A Data-Bearing Approach for Pilot-Embedding Frameworks in Space-Time Coded MIMO Systems

Chaiyod Pirak, *Member, IEEE*, Z. Jane Wang, *Member, IEEE*, K. J. Ray Liu, *Fellow, IEEE*, and Somchai Jitapunkul, *Member, IEEE*

***Abstract*—Space-time (ST) coded MIMO systems employing coherent detectors crucially require channel state information. This paper presents a novel pilot-embedding framework for channel estimation and data detection by exploiting the null-space property and the orthogonality property of the data-bearer and pilot matrices. The ST data matrix is firstly projected onto the data bearer matrix, which is a null-space of the pilot matrix, and the resulting matrix and the pilot matrix are combined for transmitting. The data and pilot extractions are achieved independently through linear transformations by exploiting the null-space property. The unconstrained maximum-likelihood (ML) and linear minimum mean-squared error (lmmse) estimators are explored for channel estimation. Then the ML approach for data detection is developed by exploiting the orthogonality property. The mean-squared error (mse) of channel estimation, Cramer-Rao lower bound (CRLB), and the Chernoff's bound of a pair-wise error probability for ST codes are analyzed for examining the performance of the proposed scheme. The optimum power allocation scheme for data and pilot parts is also considered. Three data-bearer and pilot structures, including time-multiplexing (TM)-based, ST-block-code (STBC)-based, and code-multiplexing (CM)-based, are proposed. Simulation results show that the CM-based structure provides superior performance for nonquasi-static flat Rayleigh fading channels, while these three structures yield similar performances for quasi-static flat Rayleigh fading channels.**

***Index Terms*—Channel estimation, code-multiplexing based training, multiple-input multiple-output (MIMO), pilot embedding, space-time (ST).**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) communication systems provide prominent benefits to wireless communications due to the high capacity and reliability they can offer [1], [2]. Recently, the space-time (ST) codes have been proposed in [3] and [4] for MIMO communications, in which the bit error rate (BER) of the systems is significantly improved without increasing transmission power by exploiting transmit diversity [3].

A major challenge in wireless ST communications employing a coherent detector is the channel state information acquisition [3], [4]. Typically, the channel state information is acquired or estimated by using a pilot or training signal, a known signal transmitted from the transmitter to the receiver. This technique has been widely applied because of its feasibility for implementation with low computational complexity [5].

Two main pilot-aided channel estimation techniques have been proposed in both single-input single-output (SISO) and MIMO systems: the pilot symbol assisted modulation (PSAM) technique and the pilot-embedding technique. In the SISO system, the PSAM technique has been intensively studied in [5] for frequency-nonselective fading channels, and was recently extended to MIMO systems [6]–[11]. In this technique, first, a pilot signal is time-multiplexed into a transmit data stream, and then, at the receiver side, this pilot signal is extracted from the received signal to acquire the channel state information. Furthermore, an interpolation technique by averaging channel estimates over a certain time period is employed in order to improve the accuracy of the channel estimates. The disadvantage of this technique is the sparse pilot arrangement that results in poor tracking of channel variations. In addition, the denser the pilot signals, the poorer the bandwidth efficiency.

The pilot-embedding, also referred as pilot-superimposed technique, has been proposed for the SISO systems [12] and for the MIMO systems [13]–[15], where a sequence of pilot signals is added directly to the data stream. Some soft-decoding methods, such as Viterbi algorithm [12], [14] are employed for channel estimation and data detection. This technique yields better bandwidth efficiency, since it does not sacrifice any separate time slots for transmitting the pilot signal. The disadvantages of this technique lie in the higher computational complexity of the decoder and the longer delay in channel estimation process.

Our purpose is to design a novel pilot-embedding approach for ST coded MIMO systems with affordable computational cost and better fast-fading channel acquisition. The basic idea is to simplify channel estimation and data detection processes by taking advantages of the null-space and orthogonality properties of the data-bearer and pilot matrices. The data-bearer matrix is used for projecting the ST data matrix onto the orthogonal subspace of the pilot matrix. By the virtue of the null-space and

Manuscript received March 10, 2005; revised November 14, 2005; accepted November 16, 2005. This work was supported in part by a Ph.D. scholarship from Commission on Higher Education, Ministry of Education, Thai Government and a Grant from the Cooperation Project between the Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University, Thailand. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mats Viberg.

C. Pirak is with the Electrical and Computer Engineering Department, University of Maryland College Park, MD 20742 USA, and also with the Electrical Engineering Department, Chulalongkorn University, Bangkok 10330, Thailand (e-mail: chaiyod.p@gmail.com).

Z. J. Wang is with the Electrical and Computer Engineering Department, University of British Columbia, BC V6T 1Z4, Canada (e-mail: zjanew@ece.ubc.ca).

K. J. R. Liu is with the Electrical and Computer Engineering Department and the Institute for Systems Research, University of Maryland College Park, MD 20742 USA (e-mail: kjrliu@isr.umd.edu).

S. Jitapunkul is with the Electrical Engineering Department, Chulalongkorn University, Bangkok 10330, Thailand (e-mail: Somchai.J@chula.ac.th).

Digital Object Identifier 10.1109/TSP.2006.880046

orthogonality properties, in our proposed data-bearing approach for pilot-embedding, a block of data matrix is added into a block of pilot matrix, that are mutually orthogonal to each other. The benefit that we are able to expect from this approach is better channel estimation performance, since the estimator can take into account the channel variation in the transmitted data block. In addition, a low computational complexity channel estimator is also expected.

Now let us briefly describe the MIMO channel and system model. We consider the MIMO communication system with L_t transmit antennas and L_r receive antennas. In general, for a given block index t, a ST symbol matrix $\mathbf{U}(t)$ is an $L_t \times M$ codeword matrix transmitted across the transmit antennas in M time slots. The received symbol matrix $\mathbf{Y}(t)$ at the receiver front-end can be expressed as follows [14]:

$$\mathbf{Y}(t) = \mathbf{H}(t)\mathbf{U}(t) + \mathbf{N}(t) \tag{1}$$

where $\mathbf{H}(t)$ is the $L_r \times L_t$ channel coefficient matrix and the $L_r \times M$ additive noise matrix $\mathbf{N}(t)$ is complex white Gaussian distributed with zero mean and variance $(\sigma^2/2)\mathbf{I}_{(L_r M \times L_r M)}$ per real dimension. The elements of $\mathbf{H}(t)$ are assumed to be independent complex Gaussian random variables with zero mean and variance 0.5 per real dimension. Or equivalently, an independent Rayleigh fading channel is assumed. In this paper, we first examine a quasi-static flat Rayleigh fading channel, where $\mathbf{H}(t)$ remains constant over each symbol block but it changes block-by-block independently. Then, we extend our proposed scheme in a nonquasi-static flat Rayleigh fading channel, where $\mathbf{H}(t)$ is not constant over each symbol block. Our problems are to estimate the channel coefficient matrix $\mathbf{H}(t)$ and the ST symbol matrix $\mathbf{U}(t)$ by using the pilot or training signal embedded in $\mathbf{U}(t)$.

The rest of this paper is organized as follows. We present the proposed data-bearing approach for pilot-embedding frameworks in Section II, including general properties needed, channel estimation process, possible data bearer and pilot matrices, and data detection process. Performance analysis for the proposed scheme is carried out in Section III, in terms of channel estimation and data detection. In Section IV, we address the issue of optimum block power allocation for data and pilot parts. The simulation results are given in Section V, and we conclude this paper in Section VI.

II. The Proposed Data-Bearing Approach for Pilot-Embedding

In this section, we present the proposed data-bearing approach for pilot-embedding, including the pilot and data extraction procedures, channel estimation, possible data bearer and pilot matrices, and data detection. Our motivation of pursuing pilot-embedding by distributing the pilot signal onto the ST data is to capture the variation of the channel at every instant for achieving a better channel estimate. Without loss of generality, we describe our data matrix $\mathbf{Z}(t) \in \mathbb{C}^{L_t \times M}$ as follows:

$$\mathbf{Z}(t) = \mathbf{D}(t)\mathbf{A} \tag{2}$$

where $\mathbf{D}(t) \in \mathbb{C}^{L_t \times N}$ is the ST data matrix, and $\mathbf{A} \in \mathbb{R}^{N \times M}$ is the data-bearer matrix with N being the number of data time slots. It is worth mentioning that $N < M$ because the excessive time slot, i.e., $M - N$, will be used for providing a room to embed pilot signals. In our implementation, the ST data matrix $\mathbf{D}(t)$ is assumed to maintain the energy constraint $\mathrm{E}[\|\mathbf{D}(t)\|^2] = L_t$ with $\|\cdot\|$ being the Frobenius norm. The proposed pilot-embedded ST symbol matrix $\mathbf{U}(t)$ can be expressed as follows:

$$\mathbf{U}(t) = \mathbf{Z}(t) + \mathbf{P} = \mathbf{D}(t)\mathbf{A} + \mathbf{P} \tag{3}$$

where $\mathbf{P} \in \mathbb{R}^{L_t \times M}$ is the pilot matrix. Unlike the pilot-embedding technique previously proposed in [14] where the pilot-embedded ST symbol matrix $\mathbf{U}(t)$ is expressed as $\mathbf{U}(t) = \mathbf{D}(t) + \mathbf{P}$, the major difference of our proposed scheme in (3) is the exploitation of the data-bearer matrix $\mathbf{A}$, which plays the major role along with the pilot matrix $\mathbf{P}$ in the channel estimation and data detection processes.

By the data-bearing approach for pilot-embedding, we require that the data bearer matrix $\mathbf{A}$ and the pilot matrix $\mathbf{P}$ satisfy the following properties:

$$\mathbf{A}\mathbf{P}^T = \mathbf{0} \in \mathbb{R}^{N \times L_t} \tag{4}$$

$$\mathbf{P}\mathbf{A}^T = \mathbf{0} \in \mathbb{R}^{L_t \times N} \tag{5}$$

$$\mathbf{A}\mathbf{A}^T = \beta\mathbf{I} \in \mathbb{R}^{N \times N} \tag{6}$$

$$\mathbf{P}\mathbf{P}^T = \alpha\mathbf{I} \in \mathbb{R}^{L_t \times L_t} \tag{7}$$

where β is a real-valued data-power factor for controlling the value of data-part power, α is a real-valued pilot-power factor for controlling the value of pilot-part power, $\mathbf{0}$ stands for an all-zero-element matrix, and $\mathbf{I}$ stands for an identity matrix. The key concept of our pilot-embedding approach is the exploitation of the null-space properties [16], i.e., the properties (4) and (5), and the orthogonality properties [16], i.e., the properties (6) and (7), about the data-bearer matrix $\mathbf{A}$ and the pilot matrix $\mathbf{P}$. Obviously, in (3), the data-bearer matrix $\mathbf{A}$ plays a major role in projecting the ST data matrix $\mathbf{D}(t)$ onto the orthogonal subspace of the pilot matrix $\mathbf{P}$. From (6) and (7), we can imply that $\mathrm{Rank}(\mathbf{A}) = N$ and $\mathrm{Rank}(\mathbf{P}) = L_t$ with $\mathrm{Rank}(\cdot)$ being the rank of a matrix. In order to satisfy the null-space properties in (4) and (5), the minimum number of the column in $\mathbf{A}$ and $\mathbf{P}$ must be equal to the sum of the rank of $\mathbf{A}$ and $\mathbf{P}$ [17]. Consequently, the number of time slots M of the pilot-embedded ST symbol matrix $\mathbf{U}(t)$ must satisfy the following equality:

$$\mathrm{Rank}(\mathbf{A}) + \mathrm{Rank}(\mathbf{P}) = M. \tag{8}$$

The proposed pilot-embedded ST symbol block structure is demonstrated in Fig. 1. The proposed pilot-embedded ST symbol block $\mathbf{U}(t)$ consists of two main parts: data sequences $\{\mathbf{Z}(t)\}_i$ and pilot sequences $\{\mathbf{P}\}_i$, where i stands for a row index, $i = 1, \ldots, L_t$. Substituting (3) into (1), the received symbol matrix $\mathbf{Y}(t)$ in (1) can be rewritten as follows:

$$\mathbf{Y}(t) = \mathbf{H}(t)(\mathbf{D}(t)\mathbf{A} + \mathbf{P}) + \mathbf{N}(t). \tag{9}$$

There are at least three possible structures of data-bearer and pilot matrices, in which the elements of these matrices are real numbers, that satisfy the properties (4)–(7) as follows.

