Performance of Noncoherent Maximum-Likelihood Sequence Detection for Differential OFDM Systems with Diversity Reception

Ding-Bing Lin\textsuperscript{1}, Ping-Hung Chiang\textsuperscript{2}, Hsueh-Jyh Li\textsuperscript{2}

\textsuperscript{1}Institute of Computer and Communication, National Taipei University of Technology, Taiwan, R.O.C.
\textsuperscript{2}Graduate Institute of Communication Engineering, National Taiwan University, Taiwan, R.O.C.

Abstract—For the single-carrier M-ary differential phase-shift keying (MDPSK), the multiple-symbol differential detector, or the noncoherent maximum-likelihood sequence detector (NSD), and its three special cases, namely, the noncoherent one-shot detector, the linearly predictive decision-feedback (DF) detector, and the linearly predictive Viterbi receiver are reviewed based on a hierarchical interpretation. For the multicarrier transmission, the differential OFDM systems with diversity reception are discussed. It is well known that there are two types of differential OFDM systems, namely, the \textit{time domain} differentially encoded OFDM (TD-OFDM) and the \textit{frequency domain} differentially encoded OFDM (FD-OFDM). In this paper, the NSD and its special cases are incorporated to the differential OFDM systems. Furthermore, a simple closed-form BER expression for the differential OFDM systems utilizing the noncoherent one-shot detector with diversity reception in the time-varying multipath Rayleigh fading channels is given. Numerical results have revealed that the noncoherent one-shot detector with diversity reception can obtain significant performance improvement. However, when only one or two receiving branches are available, the implementation of the linearly predictive DF detector or the linearly predictive Viterbi receiver is necessary for achieving better performance.

\textbf{Index Terms}—Noncoherent ML sequence detection, multiple-symbol differential detection, noncoherent one-shot detection, decision-feedback differential detection, linear prediction, estimator-detector, Viterbi algorithm, differential OFDM, ICI, diversity reception.
I. INTRODUCTION

For the single-carrier M-ary differential phase-shift keying (MDPSK) [1] [2], Divsalar et al. [3] and Ho et al. [4] proposed the multiple-symbol differential detector, or the noncoherent maximum-likelihood (ML) sequence detector, to improve the conventional product detector [1], or the noncoherent one-shot detector. Since the noncoherent one-shot detector is built assuming that the channel is constant over two consecutive symbol durations, its bit-error-rate (BER) performance exhibits an error floor in the time-selective fading channel. It is proved theoretically that the noncoherent ML sequence detector (NSD) can lower this error floor significantly by detecting a sequence of symbols jointly [4]. Although the NSD is optimal, its complexity increases exponentially with the number of symbols being jointly detected. To reduce its complexity and obtain the satisfactory performance simultaneously, the linearly predictive decision-feedback (DF) detector [5] and the linearly predictive Viterbi receiver [6] [7] [8] were proposed. Generally, the metrics used by the above two detection schemes contain a linear predictor of the fading-plus-noise processes. Therefore, they have an attractive estimator-detector structure.

Orthogonal frequency division multiplexing (OFDM) is an excellent technique that is capable of reducing the frequency-selective fading into the frequency-flat fading [9] [10]. For this multicarrier scheme, the differential modulation can be applied on each subcarrier, and the corresponding differential encoding and detection can be performed in either the time domain or the frequency domain. According to the direction of the differential encoding and detection (see Fig. 1), the differential OFDM systems are classified into two categories: 1) the time domain differential modulation OFDM (TD-OFDM) [10]-[15]; 2) the frequency domain differential modulation OFDM (FD-OFDM) [10] [13]. The former has been standardized in the terrestrial digital audio broadcasting (DAB) system [16].

Nevertheless, only the noncoherent one-shot detector was considered in the above cited papers regarding the differential OFDM systems. In this paper, the NSD and its three special cases, namely, the noncoherent one-shot detector, the linearly predictive DF detector, and the linearly predictive Viterbi receiver are incorporated to the differential OFDM systems with diversity reception. Furthermore, a simple closed-form
BER expression for the differential OFDM systems utilizing the noncoherent one-shot detector with diversity reception in the time-varying multipath Rayleigh fading channels is given.

The rest of this paper is organized as follows. In Section II, a hierarchical interpretation of the NSD and its three special cases are presented. In Section III, the preliminaries of the differential OFDM systems with diversity reception are described. In Section IV, the receiver designs of the differential OFDM systems according to the NSD and its three special cases are illustrated. Also, a simple BER expression for the differential OFDM systems employing the noncoherent one-shot detector is provided. Numerical results are shown in Section V, whereas the conclusions are drawn in Section VI.

II. NONCOHERENT ML SEQUENCE DETECTOR

In this section, the NSD with diversity reception is reviewed. Then, three special cases of the NSD are introduced, namely, the noncoherent one-shot detector, the linearly predictive DF detector, and the linearly predictive Viterbi receiver. At the end of this section, a hierarchical interpretation of the NSD and its special cases are also given.

For the single-carrier transmission, an $N_T$-symbol sequence is assumed to be transmitted and received over a single-input-multiple-output (SIMO) time-varying flat Rayleigh fading channel. Let $v_n$ denote the $n^{\text{th}}$ information symbol, which is a $L$-PSK symbol, i.e., $v_n \in \Omega_v$ and $\Omega_v = \{\exp(j2\pi l/L), \ l = 0,1,\cdots, L-1\}$. Each information symbol is differentially encoded as follows [1] [2].

$$x_n = v_n \cdot x_{n-1}, \ x_0 = 1.$$  \hspace{1cm} (1)

