}(\alpha, \alpha_1)$. For the method used to numerically evaluate the mutual information of USTM inputs, we refer the reader to [10].

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 0.4 dB penalty. These results are very close to the asymptotic SNR loss given by Proposition 1 and Theorem 1: $\rho_{\mathrm{loss}}(0.1) = 1.598$ dB, $\rho'_{\mathrm{loss}}(0.1, 0.4) = 0.438$ dB.

Fig. 7 depicts the result for $M = N = 2$, $T = 10$ and $T_\tau = 5$. Compared with USTM, training with known pilot symbols suffers

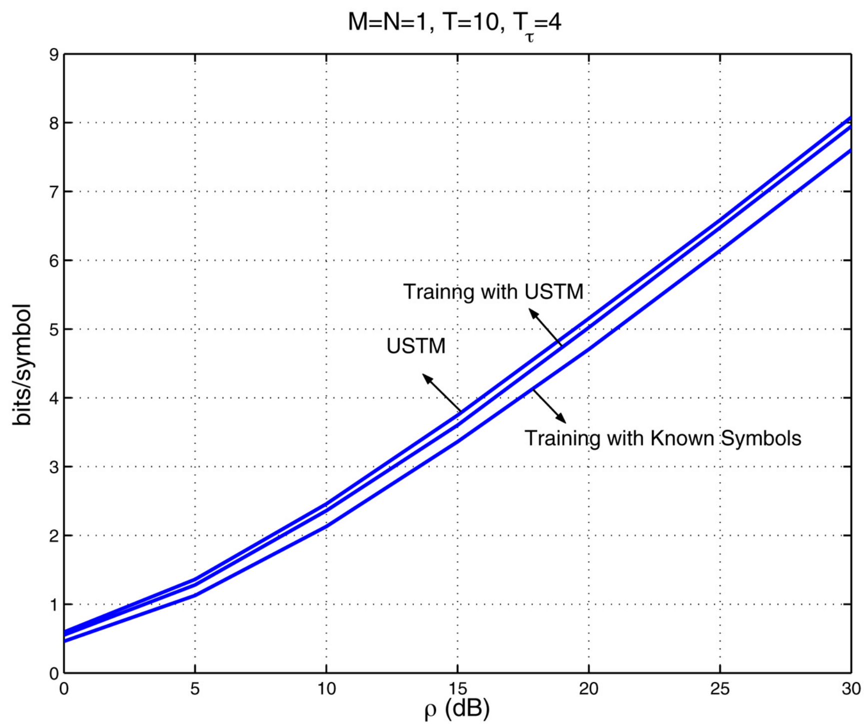

Fig. 6. $M = N = 1, T = 10, T_\tau = 4$.

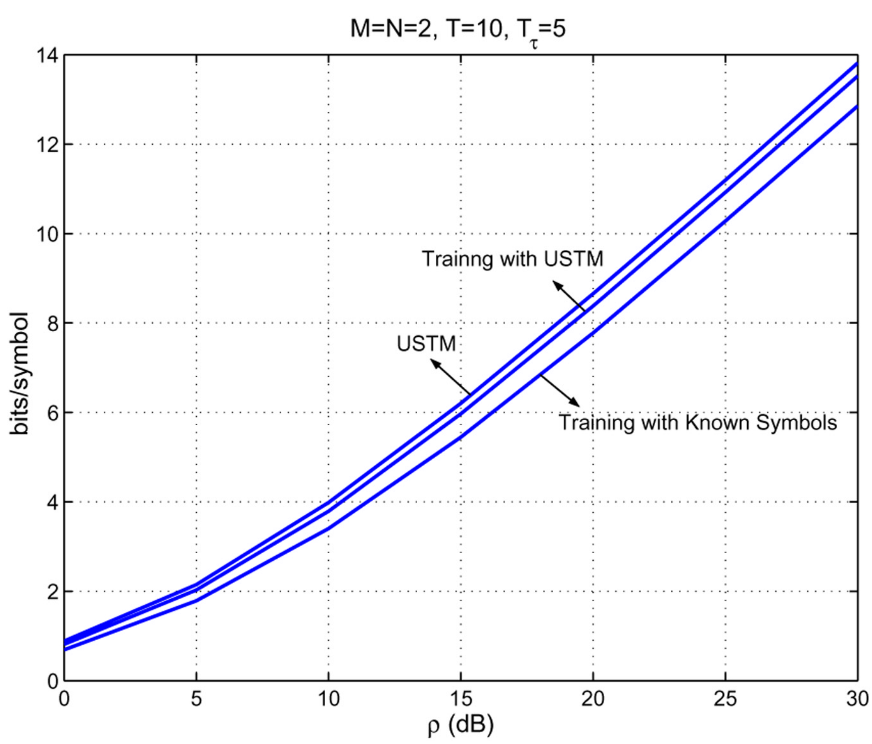

Fig. 7. $M = N = 2, T = 10, T_\tau = 5$.

about 1.8 dB SNR loss at $\rho = 25$ dB, while training with USTM incurs only about 0.55 dB loss. Similarly, these results are very close to the asymptotic SNR loss given by Proposition 1 and Theorem 1: $\rho_{\mathrm{loss}}(0.2) = 1.912$ dB, $\rho'_{\mathrm{loss}}(0.2, 0.5) = 0.580$ dB.