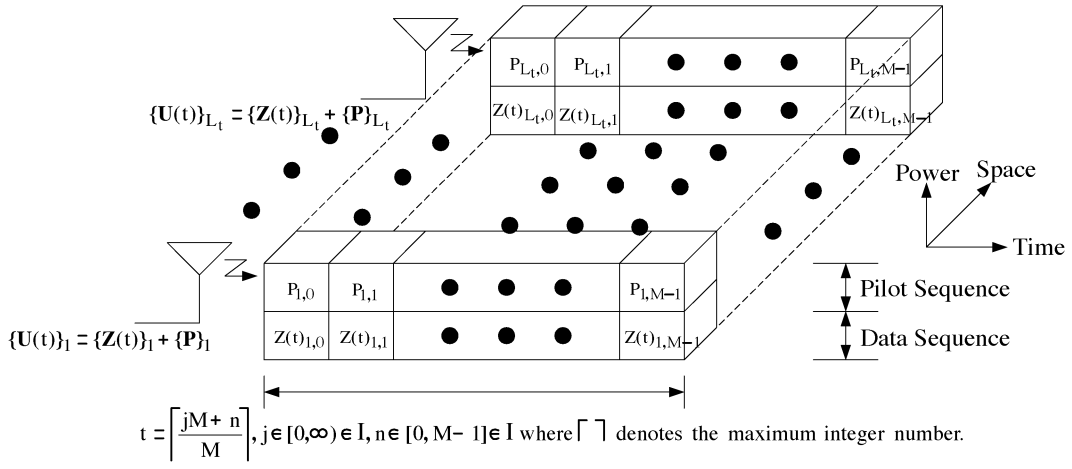


Fig. 1. The proposed pilot-embedded ST symbol block structure.

1) Time-Multiplexing (TM)-Based Data-Bearer and Pilot Matrices: The structures of these matrices are given as

$$\mathbf{A} = \sqrt{\beta}\left[\mathbf{0}_{(N\times L_t)}; \mathbf{I}_{(N\times N)}\right]$$
$$\mathbf{P} = \sqrt{\alpha}\left[\mathbf{I}_{(L_t\times L_t)}; \mathbf{0}_{(L_t\times N)}\right],\ M = N + L_t \qquad (10)$$

where ; denotes matrix combining. In this structure, the $L_t \times L_t$ identity matrix $\mathbf{I}$ is used as a pilot or training symbol. In addition, PSAM belongs to this category [5], because it employs the time-multiplexing structure for pilot and data allocation, and has been used in [7]–[10]. Therefore, the existing PSAM technique is subsumed in the proposed general idea in (3).

2) ST-Block-Code (STBC)-Based Data-Bearer and Pilot Matrices: The structures of these matrices are given as

$$\mathbf{A} = \sqrt{\beta}\left[\mathbf{0}_{(N\times \tau)}; \mathbf{I}_{(N\times N)}\right]$$
$$\mathbf{P} = \sqrt{\alpha}\left[\mathrm{STBC}_{(L_t\times \tau)}; \mathbf{0}_{(L_t\times N)}\right],\ M = N + \tau \qquad (11)$$

where τ is the number of time slots used for transmitting one ST block code. In addition, τ is greater than or equal to L_t, i.e., $\tau \geq L_t$, depending on the structure of the chosen ST block code. In this structure, the major difference from the TM-based structure is that it employs the normalized known ST block code [4] as the pilot symbol instead of using the identity matrix. It also inherits the time-multiplexing structure in pilot and data allocation. This kind of data bearer and pilot matrices have been used in [6], for instance.

3) Code-Multiplexing (CM)-Based Data-Bearer and Pilot Matrices: The structures of these matrices are given as

$$\mathbf{A} = \sqrt{\beta}\mathrm{WH}[1:N]_{(N\times M)},$$
$$\mathbf{P} = \sqrt{\alpha}\mathrm{WH}[N+1:M]_{(L_t\times M)},\ M = N + L_t \qquad (12)$$

where $\mathrm{WH}[x:y]$ denotes a submatrix created by splitting the $M \times M$ normalized Walsh-Hadamard matrix [18] starting from x^{th}-row to y^{th}-row. Because the power is evenly distributed over all entries of these matrices, i.e., no zero-entry is contained in these matrices, we expect that their channel estimation performance is superior to the other two structures. The disadvantage of this structure is the limitation of dimensionality of Walsh-Hadamard matrix, which has a dimension proportionally to 2^n, $n \in \mathbb{I}$. In addition, this structure provides an instructive example of the proposed general idea in (3) for pilot-embedding.

It is worth mentioning that our proposed data-bearing approach for pilot-embedding frameworks subsumes the general idea of the existing pilot-based techniques, i.e., PSAM and pilot-embedding techniques. Furthermore, our designed criteria in (6) and (7), and the above three examples satisfy the optimal designed criteria in [10], i.e., the optimal training data and the optimal training interval length, respectively. In addition, the property in (7) is optimal in the sense that the bandwidth efficiency loss due to the pilot transmission is proportional to the factor $(M - L_t)/M$ [10], so that all three structures yield the same loss, provided that the case of $M = N + L_t$ is considered. In what follows, we further consider the problems of channel estimation and ST data detection by using the aforementioned data-bearing approach.

A. Channel Estimation

The channel estimation of our proposed data-bearing approach for pilot-embedding frameworks can be achieved by first simply postmultiplying the received symbol matrix $\mathbf{Y}(t)$ in (9) by the transpose of the pilot matrix $\mathbf{P}^T$ for extracting the pilot part. Using (4) and (7), and dividing the result by α, thus, yielding

$$\frac{\mathbf{Y}(t)\mathbf{P}^T}{\alpha} = \mathbf{H}(t) + \frac{\mathbf{N}(t)\mathbf{P}^T}{\alpha}. \qquad (13)$$

Let us denote $\mathbf{y}(t) = \mathbf{vec}(\mathbf{Y}(t)\mathbf{P}^T/\alpha)$, $\mathbf{n}(t) = \mathbf{vec}(\mathbf{N}(t)\mathbf{P}^T/\alpha)$, and $\mathbf{h}(t) = \mathbf{vec}(\mathbf{H}(t))$ with $\mathbf{vec}(\cdot)$ being the vectorization conversion [19], hence, (13) can be rewritten as follows:

$$\mathbf{y}(t) = \mathbf{h}(t) + \mathbf{n}(t). \qquad (14)$$

For the pilot-projected noise vector $\mathbf{n}(t)$, using the following relationship [19]:

$$\mathbf{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathbf{vec}(\mathbf{B}) \qquad (15)$$

where $\otimes$ is the Kronecker product, we have

$$\mathbf{n}(t) = \frac{1}{\alpha}\left(\mathbf{I} \otimes \mathbf{N}(t)\right)\mathbf{vec}(\mathbf{P}^T). \tag{16}$$

From the white Gaussian-distributed assumption of $\mathbf{N}(t)$, the mean vector and the covariance matrix of the pilot-projected noise vector $\mathbf{n}(t)$ are determined as follows:

$$\boldsymbol{\mu}_{\mathbf{n}(t)} = \frac{1}{\alpha}\mathrm{E}\left[\left(\mathbf{I} \otimes \mathbf{N}(t)\right)\mathbf{vec}(\mathbf{P}^T)\right] = \mathbf{0}_{(L_tL_r\times 1)} \tag{17}$$

$$\begin{aligned}\mathbf{V}_{\mathbf{n}(t)} &= \frac{1}{\alpha^2}\mathrm{E}\Big[\left(\left(\mathbf{I} \otimes \mathbf{N}(t)\right)\mathbf{vec}(\mathbf{P}^T)\right) \\ &\qquad \times \left(\left(\mathbf{I} \otimes \mathbf{N}(t)\right)\mathbf{vec}(\mathbf{P}^T)\right)^H\Big] \\ &= \frac{\sigma^2}{2\alpha^2}\mathrm{Diag}(B_i) \\ B_i &= \sum_{j=1}^{M}|P_{i,j}|^2\mathbf{I}_{(L_r\times L_r)},\ i \in \{1,\ldots,L_t\} \\ &\qquad \text{per real dimension}\end{aligned} \tag{18}$$

where $P_{i,j}$ is the $(i,j)^{\text{th}}$ element of the pilot matrix $\mathbf{P}$, $\mathrm{Diag}(\cdot)$ stands for the diagonal matrix created by concatenating submatrices $B_i, i \in 1,\ldots,L_t$ into the diagonal elements.

From (7), it can be shown that $\sum_{j=1}^{M}|P_{i,j}|^2 = \alpha, \forall i$. Hence, we can rewrite (18) as follows:

$$\mathbf{V}_{\mathbf{n}(t)} = \frac{\sigma^2}{2\alpha}\mathbf{I}_{(L_tL_r\times L_tL_r)},\ \text{per real dimension.} \tag{19}$$

Obviously, the pilot-projected noise vector $\mathbf{n}(t)$ is a complex white Gaussian vector, hence, the log-likelihood function $\ln(p(\mathbf{y}(t)|\mathbf{h}(t)))$ is given by [20]

$$\begin{aligned}\ln\left(p\left(\mathbf{y}(t)|\mathbf{h}(t)\right)\right) &= \ln\left(\frac{1}{\pi^{L_tL_r}det(\mathbf{V}_{\mathbf{n}(t)})}\right) \\ &\quad -\left(\mathbf{y}(t)-\mathbf{h}(t)\right)^H\mathbf{V}_{\mathbf{n}(t)}^{-1}\left(\mathbf{y}(t)-\mathbf{h}(t)\right).\end{aligned} \tag{20}$$

1) Unconstrained Maximum-Likelihood (ML) Channel Estimator: It is straightforward to show that the maximum-likelihood estimator [20] maximizing the log-likelihood function $\ln(p(\mathbf{y}(t)|\mathbf{h}(t)))$ is as follows:

$$\hat{\mathbf{h}}(t) = \max_{\mathbf{h}(t)}\left\{\ln\left(p\left(\mathbf{y}(t)|\mathbf{h}(t)\right)\right)\right\} = \mathbf{y}(t) \text{ or } \hat{\mathbf{H}}(t) = \frac{\mathbf{Y}(t)\mathbf{P}^T}{\alpha} \tag{21}$$

meaning that the unconstrained ML estimator is the pilot-projected received vector $\mathbf{y}(t)$ itself.

2) Linear Minimum Mean-Squared Error (lmmse) Channel Estimator: We further improve the performance of the unconstrained ML channel estimator in (21) by employing the L-tap lmmse channel interpolation. The L-tap lmmse channel interpolation interpolates the last L channel estimates estimated in the last L ST symbol blocks. The L-tap lmmse channel estimator can be expressed as follows:

$$h_{j,i}^{\text{lmmse}}(t) = \mathbf{w}_{j,i}^H\hat{\mathbf{h}}_{j,i}^L(t) \tag{22}$$

where $h_{j,i}^{\text{lmmse}}(t)$ denotes the $(j,i)^{\text{th}}$ element of the lmmse-estimated channel matrix, $\mathbf{w}_{j,i} = [w_{j,i}(0)\cdots w_{j,i}(L-1)]^T$ denotes the L-tap finite impulse response (FIR) linear filter's weight vector, and $\hat{\mathbf{h}}_{j,i}^L(t) = [\hat{h}_{j,i}(t)\cdots\hat{h}_{j,i}(t-L+1)]^T$ denotes the L-element input vector constructed from the $(j,i)^{th}$ element of the ML-estimated channel matrix in (21) taking values corresponding to the block indices $[t-L+1,t]$, or equivalently corresponding to the time interval $[(t-L+1)M, Mt]$. The optimization criterion, assuming the channels are wide-sense stationary (WSS), for the L-tap lmmse channel estimator is given by

$$J(\mathbf{w}_{j,i}) = \arg\min_{\mathbf{w}_{j,i}}\mathrm{E}\left[\left\|h_{j,i}(t) - \mathbf{w}_{j,i}^H\hat{\mathbf{h}}_{j,i}^L(t)\right\|^2\right] \tag{23}$$

where $h_{j,i}(t)$ denotes the $(j,i)^{\text{th}}$ element of the true channel matrix $\mathbf{H}(t)$ in (1).

