Author(s): Chen, Yunfei, Beaulieu, Norman C.

Article Title: Maximum likelihood receivers for space-time coded MIMO systems with gaussian estimation errors

Year of publication: 2009
Link to published version:
http://dx.doi.org/ 10.1109/TCOMM.2009.06.070241
Publisher statement: None

Maximum Likelihood Receivers for Space-Time Coded MIMO Systems with Gaussian Estimation Errors

Yunfei Chen , *Member, IEEE,* Norman C. Beaulieu, *Fellow, IEEE*

Corresponding Address:

Yunfei Chen

School of Engineering

University of Warwick, Coventry, U.K. CV4 7AL

Tel: +44 (0)24 765 23105

e-mail: `Yunfei.Chen@warwick.ac.uk`

Yunfei Chen is with the School of Engineering, University of Warwick, Coventry, U.K. CV4 7AL (e-mail: Yunfei.Chen@warwick.ac.uk)

Norman C. Beaulieu is with the Department of Electrical & Computer Engineering, University of Alberta, Edmonton, Alberta, CANADA T6G 2V4 (e-mail: beaulieu@ece.ualberta.ca)

Abstract

Maximum likelihood (ML) receivers for space-time coded multiple-input multiple-output (MIMO) systems with Gaussian channel estimation errors are proposed. Two different cases are considered. In the first case, the conditional probablity density function (PDF) of the channel estimate is assumed Gaussian and known. In the second case, the joint PDF of the channel estimate and the true channel gain is assumed Gaussian and known. In addition to ML signal detection for space-time coded MIMO with ML and minimum mean squared error channel estimation, ML signal detection without channel estimation is also studied. Two suboptimal structures are derived. The Alamouti space-time codes are used to examine the performances of the new receivers. Simulation results show that the new receivers can reduce the gap between the conventional receiver with channel estimation errors and the receiver with perfect channel knowledge at least by half in some cases.

Index Terms

channel estimation, imperfect, maximum likelihood, MIMO.

I. Introduction

Multiple-input multiple-output (MIMO) techniques have been well recognized as effective methods for increasing the reliability and the data rate of a wireless communication system [1]- [4]. The results in [1]- [4] are based on the assumption of perfect channel knowledge. In practice, however, perfect channel knowledge is never available. Instead, one has to estimate the channel. When the channel is estimated, estimation errors will occur. These estimation errors cause performance degradations. Therefore, the system performances reported in [1]- [4] are only upper limits, and the exact performances of MIMO systems with channel estimation errors are yet to be determined. Inspired by this, many researchers have examined the effect of channel estimation errors on the performances of MIMO systems. For example, in [5]- [9], the effect of channel estimation errors on the capacities of MIMO systems has been evaluated. These results give the maximum achievable transmission rates or the multiplexing gains of MIMO systems when the channel knowledge is not perfectly known. In [10]- [14], the authors analyzed the error rates of MIMO systems with imperfect channel estimation. The loss in terms of diversity gain due to imperfect channel estimation can be determined from these results.

In addition to the analyses of the channel capacities and the error rates of MIMO systems with channel estimation errors in [5]- [14], several researchers have also studied ways of improving the performances of MIMO systems when channel estimation errors occur. For example, in [15]- [18], several methods were proposed to improve the performances of MIMO systems by optimizing the pilot powers and the pilot positions. In [19], by assuming a Gaussian channel estimation error and using the correlation between the channel estimate and the true channel gain, the authors derived the optimum maximum likelihood receiver for space-time coded signals in the presence of channel estimation errors. This receiver design is valid for orthogonal space-time codes. In [20], the authors derived the optimum maximum likelihood receiver for any space-time code, which can be regarded as a generalization of the receiver in [19]. The results in [19] and [20] suggest that one can improve the performances of MIMO systems by using additional knowledge of the joint statistics of the channel estimate and the true channel gain. This conclusion agrees with those made in [21], where a single-input and multiple-output diversity system was considered. Motivated by this observation, in this paper we extend the results in [20] to two more general cases by using methods similar to those in [21].

Specifically, in this paper, we derive the maximum likelihood (ML) receivers for space-time coded MIMO systems when channel estimation error is Gaussian and particular additional knowledge of statistics of the channel estimate and/or the true channel gain is available. The assumption of Gaussian estimation error is justfied by the fact that many channel estimation errors are determined by Gaussian noise in the estimation, as can be seen from [20] as well as (6) and (8) in the next section. It is also justified by the fact that many well-designed estimators are asymptotically Gaussian when the sample size is large [22]. We assume a block-fading channel, where the length of a data packet is chosen to be smaller than the channel coherence time, to simplify the receiver design, similar to [19] and [20]. Two different cases are discussed. In the first case, the conditional probability density function (PDF) of the channel estimate, conditioned on the true value of the channel gain, is assumed Gaussian and known. The conditional Gaussian PDF of the channel estimate can be obtained by analyzing or simulating the mean and the variance of the channel estimates. In the second case, in addition to the conditional PDF of the channel estimate, conditioned on the true channel gain, the PDF of the true channel gain is also assumed Gaussian and known, which is the case when the MIMO channels are Rayleigh or Ricean faded. Therefore, we assume a joint Gaussian PDF for the channel estimate and the true channel gain. We derive the general structures of the ML receivers in both cases. Based on these general structures, we then study two special cases when the ML channel estimator and the minimum mean squared error (MMSE) channel estimator are used. These receivers presumably work in two steps: a first step of using the pilot symbols for channel estimation and a second step of using the data symbols and the channel estimates for signal detection. To make this study fully comprehensive, we also propose ML receivers without channel estimation, where the pilot symbols are used directly in the signal detection. Finally, we present two suboptimal receivers with simplified structures and compare their performances with the conventional receivers by simulation.

The remainder of this paper is organized as follows. In Section II, the system model is introduced. In Section III, the ML receivers for the first case are presented where only the conditional PDF of the channel estimate is known. Section IV discusses the ML receivers in the second case where the joint PDF of the channel estimate and the true channel gain is known. Numerical results are shown in Section V.

II. System Model

A. Channel Model

Consider a MIMO system with t transmitter antennas and r receiver antennas. The transmitter sends data packets with N data symbols and M pilot symbols to the receiver. For simplicity, we assume that the first N symbols in the data packet are data symbols and the following M symbols in the data packet are pilot symbols. Assume a block-fading channel where the channel gain remains approximately the same during the transmission of the whole data packet, similar to [19] and [20]. The received data symbols can be expressed as

$$\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{Z} \tag{1}$$

where $\mathbf{Y}$ is a $r \times N$ matrix representing the received data symbols, $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_r]^T$ with the i-th row $\mathbf{Y}_i = [Y_{i1} \quad Y_{i2} \quad \cdots \quad Y_{iN}]$, $i = 1, 2, \cdots, r$, T denotes the transpose operation, $\mathbf{C}$ is a $r \times t$ matrix representing the MIMO channel gains, $\mathbf{C} = [\mathbf{C}_1 \quad \mathbf{C}_2 \quad \cdots \quad \mathbf{C}_r]^T$ with the i-th row $\mathbf{C}_i = [C_{i1} \quad C_{i2} \quad \cdots \quad C_{it}]$, $i = 1, 2, \cdots, r$, $\mathbf{X}$ is a $t \times N$ matrix representing the transmitted space-time coded signals, $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_N]$ with the j-th column $\mathbf{X}_j = [X_{1j} \quad X_{2j} \quad \cdots \quad X_{tj}]^T$, $j = 1, 2, \cdots, N$, and $\mathbf{Z}$ is a $r \times N$ matrix representing the noise, $\mathbf{Z} = [\mathbf{Z}_1 \quad \mathbf{Z}_2 \quad \cdots \quad \mathbf{Z}_r]^T$ with the i-th row $\mathbf{Z}_i = [Z_{i1} \quad Z_{i2} \quad \cdots \quad Z_{iN}]$, $i = 1, 2, \cdots, r$.

Denote $\tilde{\mathbf{C}} = [\mathbf{C}_1 \quad \mathbf{C}_2 \quad \cdots \quad \mathbf{C}_r]$ as the $1 \times rt$ channel gain vector. Assume a separable Kronecker correlation model. In a Ricean fading channel, the elements of $\tilde{\mathbf{C}}$ are assumed to be circularly symmetric complex Gaussian random variables with mean $E\{\mathbf{C}_i\} = \mathbf{m}_i$, $i = 1, 2, \cdots, r$, and $rt \times rt$ covariance matrix $2\alpha^2\mathbf{R} \otimes \mathbf{T}$, where $2\alpha^2$ is the mean fading power of the scattering component, $\mathbf{R}$ represents the $r \times r$ covariance matrix of the receiver antennas, $\mathbf{T}$ represents the $t \times t$ covariance matrix of the transmitter antennas, and $\otimes$ represents the Kronecker product. In [23] and [24], a Bessel model and an exponential model have been proposed for the antenna correlations, respectively. We assume equi-spaced antennas and use the Bessel model in this paper. Therefore, the (i, j)-th element of $\mathbf{R}$ satisfies $\mathbf{R}(i, j) = J_0\left(2\pi\frac{d_r}{\lambda}|i - j|\right)$, $i, j = 1, 2, \cdots, r$, and the (i, j)-th element of $\mathbf{T}$ satisfies $\mathbf{T}(i, j) = J_0\left(2\pi\frac{d_t}{\lambda}|i - j|\right)$, $i, j = 1, 2, \cdots, t$, where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind, λ is the wavelength, $d_r|i - j|$ is the distance between the i-th receiver antenna and the j-th receiver antenna, and $d_t|i - j|$ is the distance between the i-th transmitter antenna and the j-th transmitter antenna. For convenience, we let $\tilde{\mathbf{m}} = [\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_r]$ be the $1 \times rt$ mean channel vector and $\mathbf{m} = [\mathbf{m}_1 \quad \mathbf{m}_2 \quad \cdots \quad \mathbf{m}_r]^T$ be the $r \times t$ mean channel matrix. The elements of the channel noise matrix $\mathbf{Z}$ are assumed to be independent, circularly symmetric complex Gaussian random variables each with mean zero and variance $2\sigma^2$. Further, $\mathbf{C}$ is independent of $\mathbf{Z}$.

Using (1), the likelihood function can be expressed as

$$f(\mathbf{Y}|\mathbf{C},\mathbf{X}) = \frac{1}{(2\pi\sigma^2)^{rN}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{r}(\mathbf{Y}_i-\mathbf{C}_\mathbf{i}\mathbf{X})(\mathbf{Y}_i-\mathbf{C}_\mathbf{i}\mathbf{X})^H} \quad (2)$$

where H represents the Hermitian transpose. When the channel gain $\mathbf{C}$ is perfectly known, one has the ML receiver as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \{tr\left((\mathbf{Y}-\mathbf{C}\mathbf{X})(\mathbf{Y}-\mathbf{C}\mathbf{X})^H\right)\} \quad (3)$$

where $tr(\cdot)$ denotes the trace of a matrix. We denote (3) as the genie receiver. In practice, it is impossible to know the channel gain matrix $\mathbf{C}$ perfectly. Instead, one has to use the pilot symbols in the data packet to estimate it.