Obviously, the differentially encoded symbol, or the channel symbol, $x_n$ still belongs to the $L$-ary constellation $\Omega_v$. Then, the $1 \times Q$ received signal vector for the $n^{\text{th}}$ symbol interval from all $Q$ receive antennas is

$$r_n = x_n h_n + w_n,$$  \hspace{1cm} (2)
where \( \mathbf{r}_n = \left[ r_n^{(0)} \ r_n^{(1)} \ \cdots \ r_n^{(Q-1)} \right] \), \( \mathbf{h}_n = \left[ h_n^{(0)} \ h_n^{(1)} \ \cdots \ h_n^{(Q-1)} \right] \), and \( \mathbf{w}_n = \left[ w_n^{(0)} \ w_n^{(1)} \ \cdots \ w_n^{(Q-1)} \right] \). \( r_n^{(q)} \) is the received signal from the \( q \)th receive antenna in the \( n \)th symbol interval. \( h_n^{(q)} \) is the path gain with mean zero and autocorrelation \( \phi_{mm} \triangleq E \left[ h_n^{(q)} \left( h_n^{(q)} \right)^H \right] \) assuming \( Q \) identical and independent channels. \( w_n^{(q)} \) is the AWGN with mean zero and variance \( \sigma_w^2 \). Suppose that the transmission begins at time \( n = 0 \) and ends at time \( n = N_T - 1 \). The idea of the NSD is the optimum block detection for the entire \( N_T - 1 \) information symbols as a whole. Stacking up the variables involved in the detection yields the signal model as

\[
\mathbf{R} = \mathbf{XH} + \mathbf{W},
\]

(3)

where \( \mathbf{R} = \left[ \mathbf{r}_0^T \ \mathbf{r}_1^T \ \cdots \ \mathbf{r}_{N_T-1}^T \right]^T \), \( \mathbf{X} = \text{diag} \{ x_0, x_1, \cdots, x_{N_T-1} \} \), \( \mathbf{H} = \left[ \mathbf{h}_0 \ \mathbf{h}_1 \ \cdots \ \mathbf{h}_{N_T-1} \right]^T \), and \( \mathbf{W} = \left[ \mathbf{w}_0^T \ \mathbf{w}_1^T \ \cdots \ \mathbf{w}_{N_T-1}^T \right]^T \). From (1), it is clear that there exists a one-to-one correspondence between the \((N_T - 1) \times 1\) information symbol vector \( \mathbf{v} = \left[ v_1 \ v_2 \ \cdots \ v_{N_T-1} \right]^T \) and the \( N_T \times 1 \) channel symbol vector \( \mathbf{x} = \left[ x_0 \ x_1 \ \cdots \ x_{N_T-1} \right]^T \). Then the ML estimate of \( \mathbf{v} \) is obtained through

\[
\hat{\mathbf{v}} = \arg \max_{\mathbf{v}} f \left( \mathbf{R} | \mathbf{v} \right),
\]

(4)

where the likelihood function is given by

\[
f \left( \mathbf{R} | \mathbf{v} \right) = \left[ \pi^{N_T} \det \left( \mathbf{C}_r \right) \right]^{-Q} \exp \left[ - \text{tr} \left( \mathbf{R}^H \mathbf{C}_r^{-1} \mathbf{R} \right) \right],
\]

(5)

since each column of the \( N_T \times Q \) matrix \( \mathbf{R} \) conditioned on \( \mathbf{v} \) is Gaussian distributed with mean zero and covariance matrix \( \mathbf{C}_r = \mathbf{X} \mathbf{\Phi} \mathbf{X}^H \), where \( \mathbf{\Phi} = \mathbf{C}_h + \sigma_w^2 \mathbf{I}_{N_T} \) and \( \mathbf{C}_h \) is the covariance matrix of \( \mathbf{h}^{(q)} = \left[ h_0^{(q)} \ h_1^{(q)} \ \cdots \ h_{N_T-1}^{(q)} \right]^T \). Note that \( \mathbf{X}^H \mathbf{X} = \mathbf{I}_{N_T} \), then \( \det \left( \mathbf{C}_r \right) = \det \left( \mathbf{\Phi} \right) \) dose not depend on \( \mathbf{v} \). Therefore, (4) is simplified to

\[
\hat{\mathbf{v}} = \arg \min_{\mathbf{v}} \left\{ \text{tr} \left( \mathbf{R}^H \mathbf{C}_r^{-1} \mathbf{R} \right) \right\} = \arg \min_{\mathbf{v}} \left\{ \text{tr} \left( \mathbf{R}^H \mathbf{X} \mathbf{\Phi}^{-1} \mathbf{X}^H \mathbf{R} \right) \right\},
\]

(6)
which is the noncoherent sequence detector that optimum in the ML sense [4]. As a result that its complexity grows exponentially with \( N_r \), the direct implementation of this detector is not recommended. In the subsequence, we introduce three special cases of this NSD, which can be carried out in practice.

### A. The Noncoherent One-Shot Detector

The noncoherent one-shot detector operates in a symbol-by-symbol manner (i.e. \( N_r = 2 \)). Thus, the ML decision rule for \( v_n \) is as

\[
\hat{v}_n = \arg \max_{v_n} f(\mathcal{R}|v_n),
\]  

(7)

where

\[
\mathcal{R} = \mathcal{X} \mathcal{H} + \mathcal{W}
\]

(8)

Performing similar derivation to that of NSD yields

\[
\hat{v}_n = \arg \min_{v_n} \{ \text{tr}(\mathcal{R}^H \mathcal{X} \Phi^{-1} \mathcal{X}^H \mathcal{R}) \},
\]  

(9)

where \( \Phi = \mathcal{C}_n + \sigma_w^2 \mathcal{I}_2 \) and \( \mathcal{C}_n \), the covariance matrix of \( \mathcal{H}^{(q)} = [h_{n-1}^{(q)} h_n^{(q)}]^T \), is given by

\[
\mathcal{C}_n = \begin{bmatrix}
\phi_0 & \phi_1^*
\phi_1 & \phi_0^*
\end{bmatrix}.
\]