VI. Conclusion

We developed a new training scheme that uses information-bearing USTM symbols as "pilots" instead of known symbols utilized by 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 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 the 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 information-theoretic aspects, our future work will pursue practical coding schemes for the proposed approach.

REFERENCES

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[2] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 2358–2368, Dec. 2001.

[3] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of noncoherent space-time constellations," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2332–2351, Oct. 2003.

[4] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[5] P. Dayal, M. Brehler, and M. K. Varanasi, "Leveraging coherent space-time codes for noncoherent communication via training," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2058–2080, Sep. 2004.

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

[7] 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.

[8] B. Hassibi and B. Hochwald, "Cayley differential unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1485–1503, Jun. 2002.

[9] ——, "How much training is needed in a multiple-antenna wireless link," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–964, Apr. 2003.

[10] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1473–1484, Jun. 2002.

[11] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[12] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1962–1973, Sep. 2000.

[13] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2041–2052, Dec. 2000.

[14] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[15] Y. Jing and B. Hassibi, "Cayley codes for unitary space-time modulation," *IEEE Trans. Signal Process.*, vol. 51, pp. 2891–2904, Nov. 2003.

[16] 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.

[17] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inf. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[18] 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. Inf. Theory*, vol. 48, pp. 1186–1194, May 2002.

[19] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, pp. 933–946, May 2000.

[20] A. L. Moustakas, S. H. Simon, and T. L. Marzetta, "Capacity of differential versus nondifferential unitary space-time modulation for mimo channels," *IEEE Trans. Inf. Theory*, 2004, submitted for publication.

[21] S. Ohno and G. B. Giannakis, "Capacity maximizing mmse-optimal pilots and precoders for wireless OFDM over rapidly fading channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2138–2145, Sep. 2004.

[22] A. Sabharwal, E. Erkip, and B. Aazhang, "On channel state information in multiple antenna block fading channels," in *IEEE Symp. Inf. Theory*, Honolulu, HI, Nov. 2000, pp. 116–119.

[23] S. Sampei and T. Sunaga, "Rayleigh fading compensation for QAM in land mobile radio communications," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 137–147, May 1993.

[24] V. Tarokh and I. Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, pp. 3112–3117, Dec. 2002.

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

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

[27] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, vol. 52, pp. 2572–2583, Sep. 2004.

[28] W. Zhao, G. Leus, and G. B. Giannakis, "Orthogonal design of unitary constellations for uncoded and trellis coded noncoherent space-time systems," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1319–1327, Jun. 2004.

[29] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383, Feb. 2002.

On the Elementwise Convergence of Continuous Functions of Hermitian Banded Toeplitz Matrices

Pedro M. Crespo, *Senior Member, IEEE*, and Jesús Gutiérrez-Gutiérrez

***Abstract*—Toeplitz matrices and functions of Toeplitz matrices (such as the inverse of a Toeplitz matrix, powers of a Toeplitz matrix or the exponential of a Toeplitz matrix) arise in many different theoretical and applied fields. They can be found in the mathematical modeling of problems where some kind of shift invariance occurs in terms of space or time. R. M. Gray's excellent tutorial monograph on Toeplitz and circulant matrices has been, and remains, the best elementary introduction to the Szegö distribution theory on the asymptotic behavior of continuous functions of Toeplitz matrices. His asymptotic results, widely used in engineering due to the simplicity of its mathematical proofs, do not concern individual entries of these matrices but rather, they describe an "average" behavior. However, there are important applications where the asymptotic expressions of interest are directly related to the convergence of a single entry of a continuous function of a Toeplitz matrix. Using similar mathematical tools and to gain insight into the solutions of this sort of problems, the present correspondence derives new theoretical results regarding the convergence of these entries.**

***Index Terms*—Circulant matrices, covariance matrices, elementwise convergence, functions of matrices, minimum mean square error (MMSE), stationary stochastic time series, Szegö's theorem, Toeplitz matrices.**

I. INTRODUCTION

A square Toeplitz matrix is an $n \times n$ matrix $T_n = \left(t_{i,j}^{(n)}\right)$ where $t_{i,j}^{(n)} = t_{i-j}$ are complex numbers. Toeplitz matrices arise in many theoretical and applied fields. In particular, there are many signal processing applications, where the Toeplitz matrix T_n is often Hermitian. Generally speaking, what is really relevant in many of the associated problems is a continuous function of the Toeplitz matrix

$$g(T_n) := U_n \mathrm{diag}(g(\lambda_1(T_n)), \ldots, g(\lambda_n(T_n)))U_n^{-1}$$

Manuscript received May 13, 2006. This work was supported in part by the Spanish Ministry of Education and Science, by the European Regional Development Fund and by the European Social Fund through the MIMESIS project (no. TEC2004-06451-C05-04/TCM) and the Torres-Quevedo program.

The authors are with CEIT and Tecnun (University of Navarra), 20018 San Sebastián, Spain (e-mail: pcrespo@ceit.es; jgutierrez@ceit.es).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2006.890697