B. Channel Estimation

The received signals of the pilot symbols can be written as

$$\mathbf{Q} = \mathbf{C}\mathbf{P} + \mathbf{W} \quad (4)$$

where $\mathbf{Q}$ is a $r \times M$ matrix representing the received signals of the pilot symbols, $\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2 \quad \cdots \quad \mathbf{Q}_r]^T$ with the i-th row $\mathbf{Q}_i = [Q_{i1} \quad Q_{i2} \quad \cdots \quad Q_{iM}]$, $i = 1, 2, \cdots, r$, $\mathbf{P}$ is a $t \times M$ matrix representing the transmitted pilot symbols, $\mathbf{P} = [\mathbf{P}_1 \quad \mathbf{P}_2 \quad \cdots \quad \mathbf{P}_M]$ with the j-th column $\mathbf{P}_j = [P_{1j} \quad P_{2j} \quad \cdots \quad P_{tj}]^T$, $j = 1, 2, \cdots, M$, and $\mathbf{W}$ is a $r \times M$ matrix representing the noise corrupting the pilot symbols, $\mathbf{W} = [\mathbf{W}_1 \quad \mathbf{W}_2 \quad \cdots \quad \mathbf{W}_r]^T$ with the i-th row $\mathbf{W}_i = [W_{i1} \quad W_{i2} \quad \cdots \quad W_{iM}]$, $i = 1, 2, \cdots, r$. Similar to [20], we assume that $M \geq t$, $\mathbf{P}$ is known, $(\mathbf{P}\mathbf{P}^H)^{-1}$ exists and $\mathbf{P}\mathbf{P}^H$ is real.

Using (4), the ML channel estimator for $\mathbf{C}$ can be derived by finding the $r \times t$ matrix $\hat{\mathbf{C}}$ that minimizes $||\mathbf{Q} - \mathbf{C}\mathbf{P}||^2$, where $||\cdot||^2$ is the sum of the squares of all elements in the matrix. It was derived in [20] that the ML channel estimator is given by

$$\hat{\mathbf{C}} = \mathbf{Q}\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1} \quad (5)$$

where $\hat{\mathbf{C}}$ is the $r \times t$ matrix representing the channel gain estimates, $\hat{\mathbf{C}} = [\hat{\mathbf{C}}_1 \quad \hat{\mathbf{C}}_2 \quad \cdots \quad \hat{\mathbf{C}}_r]^T$ with the i-th row $\hat{\mathbf{C}}_i = [\hat{C}_{i1} \quad \hat{C}_{i2} \quad \cdots \quad \hat{C}_{it}]$, $i = 1, 2, \cdots, r$. For later use, denote $\hat{\tilde{C}} = [\hat{\mathbf{C}}_1 \quad \hat{\mathbf{C}}_2 \quad \cdots \quad \hat{\mathbf{C}}_r]$ as the channel estimate vector. Using (4) in (5), one has

$$\hat{\mathbf{C}} = \mathbf{C} + \mathbf{W}\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}. \tag{6}$$

Therefore, the ML channel estimator $\hat{\mathbf{C}}$ gives an unbiased estimate of $\mathbf{C}$ with a Gaussian estimation error of $\mathbf{W}\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}$.

When the covariance matrix of the channel gains and the mean channel matrix are known, the MMSE channel estimator for $\mathbf{C}$ can also be derived by finding the $M \times t$ matrix $\hat{\mathbf{F}}$ that minimizes $E\{||\mathbf{Q}\mathbf{F} - \mathbf{C}||^2\}$. After some manipulations, the MMSE channel estimator can be derived as

$$\hat{\mathbf{C}} = \mathbf{Q}\hat{\mathbf{F}} \tag{7}$$

where $\hat{\mathbf{F}} = (\mathbf{P}^H\mathbf{T}\mathbf{P} + \frac{\sigma^2}{\alpha^2}\mathbf{I}_{M\times M} + \mathbf{P}^H\frac{\mathbf{m}^H\mathbf{m}}{2\alpha^2 r}\mathbf{P})^{-1}\mathbf{P}^H[\mathbf{T} + \frac{\mathbf{m}^H\mathbf{m}}{2\alpha^2 r}]$ and $\mathbf{I}_{M\times M}$ is the $M \times M$ identity matrix. Compare (7) with [20, eq. (12)], one sees that [20, eq. (12)] is a special case of (7) when $\mathbf{T} = \mathbf{I}_{t\times t}$ and $\mathbf{m} = \mathbf{0}$. Further, using (4) in (7), one has

$$\hat{\mathbf{C}} = \mathbf{C}\mathbf{P}\hat{\mathbf{F}} + \mathbf{W}\hat{\mathbf{F}} \tag{8}$$

which gives a biased estimate of $\mathbf{C}$. This bias can be removed by multiplying both sides of (8) with $(\mathbf{P}\hat{\mathbf{F}})^{-1}$ from the right, when $(\mathbf{P}\hat{\mathbf{F}})^{-1}$ exists. Note that the MMSE channel estimator in (7) can only be used when the covariance matrix of the transmitter antennas $\mathbf{T}$ and the mean channel matrix $\mathbf{m}$ are known.

Using these channel estimates, the receiver decision rule

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{tr\left((\mathbf{Y} - \hat{\mathbf{C}}\mathbf{X})(\mathbf{Y} - \hat{\mathbf{C}}\mathbf{X})^H\right)\} \tag{9}$$

has been widely used in current systems. In this paper, we will design new receivers that improve upon the performance of (9). These new receivers can be obtained by processing the channel estimates or the pilot symbols used to estimate the channels in a better way. No extra knowledge of the true channel gain is needed. They can also be obtained by using additional knowledge of the statistics of the true channel gain, similar to [20]. However, [20] only considered the case when the MIMO channels are independent and identically distributed. Here, we obain results for the case of correlated MIMO channels.

III. Conditional PDF of Channel Estimate

In the first case, no extra knowledge of the true channel gain is available. One only knows the conditional Gaussian PDF of the channel estimate, conditioned on the true channel gain. This knowledge is available for many receivers by analysis or simulation of the estimator performance. Therefore, one has

$$f(\hat{\mathbf{C}}|\mathbf{C}) = \frac{1}{(2\pi)^{rt}|\mathbf{\Delta}_1|} e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\hat{\mathbf{C}}_{i_1}-\mathbf{C}_{i_1}\mathbf{A}(i_1)-\mathbf{B}(i_1))\mathbf{\Delta}_1^{-1}(i_1,i_2)(\hat{\mathbf{C}}_{i_2}-\mathbf{C}_{i_2}\mathbf{A}(i_2)-\mathbf{B}(i_2))^H} \tag{10}$$

where $\mathbf{\Delta}_1$ is the $rt \times rt$ covariance matrix of $Re\{\hat{\hat{C}}\}$ or $Im\{\hat{\hat{C}}\}$, $Re\{\cdot\}$ and $Im\{\cdot\}$ give the real and imaginary part of a complex number, respectively, $\mathbf{\Delta}_1^{-1}(i_1, i_2)$ is the (i_1, i_2)-th submatrix of $\mathbf{\Delta}_1^{-1}$ obtained by evenly partitioning $\mathbf{\Delta}_1^{-1}$ into a $r \times r$ block matrix, and $E\{\hat{\mathbf{C}}_i\} = \mathbf{C}_i\mathbf{A}(i) - \mathbf{B}(i)$. Using (10) and (2), it can be shown that

$$\begin{aligned} f(\mathbf{Y}, \hat{\mathbf{C}}|\mathbf{X}) &= \int\cdots\int f(\mathbf{Y}|\mathbf{C},\mathbf{X})f(\hat{\mathbf{C}}|\mathbf{C})d\mathbf{C} \\ &= \frac{D_1}{|\tilde{\mathbf{\Delta}}_1|} e^{\frac{1}{2}\mathbf{u}\tilde{\mathbf{\Delta}}_1^{-1}\mathbf{u}^H} \end{aligned} \tag{11}$$

where $\int\cdots\int$ represents a rt-dimensional integral, $d\mathbf{C} = dC_{11}\cdots dC_{rt}$, D_1 is a constant independent of $\mathbf{X}$, $\tilde{\mathbf{\Delta}}_1$ is a $rt \times rt$ matrix which can be partitioned into a $r \times r$ block matrix with the (i_1, i_2)-th submatrix $\tilde{\mathbf{\Delta}}_1(i_1, i_2) = \mathbf{A}(i_1)\mathbf{\Delta}_1^{-1}(i_1, i_2)\mathbf{A}^H(i_2) + \frac{\mathbf{X}\mathbf{X}^H}{\sigma^2}$, $|\cdot|$ denotes the determinant of a matrix, $\mathbf{u} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_r]$ and $\mathbf{u}_i = \frac{\mathbf{Y}_i\mathbf{X}^H}{\sigma^2} + \sum_{i_1=1}^{r}(\hat{\mathbf{C}}_{i_1} - \mathbf{B}(i_1))\mathbf{\Delta}_1^{-1}(i_1, i)\mathbf{A}^H(i)$ with $i = 1, 2, \cdots, r$. The optimum ML receiver in this case can be derived from (11) as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{\ln|\tilde{\mathbf{\Delta}}_1| - \frac{1}{2}\mathbf{u}\tilde{\mathbf{\Delta}}_1^{-1}\mathbf{u}^H\}. \tag{12}$$

Comparing (12) with (9), one sees that there is an additional bias term of $\ln|\tilde{\mathbf{\Delta}}_1|$ in the new receiver. In general, (12) is not equivalent to (9). The receiver in (12) applies to all fading channel models, including Ricean, Nakagami-m and Laplacian channels, as no knowledge of the statistics of the true channel gain is assumed in the derivation. It also applies to any channel estimators satisfying (10).

A special case occurs when the ML channel estimator in (5) is used. In this case, one further has

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \Big\{ r\ln|\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H| \\ &\quad - \frac{tr\left((\mathbf{Y}\mathbf{X}^H + \hat{\mathbf{C}}\mathbf{P}\mathbf{P}^H)[\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H]^{-1}(\mathbf{Y}\mathbf{X}^H + \hat{\mathbf{C}}\mathbf{P}\mathbf{P}^H)^H\right)}{2\sigma^2}\Big\} \end{aligned} \quad (13)$$

since from (6), one has $\mathbf{A}(i) = \mathbf{I}_{t\times t}$, $\mathbf{B}(i) = \mathbf{0}$ and $\mathbf{\Delta}_1 = \mathbf{I}_{r\times r} \otimes [\sigma^2(\mathbf{P}\mathbf{P}^H)^{-1}]$. Note that the optimum receiver in (13) is equivalent to the conventional receiver in (9) when $\mathbf{X}\mathbf{X}^H$ is constant for all $\mathbf{X}$. Note also that the MMSE channel estimator cannot be used here, as no knowledge of the covariance matrix $\mathbf{T}$ and the mean channel matrix $\mathbf{m}$ is available in this case.