Then, the inverse of \( \Phi \) is

\[
\Phi^{-1} = \frac{1}{(\phi_0 + \sigma_w^2)^2 - |\phi|^2} \begin{bmatrix}
\phi_0 + \sigma_w^2 & -\phi_1^*
-\phi_1 & \phi_0 + \sigma_w^2
\end{bmatrix}.
\]  

(10)

Define \( \Xi = \text{diag} \{1, v_n\} \), then \( \mathcal{X} = x_{n,1} \Xi \). Thereupon, \( \mathcal{X} \Phi^{-1} \mathcal{X}^H = \Xi \Phi^{-1} \Xi^H \) and hence (9) is equivalent to

\[
\hat{v}_n = \arg \min_{v_n} \{ \text{tr}(\mathcal{R}^H \Xi \Phi^{-1} \Xi^H \mathcal{R}) \}.
\]  

(11)

Then, plugging (10) into (11) yields
\[ \hat{v}_n = \arg \max_{v_n} \Re \left[ \text{tr} \left( \phi \begin{bmatrix} r_n \\ r_n^{H} \end{bmatrix} v_n^* \right) \right] = \arg \max_{v_n} \Re \left[ \phi^* v_n^* \sum_{q=0}^{Q-1} r_n^{(q)} \begin{bmatrix} r_n^{(q)} \end{bmatrix}^* \right] \]

which is the noncoherent one-shot detector that optimum in the ML sense and is dubbed the conventional receiver (CR), subsequently.

Ignoring the fading correlation \( \phi \) in (12) results in a suboptimum detector as

\[ \hat{v}_n = \arg \max_{v_n} \Re \left[ v_n^* \sum_{q=0}^{Q-1} r_n^{(q)} \begin{bmatrix} r_n^{(q)} \end{bmatrix}^* \right], \]

which is the well-known product detector [1]. For this suboptimum detector, Kam derived a closed-form BER expression as follows [2].

\[ P_b = \left( \frac{1-\mu}{2} \right)^Q \sum_{q=0}^{Q-1} \binom{Q-1+q}{q} \left( \frac{1+\mu}{2} \right)^q, \]

where \( \mu = |\phi| \Gamma/(1+\Gamma) \) for differential BPSK (i.e. \( L=2 \)) and \( \mu = \left\{ \frac{2[(1+\Gamma)/(|\phi| \Gamma)]^2-1}{1-\mu} \right\}^{1/2} \) for differential QPSK (i.e. \( L=4 \)). \( \Gamma = \phi_0^2/\sigma_w^2 \) is the average signal-to-noise-ratio (SNR) per receiving branch.

Then, (14) can be an upper bound for the BER of the CR given by (12).

### B. The Linearly Predictive Viterbi Receiver

The NSD given in (6) brings a linear-prediction interpretation of its structure. To avoid that its complexity increases exponentially with \( N_r \), the minimization problem in (6) is solved by using the innovations-based approach [6], which is mathematically equivalent to the Cholesky factorization approach [7] [17, Ch. 3.7] as follows. Applying the Cholesky factorization to the \( N_r \times N_r \) matrix \( \Phi \) produces

\[ \Phi = L \Sigma L^H, \]

where \( L \) is a \( N_r \times N_r \) lower triangular matrix and \( \Sigma = \text{diag} \{ \epsilon_0, \epsilon_1, \cdots, \epsilon_{N_r-1} \} \). The inverse of \( L \) is denoted as
Note that the element $p_0^n = -1$ and $p_k^n$ is the $k$th coefficient of the $n$th order one-step forward linear predictor for the fading-plus-noise process $\{h_n^{(q)} + x_n^{*}w_n^{(q)}\}$ and $\varepsilon_n$ is the corresponding mean square prediction error. 

Since the processes $\{h_n^{(q)} + x_n^{*}w_n^{(q)}\}$, for $q = 0, 1, \ldots, Q - 1$ are independently and identically distributed (i.i.d.), the prediction coefficients are identical for all $Q$ processes. Then, substituting (15) and (16) into (6) results in a linearly predictive sequence detector as

$$\hat{\mathbf{v}} = \arg \min_{\mathbf{v}} \sum_{n=0}^{N_t-1} \|\mathbf{y}_n - \mathbf{\tilde{y}}\|^2_{\varepsilon_n},$$

(17)

where $\mathbf{\tilde{y}} = \sum_{k=1}^{n} p_k^n x_{n-k}^* \mathbf{r}_{n-k}$ is the $n$th order prediction of the fading-plus-noise vector $\mathbf{y}_n = x_n^* \mathbf{r}_n = \mathbf{h}_n + x_n^* \mathbf{w}_n$.

As a result of the infinite memory inherent in this estimator-detector structure, its complexity still increases exponentially with $n$. To limit the complexity, a truncation of the memory is applied. That is, only the recent $K$ observations $\mathbf{r}_{n-K}, \mathbf{r}_{n-K+1}, \ldots, \mathbf{r}_{n-1}$ are used in the estimator. Accordingly, an approximate ML sequence detector containing a $K$th order linear predictor is

$$\hat{\mathbf{v}} = \arg \min_{\mathbf{v}} \sum_{n=0}^{N_t-1} \|\mathbf{x}_n^{*} - \sum_{k=1}^{K} p_k^K x_{n-k}^* \mathbf{r}_{n-k}\|^2_{\varepsilon_K} = \arg \min_{\mathbf{v}} \sum_{n=0}^{N_t-1} \sum_{q=0}^{Q-1} x_n^{(q)} - \sum_{k=1}^{K} p_k^K x_{n-k}^{(q)}\|^2,$$

(18)

where $\varepsilon_K$, a constant not affecting the detection, is dropped. Based on (18), a trellis with $L^K$ states is defined as follows. Each state of the trellis in the $n$th symbol interval is represented as $\mathbf{\Gamma}_n = [x_{n-K}, x_{n-K+1}, \ldots, x_{n+1}]$. There are $L$ transitions emerging from each state and terminating in $L$ different states $\mathbf{\Gamma}_{n+1} = [x_{n+1}, x_{n+2}, \ldots, x_{n+1}]$, where $x_n$ is determined by $\mathbf{v}_n$ via the differential encoding given in (1). The branch metric associated with each transition is defined as
\[ \Delta(\Gamma_n, v_n) = \left\| x_n^r - \sum_{k=1}^{K} p_k^* x_{n-k}^r \right\|^2 = \sum_{q=0}^{Q-1} \left\| x_n^{r(q)} - \sum_{k=1}^{K} p_k^* x_{n-k}^{r(q)} \right\|^2. \] (19)

Then the Viterbi algorithm [18, Ch. 19] with a fixed decision delay \( D \) can be applied to this trellis for solving the minimization problem in (18).