The optimum lmmse weight vector $\mathbf{w}_{j,i}^{\text{opt}}$ is given by

$$\mathbf{w}_{j,i}^{\text{opt}} = \mathbf{R}_{\hat{\mathbf{h}}_{j,i}^L(t)}^{-1}\mathbf{p}_{\hat{\mathbf{h}}_{j,i}^L(t)} \tag{24}$$

where $\mathbf{R}_{\hat{\mathbf{h}}_{j,i}^L(t)} = \mathrm{E}[\hat{\mathbf{h}}_{j,i}^L(t)\hat{\mathbf{h}}_{j,i}^{H_L}(t)]$ and $\mathbf{p}_{\hat{\mathbf{h}}_{j,i}^L(t)} = \mathrm{E}[h_{j,i}^*(t)\,\hat{\mathbf{h}}_{j,i}^L(t)]$. According to (14), (19), (21), and the uncorrelatedness of the channel and noise coefficients, the L-tap lmmse channel estimator can be further rewritten as

$$h_{j,i}^{\text{lmmse}}(t) = \left[\left(\mathbf{R}_{\mathbf{h}_{j,i}^L(t)} + \frac{\sigma^2}{\alpha}\mathbf{I}_{L\times L}\right)^{-1}\mathbf{p}_{\mathbf{h}_{j,i}^L(t)}\right]^H\hat{\mathbf{h}}_{j,i}^L(t) \tag{25}$$

where $\mathbf{R}_{\mathbf{h}_{j,i}^L(t)} = \mathrm{E}[\mathbf{h}_{j,i}^L(t)\mathbf{h}_{j,i}^{H_L}(t)]$ and $\mathbf{p}_{\hat{\mathbf{h}}_{j,i}^L(t)} = \mathrm{E}[h_{j,i}^*(t)\,\mathbf{h}_{j,i}^L(t)] = \mathbf{p}_{\mathbf{h}_{j,i}^L(t)}$. The performance analysis for the unconstrained ML channel estimator will be considered in Section III-A-1). In addition, the performance analysis for the lmmse channel estimator can be found in [10].

B. Data Detection

We further describe the data detection procedure. First, the data part in the received symbol matrix $\mathbf{Y}(t)$ is extracted by postmultiplying the received symbol matrix $\mathbf{Y}(t)$ by the transpose of the data-bearer matrix $\mathbf{A}^T$. Using (5) and (6), we have

$$\frac{\mathbf{Y}(t)\mathbf{A}^T}{\beta} = \mathbf{H}(t)\mathbf{D}(t) + \frac{\mathbf{N}(t)\mathbf{A}^T}{\beta}. \tag{26}$$

Let us define $\mathbf{n}'(t) = \mathbf{vec}(\mathbf{N}(t)\mathbf{A}^T/\beta)$. From (6), it can be shown that $\sum_{j=1}^{M}|A_{i,j}|^2 = \beta, \forall i$. Then similar to (17) and (18), the mean vector and the covariance matrix of the data-bearer-projected noise vector $\mathbf{n}'(t)$ are determined as follows:

$$\boldsymbol{\mu}_{\mathbf{n}'(t)} = \mathbf{0}_{(L_rN\times 1)} \tag{27}$$

$$\mathbf{V}_{\mathbf{n}'(t)} = \frac{\sigma^2}{2\beta}\mathbf{I}_{(L_rN\times L_rN)}\ \text{per real dimension.} \tag{28}$$

The ML receiver is employed for decoding the transmitted ST data matrix $\mathbf{D}(t)$ by using the estimated channel coefficient matrix $\hat{\mathbf{H}}(t)$ obtained in either (21) or (25) as the channel state in-

formation. Due to the i.i.d. white Gaussian distribution of $\mathbf{n}'(t)$, the ML receiver computes the decision metric and decides the codeword that minimizes this decision metric as in [3]

$$\left\{\hat{d}_k^i\right\} = \min_{\{d_k^i\}} \left\{ \sum_{k=1}^{N} \sum_{j=1}^{L_r} \left| y_k^j - \sum_{i=1}^{L_t} \hat{h}_{j,i} d_k^i \right|^2 \right\} \quad \forall d_k^i,\ i \in \{1,\ldots,L_t\},\ k \in \{1,\ldots,N\} \tag{29}$$

where y_k^j denotes the $(j,k)^{\text{th}}$ element of the data-bearer-projected received symbol matrix $\mathbf{Y}(t)\mathbf{A}^T/\beta$, $\hat{h}_{j,i}$ denotes the $(j,i)^{\text{th}}$ element of the estimated channel coefficient matrix $\hat{\mathbf{H}}(t)$, and $\hat{d}_k^i$ denotes the $(i,k)^{\text{th}}$ element of the estimated ST data matrix $\hat{\mathbf{D}}(t)$.

The performance analysis for the ST data detection is discussed in Section III-A-2). Note that the null-space and orthogonality properties of the data-bearer matrix $\mathbf{A}$ and the pilot matrix $\mathbf{P}$ play the major role in the pilot and the data extraction for channel estimation and data detection, respectively. In addition, the ranks of the data-bearer matrix $\mathbf{A}$ and the pilot matrix $\mathbf{P}$ also determine the minimum number of time slots, obtained in (8), of the pilot-embedded ST symbol matrix $\mathbf{U}(t)$.

III. The Performance Analysis for The Proposed Scheme

In this section, we analyze the performances of our data-bearing approach for pilot-embedding frameworks, including both the unconstrained ML channel estimation and data detection performance, under two different scenarios, i.e., quasi-static and nonquasi-static flat Rayleigh fading channels. We use the analysis as the theoretical benchmarks for later comparisons in Section V.

A. Quasi-Static Flat Rayleigh Fading Channels

1) Channel Estimation Performance Analysis: We analyze the channel estimation error first, and then compute the Cramer-Rao lower bound (CRLB), which is a lower bound of the conditional variance of the channel estimation error. A channel estimation error vector can be evaluated as follows:

$$\tilde{\mathbf{h}}(t) = \mathbf{h}(t) - \hat{\mathbf{h}}(t). \tag{30}$$

Substituting (21) into (30) and using the fact that $\mathbf{n}(t)$ is the AWGN with zero-mean and variance expressed in (19), the covariance matrix of the channel estimation error is given by

$$\begin{aligned}\text{Cov}\left[\tilde{\mathbf{h}}(t)\right] &= \text{E}\left[\left(\mathbf{h}(t)-\hat{\mathbf{h}}(t)-\mathbf{n}(t)\right)\left(\mathbf{h}(t)-\hat{\mathbf{h}}(t)-\mathbf{n}(t)\right)^H\right] \\ &= \mathbf{V}_{\mathbf{n}(t)}.\end{aligned} \tag{31}$$

The mean-squared error (mse) of the channel estimation is given by

$$\text{mse} = \text{tr}\left\{\text{Cov}\left[\tilde{\mathbf{h}}(t)\right]\right\} = \frac{\sigma^2 L_t L_r}{\alpha} \tag{32}$$

where $\text{tr}\{\cdot\}$ stands for the trace operator of a matrix. It is worth noticing that the mse of the ML channel estimation is inversely proportional to the pilot-power factor α; as a result, more power allocated to the pilot part resulting in lower mse of the channel estimation. Since the pilot-projected noise vector $\mathbf{n}(t)$ is the Gaussian distributed random vector with zero-mean and variance expressed in (19), the ML channel estimator is efficient and unbiased, and it achieves the CRLB [20], [21]. It can be shown that the CRLB for an unbiased estimator is given by [20]

$$\begin{aligned}\text{Cov}\left[\hat{\mathbf{h}}(t) - \mathbf{h}(t)|\mathbf{h}(t)\right] &= \left[-\text{E}\left[\frac{\partial^2 \ln\left(p\left(\mathbf{y}(t)|\mathbf{h}(t)\right)\right)}{\partial \mathbf{h}^2(t)}\right]\right]^{-1} \\ &= \mathbf{V}_{\mathbf{n}(t)}.\end{aligned} \tag{33}$$

The trace of the CRLB matrix in (33) is given by

$$\text{tr}\left\{\text{Cov}\left[\hat{\mathbf{h}}(t) - \mathbf{h}(t)|\mathbf{h}(t)\right]\right\} = \frac{\sigma^2 L_t L_r}{\alpha}. \tag{34}$$

Therefore, one can see that the channel estimator achieves the desired properties of a good estimator that is unbiased, and achieves the CRLB.

2) Data Detection Performance Analysis: We further analyze the probability of error of the proposed scheme in data detection. First, the data-bearer-projected received signal $\mathbf{Y}(t)\mathbf{A}^T/\beta$ in (26) can be alternatively represented by

$$\frac{\mathbf{Y}(t)\mathbf{A}^T}{\beta} = \hat{\mathbf{H}}(t)\mathbf{D}(t) + \left(\mathbf{H}(t) - \hat{\mathbf{H}}(t)\right)\mathbf{D}(t) + \frac{\mathbf{N}(t)\mathbf{A}^T}{\beta}. \tag{35}$$

Let us define $\mathbf{N}_A(t) = \tilde{\mathbf{H}}(t)\mathbf{D}(t) + (\mathbf{N}(t)\mathbf{A}^T/\beta)$ an additive noise taking into account both channel estimation error and noise, where $\tilde{\mathbf{H}}(t) = \mathbf{H}(t) - \hat{\mathbf{H}}(t)$ denotes the matrix-based channel estimation error. It can be shown that the ML channel estimation error can be expressed as follows:

$$\tilde{\mathbf{H}}(t) = -\frac{\mathbf{N}(t)\mathbf{P}^T}{\alpha} \tag{36}$$

by substituting $\hat{\mathbf{H}}(t)$ in (21). Hence, we have

$$\mathbf{N}_A(t) = \mathbf{N}(t)\left(\frac{\mathbf{A}^T}{\beta} - \frac{\mathbf{P}^T\mathbf{D}(t)}{\alpha}\right). \tag{37}$$

Next, we find the statistics of $\mathbf{N}_A(t)$. Since $\mathbf{N}(t)$ is the AWGN with zero-mean and variance $(\sigma^2/2)\mathbf{I}_{(L_r M \times L_r M)}$ per real dimension, it can be shown that $\mathbf{N}_A(t)$ has a zero-mean. The variance of the element of $\mathbf{N}_A(t)$ can be computed by

$$N_0 = \frac{1}{L_r N}\text{tr}\left\{\text{E}\left[\mathbf{N}_A(t)\mathbf{N}_A^H(t)\right]\right\} \tag{38}$$

where $L_r N$ is the number of elements of $\mathbf{N}_A(t)$. Substituting $\mathbf{N}_A(t)$ into (38); and using the fact that $\text{tr}\{\text{E}[\mathbf{N}(t)\mathbf{A}^T\mathbf{A}\mathbf{N}(t)^H/\beta^2]\} = \sigma^2 L_r N/\beta$ (see also (28)); and the channel estimation error matrix $\tilde{\mathbf{H}}(t)$, the ST coded data

matrix $\mathbf{D}(t)$, and the noise matrix $\mathbf{N}(t)\mathbf{A}^T/\beta$ are statistically independent, N_0 can be expressed as follows:

$$N_0 = \frac{1}{L_r N}\mathrm{tr}\left\{\mathrm{E}\left[\tilde{\mathbf{H}}(t)^H\tilde{\mathbf{H}}(t)\right]\mathrm{E}\left[\mathbf{D}(t)\mathbf{D}(t)^H\right]\right\} + \frac{\sigma^2}{\beta}. \tag{39}$$