Observe that both the receiver in (9) and the receiver in (12) involve a two-step procedure. In the first step, channel estimation is performed by using $\mathbf{Q}$ to obtain an estimate $\hat{\mathbf{C}}$. In the second step, signal detection is performed by using $\mathbf{Y}$ and $\hat{\mathbf{C}}$. The only difference is in the second step where $\mathbf{Y}$ and $\hat{\mathbf{C}}$ are processed using (9) in the conventional receiver, while they are processed using (12) in the optimum receiver. As a further study, it is also of interest to examine the detection of $\mathbf{X}$ without using channel estimation [20]. From (4), the likelihood function of the pilot symbols can be expressed as

$$f(\mathbf{Q},\mathbf{P}|\mathbf{C}) = \frac{1}{(2\pi\sigma^2)^{rM}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{r}(\mathbf{Q}_i - \mathbf{C_i}\mathbf{P})(\mathbf{Q}_i - \mathbf{C_i}\mathbf{P})^H}. \quad (14)$$

Using (14) and (4), one has

$$\begin{aligned} f(\mathbf{Y},\mathbf{Q},\mathbf{P}|\mathbf{X}) &= \int\cdots\int f(\mathbf{Y}|\mathbf{C},\mathbf{X})f(\mathbf{Q},\mathbf{P}|\mathbf{C})d\mathbf{C} \\ &= \frac{D_2}{|\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H|^r} e^{\frac{1}{2\sigma^2}\sum_{i=1}^{r}[\mathbf{Y}_i\mathbf{X}^H + \mathbf{Q}_i\mathbf{P}^H][\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H]^{-1}[\mathbf{Y}_i\mathbf{X}^H + \mathbf{Q}_i\mathbf{P}^H]^H} \end{aligned} \quad (15)$$

where D_2 is a constant independent of $\mathbf{X}$. Therefore, the optimum ML receiver without channel estimation can be obtained from (15) as

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \Big\{ r \ln |\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H| \\ &\quad - \frac{tr\left((\mathbf{Y}\mathbf{X}^H + \mathbf{Q}\mathbf{P}^H)[\mathbf{P}\mathbf{P}^H + \mathbf{X}\mathbf{X}^H]^{-1}(\mathbf{Y}\mathbf{X}^H + \mathbf{Q}\mathbf{P}^H)^H\right)}{2\sigma^2} \Big\}. \end{aligned} \tag{16}$$

Comparing (16) with (13), one sees that they are actually the same, as the ML channel estimator satisfies $\hat{\mathbf{C}} = \mathbf{Q}\mathbf{P}^H(\mathbf{P}\mathbf{P}^H)^{-1}$ in (13). Therefore, the optimum receiver without channel estimation in (16) can be treated as a special case of the optimum receiver with channel estimation in (12) when ML channel estimation is performed.

To make a fair comparison, in the following, we assume that both the conventional receiver in (9) and the optimum receiver in (12) use ML channel estimation. Thus, we will focus on (13) or (16). The receiver in (13) (or (16)) requires six matrix multiplifications, two matrix additions, one matrix inversion, and one matrix determinant. Most of them have to be done for each possible sequence of $\mathbf{X}$. On the other hand, the conventional receiver in (9) requires five matrix multiplifications, one matrix addition and one matrix inversion, and most of them are done only once for all possible sequences of $\mathbf{X}$. Therefore, the new receiver is more complex than the conventional receiver. We propose two simpler suboptimal structures that are based on each space-time coded symbol (STCS) in the following for later comparison.

In the first suboptimal structure, the detection considers each STCS separately. Assume that one STCS spans a period of h data symbol intervals and that N is a multiple of h. Thus, one has

$$\mathbf{X} = [\mathbf{S}_1 \quad \mathbf{S}_2 \quad \cdots \quad \mathbf{S}_{N'}] \tag{17}$$

where $\mathbf{S}_n = [\mathbf{X}_{(n-1)h+1} \quad \cdots \quad \mathbf{X}_{(n-1)h+h}]$ is the n-th space-time coded symbol with $n = 1, 2, \cdots, N'$ and $N' = \frac{N}{h}$. The sequence detector in (16) can then be simplified to

$$\begin{aligned}\hat{\mathbf{S}}_n &= \arg\min_{\mathbf{S}_n} \Big\{ r \ln |\mathbf{P}\mathbf{P}^H + \mathbf{S}_n\mathbf{S}_n^H| \\ &\quad - \frac{tr\left((\tilde{\mathbf{Y}}_n\mathbf{S}_n^H + \mathbf{Q}\mathbf{P}^H)[\mathbf{P}\mathbf{P}^H + \mathbf{S}_n\mathbf{S}_n^H]^{-1}(\tilde{\mathbf{Y}}_n\mathbf{S}_n^H + \mathbf{Q}\mathbf{P}^H)^H\right)}{2\sigma^2} \Big\} \end{aligned} \tag{18}$$

where $\tilde{\mathbf{Y}}_n = [\mathbf{Y}^{(n-1)h+1} \quad \cdots \quad \mathbf{Y}^{(n-1)h+h}]$ is the received signal of $\mathbf{S}_n$ and $\tilde{\mathbf{Y}}^j$ represents the j-th column of $\mathbf{Y}$, $j = 1, 2, \cdots, N$. Assume that the signalling constellation size is J. The sequence detector in (16) has a time complexity of J^{Nt} and a space complexity of Nt to store the decoded sequence, the detector using the Viterbi algorithm has a time complexity of $Nt * J^2$ and a space complexity of $Nt * (J+1)$ to store the survivor paths and the decoded sequence, while the STCS-based detector has a time complexity of $N' * J^{ht}$ and a space complexity of Nt to store the decoded sequence. The larger the value of h is, the higher the complexity of the STCS-based detector will be, but the better the performance of the STCS-based detector can be expected to have. When $h = N$, the STCS-based detector becomes the sequence detector. In the special case when the Alamouti space-time coding scheme is used, one further has $t = 2$, $h = 2$, $\mathbf{S}_n = \begin{bmatrix} X_{1(2n-1)} & X_{1(2n)} \\ X_{2(2n-1)} & X_{2(2n)} \end{bmatrix} = \begin{bmatrix} x_{n_1} & -x_{n2}^* \\ x_{n_2} & x_{n_1}^* \end{bmatrix}$ where $(\cdot)^*$ denotes the conjugate operation. The STCS-based detector becomes

$$(\hat{x}_{n_1}, \hat{x}_{n_2}) = \arg\min_{x_{n_1}, x_{n_2}} \Bigg\{ 2r \ln(1 + \frac{|x_{n_1}|^2 + |x_{n_2}|^2}{d_P^2}) - \frac{\frac{(|x_{n_1}|^2 + |x_{n_2}|^2)}{d_P^2} tr(\tilde{\mathbf{Y}}_n \tilde{\mathbf{Y}}_n^H) + tr(\mathbf{Q}\mathbf{Q}^H)}{2\sigma^2(1 + \frac{|x_{n_1}|^2 + |x_{n_2}|^2}{d_P^2})} - \frac{tr\left((\tilde{\mathbf{Y}}_n \mathbf{S}_n^H \mathbf{P} \mathbf{Q}^H + \mathbf{Q}\mathbf{P}^H \mathbf{S}_n \tilde{\mathbf{Y}}_n^H)\right)}{2\sigma^2(d_P^2 + |x_{n_1}|^2 + |x_{n_2}|^2)} \Bigg\} \quad (19)$$

where $d_P^2 = tr(\mathbf{P}\mathbf{P}^H)/t$. Note that the receiver in (19) is equivalent to the conventional receiver when phase shift keying (PSK) signals are used.

In the second suboptimal structure, the detection is performed based on the decisions of the previous data symbols. Specifically, the decision-based detector is given by

$$\begin{aligned} \hat{\mathbf{S}_n} = \arg\min_{\mathbf{S}_n} \Big\{ & r \ln |\mathbf{P}\mathbf{P}^H + \hat{\mathbf{X}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{S}_n \mathbf{S}_n^H| \\ & - \frac{1}{2\sigma^2} tr\Big((\tilde{\mathbf{Y}}_n \mathbf{S}_n^H + \tilde{\mathbf{Y}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{Q}\mathbf{P}^H) \\ & [\mathbf{P}\mathbf{P}^H + \hat{\mathbf{X}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{S}_n \mathbf{S}_n^H]^{-1} \\ & (\tilde{\mathbf{Y}}_n \mathbf{S}_n^H + \tilde{\mathbf{Y}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{Q}\mathbf{P}^H)^H \Big) \Big\} \end{aligned} \quad (20)$$

where $n = 2, 3, \cdots, N'$, $\hat{\mathbf{X}}(n-1) = [\hat{\mathbf{S}}_1 \quad \hat{\mathbf{S}}_2 \quad \cdots \quad \hat{\mathbf{S}}_{n-1}]$ represents the data decisions of the previous $n-1$ space-time coded symbols, $\tilde{\mathbf{Y}}(n-1) = [\tilde{\mathbf{Y}}_1 \quad \cdots \quad \tilde{\mathbf{Y}}_{n-1}]$ represents the received signals of the previous $n-1$ space-time coded symbols, and the initial condition is given by

$$\hat{\mathbf{S}}_1 = \arg\min_{\mathbf{S}_1} \Big\{ r \ln |\mathbf{P}\mathbf{P}^H + \mathbf{S}_1\mathbf{S}_1^H| - \frac{tr\left((\tilde{\mathbf{Y}}_1\mathbf{S}_1^H + \mathbf{Q}\mathbf{P}^H)[\mathbf{P}\mathbf{P}^H + \mathbf{S}_1\mathbf{S}_1^H]^{-1}(\tilde{\mathbf{Y}}_1\mathbf{S}_1^H + \mathbf{Q}\mathbf{P}^H)^H \right)}{2\sigma^2} \Big\}. \quad (21)$$

The decision-based receiver in (20) has the same time complexity as the STCS-based detector. It needs two additional memory units to store the values of $\hat{\mathbf{X}}(n-1)\hat{\mathbf{X}}^H(n-1)$ and $\tilde{\mathbf{Y}}(n-1)\hat{\mathbf{X}}^H(n-1)$. Thus, its space complexity is slightly higher than the STCS-based detector. When the Alamouti space-time code is used, (20) can be simplified as

$$\begin{aligned}(\hat{x}_{n_1}, \hat{x}_{n_2}) = \arg\min_{x_{n_1}, x_{n_2}} \Big\{ & 2r \ln\Big(1 + \frac{d_X^2(n-1)}{d_P^2} + \frac{|x_{n_1}|^2 + |x_{n_2}|^2}{d_P^2}\Big) \\ & - \frac{\frac{|x_{n_1}|^2+|x_{n_2}|^2}{d_P^2} tr(\tilde{\mathbf{Y}}_n\tilde{\mathbf{Y}}_n^H) + \frac{d_X^2(n-1)}{d_P^2} tr(\tilde{\mathbf{Y}}(n-1)\tilde{\mathbf{Y}}^H(n-1)) + tr(\mathbf{Q}\mathbf{Q}^H)}{2\sigma^2\big(1 + \frac{d_X^2(n-1)}{d_P^2} + \frac{|x_{n_1}|^2+|x_{n_2}|^2}{d_P^2}\big)} \\ & - \frac{tr\left(\tilde{\mathbf{Y}}_n\mathbf{S}_n^H(\mathbf{P}\mathbf{Q}^H + \hat{\mathbf{X}}(n-1)\tilde{\mathbf{Y}}^H(n-1))\right)}{2\sigma^2(d_P^2 + |x_{n_1}|^2 + |x_{n_2}|^2 + d_X^2(n-1))} \\ & - \frac{tr\left((\tilde{\mathbf{Y}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{Q}\mathbf{P}^H)\mathbf{S}_n\tilde{\mathbf{Y}}_n^H\right)}{2\sigma^2(d_P^2 + |x_{n_1}|^2 + |x_{n_2}|^2 + d_X^2(n-1))} \Big\} \end{aligned} \quad (22)$$

where $d_X^2(n-1) = tr(\hat{\mathbf{X}}(n-1)\hat{\mathbf{X}}^H(n-1))/t$. We will compare the performances of (18) and (20) with that of the conventional receiver with ML channel estimation in Section V.