The algorithmic complexity of the above trellis-based sequence detector is proportional to the number of states \( L^K \). Even for moderate values of \( L \) and \( K \), the complexity increases drastically with \( K \). For example, a trellis will contain \( L^K = 1024 \) states if \( L = 4 \) (i.e. differential QPSK) and \( K = 5 \). It might difficult to implement a real time sequence detector with such a large-size trellis in practice. Accordingly, the reduced-state sequence detection [8] is incorporated to overcome the implementation complexity. For this, a reduced state \( \tilde{\Gamma}_n \) is defined with the most recent \( U \) channel symbols, namely, \( \tilde{\Gamma}_n = [x_{n-U}, x_{n-U+1}, \ldots, x_{n-1}] \), and \( U < K \). The calculation of the branch metric \( \Delta(\tilde{\Gamma}_n, v_n) \) involves \( K + 1 \) observations and \( K + 1 \) trial channel symbols. Thus, there are still \( K - U \) channel symbols \( x_{n-K}, x_{n-K+1}, \ldots, x_{n-U-1} \) unavailable in the state \( \tilde{\Gamma}_n \). However, one can extract these unavailable channel symbols from the survivor history according to the per-survivor processing (PSP) technique [8]. More specifically, these unavailable symbols are extracted from the survivor (i.e. the remaining path) entering the state \( \tilde{\Gamma}_n \). Since only one path enters each state, then each state has its own feedback-decision for the estimator contained in the branch metric. Thereupon, the branch metric for this reduced-state approach is written as

\[ \Delta(\tilde{\Gamma}_n, v_n) = \sum_{q=0}^{Q-1} \left\| x_n^{r(q)} - \sum_{k=1}^{U} p_k^* x_{n-k}^{r(q)} - \sum_{k=U+1}^{K} p_k^* \tilde{x}_{n-k}^{r(q)} \right\|^2, \] (20)

where \( \tilde{x}_{n-K}, \tilde{x}_{n-K+1}, \ldots, \tilde{x}_{n-U-1} \) are determined by the survivor entering the state \( \tilde{\Gamma}_n \). Subsequently, this reduced-state trellis-based sequence detector [7] is named the Viterbi receiver (VR).
C. The Linear Predictive Decision-Feedback Detector

Here, we consider an extreme case of $U = 0$ for the aforementioned reduced-state trellis-based sequence detector. In this case, only a single path is allowed to survive, then all the $K$ channel symbols required by the estimator are found from the history of this survivor and hence the sequence detector degenerates into a DF detector performing symbol-by-symbol detection as

$$\hat{v}_n = \arg \min_{v_n} \left\{ x_n^* r_n - \sum_{k=1}^{K} p_k^* x_{n-k}^* r_{n-k} \right\}^2 = \arg \min_{v_n} \sum_{q=0}^{Q-1} x_n^* r_n^{(q)} - \sum_{k=1}^{K} p_k^* x_{n-k}^* r_{n-k}^{(q)} \right\}^2,$$

which is exactly the linearly predictive DF detector [5]. Subsequently, this detector is called the decision-feedback receiver (DR).

D. The Hierarchy of the NSD and its Special Cases

Considering the DR with the prediction order $K = 1$, (21) is reduced to

$$\hat{v}_n = \arg \min_{v_n} \left\{ x_n^* r_n^{(q)} - p_1^* x_{n-1}^* r_{n-1}^{(q)} \right\}^2 = \arg \max_{v_n} \Re \left[ \left( p_1^* x_n^* r_n^{(q)} \sum_{q=0}^{Q-1} r_n^{(q)} \left( r_{n-1}^{(q)} \right)^* \right) \right].$$

In light of (10), (15), and (16), it is easy to find that $p_1^* = \phi_1 / \left( \phi_0 + \sigma_w^2 \right)$. Thus, it is readily to see that (22) is exactly the CR given in (12), provided that $\phi_0 + \sigma_w^2$ is real-valued.

Subsequently, we interpret the hierarchy of the NSD (17) and its three special cases, namely, the CR (22), the DR (21), and the VR (20). Firstly, the VR is a suboptimum version of the NSD; secondly, the DR is the one-survivor version of the VR; finally, the CR is the one-order-prediction version of the DR. Moreover, since the NSD and its special cases are inherently of the estimator-detector structure, their performances are dominated by the qualities of their estimates regarding the fading-plus-noise processes. More specifically, the VR employs the PSP technique so that each state in the trellis has its own survivor. Then the estimates contained in the branch metrics are different from state to state or from path to path. Indeed, the VR benefits from the path diversity and thus performs better than its one-survivor version, namely, the DR. Furthermore,
with a larger prediction order and hence better estimates, the DR performs better than its one-order-prediction version, i.e. the CR.

III. PRELIMINARIES FOR DIFFERENTIAL OFDM SYSTEMS WITH DIVERSITY RECEIPTION

It is well known that there are two types of differential OFDM systems, namely, the TD-OFDM and the FD-OFDM. The comparison is illustrated in Fig. 1. In this section, the system model and some statistical properties of the differential OFDM systems with diversity reception are provided. Then, the receiver designs will be discussed later in Section IV.