Since the transmit antennas are used independently with equal power, which is reasonable because no channel knowledge is assumed at the transmitter, we then have $\mathrm{E}[\mathbf{D}(t)\mathbf{D}(t)^H] = \mathbf{I}_{(L_t\times L_t)}$. In fact, this property is also true for an orthogonal STBC [4], and it has been shown in [10] that this property can maximize the mutual information of the ST coded MIMO systems given $\hat{\mathbf{H}}(t)$. From (19), it can be shown that $\mathrm{tr}\{\mathrm{E}[\tilde{\mathbf{H}}(t)^H\tilde{\mathbf{H}}(t)]\} = \mathrm{tr}\{\mathrm{E}[\mathbf{N}(t)\mathbf{P}^T\mathbf{P}\mathbf{N}(t)^H/\alpha^2]\} = \sigma^2 L_t L_r/\alpha$. Hence, N_0 can be expressed as, by substituting all above derivations

$$N_0 = \frac{\sigma^2}{N}\left(\frac{L_t}{\alpha} + \frac{N}{\beta}\right). \tag{40}$$

In summary, $\mathbf{N}_A(t)$ is the AWGN matrix whose element has zero-mean and variance $N_0/2$ per real dimension. For the data-bearer-projected received signal in (35), we can derive Chernoff's upper bound of the probability of transmitting a codeword $\mathbf{d} \triangleq (d_1^1 d_1^2 \cdots d_1^{L_t} \cdots d_N^1 d_N^2 \cdots d_N^{L_t})^T$ and deciding in favor of a different codeword $\mathbf{e} \triangleq (e_1^1 e_1^2 \cdots e_1^{L_t} \cdots e_N^1 e_N^2 \cdots e_N^{L_t})^T$ at the maximum-likelihood receiver in a similar way to [22] as follows:

$$P\left(\mathbf{d}\to\mathbf{e}|\hat{h}_{j,i},\ j=1,\ldots,L_r,\ i=1,\ldots,L_t\right) \le \exp\left(\frac{-m^2(\mathbf{d},\mathbf{e})}{4N_0}\right) \tag{41}$$

where $m^2(\mathbf{d},\mathbf{e}) = \sum_{j=1}^{L_r}\sum_{t=1}^{N}|\sum_{i=1}^{L_t}\hat{h}_{j,i}(d_t^i - e_t^i)|^2$.

In our analysis, the Chernoff's upper bound of the average probability of error with respect to independent Rayleigh distributions of the channel is expressed as (see also Appendix I)

$$P(\mathbf{d}\to\mathbf{e})_{\hat{\mathbf{H}}(t)} \le \left(\prod_{i=1}^{L_\Delta}\lambda_i\right)^{-L_r}\left(\frac{\sigma_Q^2}{\frac{4}{N}\left(\frac{N}{\beta}+\frac{L_t}{\alpha}\right)\sigma^2}\right)^{-L_\Delta L_r} \tag{42}$$

where λ_is are the eigenvalues of the code-error matrix $\mathbf{C}(\mathbf{d},\mathbf{e})$, defined as $C_{p,q} = \mathbf{x}_q^H\mathbf{x}_p$ where $\mathbf{x}_p = (d_1^P - e_1^P,\ldots,d_N^P - e_N^P)^T$, $\sigma_Q^2 = 1+(\sigma^2/\alpha)$ is the variance of the element of the estimated channel coefficient vector $\hat{\mathbf{h}}(t)$, and L_Δ is the rank of ST codes, whose maximum achievable rank is L_t.

In comparison to the case that the channel coefficient matrix $\mathbf{H}(t)$ is exactly available to the maximum-likelihood receiver, the average probability of error is given in [3] as,

$$P(\mathbf{d}\to\mathbf{e})_{\mathbf{H}(t)} \le \left(\prod_{i=1}^{L_\Delta}\lambda_i\right)^{-L_r}\left(\frac{P_s}{4\sigma^2}\right)^{-L_\Delta L_r} \tag{43}$$

where P_s is the normalized power allocated to the data part when the channel coefficients are known.

Notice that the noise variance N_0 is affected by the variances of the channel estimation error, i.e. σ^2/α, and the data-bearer-projected noise, i.e., σ^2/β; therefore, it completely reveals the underlined effects of pilot- and data-power factors in the probability of error. Hence, this probability of error can be reasonably used as a cost function for optimum power allocation purpose.

Let us define the probability of error upper bound (PEUB) mismatch factor as follows:

$$\eta = \ln\left(\frac{P(\mathbf{d}\to\mathbf{e})_{\hat{\mathbf{H}}(t)}}{P(\mathbf{d}\to\mathbf{e})_{\mathbf{H}(t)}}\right) = L_\Delta L_r \ln\left(\frac{P_s\left(\frac{N}{\beta}+\frac{L_t}{\alpha}\right)}{N\left(1+\frac{\sigma^2}{\alpha}\right)}\right). \tag{44}$$

This PEUB mismatch factor can be used as an optimization objective for optimum power allocation for the data and pilot parts. In other words, this factor is minimized when the power is allocated optimally. The advantage of using this PEUB mismatch factor as the objective function inherits directly from its expression that takes both the channel estimation error and the effect of the data-bearer-projected noise into account. In addition, the use of this factor as the objective function is better than using the channel estimation error as the cost function merely, because, under the constant power constraint, despite the fact that assigning a larger power to the pilot part yields better channel coefficient estimates, i.e., a lower channel estimation error; the remaining smaller amount of power given to the data part yields a poorer probability of error in decoding. Hence, this power tradeoff is essential for the overall performances of the pilot-embedded MIMO system, e.g., channel estimation error and the probability of detection error.

B. Nonquasi-Static Flat Rayleigh Fading Channels

When the channel changes rapidly, the assumption of quasi-static fading channels does not hold anymore. Appropriate channel estimation approaches have to be designed and analyzed for combatting such channel situations. In what follows, we investigate the performance of the proposed scheme for nonquasi-static flat Rayleigh fading channels. For the sake of exposition, we study a half-block fading channel model in which the channel coefficient matrix $\mathbf{H}(t)$ symmetrically changes once within one ST symbol block, i.e., there exists $\mathbf{H}_1(t)$ and $\mathbf{H}_2(t)$ in the t^{th}-block ST symbol matrix. With $\mathbf{P} = [\mathbf{P}_1;\mathbf{P}_2]$ and $\mathbf{A} = [\mathbf{A}_1;\mathbf{A}_2]$, the received symbol matrix in (9) can be rewritten as follows:

$$\mathbf{Y}(t) = [\mathbf{H}_1(t)\,(\mathbf{D}(t)\mathbf{A}_1 + \mathbf{P}_1);\ \mathbf{H}_2(t)\,(\mathbf{D}(t)\mathbf{A}_2 + \mathbf{P}_2)] + \mathbf{N}(t) \tag{45}$$

where $\mathbf{H}_1(t)$, $\mathbf{A}_1$ and $\mathbf{P}_1$ denote the first part of the channel coefficient, the data bearer, and the pilot matrices, respectively; $\mathbf{H}_2(t)$, $\mathbf{A}_2$ and $\mathbf{P}_2$ denote the second part of the channel coefficient, the data bearer, and the pilot matrices, respectively. In addition, we would like to remind readers about the properties of matrices $\mathbf{A}$ and $\mathbf{P}$ in (4)–(7). First, we compute the ML channel estimation as in (21), i.e., $\hat{\mathbf{H}}(t) = \mathbf{Y}(t)\mathbf{P}^T/\alpha$. To gain

insight into the statistical property of $\hat{\mathbf{H}}(t)$, by substituting $\mathbf{Y}(t)$ in (45), we notice that the underlying structure of $\hat{\mathbf{H}}(t)$ can be explained as follows:

$$\hat{\mathbf{H}}(t) = \frac{1}{\alpha}\left[\mathbf{H}_1(t)\mathbf{D}(t)\mathbf{A}_1\mathbf{P}_1^T + \mathbf{H}_2(t)\mathbf{D}(t)\mathbf{A}_2\mathbf{P}_2^T\right] + \frac{1}{\alpha}\left[\mathbf{H}_1(t)\mathbf{P}_1\mathbf{P}_1^T + \mathbf{H}_2(t)\mathbf{P}_2\mathbf{P}_2^T\right] + \mathbf{N}_1(t) \quad (46)$$

where $\mathbf{N}_1(t) = \mathbf{N}(t)\mathbf{P}^T/\alpha$. Next, we compute the data extraction as in (26), i.e., $\mathbf{Y}_1(t) = \mathbf{Y}(t)\mathbf{A}^T/\alpha$. Likewise, to gain insight into the statistical property of $\mathbf{Y}_1(t)$, by substituting $\mathbf{Y}(t)$ in (45), we notice that the underlying structure of $\mathbf{Y}_1(t)$ can be explained as follows:

$$\mathbf{Y}_1(t) = \frac{1}{\beta}\left[\mathbf{H}_1(t)\mathbf{D}(t)\mathbf{A}_1\mathbf{A}_1^T + \mathbf{H}_2(t)\mathbf{D}(t)\mathbf{A}_2\mathbf{A}_2^T\right] + \frac{1}{\beta}\left[\mathbf{H}_1(t)\mathbf{P}_1\mathbf{A}_1^T + \mathbf{H}_2(t)\mathbf{P}_2\mathbf{A}_2^T\right] + \mathbf{N}_2(t) \quad (47)$$

where $\mathbf{N}_2(t) = \mathbf{N}(t)\mathbf{A}^T/\beta$. In a similar way to (29), the estimators $\mathbf{Y}_1(t)$ and $\hat{\mathbf{H}}(t)$ are used to detect the ST coded data matrix $\mathbf{D}(t)$.

As an illustrative example, we are investigating the case where $L_t = 2$, $L_r = 2$, $\tau = 2$, $N = 2$, and $M = N + L_t = 4$.