IV. Joint PDF of Channel Estimate and True Channel Gain

In the second case, one has extra knowledge of the statistics of the true channel gain. We assume that the channel estimate and the true channel gain are jointly Gaussian distributed with

PDF

$$
\begin{aligned}
f(\hat{\mathbf{C}}, \mathbf{C}) &= \frac{1}{(2\pi)^{2rt}|\boldsymbol{\Delta}_2|} e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\hat{\mathbf{C}}_{i_1}-\hat{\mathbf{m}}(i_1))\boldsymbol{\Delta}_{11}(i_1,i_2)(\hat{\mathbf{C}}_{i_2}-\hat{\mathbf{m}}(i_2))^H} \\
&\cdot \quad e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\mathbf{C}_{i_1}-\mathbf{m}(i_1))\boldsymbol{\Delta}_{22}(i_1,i_2)(\mathbf{C}_{i_2}-\mathbf{m}(i_2))^H} \\
&\cdot \quad e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\hat{\mathbf{C}}_{i_1}-\hat{\mathbf{m}}(i_1))\boldsymbol{\Delta}_{12}(i_1,i_2)(\mathbf{C}_{i_2}-\mathbf{m}(i_2))^H} \\
&\cdot \quad e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\mathbf{C}_{i_1}-\mathbf{m}(i_1))\boldsymbol{\Delta}_{21}(i_1,i_2)(\hat{\mathbf{C}}_{i_2}-\hat{\mathbf{m}}(i_2))^H} \qquad (23)
\end{aligned}
$$

where $\boldsymbol{\Delta}_2$ is the $2rt \times 2rt$ covariance matrix of $[Re\{\hat{\mathbf{C}}\}Re\{\mathbf{C}\}]$ or $[Im\{\hat{\mathbf{C}}\}Im\{\mathbf{C}\}]$ with $\boldsymbol{\Delta}_2 = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, $\boldsymbol{\Sigma}_{11}$ is the $rt \times rt$ covariance matrix of $Re\{\hat{\mathbf{C}}\}$ or $Im\{\hat{\mathbf{C}}\}$, $\boldsymbol{\Sigma}_{22}$ is the $rt \times rt$ covariance matrix of $Re\{\mathbf{C}\}$ or $Im\{\mathbf{C}\}$, $\boldsymbol{\Sigma}_{12}$ is the $rt \times rt$ cross-covariance matrix between $Re\{\hat{\mathbf{C}}\}$ and $Re\{\mathbf{C}\}$ or $Im\{\hat{\mathbf{C}}\}$ and $Im\{\mathbf{C}\}$, $\boldsymbol{\Sigma}_{21}$ is the $rt \times rt$ cross-covariance matrix between $Re\{\mathbf{C}\}$ and $Re\{\hat{\mathbf{C}}\}$ or $Im\{\mathbf{C}\}$ and $Im\{\hat{\mathbf{C}}\}$, $\boldsymbol{\Delta}_{11} = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Delta}_{22} = \boldsymbol{\Phi}^{-1}$, $\boldsymbol{\Delta}_{12} = -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Phi}^{-1}$, $\boldsymbol{\Delta}_{21} = -\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Delta}_{11}(i_1, i_2)$, $\boldsymbol{\Delta}_{22}(i_1, i_2)$, $\boldsymbol{\Delta}_{12}(i_1, i_2)$, $\boldsymbol{\Delta}_{21}(i_1, i_2)$ are the (i_1, i_2)-th submatrices of $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{22}$, $\boldsymbol{\Delta}_{12}$, $\boldsymbol{\Delta}_{21}$ obtained by evenly partitioning $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{22}$, $\boldsymbol{\Delta}_{12}$, $\boldsymbol{\Delta}_{21}$ into $r \times r$ block matrices, respectively, and $\hat{\mathbf{m}}_i = E\{\hat{\mathbf{C}}_i\}$, $i = 1, 2, \cdots, r$. Using (23) and (2), it can be shown that

$$
\begin{aligned}
f(\mathbf{Y}, \hat{\mathbf{C}}|\mathbf{X}) &= \int \cdots \int f(\mathbf{Y}|\mathbf{C}, \mathbf{X}) f(\hat{\mathbf{C}}, \mathbf{C}) d\mathbf{C} \\
&= \frac{D_3}{|\tilde{\boldsymbol{\Delta}}_2|} e^{\frac{1}{2}\mathbf{v}\tilde{\boldsymbol{\Delta}}_2^{-1}\mathbf{v}^H} \qquad (24)
\end{aligned}
$$

where D_3 is a constant independent of $\mathbf{X}$, $\tilde{\boldsymbol{\Delta}}_2$ is a $rt \times rt$ matrix with $\tilde{\boldsymbol{\Delta}}_2 = \boldsymbol{\Delta}_{22} + \mathbf{I}_{r\times r} \otimes \frac{\mathbf{X}\mathbf{X}^H}{\sigma^2}$, $\mathbf{v} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_r]$ and $\mathbf{v}_i = \frac{\mathbf{Y}_i\mathbf{X}^H}{\sigma^2} + \sum_{i_1=1}^{r} \mathbf{m}_{i_1}\boldsymbol{\Delta}_{22}(i_1, i) - \sum_{i_1=1}^{r}(\hat{\mathbf{C}}_{i_1} - \hat{\mathbf{m}}_{i_1})\boldsymbol{\Delta}_{12}(i_1, i)$ with $i = 1, 2, \cdots, r$. The optimum ML receiver in this case can be derived from (24) as

$$
\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{\ln|\tilde{\boldsymbol{\Delta}}_2| - \frac{1}{2}\mathbf{v}\tilde{\boldsymbol{\Delta}}_2^{-1}\mathbf{v}^H\}. \qquad (25)
$$

The receiver in (25) has a similar form to that in (12). However, (25) uses additional knowledge of the statistics of the channel gain such as $\boldsymbol{\Sigma}_{22}$ and $\mathbf{m}$. Therefore, unlike (12), the optimum receiver in (25) can only be applied to Ricean fading channels, as the Gaussian PDF of the true channel gain is assumed in (23). Also, the receiver in (25) applies to any channel estimators where the channel estimate and the true channel gain satisfy (23). In the following, we discuss two special cases when the ML channel estimator and the MMSE channel estimator are used.

When the ML channel estimator is used, from (6), it can be derived that $\hat{\mathbf{m}} = \mathbf{m}$, $\mathbf{\Sigma}_{11} = \alpha^2\mathbf{R} \otimes \mathbf{T} + \mathbf{I}_{r\times r} \otimes [\sigma^2(\mathbf{P}\mathbf{P}^H)^{-1}]$ and $\mathbf{\Sigma}_{22} = \mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21} = \alpha^2\mathbf{R} \otimes \mathbf{T}$. These give

$$\tilde{\mathbf{\Delta}}_2 = \mathbf{\Sigma}_{22}^{-1} + \mathbf{I}_{r\times r} \otimes \frac{\mathbf{X}\mathbf{X}^H + \mathbf{P}\mathbf{P}^H}{\sigma^2} \tag{26}$$

and

$$\mathbf{v}_i = \frac{\mathbf{Y}_i\mathbf{X}^H}{\sigma^2} + \sum_{i_1=1}^{r} \mathbf{m}_{i_1}\mathbf{\Sigma}_{22}^{-1}(i_1, i) + \frac{\mathbf{Q}_i\mathbf{P}^H}{\sigma^2}. \tag{27}$$

where the ML channel estimate $\hat{\mathbf{C}} = \mathbf{Q}\mathbf{P}^H(\mathbf{P}\mathbf{P})^{-1}$ has been used. Using (26) and (27) in (25), the optimum receiver with ML channel estimation can be derived.

When the MMSE channel estimator is employed, using (8), one can also derive $\hat{\mathbf{m}} = \mathbf{m}(\mathbf{P}\hat{\mathbf{F}})$, $\mathbf{\Sigma}_{11} = \mathbf{I}_{r\times r} \otimes [\sigma^2\hat{\mathbf{F}}^H\hat{\mathbf{F}}] + \alpha^2\mathbf{R} \otimes [(\mathbf{P}\hat{\mathbf{F}})^H\mathbf{T}(\mathbf{P}\hat{\mathbf{F}})]$, $\mathbf{\Sigma}_{22} = \alpha^2\mathbf{R} \otimes \mathbf{T}$, $\mathbf{\Sigma}_{12} = \alpha^2\mathbf{R} \otimes [(\mathbf{P}\hat{\mathbf{F}})^H\mathbf{T}]$ and $\mathbf{\Sigma}_{21} = \alpha^2\mathbf{R} \otimes [\mathbf{T}(\mathbf{P}\hat{\mathbf{F}})]$. In the derivation, we assume that $\hat{\mathbf{F}}^H\hat{\mathbf{F}}$ and $\mathbf{P}\hat{\mathbf{F}}$ are real in order to make $[Re\{\hat{\mathbf{C}}\}Re\{\mathbf{C}\}]$ and $[Im\{\hat{\mathbf{C}}\}Im\{\mathbf{C}\}]$ circularly symmetric. Note that $(\mathbf{P}\hat{\mathbf{F}})^H \neq \hat{\mathbf{F}}^H\mathbf{P}^H$ in this case, as $\mathbf{P}$ and $\hat{\mathbf{F}}$ may not be square matrices. Based on these results, one can show that

$$\tilde{\mathbf{\Delta}}_2 = \mathbf{\Sigma}_{22}^{-1} + \mathbf{I}_{r\times r} \otimes \frac{\mathbf{X}\mathbf{X}^H + \mathbf{P}\hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H}{\sigma^2} \tag{28}$$

and

$$\mathbf{v}_i = \frac{\mathbf{Y}_i\mathbf{X}^H}{\sigma^2} + \sum_{i_1=1}^{r} \mathbf{m}_{i_1}\mathbf{\Sigma}_{22}^{-1}(i_1, i) + \frac{\mathbf{Q}_i\hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H}{\sigma^2}. \tag{29}$$

where the MMSE channel estimate $\hat{\mathbf{C}} = \mathbf{Q}\hat{\mathbf{F}}$ has been used. Thus, the optimum receiver with MMSE channel estimation can be obtained by using (28) and (29) in (25). Comparing (26) and (27) with (28) and (29), one observes that the optimum receiver with ML channel estimation is equivalent to the optimum receiver with MMSE channel estimation when $\mathbf{P}\mathbf{P}^H = \mathbf{P}\hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$ and $\mathbf{P}^H = \hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$.