A. System Model

As shown in Fig. 2, we consider the DFT-based OFDM transmission over the SIMO WSSUS Rayleigh fading channel and assume sufficient cyclic prefix (CP) is inserted such that the inter-block-interference (IBI) is eliminated completely [9]. Let $V_{i,k}$ denote the $L$-PSK information symbol for the $k$th subcarrier in the $i$th OFDM block interval. Then, $V_{i,k}$ is differentially encoded via [10] [13]

$$X_{i,k} = \begin{cases} V_{i,k} X_{i-1,k}, & X_{0,k} = 1 \text{ (TD-OFDM)} \\ V_{i,k} X_{i-1,k}, & X_{1,k} = 1 \text{ (FD-OFDM)} \end{cases}$$

(23)

Let $N$, $G$, and $M$ represent the number of subcarriers, guard samples, and taps, respectively. Also, let $h_{m,i,j}$ be the $m$th tap coefficient and $n_{i,j}^{(q)}$ be the AWGN in the $\left[(N+G)(i-1)+G+i\right]$th symbol interval for the $q$th receiving branch. Assume they are complex Gaussian distributed with mean zero and variances $\sigma_m^2$ and $N_0$, respectively. Then, the discrete-time baseband representation of the received signal for the $k$th subcarrier in the $i$th block interval from the $q$th receive antenna is [9]

$$R_{i,k}^{(q)} = H_{i,k}^{(q)} X_{i,k} + C_{i,k}^{(q)} + N_{i,k}^{(q)}, \text{ for } 0 \leq k \leq N-1,$$

(24)

where
\begin{align}
H_{i,k}^{(q)} &= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{m=0}^{M-1} h_{m,i,j}^{(q)} e^{-j2\pi j m}\frac{2\pi i n}{N}, \quad (25) \\
C_{i,k}^{(q)} &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{i=0}^{N-1} \sum_{m=0}^{M-1} h_{m,i,j}^{(q)} e^{-j2\pi j m}\frac{2\pi i (n-i)}{N} X_{i,n}, \quad (26) \\
N_{i,k}^{(q)} &= \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} n_{i,j}^{(q)} e^{-j2\pi j l}\frac{2\pi i l}{N}. \quad (27)
\end{align}

Here, \( H_{i,k}^{(q)} \) and \( C_{i,k}^{(q)} \) are the multiplicative distortion (MD) and the ICI, respectively. As a linear combination of i.i.d. Gaussian random variables, \( N_{i,k}^{(q)} \) is still the AWGN with mean zero and variance \( N_0 \).

For both the TD- and the FD-OFDM with the assumption that \( \{V_{i,k}\} \) is a set of i.i.d. symbols, the channel symbols \( \{X_{i,k}\} \) are uncorrelated with mean zero and variance unity (i.e. symbol energy \( E_s = 1 \)). Therefore, the ICI \( \{C_{i,k}^{(q)}\} \) are at least uncorrelated with mean zero and variance \( \sigma_C^2 \) even though the distribution of the ICI is not easy to verify. However, since independent Gaussian noise results in the smallest capacity, it is reasonable to model \( \{C_{i,k}^{(q)}\} \) as Gaussian random variables and hence to achieve the performance bound [19].

Indeed, (24) is rewritten as
\[
R_{i,k}^{(q)} = H_{i,k}^{(q)} X_{i,k} + W_{i,k}^{(q)}, \quad (28)
\]

where \( W_{i,k}^{(q)} = C_{i,k}^{(q)} + N_{i,k}^{(q)} \) is the equivalent AWGN with mean zero and variance \( \sigma_W^2 = \sigma_C^2 + N_0 \).

\section{Statistical Properties}

For the ease of demonstrating the receiver designs later in Section IV, the statistical properties of the channel tap coefficient \( h_{m,i,j}^{(q)} \), the MD \( H_{i,k}^{(q)} \), and the ICI \( C_{i,k}^{(q)} \) are clarified. Since the channels are assumed to be spatially independent and identical WSSUS Rayleigh fading channels with the classical Doppler spectrum, the correlation between the tap coefficients is
\[
E \left[ h_{m,i,j}^{(q)} \left( h_{m',i',p}^{(q)} \right)^* \right] = \sigma_m^2 \mathcal{J}_0 \left[ 2\pi f_D T \left[ (1-i') + (i-i')(N+G) \right] \right] \delta_{mm'} \delta_{qq'}, \quad (29)
\]
where $\sigma_m^2$ is the fading power of the $m$th tap, $J_0(\cdot)$ is the zero-order Bessel function of the first kind, $f_D$ is the maximum Doppler frequency, $T$ is the symbol duration or the reciprocal of the system bandwidth, and $\delta$ is the Kronecker delta. Considering the exponential power delay profile \cite{20} with the constraint $\sum_{m=0}^{M-1} \sigma_m^2 = 1$, we have

$$\sigma_m^2 = \frac{1 - e^{-1/d}}{1 - e^{-M/d}} e^{-m/d} = \frac{1 - \lambda}{1 - \lambda^M} \lambda^m,$$

where $\lambda = e^{-1/d}$ and the delay control $d$ dominates the normalized root-mean-square (RMS) delay spread $\tau_{rms}/T$.