- TM- and STBC-Based Data-Bearer and Pilot Matrices
  According to (10) and (11), respectively, for the above illustrative example, we have the data-bearer and pilot matrices as follows:

$$\mathbf{A}_{\mathrm{TM}} = \sqrt{\beta}\begin{pmatrix} 0 & 0 & 1 & 0 \\ \underbrace{0 & 0}_{\mathbf{A}_{T_1}} & \underbrace{0 & 1}_{\mathbf{A}_{T_2}} \end{pmatrix},\ \mathbf{P}_{\mathrm{TM}} = \sqrt{\alpha}\begin{pmatrix} 1 & 0 & 0 & 0 \\ \underbrace{0 & 1}_{\mathbf{P}_{T_1}} & \underbrace{0 & 0}_{\mathbf{P}_{T_2}} \end{pmatrix}$$

  and

$$\mathbf{A}_{\mathrm{STBC}} = \sqrt{\beta}\begin{pmatrix} 0 & 0 & 1 & 0 \\ \underbrace{0 & 0}_{\mathbf{A}_{S_1}} & \underbrace{0 & 1}_{\mathbf{A}_{S_2}} \end{pmatrix}$$

$$\mathbf{P}_{\mathrm{STBC}} = \sqrt{\alpha}\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \underbrace{\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}}_{\mathbf{P}_{S_1}} & \underbrace{0 & 0}_{\mathbf{P}_{S_2}} \end{pmatrix}. \quad (48)$$

  From the matrix design in (48), we then have

$$\mathbf{A}_{T_1}\mathbf{P}_{T_1}^T = \mathbf{A}_{T_2}\mathbf{P}_{T_2}^T = \mathbf{P}_{T_2}\mathbf{P}_{T_2}^T = \mathbf{A}_{T_1}\mathbf{A}_{T_1}^T = \mathbf{0}_{(2\times 2)}$$
$$\mathbf{P}_{T_1}\mathbf{P}_{T_1}^T = \alpha\mathbf{I}_{(2\times 2)},\ \text{and}\ \mathbf{A}_{T_2}\mathbf{A}_{T_2}^T = \beta\mathbf{I}_{(2\times 2)}. \quad (49)$$

  Similarly, for the STBC-based matrices, the derivation in (49) is also applied, except the notation. Substituting (49) into (46) and (47), thus yielding, respectively

$$\hat{\mathbf{H}}_{\mathrm{TM\&STBC}}(t) = \mathbf{H}_1(t) + \mathbf{N}_1(t). \quad (50)$$
$$\mathbf{Y}_{1_{\mathrm{TM\&STBC}}}(t) = \mathbf{H}_2(t)\mathbf{D}(t) + \mathbf{N}_2(t). \quad (51)$$

- CM-Based Data-Bearer and Pilot Matrices
  According to (12), we are able to design the data-bearer and pilot matrices as follows:

$$\mathbf{A}_{\mathrm{CM}} = \sqrt{\beta}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \underbrace{\frac{1}{2} & -\frac{1}{2}}_{\mathbf{A}_{C_1}} & \underbrace{\frac{1}{2} & -\frac{1}{2}}_{\mathbf{A}_{C_2}} \end{pmatrix}$$

$$\mathbf{P}_{\mathrm{CM}} = \sqrt{\alpha}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \underbrace{\frac{1}{2} & -\frac{1}{2}}_{\mathbf{P}_{C_1}} & \underbrace{-\frac{1}{2} & \frac{1}{2}}_{\mathbf{P}_{C_2}} \end{pmatrix}. \quad (52)$$

  In a similar way to (49), we have

$$\mathbf{P}_{C_1}\mathbf{P}_{C_1}^T = \mathbf{P}_{C_2}\mathbf{P}_{C_2}^T = \frac{\alpha}{2}\mathbf{I}_{(2\times 2)}$$
$$\mathbf{A}_{C_1}\mathbf{A}_{C_1}^T = \mathbf{A}_{C_2}\mathbf{A}_{C_2}^T = \frac{\beta}{2}\mathbf{I}_{(2\times 2)}$$
$$\mathbf{A}_{C_1}\mathbf{P}_{C_1}^T = \frac{\sqrt{\beta\alpha}}{2}\mathbf{I}_{(2\times 2)},\ \text{and}$$
$$\mathbf{A}_{C_2}\mathbf{P}_{C_2}^T = -\frac{\sqrt{\beta\alpha}}{2}\mathbf{I}_{(2\times 2)}. \quad (53)$$

  Substituting (53) into (46) and (47), thus yielding, respectively

$$\hat{\mathbf{H}}_{\mathrm{CM}}(t) = \frac{1}{2}\left[\mathbf{H}_1(t) + \mathbf{H}_2(t)\right] + \frac{1}{2}\sqrt{\frac{\beta}{\alpha}}\left[\mathbf{H}_1(t) - \mathbf{H}_2(t)\right]\mathbf{D}(t) + \mathbf{N}_1(t). \quad (54)$$
$$\mathbf{Y}_{1_{\mathrm{CM}}}(t) = \frac{1}{2}\left[\mathbf{H}_1(t) + \mathbf{H}_2(t)\right]\mathbf{D}(t) + \frac{1}{2}\sqrt{\frac{\alpha}{\beta}}\left[\mathbf{H}_1(t) - \mathbf{H}_2(t)\right] + \mathbf{N}_2(t). \quad (55)$$

1) Channel Estimation Performance Analysis: In the following analysis, the channel estimation error for the TM-, STBC-, CM-based matrices are analyzed and compared to one another.

a) TM- and STBC-Based Data-Bearer and Pilot Matrices: According to (50) and (51), we are going to use this channel estimate in (50) to decode the ST data matrix $\mathbf{D}(t)$ in (51). Therefore, the channel estimation error can be expressed by, in the matrix form

$$\tilde{\mathbf{H}}_{\mathrm{TM\&STBC}}(t) = \mathbf{H}_2(t) - \hat{\mathbf{H}}_{\mathrm{TM\&STBC}}(t). \quad (56)$$

If we model $\mathbf{H}_2(t)$ as a linear combination of $\mathbf{H}_1(t)$ and the increment channel matrix $\Delta\mathbf{H}(t)$, i.e., $\mathbf{H}_2(t) = \mathbf{H}_1(t) + \Delta\mathbf{H}(t)$, then substituting this linear channel model into (50) and (56) yielding

$$\tilde{\mathbf{H}}_{\mathrm{TM\&STBC}}(t) = \Delta\mathbf{H}(t) - \mathbf{N}_1(t). \quad (57)$$

The mse of the ML channel estimation in (50) can be computed by

$$\begin{aligned}\mathrm{mse}_{\mathrm{TM\&STBC}} &= \mathrm{E}\left[\left\|\tilde{\mathbf{H}}_{\mathrm{TMandSTBC}}(t)\right\|^2\right] \\ &= \mathrm{E}\left[\|\Delta\mathbf{H}(t)-\mathbf{N}_1(t)\|^2\right]. \end{aligned} \quad (58)$$

Since $\Delta\mathbf{H}(t)$ and $\mathbf{N}_1(t)$ are statistically independent, and $\mathbf{N}_1(t)$ is the AWGN with zero-mean and variance expressed in (19), $\mathrm{mse}_{\mathrm{TMandSTBC}}$ can be expressed as follows,

$$\mathrm{mse}_{\mathrm{TMandSTBC}} = \mathrm{mse}_{\mathrm{quasi}} + \mathrm{E}\left[\|\Delta\mathbf{H}(t)\|^2\right] \quad (59)$$

where $\mathrm{mse}_{\mathrm{quasi}} = \sigma^2 L_t L_r/\alpha$ in (32).

b) CM-Based Data-Bearer and Pilot Matrices: According to (54) and (55), similarly we use the channel estimate in (54) to decode the data matrix $\mathbf{D}(t)$ in (55). Therefore, the channel estimation error can be expressed by, in the matrix form

$$\tilde{\mathbf{H}}_{\mathrm{CM}}(t) = \frac{1}{2}\left[\mathbf{H}_1(t)+\mathbf{H}_2(t)\right] - \hat{\mathbf{H}}_{\mathrm{CM}}(t). \quad (60)$$

Substituting $\hat{\mathbf{H}}_{\mathrm{CM}}(t)$ in (54) into (60), the mse of the ML channel estimation in (54) can be computed by

$$\begin{aligned}\mathrm{mse}_{\mathrm{CM}} &= \mathrm{E}\left[\left\|\tilde{\mathbf{H}}_{\mathrm{CM}}(t)\right\|^2\right] \\ &= \mathrm{E}\left[\left\|\frac{\xi}{2}\Delta\mathbf{H}(t)\mathbf{D}(t)-\mathbf{N}_1(t)\right\|^2\right] \end{aligned} \quad (61)$$

where $\xi = \sqrt{\beta/\alpha}$. For the same reason described in Section III-B-1)-a), $\mathrm{mse}_{\mathrm{CM}}$ can be expressed as follows:

$$\mathrm{mse}_{\mathrm{CM}} = \mathrm{mse}_{\mathrm{quasi}} + \frac{\xi^2}{4}\mathrm{E}\left[\|\Delta\mathbf{H}(t)\mathbf{D}(t)\|^2\right]. \quad (62)$$

For the orthogonal STBC which is normalized to have $\mathrm{E}[\|\mathbf{D}(t)\|^2] = L_t$, $\mathrm{mse}_{\mathrm{CM}}$ can be expressed as follows:

$$\mathrm{mse}_{\mathrm{CM}} = \mathrm{mse}_{\mathrm{quasi}} + \frac{\xi^2}{4}\mathrm{E}\left[\|\Delta\mathbf{H}(t)\|^2\right]. \quad (63)$$

Notice that, in high SNR regimes where $\mathrm{mse}_{\mathrm{quasi}} = 0$, if the equal power allocation (i.e., $\xi = 1$) the $\mathrm{mse}_{\mathrm{CM}}$ in (63) is four times less than the $\mathrm{mse}_{\mathrm{TMandSTBC}}$ in (59). In comparison, the ratio between $\mathrm{mse}_{\mathrm{CM}}$ and $\mathrm{mse}_{\mathrm{TMandSTBC}}$ can be shown as follows:

$$10\log\left(\frac{\mathrm{mse}_{\mathrm{CM}}}{\mathrm{mse}_{\mathrm{TMandSTBC}}}\right) = 10\log\left(\frac{1}{4}\right) = -6.02\ \mathrm{dB} \quad (64)$$

which indicates that the mse of the channel estimation of CM-based matrices is 6.02-dB superior to that of TM- and STBC-based matrices, in the half-block fading channel model.

2) Data Detection Performance Analysis: In the following analysis, we provide the closed form expression, in a matrix form for the sake of convenience, for the conditional pair-wise probability of transmitting a codeword $\mathbf{D}(t)$ and deciding in favor of a different codeword $\mathbf{E}(t)$ at the maximum-likelihood receiver. By using the linear channel model described in the channel estimation performance analysis, the pair-wise probability of error, given $\hat{\mathbf{H}}(t)$ and $\Delta\mathbf{H}(t)$, is given by [22]

$$\begin{aligned}P\left(\mathbf{D}(t)\to\mathbf{E}(t)|\hat{\mathbf{H}}(t),\Delta\mathbf{H}(t)\right) = P\Bigg(&\left\|\mathbf{Y}_1(t)-\hat{\mathbf{H}}(t)\mathbf{E}(t)\right\|^2 \\ &< \left\|\mathbf{Y}_1(t)-\hat{\mathbf{H}}(t)\mathbf{D}(t)\right\|^2 |\hat{\mathbf{H}}(t),\Delta\mathbf{H}(t)\Bigg). \end{aligned} \quad (65)$$

For the sake of convenience, we drop the block index t in all parameters in this section.

a) TM- and STBC-Based Data-Bearer and Pilot Matrices: By the virtue of the AWGN assumption, substituting (50) and (51) into (65) to arrive at, after some algebraic manipulation

$$\begin{aligned}&P(\mathbf{D}\to\mathbf{E}|\hat{\mathbf{H}}_{\mathrm{TMandSTBC}},\Delta\mathbf{H}) \\ &= Q\left(\frac{\left\|\hat{\mathbf{H}}_{\mathrm{TMandSTBC}}(\mathbf{D}-\mathbf{E})+\Delta\mathbf{HD}\right\|^2-\|\Delta\mathbf{HD}\|^2}{\sqrt{\frac{2}{N}\left(\frac{N}{\beta}+\frac{L_t}{\alpha}\right)\sigma^2}\left\|\hat{\mathbf{H}}_{\mathrm{TMandSTBC}}(\mathbf{D}-\mathbf{E})\right\|}\right) \end{aligned} \quad (66)$$

where $Q(\cdot)$ is the Q-function defines as $Q(x) = \int_x^\infty (1/\sqrt{2\pi})e^{-(y^2/2)}dy$.

It can be shown that a Chernoff's upper bound for (66) can be computed using the inequality $Q(x) \le e^{-x^2/2}$, given by (67) at the bottom of the page.