Similar to before, it is also of interest to derive the optimum ML receiver without channel estimation. From (23), one has the PDF of $\mathbf{C}$ as

$$f(\mathbf{C}) = \frac{1}{(2\pi)^{rt}|\mathbf{\Sigma}_{22}|} e^{-\frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}(\mathbf{C}_{i_1}-\mathbf{m}(i_1))\mathbf{\Sigma}_{22}^{-1}(i_1,i_2)(\mathbf{C}_{i_2}-\mathbf{m}(i_2))^H}. \quad (30)$$

Using (2), (14) and (30), one has

$$\begin{aligned} f(\mathbf{Y},\mathbf{Q},\mathbf{P}|\mathbf{X}) &= \int\cdots\int f(\mathbf{Y}|\mathbf{C},\mathbf{X})f(\mathbf{Q},\mathbf{P}|\mathbf{C})f(\mathbf{C})d\mathbf{C} \\ &= \frac{D_4}{|\tilde{\mathbf{\Delta}}_3|}e^{\frac{1}{2}\mathbf{w}\tilde{\mathbf{\Delta}}_3^{-1}\mathbf{w}^H} \end{aligned} \quad (31)$$

where D_4 is a contant independent of $\mathbf{X}$, $\tilde{\mathbf{\Delta}}_3 = \mathbf{I}_{r\times r} \otimes \frac{\mathbf{X}\mathbf{X}^H+\mathbf{P}\mathbf{P}^H}{\sigma^2} + \mathbf{\Sigma}_{22}^{-1}$, the vector $\mathbf{w} = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_r]$ with $\mathbf{w}_i = \frac{\mathbf{Y}_i\mathbf{X}^H+\mathbf{Q}_i\mathbf{P}^H}{\sigma^2} + \sum_{i_1=1}^{r}\mathbf{m}_{i_1}\mathbf{\Sigma}_{22}^{-1}(i_1,i)$ and $i = 1,2,\cdots,r$. The optimum ML receiver without channel estimation is then given by

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{\ln|\tilde{\mathbf{\Delta}}_3| - \frac{1}{2}\mathbf{w}\tilde{\mathbf{\Delta}}_3^{-1}\mathbf{w}^H\}. \quad (32)$$

Comparing (32) with (25), one sees that the opitmum ML receiver without channel estimation can again be treated as a special case of the optimum ML receiver with channel estimation, when the ML channel estimator is used, or when the MMSE channel estimator is used and $\mathbf{P}\mathbf{P}^H = \mathbf{P}\hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$ and $\mathbf{P}^H = \hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$. It is also interesting to note that [20, eq. (26)] is a special case of (32) when $\mathbf{R} = \mathbf{I}_{r\times r}$, $\mathbf{T} = \mathbf{I}_{t\times t}$ and $\mathbf{m} = \mathbf{0}$, as expected.

Again, to make the comparison fair, we assume that both the optimum receiver and the conventional receiver use either the ML channel estimator or the MMSE channel estimator. Further, we assume that $\mathbf{P}$ satisfies $\mathbf{P}\mathbf{P}^H = \mathbf{P}\hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$ and $\mathbf{P}^H = \hat{\mathbf{F}}(\hat{\mathbf{F}}^H\hat{\mathbf{F}})^{-1}(\mathbf{P}\hat{\mathbf{F}})^H$. Then, we only need to focus on the optimum receiver in (32). The receiver in (32) requires $(r+2)r+4$ matrix multiplifications, $(r+1)r+2$ matrix additions, one matrix inversion, one matrix determinant and one Kronecker product. Most of them have to be done for each possible sequence of $\mathbf{X}$. In addition, some of the matrix multiplifications are of much larger dimension than (16). Thus, it is more complicated than (16) and (9). Similarly, two simplified suboptimal structures based on each space-time coded symbol will be derived.

In the first suboptimal structure, the detector considers each space-time coded symbol separately. Using similar methods as before, one has

$$\tilde{\mathbf{\Delta}}_3(n) = \mathbf{I}_{r\times r} \otimes \frac{\mathbf{S}_n\mathbf{S}_n^H + \mathbf{P}\mathbf{P}^H}{\sigma^2} + \mathbf{\Sigma}_{22}^{-1} \tag{33}$$

and

$$\mathbf{w}_i(n) = \frac{\tilde{\mathbf{Y}}_{in}\mathbf{S}_n^H + \mathbf{Q}_i\mathbf{P}^H}{\sigma^2} + \sum_{i_1=1}^{r} \mathbf{m}_{i_1}\mathbf{\Sigma}_{22}^{-1}(i_1, i) \tag{34}$$

where $\tilde{\mathbf{Y}}_{in} = [Y_{i((n-1)h+1)} \quad \cdots \quad Y_{i((n-1)h+h)}]$ is the i-th row of $\tilde{\mathbf{Y}}_n$ defined as before and $\mathbf{w}(n) = [\mathbf{w}_1(n) \quad \mathbf{w}_2(n) \quad \cdots \quad \mathbf{w}_r(n)]$. Then, the STCS-based receiver is given by

$$\hat{\mathbf{S}}_n = \arg\min_{\mathbf{S}_n}\{\ln|\tilde{\mathbf{\Delta}}_3(n)| - \frac{1}{2}\mathbf{w}(n)\tilde{\mathbf{\Delta}}_3^{-1}(n)\mathbf{w}^H(n)\} \tag{35}$$

where $n = 1, 2, \cdots, N'$. When the Alamouti space-time code is used, one further has $t = 2$, $h = 2$ and $\tilde{\mathbf{\Delta}}_3(n) = \frac{(|x_{n_1}|^2+|x_{n_2}|^2+d_P^2)}{\sigma^2}\mathbf{I}_{rt\times rt} + \mathbf{\Sigma}_{22}^{-1}$.

In the second suboptimal structure, the detection is based on data decisions of previous symbols. Similarly, one has

$$\tilde{\mathbf{\Delta}}_3(n, \hat{\mathbf{X}}(n-1)) = \mathbf{I}_{r\times r} \otimes \frac{\mathbf{S}_n\mathbf{S}_n^H + \hat{\mathbf{X}}(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{P}\mathbf{P}^H}{\sigma^2} + \mathbf{\Sigma}_{22}^{-1} \tag{36}$$

and

$$\mathbf{w}_i(n, \hat{\mathbf{X}}(n-1)) = \frac{\tilde{\mathbf{Y}}_{in}\mathbf{S}_n^H + \tilde{\mathbf{Y}}_i(n-1)\hat{\mathbf{X}}^H(n-1) + \mathbf{Q}_i\mathbf{P}^H}{\sigma^2} + \sum_{i_1=1}^{r} \mathbf{m}_{i_1}\mathbf{\Sigma}_{22}^{-1}(i_1, i) \tag{37}$$

where $\hat{\mathbf{X}}(n-1)$ and $\tilde{\mathbf{Y}}(n-1)$ are defined as before. Then, the decision-based receiver is given by

$$\hat{\mathbf{S}}_n = \arg\min_{\mathbf{S}_n}\{\ln|\tilde{\mathbf{\Delta}}_3(n, \hat{\mathbf{X}}(n-1))| - \frac{1}{2}\mathbf{w}(n, \hat{\mathbf{X}}(n-1))\tilde{\mathbf{\Delta}}_3^{-1}(n, \hat{\mathbf{X}}(n-1))\mathbf{w}^H(n, \hat{\mathbf{X}}(n-1))\} \tag{38}$$

where $n = 2, 3, \cdots, N'$ and the initial condition is given by

$$\hat{\mathbf{S}}_1 = \arg\min_{\mathbf{S}_1}\{\ln|\tilde{\mathbf{\Delta}}_3(1)| - \frac{1}{2}\mathbf{w}(1)\tilde{\mathbf{\Delta}}_3^{-1}(1)\mathbf{w}^H(1)\}. \tag{39}$$

In the case when the Alamouti space-time code is used, one can further simplify the expression of $\tilde{\mathbf{\Delta}}_3(n, \hat{\mathbf{X}}(n-1))$ as $\tilde{\mathbf{\Delta}}_3(n, \hat{\mathbf{X}}(n-1)) = \frac{(|x_{n_1}|^2+|x_{n_2}|^2+d_P^2+d_X^2(n-1))}{\sigma^2}\mathbf{I}_{rt\times rt} + \mathbf{\Sigma}_{22}^{-1}$. Note that the above results only apply to a separable Kronecker correlation model where $\mathbf{\Sigma}_{22} = \alpha^2\mathbf{R}\otimes\mathbf{T}$. However, these results can be easily extended to any correlation models by replacing $\mathbf{\Sigma}_{22} = \alpha^2\mathbf{R}\otimes\mathbf{T}$ with other covariance matrices in (10) and (23) in the derivation. In the next section, we compare the derived new receivers with the conventional receiver.

V. Numerical results and discussion

Consider Alamouti space-time coding. For convenience, we denote the novel STCS-based receiver as the NovSTCS receiver, the novel decision-based receiver as the NovDB receiver, the conventional receiver based on each STCS with ML channel estimation as the ConvML receiver and the conventional receiver based on each STCS with MMSE channel estimation as the ConvMMSE receiver. The average signal-to-noise ratio (SNR) is defined as

$$\gamma = \frac{tr(\mathbf{P}\mathbf{P}^H) + NtE_s}{N} \cdot \frac{2\alpha^2}{2\sigma^2} \tag{40}$$

where E_s is the average energy of the data symbol and E_s is normalized to 1 in the simulation. The definition of the SNR accounts for the energy consumed by the pilot symbols and by the multiple transmitter antennas. Two signalling schemes, 16-QAM and quanternary phase shift keying (QPSK) are studied. In 16-QAM, all M pilot symbols in the data packet are fixed to $\frac{1}{\sqrt{10}} + \frac{1}{\sqrt{10}}i$. In QPSK, all M pilot symbols in the data packet are fixed to 1. The length of the data packet is chosen as 100. We assume that $d_t = d_r$. Also, denote the Ricean factor as K. In the first case, the symbol error rates (SERs) of the NovSTCS receiver in (18), the NovDB receiver in (20), the ConvML receiver in (9) with (5), and the genie receiver in (3) are derived for 16-QAM only, since the conventional receiver and the new ML receiver are equivalent for QPSK. In the second case, the SERs of the NovSTCS receiver in (35), the NovDB receiver in (38), the ConvML receiver in (9) with (5), the ConvMMSE receiver in (9) with (7), and the genie receiver in (3) are derived for both 16-QAM and QPSK. The purpose of the simulation is to examine how much gain one can achieve by using extra knowledge of channel statistics. To see these gains clearly, one has to choose the same decoding complexity for all receivers. In our simulation, we use STCS-based detection. Thus, $\hat{\mathbf{S}_n} = \arg\min_{\mathbf{S}_n}\{tr\left((\tilde{\mathbf{Y}}_n - \mathbf{C}\mathbf{S}_n)(\tilde{\mathbf{Y}}_n - \mathbf{C}\mathbf{S}_n)^H\right)\}$ as the genie receiver and $\hat{\mathbf{S}_n} = \arg\min_{\mathbf{S}_n}\{tr\left((\tilde{\mathbf{Y}}_n - \hat{\mathbf{C}}\mathbf{S}_n)(\tilde{\mathbf{Y}}_n - \hat{\mathbf{C}}\mathbf{S}_n)^H\right)\}$ as the conventional receiver are compared with (18), (20), (35) and (38). On the other hand, one could compare the receivers using sequence-based detection, where sphere decoding is an efficient way of finding a sequence decision with reasonable accuracy. In this case, (3) and (9) should be compared directly with (16) and (32). Both ways of comparison will allow us to identify the performance gains achieved by using extra channel statistics. However, if we compare the proposed suboptimal receivers in (18), (20), (35) and (38) using STCS-based detection with the conventional receivers using sphere decoding, the performance gain due to extra knowledge of channel statistics will be compromised by the performance loss due to STCS-based detection, and we won't be able to identify the performance gain easily. Note also that decision errors may occur in $\hat{\mathbf{X}}(n-1)$ in (20) and (38). The presented simulation results take the effect of possible error propagation into account.