1) **Correlation of the Multiplicative Distortion**

For spatially identical channels, let $\rho_{H,H}^{(q)} (\Delta i, \Delta k) \triangleq \rho_{H,H} (\Delta i, \Delta k)$, for $q = 0, 1, \cdots, Q-1$, where $\rho_{H,H}^{(q)} (\Delta i, \Delta k)$ is the correlation of $H_{i,k}^{(q)}$. Consequently, the correlation of the MD is given as \cite{21}

$$\rho_{H,H} (\Delta i, \Delta k) = \mathbb{E} \left[ H_{i,k}^{(q)} H_{i,k}^{(q)*} \right] = \left( \sum_{m=0}^{M-1} \sigma_m^2 e^{-2\pi\lambda_m N} \right) \left( \frac{1}{N^2} \sum_{l=1}^{N-1} \left( N - l \right) J_0 \left( 2\pi f_D T \left( N + G \Delta i + l \right) \right) \right).$$

Thereupon, the correlation for the same subcarrier between adjacent blocks is

$$\rho_t \triangleq \rho_{H,H} (1,0) = \frac{1}{N^2} \sum_{l=1}^{N-1} \left( N - l \right) J_0 \left( 2\pi f_D T \left( N + G + l \right) \right),$$

which characterizes the time-selectivity of the channel. Similarly, the correlation for the same block between adjacent subcarriers is

$$\rho_f \triangleq \rho_{H,H} (0,1) = \left( \sum_{m=0}^{M-1} \sigma_m^2 e^{-2\pi\lambda_m N} \right) \left( \frac{1}{N^2} \left[ N + 2 \sum_{l=1}^{N-1} \left( N - l \right) J_0 \left( 2\pi f_D T l \right) \right] \right),$$

which represents the frequency-selectivity of the channel.

2) **Variances of the Multiplicative Distortion and the ICI**

According to (25) and (31), it is clear that the MD $H_{i,k}^{(q)}$ is of mean zero and hence variance
\[ \sigma^2_{\hat{H}} = \rho_{H}(0,0) = \frac{1}{N^2} \left[ N + 2 \sum_{l=1}^{N-1} (N-l)J_0(2\pi f_B T_l) \right]. \]  

(34)

In addition, the variance of the ICI \( C_{i,k}^{(q)} \) is calculated as [21]

\[ \sigma^2_{c} \triangleq \mathbb{E} \left[ C_{i,k}^{(q)}(C_{i,k}^{(q)})^* \right] = \frac{E_s}{N^2} \sum_{n=0}^{N-1} \sum_{l=-N+1}^{N-1} \left( N-|l| \right)J_0(2\pi f_B T_l)e^{-\frac{2\pi l(n-k)}{N}}. \]  

(35)

The above expression is exact; however, a tight upper bound on the variance of the ICI was derived by Li et al. [22], i.e.,

\[ \sigma^2_{c} \leq \frac{1}{24} (2\pi f_B NT)^2 E_s. \]  

(36)

This approximation is quite accurate for \( f_B NT < 0.15 \), even though it is derived based on assuming infinite number of subcarriers. Furthermore, with \( \sigma^2_{\hat{H}} + \sigma^2_{c}/E_s = 1 \) [21], a tight lower bound on the variance of the MD can be expressed as

\[ \sigma^2_{\hat{H}} \geq 1 - \frac{1}{24} (2\pi f_B NT)^2. \]  

(37)

IV. RECEIVER DESIGNS FOR DIFFERENTIAL OFDM SYSTEMS WITH DIVERSITY RECEPTION

In this section, the NSD and its special cases, i.e., the CR, the DR, and the VR, introduced in Section II, are applied to the differential OFDM systems, namely, the TD- and the FD-OFDM. The theoretical BERs for both systems employing the CR are also given.

A. TD-OFDM

For the TD-OFDM, since the differential encoding and the detection are independent and simultaneous on all \( N \) subcarriers (see Fig. 1), only the detection on the \( k^{th} \) subcarrier is illustrated subsequently. In light of (2) and (28), the \( 1 \times Q \) received signal vector for \( k^{th} \) subcarrier in the \( r^{th} \) block interval from all \( Q \) receive antennas is
\[ r_{i,k} = X_{i,k} h_{i,k} + w_{i,k}, \quad \text{(38)} \]

where \[ r_{i,k} = \begin{bmatrix} R_{i,k}^{(0)} & R_{i,k}^{(1)} & \cdots & R_{i,k}^{(Q-1)} \end{bmatrix}, \]
\[ h_{i,k} = \begin{bmatrix} H_{i,k}^{(0)} & H_{i,k}^{(1)} & \cdots & H_{i,k}^{(Q-1)} \end{bmatrix}, \]
and \[ w_{i,k} = \begin{bmatrix} W_{i,k}^{(0)} & W_{i,k}^{(1)} & \cdots & W_{i,k}^{(Q-1)} \end{bmatrix}. \]

Then, stacking up the \(1 \times Q\) received signal vectors results in the signal model for the NSD as

\[ \mathbf{R}_k = \mathbf{X}_k \mathbf{H}_k + \mathbf{W}_k, \quad \text{(39)} \]

where \[ \mathbf{R}_k = \begin{bmatrix} r_{0,k}^T & r_{1,k}^T & \cdots & r_{N_T-1,k}^T \end{bmatrix}^T, \]
\[ \mathbf{X}_k = \text{diag}\{X_{0,k}, X_{1,k}, \ldots, X_{N_T-1,k}\}, \]
\[ \mathbf{H}_k = \begin{bmatrix} h_{0,k}^T & h_{1,k}^T & \cdots & h_{N_T-1,k}^T \end{bmatrix}^T, \]
and \[ \mathbf{W}_k = \begin{bmatrix} w_{0,k}^T & w_{1,k}^T & \cdots & w_{N_T-1,k}^T \end{bmatrix}^T. \]

Then, from (6), the ML estimate of the \((N_T - 1) \times 1\) information symbol vector \[ \mathbf{v}_k = \begin{bmatrix} V_{1,k} & V_{2,k} & \cdots & V_{N_T-1,k} \end{bmatrix}^T \]
is given by

\[ \hat{\mathbf{v}}_k = \arg \min_{\mathbf{v}_k} \left\{ \text{tr} \left[ \mathbf{R}_k^H \mathbf{X}_k \left( \Phi^{(TD)} \right)^{-1} \mathbf{X}_k^H \mathbf{R}_k \right] \right\}, \quad \text{(40)} \]

where

\[ \Phi^{(TD)} = \mathbf{C}_h^{(TD)} + \sigma_h^2 I_{N_T} = \mathbf{C}_h^{(TD)} + \left( \sigma_C^2 + N_0 \right) I_{N_T}, \quad \text{(41)} \]

and \[ \mathbf{C}_h^{(TD)}, \] the covariance matrix of \[ \mathbf{h}_k^{(q)} = \begin{bmatrix} H_{0,k}^{(q)} & H_{1,k}^{(q)} & \cdots & H_{N_T-1,k}^{(q)} \end{bmatrix}^T, \]
is evaluated via (31) with \( \Delta k = 0 \).