Given the statistics of $\hat{\mathbf{H}}_{\mathrm{TM\&STBC}}$ and $\Delta\mathbf{H}$, the averaged pair-wise error probability can be computed as follows:

$$\begin{aligned}P(\mathbf{D}\to\mathbf{E}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &P(\mathbf{D}\to\mathbf{E}|\hat{\mathbf{H}}_{\mathrm{TMandSTBC}},\Delta\mathbf{H}) \\ &\times p_{x,y}(\hat{\mathbf{H}}_{\mathrm{TMandSTBC}},\Delta\mathbf{H})dxdy \end{aligned} \quad (68)$$

$$P(\mathbf{D}\to\mathbf{E}|\hat{\mathbf{H}}_{\mathrm{TM\&STBC}},\Delta\mathbf{H}) \le \exp\left(-\frac{\left(\left\|\hat{\mathbf{H}}_{\mathrm{TM\&STBC}}(\mathbf{D}-\mathbf{E})+\Delta\mathbf{HD}\right\|^2-\|\Delta\mathbf{HD}\|^2\right)^2}{\frac{4}{N}\left(\frac{N}{\beta}+\frac{L_t}{\alpha}\right)\sigma^2\left\|\hat{\mathbf{H}}_{\mathrm{TM\&STBC}}(\mathbf{D}-\mathbf{E})\right\|^2}\right). \quad (67)$$

where $p_{x,y}(\hat{\mathbf{H}}_{\mathrm{TMandSTBC}}, \Delta\mathbf{H})$ is a joint pdf of $\hat{\mathbf{H}}_{\mathrm{TMandSTBC}}$ and $\Delta\mathbf{H}$.

b) CM-Based Data-Bearer and Pilot Matrices: Similarly, by the virtue of the AWGN assumption, substituting (54) and (55) into (65) to arrive at, after some algebraic manipulation, shown in (69) at the bottom of the page, where $\hat{\mathbf{H}}_{CM_{eff}} = \hat{\mathbf{H}}_{\mathrm{TM}} + (\Delta\mathbf{H}/2) = (\mathbf{H}_1 + (\Delta\mathbf{H}/2)) + \mathbf{N}_1$.

It is straightforward to compute the Chernoff's upper bound and the averaged pair-wise error probability for (69) in a similar way as (67) and (68), respectively. Even though the comparison between (66) and (69) is difficult to get the closed form expression, we still provide the simulation performance comparison in Section V. It is worth mentioning that this analysis is valid for the fading channel model that changes in the multiple order of N, where $N \in 2^n$, $n \in \mathbb{I}$.

IV. Optimum Block Power Allocation

In this section, we address the block power allocation problem in order to optimally allocate the power to the data and the pilot parts for quasi-static flat Rayleigh fading channels. It is clear that the performances of the pilot-embedded MIMO system essentially depend on the power percentages of the data and that of the pilot. We consider the case of the constant block power, where the power of the pilot-embedded ST symbol matrix $\mathbf{U}(t)$ is constant. The normalized block power allocated to the pilot-embedded ST symbol matrix $\mathbf{U}(t)$, which is normalized by the transmit antenna numbers L_t, can be expressed as follows:

$$\begin{aligned} P_s &= \frac{\mathrm{E}\left[\|\mathbf{U}(t)\|^2\right]}{L_t} = \frac{\mathrm{E}\left[\|\mathbf{D}(t)\mathbf{A}\|^2\right]}{L_t} + \frac{\mathrm{E}\left[\|\mathbf{P}\|^2\right]}{L_t} \\ &= P_s' + P_p = \beta + \alpha \end{aligned} \tag{70}$$

where the normalized block power allocated to the data part $P_s' = \beta$, since $\mathrm{E}[\|\mathbf{D}(t)\mathbf{A}\|^2] = \mathrm{E}[\mathrm{tr}(\mathbf{D}(t)\mathbf{A}\mathbf{A}^T\mathbf{D}(t)^T)] = \mathrm{E}[\beta\mathrm{tr}(\mathbf{D}(t)\mathbf{D}(t)^T)] = \beta L_t$; and $P_p = \alpha$ is the normalized block power allocated to the pilot part.

The objective is to minimize the PEUB mismatch factor η in (44) with respect to the pilot-power factor α subject to the constraints of constant block power and acceptable mse of the channel estimation which is a threshold that indicates the acceptable channel estimation accuracy for a reliable channel estimate. Substituting $\beta = P_s - \alpha$ into (44), the problem formulation is given by

$$\min_{\alpha} \ln\left(\frac{(N - L_t)\alpha + P_s L_t}{(\alpha + \sigma^2)(P_s - \alpha)}\right) \tag{71}$$

where mse $\leq T$ with T being the acceptable threshold of the mse in channel estimation. Differentiating (71) and equating the result to zero, we have the optimum solution for the pilot-power factor α^* as follows:

$$\alpha^* = \begin{cases} \frac{P_s - \sigma^2}{2}; & N = L_t \\ \frac{P_s L_t - \sqrt{P_s N(P_s L_t + \sigma^2(L_t - N))}}{(L_t - N)}; & N \neq L_t \end{cases} \tag{72}$$

where the mse of the channel estimation obtained in (32) must satisfy the following:

$$\mathrm{mse} = \frac{\sigma^2 L_t L_r}{\alpha} \leq T. \tag{73}$$

It is worth noticing that, in the case that $N \neq L_t$, the optimum solution for the pilot-power factor α^* in (72) exists if and only if signal-to-noise ratio (SNR) SNR $\geq (N - L_t)$, where SNR $= (P_s L_t/\sigma^2)$. Since we consider the case that $N = L_t = 4$ in our simulations, we use the case that $N = L_t$ for the sake of exposition. Substituting (72) into (73), we have the feasible range of SNR, when the inequality in (73) is satisfied, and the optimum pilot-power factor $\alpha^*_{\min}$, when mse $= T$, as follows:

$$\mathrm{SNR} \geq L_t + \frac{2L_t^2 L_r}{T} \tag{74}$$

$$\alpha^*_{\min} = \frac{L_t L_r P_s}{T + 2L_t L_r}. \tag{75}$$

Accordingly, the range of the optimum pilot-power factor α^* obtained in (72), when the SNR satisfies the inequality in (74), i.e., $L_t + (2L_t^2 L_r/T) \leq \mathrm{SNR} < \infty$, is given by

$$\frac{L_t L_r P_s}{T + 2L_t L_r} \leq \alpha^* < \frac{P_s}{2}. \tag{76}$$

However, there is a case when the SNR does not satisfy the inequality in (74), i.e., SNR $< L_t + (2L_t^2 L_r/T)$, as a result, the mse of the channel estimation is not reliable, i.e., mse >T, and the probability of detection error is inevitably increased. This scenario is equivalent to the low-SNR scenario, where wireless communication is not reliable. According to the range of the optimum pilot-power factor α^* obtained in (76), we use the minimum value of α^*, e.g., $\alpha^* = L_t L_r P_s/(T + 2L_t L_r)$, in this scenario because the PEUB mismatch factor in (71) is a monotonically increasing function of α, for α within this range.

$$P(\mathbf{D} \to \mathbf{E}|\hat{\mathbf{H}}_{\mathrm{CM}}, \Delta\mathbf{H}) = Q\left(\frac{\left\|\hat{\mathbf{H}}_{CM_{eff}}(\mathbf{D} - \mathbf{E}) + \frac{\Delta\mathbf{H}}{2}\left(\xi\mathbf{D}\mathbf{E} - \frac{\mathbf{I}}{\xi}\right)\right\|^2 - \left\|\frac{\Delta\mathbf{H}}{2}\left(\xi\mathbf{D}\mathbf{D} - \frac{\mathbf{I}}{\xi}\right)\right\|^2}{\sqrt{\frac{2}{N}\left(\frac{N}{\beta} + \frac{L_t}{\alpha}\right)\sigma^2}\left\|\hat{\mathbf{H}}_{CM_{eff}}(\mathbf{D} - \mathbf{E}) + \frac{\Delta\mathbf{H}}{2}\left(\xi\mathbf{D}\mathbf{E} - \frac{\mathbf{D}\mathbf{D}}{\xi}\right)\right\|}}\right) \tag{69}$$

In summary, we propose to determine the optimum pilot-power factor α^* for the case that $N = L_t$ under different SNR scenarios as follows:

$$\alpha^* = \begin{cases} \frac{L_t L_r P_s}{T+2L_t L_r}; & \text{SNR} < L_t + \frac{2L_t^2 L_r}{T} \\ \frac{P_s - \sigma^2}{2}; & \text{Otherwise.} \end{cases} \tag{77}$$

In addition, the acceptable threshold T for the mse of the channel estimation is quite small and is determined by practice, e.g., the simulation results in Section V. It is worth noticing that, under the high-SNR scenario where $\sigma^2 \to 0$ and, hence, the Chernoff's upper bound in (42) is tight, the pilot-power factor α approaches $P_s/2$, which is an equal power allocation also reported in [10] for the case that $N = L_t$ although where the channel estimator used is the lmmse estimator. The reason of this convergence lies in the fact that, in high SNR regimes, both the ML and lmmse channel estimators yield the same effective SNR. Since the proposed scheme and [10] effectively maximize the effective SNR in order to achieve the minimum upper bound on error probability and the maximum lower bound on channel capacity, respectively, the convergence of the optimum power allocation is resulted. However, in low SNR regimes, the power allocation in both the proposed scheme and [10] are suboptimal, because the bound used in both schemes are loose and the channel estimation error is large. Nevertheless, both schemes perform fairly well in this severely unreliable scenario as shown in Section V.

V. Simulation Results

In this section, we demonstrate the performance of the proposed scheme. Without loss of generality, we examine one orthogonal ST block code introduced in [4] and [24] [see (78) at the bottom of the page] where $s_i(t)$, $i \in \{1, 2, 3\}$ are the ST symbols corresponding to the chosen modulation constellation, e.g., 4-PSK, 8-PSK. Three data bearer and pilot structures proposed in Section II are investigated for two situations: the quasi-static and nonquasi-static flat Rayleigh fading channels. Under the nonquasi-static scenario, we investigate the performances of the pilot-embedded MIMO systems for nonquasi-static flat Rayleigh fading channel with different Doppler's shifts, representing different mobility speed of the mobile unit. We use the bit error rate (BER) and the mse of the channel estimate as performance measures, in comparison with the MIMO systems employing the ideal channel coefficients for the ML receiver [see (29)]. In our simulations, for the ideal channel coefficient case, the channel matrix $\mathbf{H}(t)$ is assumed known and thus the pilot matrix $\mathbf{P}$ is not employed, in the other words, the ST symbol matrix $\mathbf{U}(t)$ in (3) is now expressed as $\mathbf{U}(t) = \mathbf{D}(t)\mathbf{A}$. In addition, the performances of the pilot-embedded MIMO systems when employ the optimum power allocation scheme in (77), and the alternative scheme proposed in [10], are compared. In order to provide the fairness when comparing different schemes, the same transmit rate and the total transmit energy are employed by different schemes in our simulations.

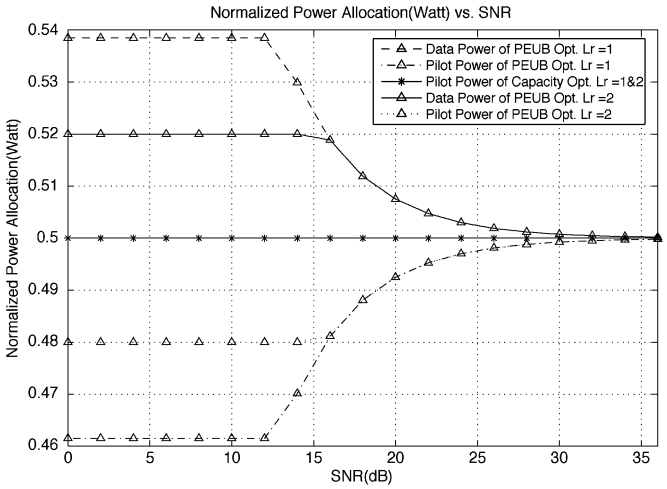


Fig. 2. The graph of the normalized power allocated to the data and pilot parts when applying the proposed and alternative optimum power allocation strategies.