Fig. 1 examines the SERs of the receivers for different values of M. One sees that the SERs of the receivers decrease as M increases, up to a certain threshold. Then, the SERs of the receivers increase as M increases. This is expected. When M increases, the receiver has a more accurate channel estimate, but it also suffers from allocating more power to the pilot symbols. At some point where the channel estimate is accurate enough, increasing M will mainly reduce useful power without achieving worthwhile improvement in the channel estimation and, thus, overall cause performance degradation. Comparing the NovSTCS receiver with the ConvML and ConvMMSE receivers, one sees that the NovSTCS receiver is only slightly better than the ConvML and ConvMMSE receivers. Also, one notes that the NovDB receiver has an obvious performance gain over the ConvML and ConvMMSE receivers. This performance gain increases as M decreases. In the following, we will use $M = 20$. This corresponds to a pilot power of

4.75% of the total power for 16-QAM and 20% of the total power for QPSK.

Fig. 2 shows the SERs of the receivers for different channel correlations. One sees that the SER curves resemble the curve of a Bessel autocorrelation function. This is because a Bessel correlation model is assumed and the receiver performs the best when the channel correlation is the smallest. The SERs of the receivers vascillate slightly as d_t/λ or d_r/λ increase. Comparing the NovSTCS receiver with the ConvML and ConvMMSE receivers, one sees that the NovSTCS receiver is slightly better. Also, the NovDB receiver outperforms the ConvML and ConvMMSE receivers as well as the NovSTCS receiver, which agrees with the previous observations from Fig. 1. We will use $d_t/\lambda = d_r/\lambda = 0.5$ next.

Fig. 3 compares the receiver performances in Case 1 for different values of r at different SNRs. The performances of the receivers improve when r increases. In all the cases, the NovSTCS receiver is slightly better than the ConvML receiver, and the NovDB receiver has an obvious performance gain over the ConvML receiver and the NovSTCS receiver. The performance gain increases as r increases. Fig. 4 examines the receiver performances in Case 2 for different values of r. In this case, the ConvML receiver performs the worst. The ConvMMSE receiver outperforms the ConvML receiver, as it uses extra knowledge of the covariance matrix and the mean channel matrix of the true channel gain. The NovSTCS receiver is slightly better than the ConvMMSE receiver. The NovDB receiver performs the best among all the practical receivers studied. Moreover, the performance gain increases when r increases or the SNR increases. Comparing Fig. 3 with Fig. 4, one sees that the NovDB receiver in Case 2 performs slightly better than that in Case 1, as expected, since Case 2 assumes more knowledge of the channel statistics.

Figs. 5 and 6 show the SERs of the receivers in Case 1 and Case 2, respectively, for different values of the Ricean K factor. From these figures, one sees that the receiver performances improve when the value of K increases. This is expected, as a larger value of K corresponds to a better channel condition. Again, the NovDB receiver outperforms all the other practical receivers. The performance gain increases when the value of K increases or the SNR increases. This implies that the performance gains of the NovDB receiver over other receivers observed in Figs. 1 to 4 are also achievable when $K > 0$. Figs. 7 and 8 show the SERs of the receivers in Case 2 for QPSK signaling. In general, the receivers using QPSK signaling perform better than those using 16-QAM, under the same conditions. Also, one sees that the performance gains of the NovDB receiver with QPSK are smaller than the corresponding gains with 16-QAM.

VI. Conclusions

Novel ML receivers for space-time coded MIMO systems with Gaussian channel estimation errors have been derived. Numerical results have shown that the overall performance of the system depends on several design parameters including the number of pilot symbols, the channel correlation, the number of antennas, the Ricean K factor and the signaling scheme. Future work includes an examination of new receivers for other MIMO systems with estimation errors.

References

[1] S.N. Diggavi, N. Al-Dhahir, A. Stamoulis and A.R. Calderbank, ”Great expectations: the value of spatial diversity in wireless networks,” *P. IEEE*, vol. 92, pp. 219-270, Feb. 2004.

[2] G. Foschini, ”Layered space-time architecture for wireless communication in a fading environment when using multiple antenna elements,” *Bell Labs Techn. J.*, vol. 1, pp. 41-59, Sept. 1996.

[3] S.M. Alamouti, ”A simple transmitter diversity scheme for wireless communications,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

[4] V. Tarokh, H. Jafarkhani and A.R. Calderbank, ”Space-time block codes from orthogonal designs,” *IEEE Trans. Info. Theo.*, vol. 45, pp. 1456-1467, July 1999.

[5] A. Lapidoth and S. Shamai, ”Fading channels: how perfect need ”perfect side information” be?,” *IEEE Trans. Info. Theo.*, vol. 48, pp. 1118-1134, May 2002.

[6] M. Medard, ”The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Info. Theo.*, vol. 46, pp. 933-946, May 2000.

[7] D. Samardzija and N. Mandayam, ”Pilot-assisted estimation of MIMO fading channel response and achievable data rates,” *IEEE Trans. Signal Processing*, vol. 51, pp. 2882-2890, Nov. 2003.

[8] T. Yoo and A. Goldsmith, ”Capacity and power allocation for fading MIMO channels with channel estimation error,” *IEEE Trans. Info. Theo.*, vol. 52, pp. 2203-2214, May 2006.

[9] P. Kyritsi, R.V. Valenzuela and D.C. Cox, ”Channel and capacity estimation errors,” *IEEE Commun. Lett.*, vol. 6, pp. 517-519, Dec. 2002.

[10] L.L. Chong and L.B. Milstein, ”The effects of channel estimation errors on a space-time spreading CDMA system with dual transmit and dual receive diversity,” *IEEE Trans. Commun.*, vol. 52, pp. 1145-1151, July 2004.

[11] R. Narasimhan, ”Error propagation analysis of V-BLAST with channel-estimation errors,” *IEEE Trans. Commun.*, vol. 53, pp. 27-31, Jan. 2005.

[12] T. Weber, A. Sklavos and M. Meurer, ”Imperfect channel-state information in MIMO transmission,” *IEEE Trans. Commun.*, vol. 54, pp. 543-552, Mar. 2006.

[13] P. Garg, R.K. Mallik and H.M. Gupta, ”Performance analysis of space-time coding with imperfect channel estimation,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 257-265, Jan. 2005.

[14] S.A. Zummo and W.E. Stark, ”Error probability of coded STBC systems in block fading environments,” *IEEE Trans. Wireless Commun.*, vol. 5, pp. 972-977, May 2006.

[15] E. Baccarelli and M. Biagi, ”Performance and optimized design of space-time codes for MIMO wireless systems with imperfect channel estimates,” *IEEE Trans. Signal Processing*, vol. 52, pp. 2911-2923, Oct. 2004.

[16] D.V. Duong and G.E. Oien, ”Optimal pilot spacing and power in rate-adaptive MIMO diversity systems with imperfect CSI,” *IEEE Trans. Wireless Commun.*, vol. 6, pp. 845-851, Mar. 2007.

[17] E. Baccarelli, M. Biagi and C. Pelizzoni, ”On the information throughput and optimized power allocation for MIMO wireless systems with imperfect channel estimation,” *IEEE Trans. Signal Processing*, vol. 53, pp. 2335-2347, July 2005.

[18] X. Wang and J. Wang, ”Effect of imperfect channel estimation on transmit diversity in CDMA systems,” *IEEE Trans. Vehi. Technol.*, vol. 53, pp. 1400-1412, Sept. 2004.

[19] V. Tarokh, A. Naguib, N. Seshadri and A.R. Calderbank, ”Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths,” *IEEE Trans. Commun.*, vol. 47, pp. 199-207, Feb. 1999.

[20] G. Taricco and E. Biglieri, ”Space-time decoding with imperfect channel estimation,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1874-1888, July 2005.

[21] Y. Chen and N.C. Beaulieu, ”Novel diversity receivers in the presence of Gaussian channel estimation errors,” *IEEE Trans. Wireless Commun.*, vol. 5, pp. 2022-2025, Aug. 2006.

[22] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[23] D.-S. Shiu, G.J. Foschini, M.J. Gans and J.M. Kahn, ”Fading correlation and its effect on the capacity of multelement antenna systems,” *IEEE Trans. Commun.*, vol. 48, pp. 502-513, Mar. 2000.

[24] T.-A. Chen, M.P. Fitz, M.P. Zoltowski, W.-Y. Kuo and J.H. Grimm, ”A space-time model for frequency nonselective Rayleigh fading channels with applications to space-time modems,” *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1175-1190, July 2000.

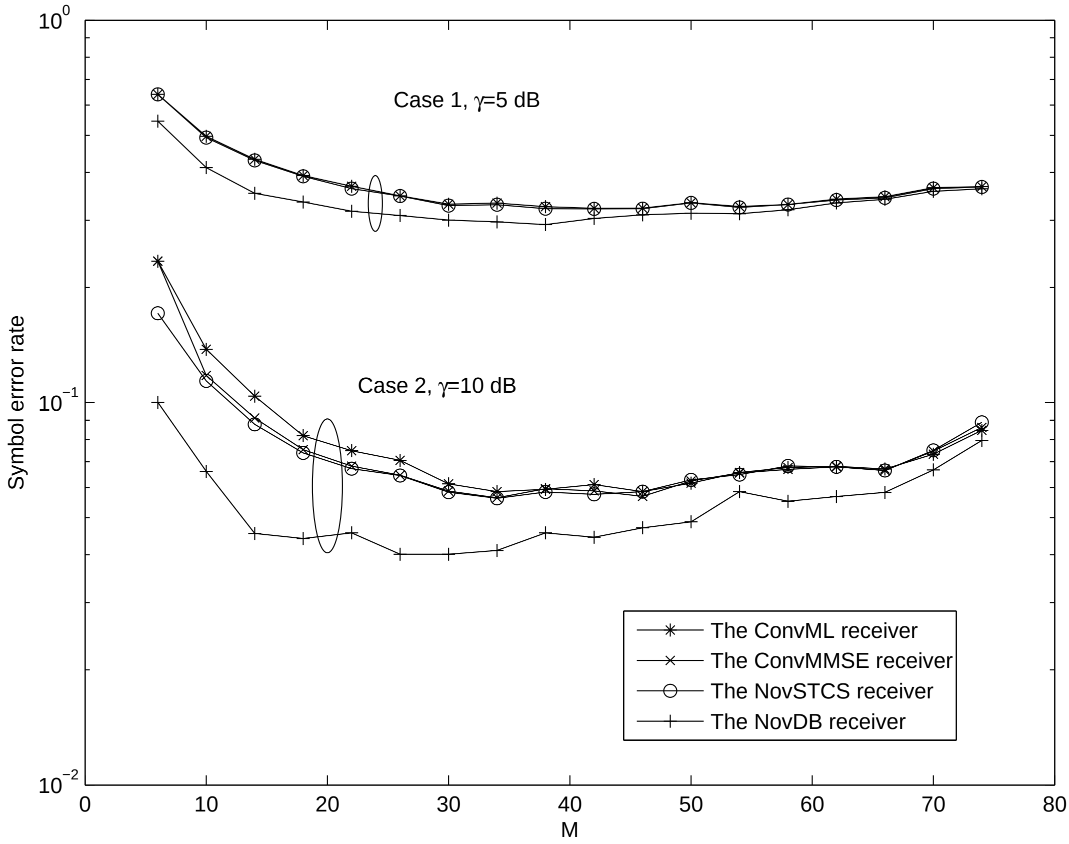

Fig. 1. Symbol error rates of receivers for different values of M when $r = 4$, $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 3$, $K = 0$, and 16-QAM is used.