1) CR

In light of (12), the noncoherent one-shot detector obtains the estimate for the information symbol \( V_{i,k} \) according to

\[ \hat{V}_{i,k} = \arg \max_{\mathbf{V}_{i,k}} \Re \left[ \rho_i^* V_{i,k} \sum_{q=0}^{Q-1} R_{i,k}^{(q)} \left( R_{i-1,k}^{(q)} \right)^* \right] = \arg \max_{\mathbf{V}_{i,k}} \Re \left[ V_{i,k}^* \sum_{q=0}^{Q-1} R_{i,k}^{(q)} \left( R_{i-1,k}^{(q)} \right)^* \right]. \quad \text{(42)} \]

The last step is due to that the fading correlation \( \rho_i \), given by (32), is real-valued. Thereupon, the BER of the TD-OFDM employing the CR is evaluated via (14) by setting \( \phi_i = \rho_i \) and \( \Gamma = \sigma_h^2 / \left( \sigma_C^2 + N_0 \right) \).
2) **DR**

Based on (21), the DF detector performs symbol-by-symbol detection through

$$
\hat{V}_{i,k} = \text{arg min}_{V_{i,k}} \left\{ \sum_{q=0}^{Q-1} \left[ X^*_i R^{(q)}_{i,k} - \sum_{j=1}^{K} p_{j,k}^{K} \tilde{X}^*_{i,j,k} R^{(q)}_{i-j,k} \right]^2 \right\},
$$

where the prediction coefficients \( \{ p_{j,k}^{K} \} \) is calculated via (15), (16), and (41).

3) **VR**

According to (20), the reduced-state trellis-based sequence detector with \( L^u \) states is defined as follows. Each state of the trellis in the \( i \)th block interval is denoted as \( \tilde{\Gamma}_{i,k} = [X_{i-U,k} \ X_{i-U+1,k} \ \cdots \ X_{i-1,k}] \), and the corresponding branch metric is expressed as

$$
\Delta(\tilde{\Gamma}_{i,k}, V_{i,k}) = \sum_{q=0}^{Q-1} \left[ X^*_i R^{(q)}_{i,k} - \sum_{j=1}^{L^u} p_{j,k}^{K} \tilde{X}^*_{i,j,k} R^{(q)}_{i-j,k} \right]^2,
$$

where the prediction coefficients \( \{ p_{j,k}^{K} \} \) are the same as the ones for the DR.

### B. **FD-OFDM**

For the FD-OFDM, since the differential encoding and the detection are independent and sequential for each OFDM blocks (see Fig. 1), only the detection in the \( i \)th block interval is described here. The NSD performs optimum block detection for the entire \( N-1 \) information symbols as a whole. Piling up the \( 1 \times Q \) received signal vectors given by (38) yields the signal model for the NSD as

$$
R_i = X_i H_i + W_i,
$$

where \( R_i = \begin{bmatrix} r_{i,0}^T & r_{i,1}^T & \cdots & r_{i,N-1}^T \end{bmatrix}^T \), \( X_i = \text{diag}\{X_{i,0}, X_{i,1}, \cdots, X_{i,N-1}\} \), \( H_i = \begin{bmatrix} h_{i,0}^T & h_{i,1}^T & \cdots & h_{i,N-1}^T \end{bmatrix}^T \), and \( W_i = \begin{bmatrix} w_{i,0}^T & w_{i,1}^T & \cdots & w_{i,N-1}^T \end{bmatrix}^T \). From (6), one can obtain the ML estimate of the \( (N-1) \times 1 \) information symbol vector \( V_i = \begin{bmatrix} V_{i,1} & V_{i,2} & \cdots & V_{i,N-1} \end{bmatrix}^T \) through
\[
\hat{v}_i = \arg \min_{\tilde{v}_i} \left\{ \text{tr} \left[ R_i H X_i \left( \Phi_i^{(FD)} \right)^{-1} X_i^H R_i \right] \right\},
\]
where
\[
\Phi_i^{(FD)} = C_h^{(FD)} + \sigma_H^2 I_N = C_h^{(FD)} + \left( \sigma_C^2 + N_0 \right) I_N,
\]
and \( C_h^{(FD)} \), the covariance matrix of \( h_i^{(q)} = \begin{bmatrix} H_{i,0}^{(q)} & H_{i,1}^{(q)} & \cdots & H_{i,N-1}^{(q)} \end{bmatrix}^T \), is calculated via (31) with \( \Delta i = 0 \).

1) **CR**

In light of (12), the CR evaluates the estimate for the information symbol \( V_{i,k} \) according to
\[
\hat{v}_{i,k} = \arg \max_{v_{i,k}} \Re \left\{ \rho_{2}^* V_{i,k} \sum_{q=0}^{Q-1} R_{i,k}^{(q)} \left( R_{i,k-1}^{(q)} \right)^* \right\}.
\]
However, based on (13), the suboptimum detector ignores the complex-valued fading correlation \( \rho_{2} \) and gets a suboptimum estimate of \( V_{i,k} \) by
\[
\hat{v}_{i,k} = \arg \max_{v_{i,k}} \Re \left\{ V_{i,k} \sum_{q=0}^{Q-1} R_{i,k}^{(q)} \left( R_{i,k-1}^{(q)} \right)^* \right\}.
\]
The BER of the FD-OFDM using this suboptimum detector is computed via (14) with the substitutions of \( \phi = \rho_{2} \) and \( \Gamma = \sigma_H^2 / \left( \sigma_C^2 + N_0 \right) \). Apparently, this BER expression is an upper bound on the performance of the FD-OFDM utilizing the CR, given by (48).