For all of three data bearer and pilot structures, the setting parameters of our experiments are: the noise elements in $\mathbf{N}(t)$ in (1) are assumed to be independent complex Gaussian random variables with zero mean and variance $\sigma^2/2$ per real dimension; the normalized ST symbol block power P_s is 1 Watt/ST symbol block; the number of time slots M is 8 time slots/ST symbol block; the number of transmit antennas L_t is 4; and the data time slots $N = M - L_t$ is 4 time slots/ST symbol block. In addition, 4-PSK modulation is employed in these experiments, the acceptable threshold of the mse of the channel estimation T is set as 0.5, and the number of lmmse channel estimator's taps is 3.

A. The Quasi-Static Flat Rayleigh Fading Channel

In this situation, the channel coefficients of $\mathbf{H}(t)$ in (1) are taken from the normalized time-varying channel which is modelled as Jakes' model [25], where $fd * T = 0.08$ (fast fading) with fd being the Doppler's shift and T being the symbol period.

In Fig. 2, the normalized power allocated to data and pilot parts of two optimum power allocation strategies derived in (77)

$$\mathbf{D}(t) = \begin{pmatrix} s_1(t) & -s_2^*(t) & \frac{s_3^*(t)}{\sqrt{2}} & \frac{s_3^*(t)}{\sqrt{2}} \\ s_2(t) & s_1^*(t) & \frac{s_3^*(t)}{\sqrt{2}} & -\frac{s_3^*(t)}{\sqrt{2}} \\ \frac{s_3}{\sqrt{2}} & \frac{s_3}{\sqrt{2}} & \frac{(-s_1(t)-s_1^*(t)+s_2(t)-s_2^*(t))}{2} & \frac{(s_2(t)+s_2^*(t)+s_1(t)-s_1^*(t))}{2} \\ \frac{s_3}{\sqrt{2}} & -\frac{s_3}{\sqrt{2}} & \frac{(-s_2(t)-s_2^*(t)+s_1(t)-s_1^*(t))}{2} & -\frac{(s_1(t)+s_1^*(t)+s_2(t)-s_2^*(t))}{2} \end{pmatrix} \tag{78}$$

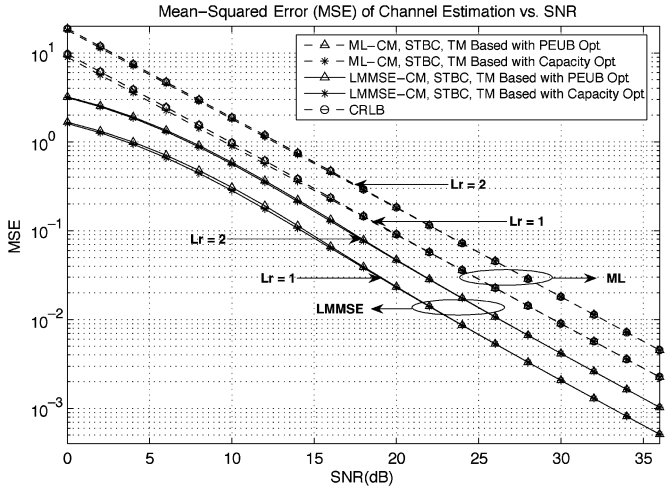


Fig. 3. The graph of mse's of the ML and lmmse channel estimations where the proposed and alternative optimum power allocation strategies are compared in the quasi-static flat Rayleigh fading channel.

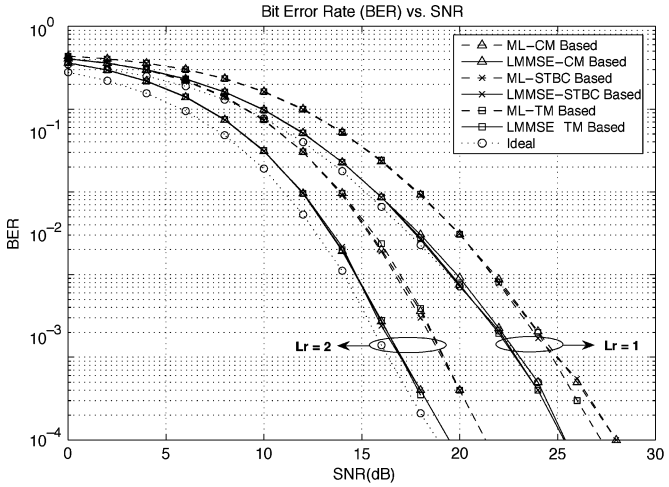


Fig. 4. The graph of BERs of the pilot-embedded optimum-power-allocated MIMO system in the quasi-static flat Rayleigh fading channel.

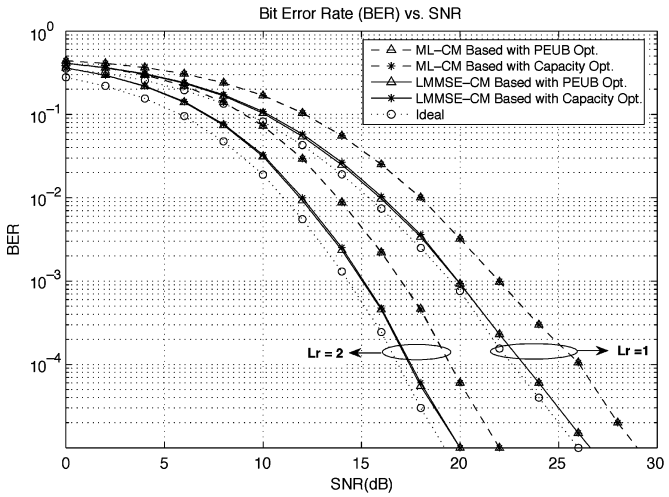


Fig. 5. The graph of BERs where the proposed and alternative optimum power allocation strategies are compared in the quasi-static flat Rayleigh fading channel.

and [10] is illustrated. It is worth noticing that these two power allocation strategies converge to 0.5 W in high SNR regimes as explained in Section IV.

In Fig. 3, we plot mse's of the channel estimation of the pilot-embedded MIMO system with applying the optimum [i.e., (77)] and the alternative optimum (i.e., [10]) power allocation strategies, when 1 and 2-receive antennas are employed. Notice that the mse's of the optimum power allocation scheme is slightly higher than that of the alternative optimum power allocation scheme in low SNR regimes. In addition, the mse's of the channels estimation of the 2-receive antenna scenario are larger than that of the 1-receive antenna scenario as explained by referring to (32), and three types of data-bearer and pilot matrices yield the same mse which coincides with the trace of the CRLB in (34). Notice that, the lmmse channel estimator outperforms the ML channel estimator, where the mse of the channel estimation is much lower in the lmmse channel estimator. In fact, the lmmse channel estimator is a Bayesian estimator in which the prior knowledge on the statistics of channels is exploited; therefore, its performance is much better than the ML channel estimator, which is a deterministic estimator, and that of CRLB. Furthermore, the lmmse channel estimator tradeoffs the bias for variance, hence, the overall mse is reduced [15]. The CRLB for Bayesian estimators including the lmmse channel estimator can be found in [11] and [15].

In Fig. 4, we plot BERs of the pilot-embedded MIMO system with applying the optimum power allocation strategy, in comparison with the ideal-channel MIMO system, when 1 and 2-receive antennas are employed. In the ideal channel case, the channel coefficients are assumed known, thus it serves as a performance bound. Notice that, at BER $= 10^{-4}$, the SNR differences between the ideal-channel and the ML channel estimator are about 2.3 dB for both the 1 and 2-receive antenna schemes, whereas the lmmse channel estimation achieves the ideal-channel error probability for the 1-receive antenna scheme, and the SNR difference between the ideal-channel and the lmmse channel estimator are about 0.5 dB for the 2-receive antenna scheme. In addition, the SNR differences between the ML and lmmse channel estimators are about 1.8 dB. It is worth noticing that the lmmse channel estimator performs better than the ML channel estimator because of the higher accurate channel estimate, as shown in Fig. 3.

In Fig. 5, the BERs are plotted in comparison between the proposed and alternative optimum power allocation strategies [10], both compared with the ideal-channel MIMO system, when 1 and 2-receive antennas are employed. For the sake of clarity, the CM-based matrices are used as the representative of all three structures that behave similarly in the experimental results. Obviously, say at BER $= 10^{-4}$, both optimum power allocation strategies are quite close resulting from the very small difference in the power allocated to the data and pilot parts in both strategies, as shown in Fig. 2.

B. The Nonquasi-Static Flat Rayleigh Fading Channel

In this situation, we consider the situation where the channel coefficient matrix $\mathbf{H}(t)$ is not kept constant over a ST symbol

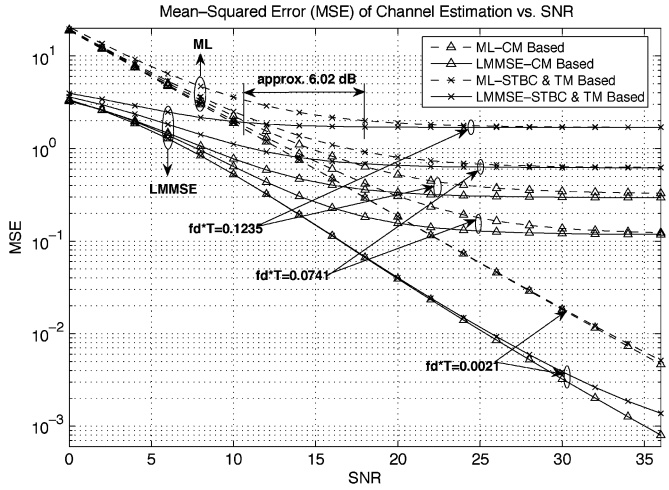


Fig. 6. The graph of mse's of the pilot-embedded optimum-power-allocated ML and lmmse channel estimations when $L_r = 2$ in the nonquasi-static flat Rayleigh fading channel.

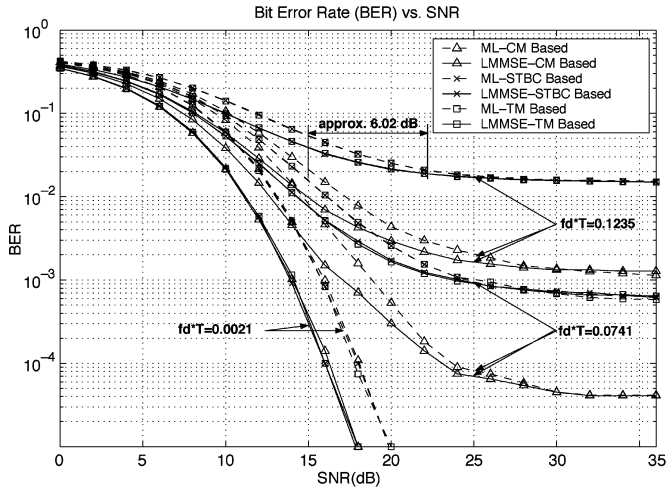


Fig. 7. The graph of BERs of the pilot-embedded optimum-power-allocated MIMO system when $L_r = 2$ in the nonquasi-static flat Rayleigh fading channel.

block. We give an example where the channel coefficient matrix symmetrically changes twice within one ST symbol block as described in Section III-B. We examine one case where 2-receive antennas are employed for the pilot-embedded optimum-power-allocated MIMO system.

In Fig. 6, the graph of mse's of the channel estimation of the pilot-embedded MIMO system when $fd * T$ are 0.0021 (slow fading), 0.0741, and 0.1235 (fast fading) is shown. Similarly to the 1-receive antenna scheme, the CM-based matrices provides the much lower mse than the TM- and STBC-based matrices. In addition, the 6.02-dB SNR difference is also observed when Doppler's shifts are fairly large, in high SNR regimes.

In Fig. 7, the graph of BERs of the pilot-embedded MIMO system when $fd * T$ are 0.0021 (slow fading), 0.0741, and 0.1235 (fast fading) is shown. Similarly to the 1-receive antenna scheme, the CM-based structure is much better than the TM- and STBC-based structures, and, in high SNR regimes, the SNR difference between the CM- and the TM- or STBC-based matrices ML channel estimators are approximately 6.02 dB, as remarked in Fig. 7.