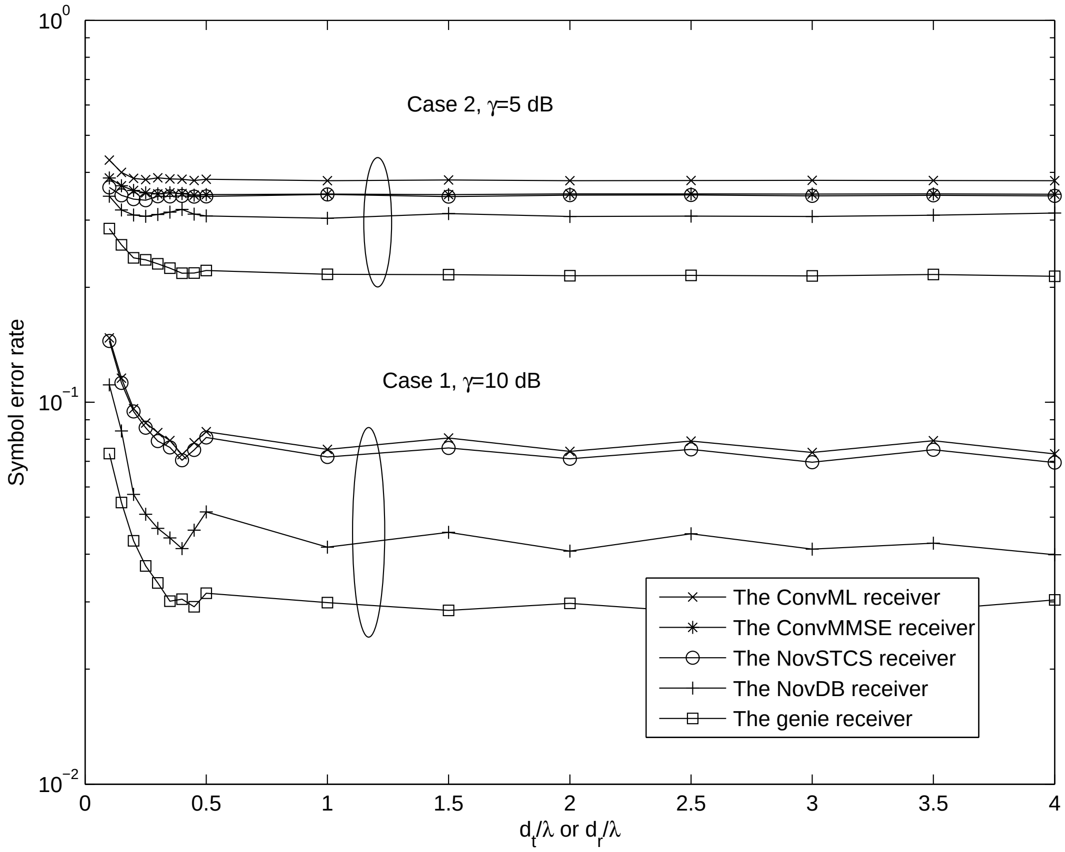

Fig. 2. Symbol error rates of receivers for different values of $\frac{d_t}{\lambda}$ or $\frac{d_r}{\lambda}$ when $r = 4$, $M = 20$, $K = 0$, and 16-QAM is used.

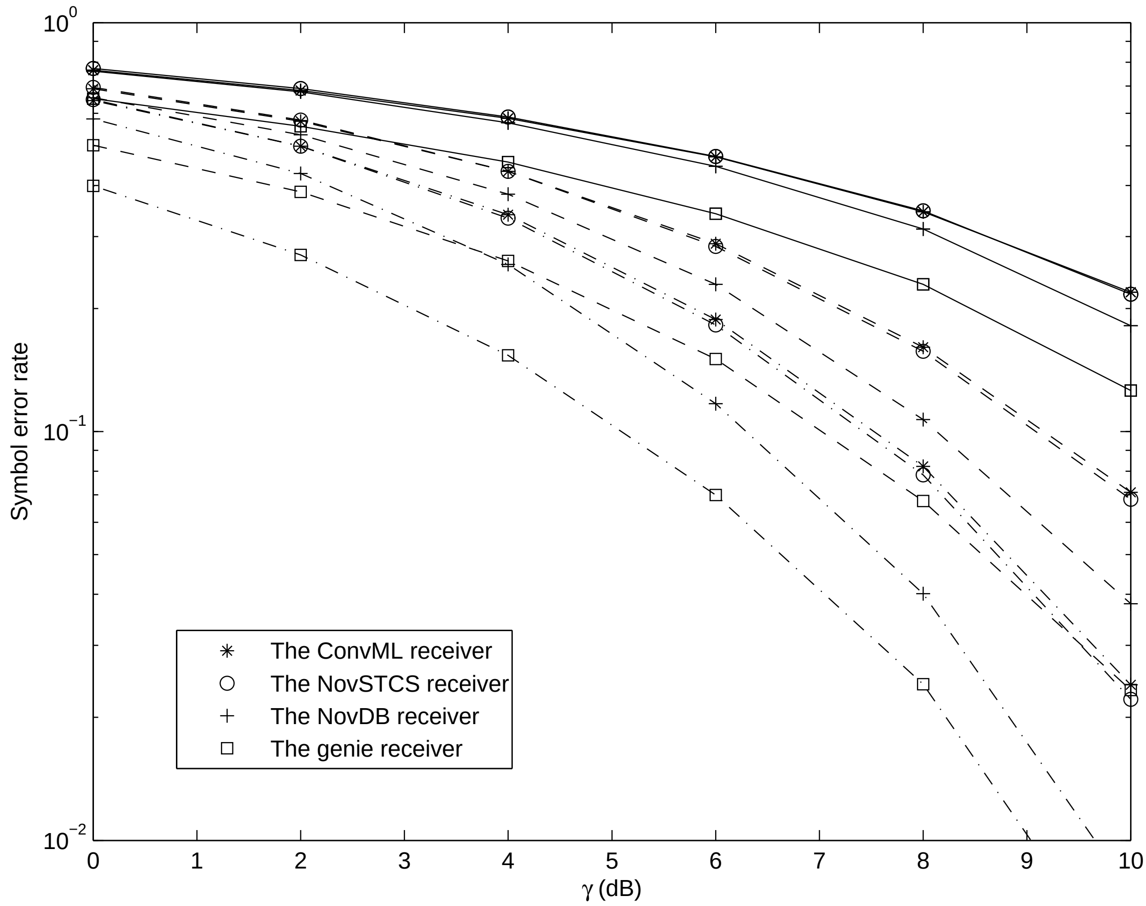

Fig. 3. Symbol error rates of receivers in Case 1 for $r = 2$ (solid line), $r = 4$ (dashed line), and $r = 6$ (dash-dotted line), when $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, $K = 0$, and 16-QAM is used.

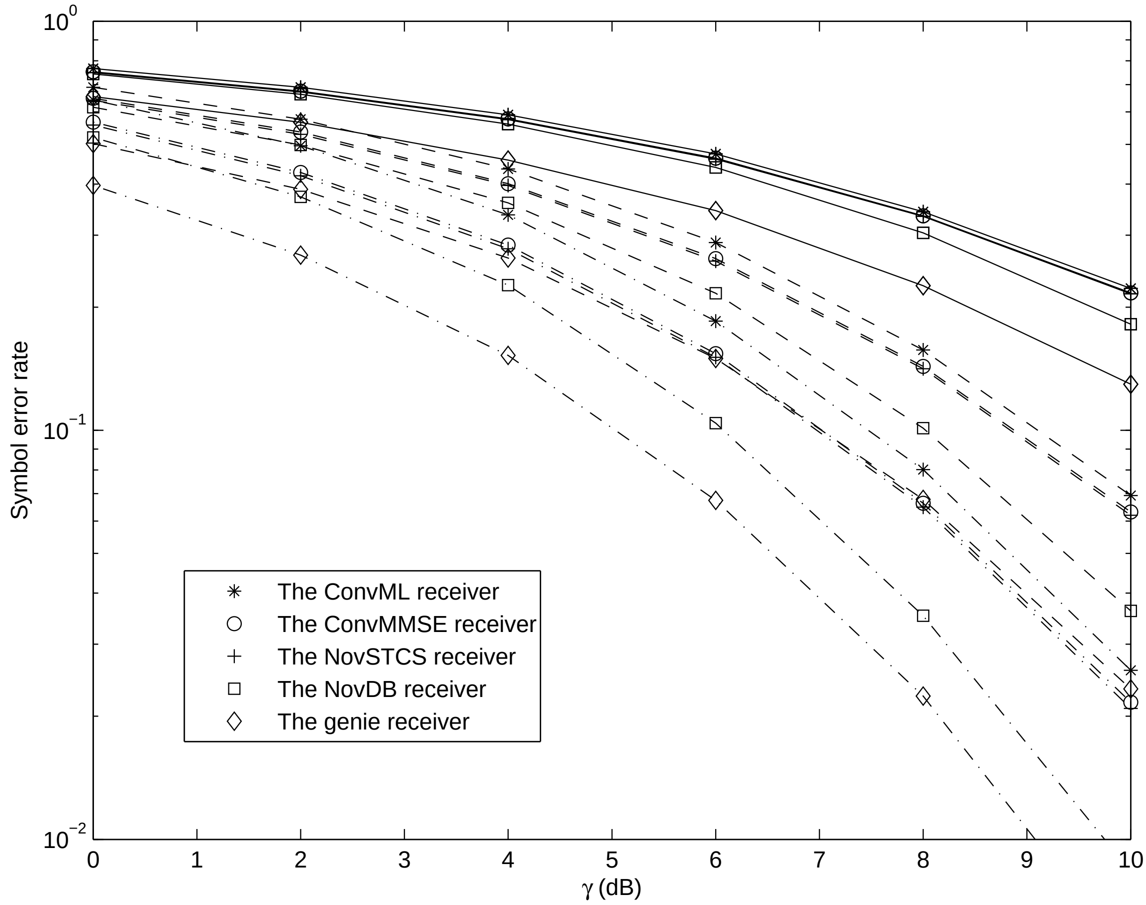

Fig. 4. Symbol error rates of receivers in Case 2 for $r = 2$ (solid line), $r = 4$ (dashed line), and $r = 6$ (dash-dotted line), when $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, $K = 0$, and 16-QAM is used.