It is worth mentioning that the CM-based structure yields better BER performances than that of the TM- and STBC-based structures, especially under the high Doppler's shift scenarios. The reason why the CM-based structure performs better than the TM-based and STBC-based structures is that it takes both of the channel coefficient matrices $\mathbf{H}_1(t)$ and $\mathbf{H}_2(t)$ into account (see (54)), whereas the other two structures exploit either some parts of $\mathbf{H}_1(t)$ or $\mathbf{H}_2(t)$ based on their structures [see (50)]. In this situation, there also exists the inevitable error floors that increase significantly as the Doppler's shift increases. These error floors result from the channel mismatch introduced as the bias in the channel estimate, thus result in a poor detection performance especially under the high Doppler's shift scenarios. Furthermore, the lmmse channel estimator performs better than the ML channel estimator in low SNR regimes, in which the AWGN is the major factor that causes the detection error; however, in high SNR regimes, the channel mismatch plays a major role in causing the detection error resulting in the comparable error floors for the lmmse and ML channel estimators.

VI. Conclusion

In this paper, we have proposed the data-bearing approach for pilot-embedding frameworks for joint data detection and channel estimation in ST coded MIMO systems. The main contributions of this paper are as follows.

- The advantages of our data-bearing approach are that it is the generalized form for pilot-embedded channel estimation and data detection in ST coded MIMO systems, in which the classical channel estimation method, e.g., PSAM, is subsumed; the low computational complexity and the efficient ML and lmmse channel estimators are achieved; and it is capable of better acquiring the channel state information in fast-fading channels.
- For the quasi-static flat Rayleigh fading channels, the error probability and the channel estimation performance of three data-bearer and pilot structures, i.e. the TM-, STBC-, and CM-based data-bearer and pilot matrices, are quite similar, where the optimum-power-allocated schemes based on the minimum upper bound on error probability and the maximum lower bound on channel capacity optimizations yield the close results. This result claims that our proposed scheme is one of the implementable scheme that achieves the maximum lower bound on channel capacity derived in [10], in high SNR regimes. In addition, the SNR differences between the optimum-power-allocated schemes and the ideal-channel schemes are about 2.3 dB when employing the unconstrained ML channel estimator and 0.5 dB for the lmmse channel estimator.
- For the case of nonquasi-static flat Rayleigh fading channels, the CM-based structure provide superior detection and channel estimation performances over the TM- and STBC-based structures. For instance, the 6.02-dB SNR difference is observed, as well as the error floors of the former are much smaller than that of the other two, under fairly high Doppler's shift scenarios, in high SNR regimes.
- In the future work, we are considering to extend the proposed data-bearing approach to MIMO orthogonal frequency division multiplexing (OFDM) systems, and

also investigating the optimum power allocation scheme for different criteria.

APPENDIX I

From (41) and the derivations in [3], the Chernoff's upper bound of the probability of error can be rewritten as follows:

$$P\left(\mathbf{d}\rightarrow\mathbf{e}|\hat{h}_{j,i},\ j=1,\ldots,L_r,\ i=1,\ldots,L_t\right) \leq \prod_{j=1}^{L_r}\exp\left\{-\frac{1}{4N_0}\sum_{i=1}^{L_t}\lambda_i|Q_{j,i}|^2\right\} \quad (1\text{-}1)$$

where λ_i is the eigenvalue of the code-error matrix $\mathbf{C}(\mathbf{d},\mathbf{e})$ defined as $C_{p,q}=\mathbf{x}_q^H\mathbf{x}_p$ where $\mathbf{x}_p=(d_1^P-e_1^P,\ldots,d_N^P-e_N^P)^T$, $Q_{j,i}\in(Q_{j,1},\ldots,Q_{j,L_t})=\mathbf{\Omega}_j^T\mathbf{V}^H$, where $\mathbf{\Omega}_j=(\hat{h}_{j,1},\ldots,\hat{h}_{j,L_t})^T$ and $\mathbf{V}$ is the eigenmatrix whose rows correspond to the eigenvectors of $\mathbf{C}(\mathbf{d},\mathbf{e})$. Since $\mathbf{V}$ is unitary, then, $Q_{j,i}$ are independent complex Gaussian random variables with zero mean and variance is given by

$$\begin{aligned}\sigma_Q^2&=\mathrm{E}\left[h_{j,i}h_{j,i}^*\right]+\mathrm{E}\left[N_{j,i}(t)N_{j,i}(t)^*\right]\\&=0.5+\frac{\sigma^2}{2\alpha}\ \text{per real dimension}\end{aligned} \quad (1\text{-}2)$$

where $N_{j,i}(t)$ is the $(j,i)^{\text{th}}$ element of the pilot-projected noise matrix $\mathbf{N}_1(t)$ in (13). Thus, $|Q_{j,i}|$ are independent Rayleigh distributions with pdf

$$p(|Q_{j,i}|)=\frac{2|Q_{j,i}|}{\sigma_Q^2}\exp\left(-\frac{|Q_{j,i}|^2}{\sigma_Q^2}\right) \quad (1\text{-}3)$$

for $|Q_{j,i}|\geq 0$.

The Chernoff's upper bound of the average probability of error can be computed by averaging (1-1) with respect to independent Rayleigh distributions of $|Q_{j,i}|$ to arrive at

$$P(\mathbf{d}\rightarrow\mathbf{e})_{\hat{\mathbf{H}}(t)}\leq\left(\prod_{i=1}^{L_\Delta}\lambda_i\right)^{-L_r}\left(\frac{\sigma_Q^2}{\frac{4}{N}\left(\frac{N}{\beta}+\frac{L_t}{\alpha}\right)\sigma^2}\right)^{-L_\Delta L_r} \quad (1\text{-}4)$$

where L_Δ is the rank of ST codes, whose maximum achievable rank is L_t.

REFERENCES

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, 1996.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Int. Tech. Memo. Jun. 1995.

[3] 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, pp. 744–765, Mar. 1998.

[4] ——, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, Jul. 1999.

[5] 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.

[6] H. Cheon and D. Hong, "Performance analysis of space-time block codes in time-varying Rayleigh fading channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2002, vol. 3, pp. 2357–2360.

[7] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[8] D. Samardzija and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2882–2890, Nov. 2003.

[9] Y. Song and S. D. Blostein, "Channel estimation and data detection for MIMO systems under spatially and temporally colored interference," in *EURASIP J. Appl. Signal Process.*, May 2004, pp. 685–695.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Thoery*, vol. 49, pp. 951–963, Apr. 2003.

[11] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[12] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. IEEE GLOBECOM*, Rio de Janeiro, Brazil, Dec. 1999, vol. 4, pp. 2162–2166.

[13] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.

[14] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, pp. 30–32, Jan. 2003.

[15] A. Vosoughi and A. Scaglione, "Everything you wanted to know about training: Guidelines derived using the affine precoding framework and the CRB," *IEEE Trans. Signal Process.*, accepted for publication.

[16] M. C. Pease, *Methods of Matrix Algebra*. New York: Academic, 1965.

[17] S. Perlis, *Theory of Matrices*. Reading, MA: Addison-Wesley, 1952.

[18] A. V. Geramita and J. Seberry, *Orthogonal Designs*. New York: Marcel Dekker, 1979.

[19] A. I. Khuri, *Advanced Calculus with Applications in Statistics*. New York: Wiley, 1993.

[20] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Cimmunications and Control*. New York: McGraw-Hill, 1971.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, vol. 1.

[22] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

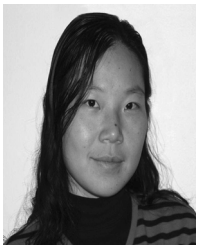
[23] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[24] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.

[25] W. C. Jakes, Jr., "Multipath interference," in *Microwave Mobile Communication*, W. C. Jakes, Jr., Ed. New York: Wiley, 1974, pp. 67–68.

Chaiyod Pirak (M'05) received the B.Eng. degree with first class honor (hons. I) in telecommunication engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2000.

Currently, he is pursuing the Ph.D. degree in electrical engineering at Chulalongkorn University, Bangkok, Thailand, in association with the University of Maryland, College Park. He received a Ph.D. scholarship from the Commission on Higher Education, Ministry of Education, Royal Thai Government, for being a faculty member of King Mongkut's Institute of Technology North Bangkok after his graduation. He was appointed as a Research Assistant with the University of Maryland, College Park, under the joint research program between Chulalongkorn University and the University of Maryland, College Park, from 2003 to 2005. His research interest is digital signal processing for wireless communications, including array signal processing, beamforming, interference cancellation techniques, channel estimation for space-time coded MIMO systems and space-frequency coded MIMO-OFDM systems, CDMA systems, and cooperative communications.



Z. Jane Wang (M'02) received the B.Sc. degree from Tsinghua University, China, in 1996, with highest honors, and the M.Sc. and Ph.D. degrees from the University of Connecticut in 2000 and 2002, respectively, all in electrical engineering.

She has been a Research Associate of Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park. Since August 1, 2004, she has been with the Department Electrical and Computer Engineering, University of British Columbia (UBC), Canada, as an Assistant Professor. Her research interests are in the broad areas of statistical signal processing, with applications to information security, biomedical imaging, genomic, and wireless communications.

Dr. Wang received the Outstanding Engineering Doctoral Student Award while at the University of Connecticut. She was a corecipient of the EURASIP *Journal on Applied Signal Processing* (JASP) Best Paper Award 2004, and the *Junior Early Career Scholar Award* from Peter Wall Institute at UBC in 2005. She coedited *Genomic Signal Processing and Statistics* and coauthored *Multimedia Fingerprinting Forensics for Traitor*. She is an Associate Editor for the EURASIP *Journal on Bioinformatics and Systems Biology*.

K. J. Ray Liu (F'03) received the B.S. degree from the National Taiwan University in 1983, and the Ph.D. degree from the University of California, Los Angeles (UCLA), in 1990, both in electrical engineering.

He is Professor and Director of Communications and Signal Processing Laboratories of Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park. His research contributions encompass broad aspects of wireless communications and networking, information forensics and security, multimedia communications and signal processing, bioinformatics and biomedical imaging, and signal processing algorithms and architectures.

Dr. Liu is the recipient of numerous honors and awards including Best Paper Awards from the IEEE Signal Processing Society, IEEE Vehicular Technology Society, and EURASIP; an IEEE Signal Processing Society Distinguished Lecturer, EURASIP Meritorious Service Award, and National Science Foundation Young Investigator Award. He also received the Poole and Kent Company Senior Faculty Teaching Award from the A. James Clark School of Engineering, and Invention of the Year Award, both from the University of Maryland. He is Vice President—Publications and on the Board of Governors of the IEEE Signal Processing Society. He was the Editor-in-Chief of the IEEE *Signal Processing Magazine* and the founding Editor-in-Chief of the EURASIP *Journal on Applied Signal Processing*.

Somchai Jitapunkul (M'90) received the B.Eng. and M.Eng. degrees in electrical engineering in 1972 and 1974, respectively, from Chulalongkorn University, Thailand. He received the D.E.A. and Dr. Ing. degrees in 1976 and 1978, respectively, in "Signaux et Systems Spatio-Temporels" from Aix-Marseille University, France.

He was appointed a Lecturer with the Department of Electrical Engineering, Chulalongkorn University, in 1972, then Assistant Professor in 1980, and Associate Professor in 1983. In 1993, he was the founder of the Digital Signal Processing Research Laboratory where he became the head of this laboratory from 1993 to 1997. From 1997 to 1999 and 1999 to 2003, he was appointed as the Head of the Communication Division and of the Department, respectively. He also held the position of Associate Dean for Information Technology, Faculty of Engineering from 1993 to 1995. His current research interests are in image and video processing, speech and character recognition, signal compression, DSP in telecommunication, software defined radio, smart antenna, and medical signal processing.