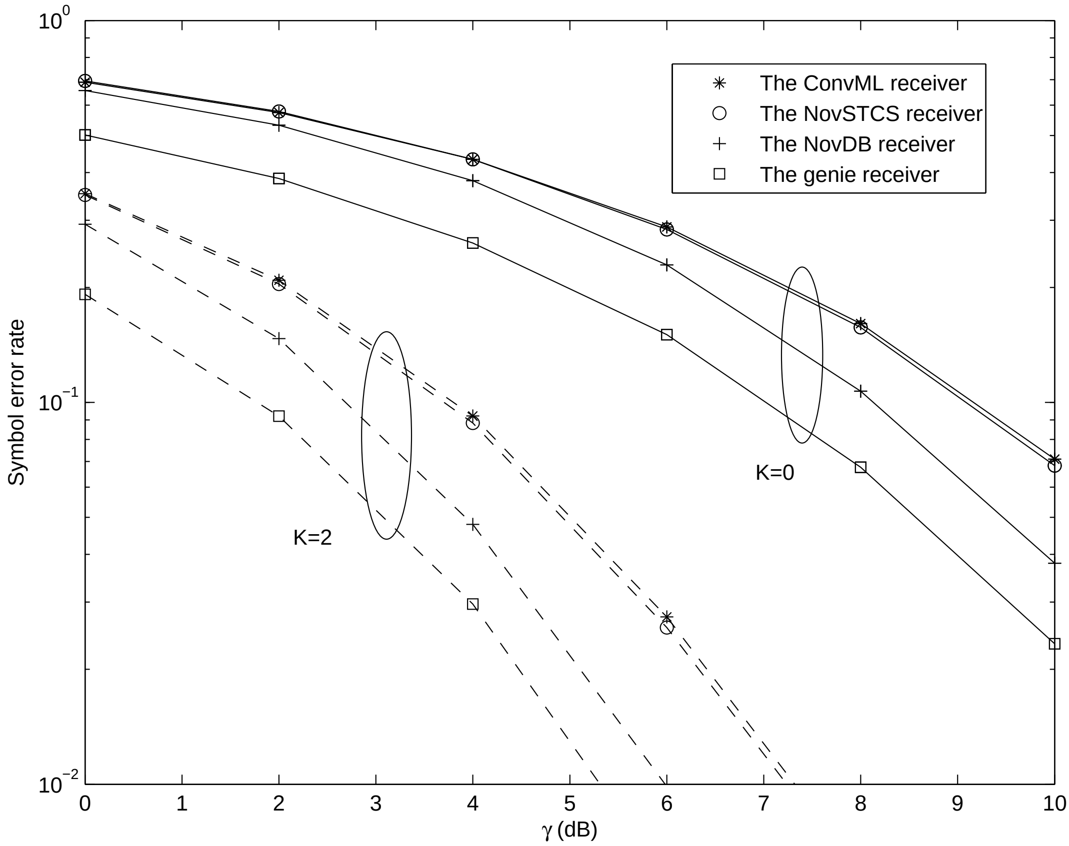

Fig. 5. Symbol error rates of receivers in Case 1 for $K = 0$ (solid line) and $K = 2$ (dashed line) when $r = 4$, $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, and 16-QAM is used.

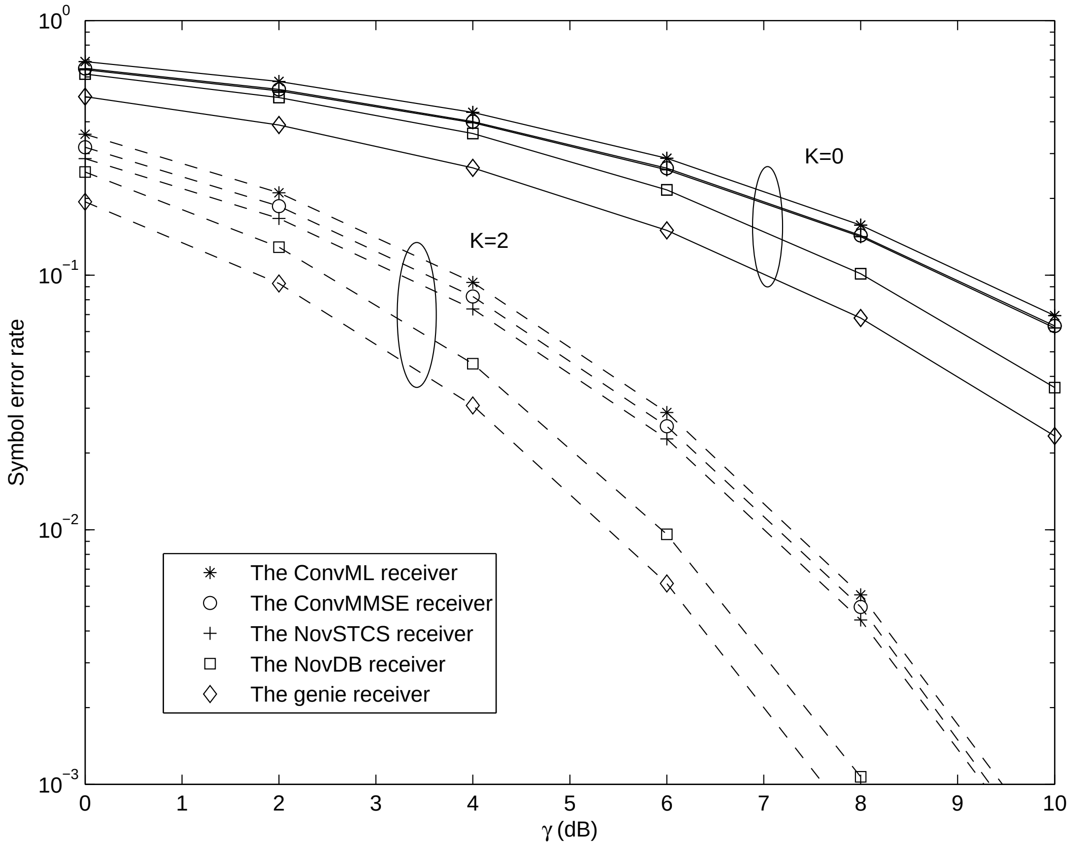

Fig. 6. Symbol error rates of receivers in Case 2 for $K = 0$ (solid line) and $K = 2$ (dashed line) when $r = 4$, $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, and 16-QAM is used.

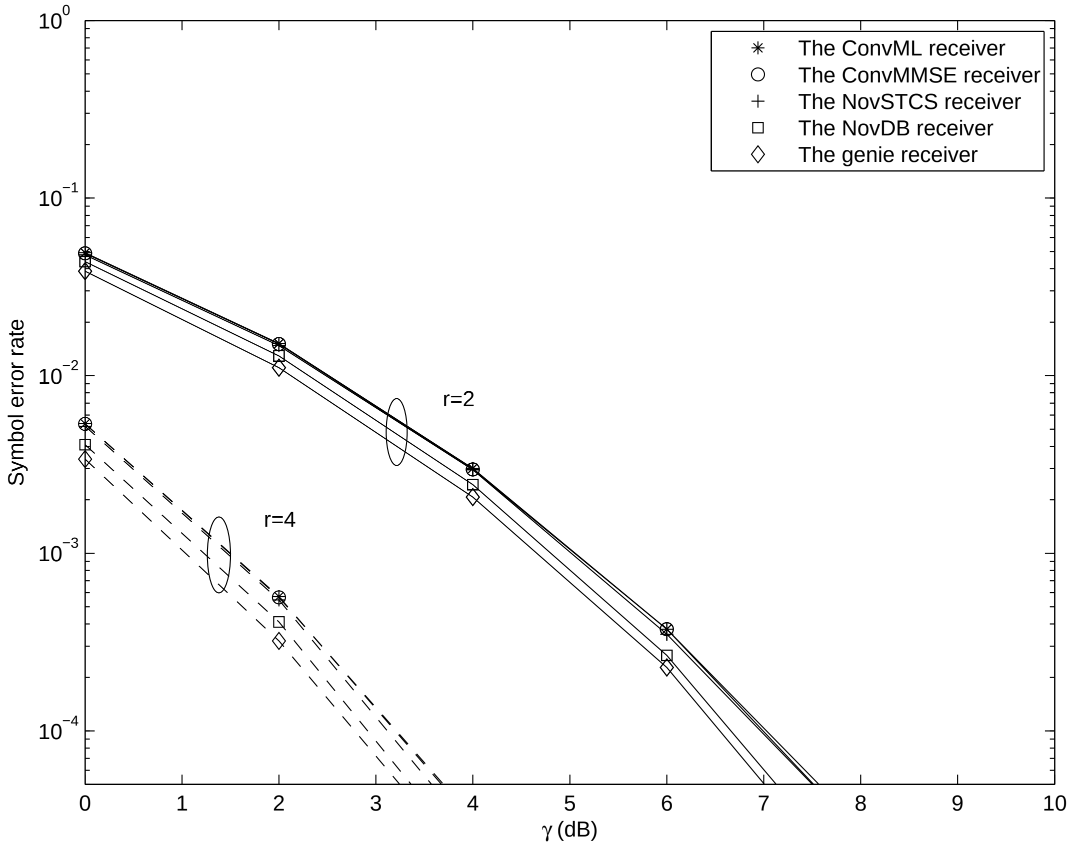

Fig. 7. Symbol error rates of receivers in Case 2 for $r = 2$ (solid line) and $r = 4$ (dashed line), when $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, $K = 2$, and QPSK is used.

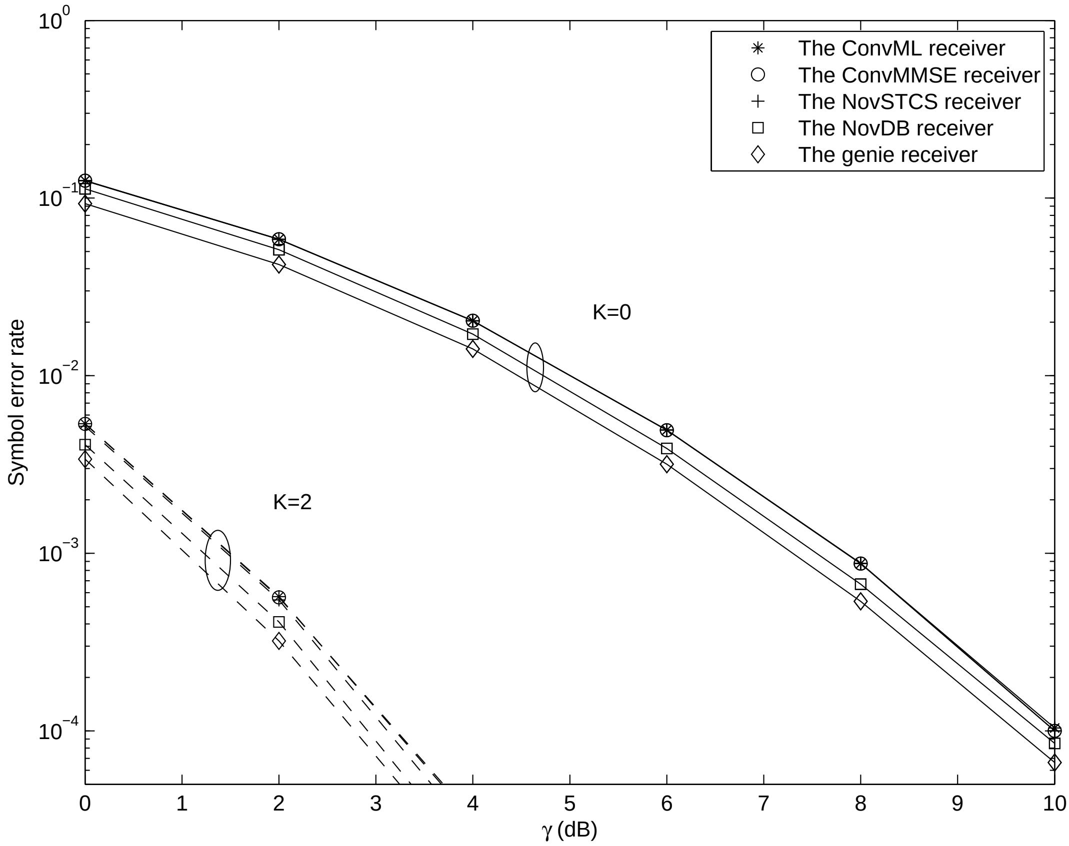

Fig. 8. Symbol error rates of receivers in Case 2 for $K = 0$ (solid line) and $K = 2$ (dashed line), when $r = 4$, $\frac{d_t}{\lambda} = \frac{d_r}{\lambda} = 0.5$, $M = 20$, and QPSK is used.