2) **DR**

Based on (21), the DF detector performs symbol-by-symbol detection through
\[
\hat{v}_{i,k} = \arg \min_{\tilde{v}_{i,k}} \left\| \sum_{q=0}^{Q-1} X_{i,k}^* R_{i,k}^{(q)} - \sum_{j=1}^{K} p_{j}^K \tilde{X}_{i,k-j}^* R_{i,k-j}^{(q)} \right\|^2,
\]
where the prediction coefficients \( \left\{ p_{j}^K \right\} \) are computed via (15), (16), and (47).
According to (20), the reduced-state trellis-based sequence detector with $L^U$ states is defined as follows. Each state of the trellis for the $k^{th}$ subcarrier is represented as $\hat{\Gamma}_{i,k} = [X_{i,k-U} \quad X_{i,k-U+1} \ldots \quad X_{i,k-1}]$, and the corresponding branch metric is expressed as

$$
\Delta(\hat{\Gamma}_{i,k}, V_{i,k}) = \sum_{q=0}^{Q-1} X_{i,k}^* R_{i,k}^{(q)} - \sum_{j=1}^{L^U} p_j^K X_{i,k-j}^* R_{i,k-j}^{(q)} - \sum_{j=U+1}^{K} p_j^K \hat{X}_{i,k-j}^* R_{i,k-j}^{(q)},
$$

(51)

where the prediction coefficients $\{p_j^K\}$ are the same as the ones for the DR.

V. NUMERICAL RESULTS

The simulation parameters are detailed as follows. 1) the carrier frequency and the system bandwidth are 1.8 GHz and 800 KHz, respectively, and thus the symbol duration is $T = 1.25 \mu$ seconds; 2) the number of subcarriers and guard samples are $N = 128$ and $G = 32$, respectively, and hence the total OFDM block duration is $(N + G)T = 200 \mu$ seconds; 3) the modulation is differential BPSK, i.e. $L = 2$; 4) the number of uncorrelated paths is $M = 12$; 5) the number of receive antennas is $Q = 1, 2, \text{ and } 4$; 6) for both the DR and the VR, the prediction order is $K = 5$; 7) for the VR, $U = 1$, i.e., the number of states is $L^U = 2$, and the decision delay is $D = 10$.

As shown in Figs. 3-6, the BERs of differential OFDM systems are compared in four extreme scenarios (see Tab. 1), namely, LTLF, HTLF, LTHF and HTHF, respectively. For instance, LTHF stands for low time-selectivity and high frequency-selectivity. In each figure, the performances of the CR, the DR, and the VR are presented from left to right. The matched filter bounds (MFBs) shown in all figures are the BERs of the single-carrier differential BPSK systems employing the noncoherent one-shot detector with one-, two-, and four-branch diversity reception, respectively, in the quasi-static Rayleigh fading channel. These bounds are produced via (14) by setting $\phi_1 = 1$. Apparently, they are the lower bounds of the BERs of differential
OFDM systems utilizing the CR. However, at low SNR level, they are slightly higher than the BERs of the systems using the DR or the VR in all channel conditions. This is due to that the DR and the VR, originated from the NSD, are inherently of better performances than the noncoherent one-shot detector (see Section II-D).

As shown in Fig. 3 (LTLF), since the channel selectivity is low, both the TD- and the FD-OFDM are of BERs similar to the corresponding MFBs disregarding which receiver structure is employed. On the other hands, as shown in Figs. 4-6, error floors appear when the channel selectivity is higher, especially for the systems utilizing the CR. The reasons for the presence of these error floors are twofold: 1) the ICI induced when the channel is not constant over an OFDM block duration; 2) the worse estimates of the fading-plus-noise processes resulted from the faster channel variation along the differential detection direction (see Section II-D). If the TD- and the FD-OFDM undergo the same channel time-selectivity, the former results in the same performance degradation. However, the later lifts the error floors further due to the estimation error increases with the channel selectivity. Thereupon, they obtain similar performance for both LTLF (Fig. 3) and HTHF (Fig. 6). In addition, the TD-OFDM performs better than the FD-OFDM for LTHF (Fig. 4), whereas the TD-OFDM performs worse than the FD-OFDM for HTLF (Fig. 5).

Moreover, observed from Fig. 5, it is evident that: 1) the VR and the DR perform better than the CR; 2) the performances of the FD-OFDM with different receiver structures are similar. The former is due to that the VR and the DR detect signals with the aid of better estimates (see Section II-D). The later is consequent on the low frequency-selectivity. In this case, the qualities of the estimates are good and similar. As seen from (28), the ICI is treated as an equivalent noise term, thus no reduction on the ICI is made in all receiver structures. Thereupon, the presence of these error floors is mainly resulted from the ICI.

Finally, viewed from Fig. 6, both the VR and the DR yield significant performance improvements over the CR when one or two receive antennas are employed. However, when four receive antennas are available, the performance of the CR is satisfactory and hence the consideration for implementing the DR and the VR is not necessary.
VI. CONCLUSIONS

In this paper, we review the NSD and its three special cases, namely, the CR, the DR, and the VR. Based on the estimator-detector structures, a hierarchical interpretation of the NSD and its special cases is also presented. Then the NSD and its special cases are applied to the differential OFDM systems with diversity reception. Moreover, assuming sufficient CP, a simple closed-form BER expression for differential OFDM systems employing the CR with diversity reception in the time-varying multipath Rayleigh fading channels is given. Numerical results have revealed that the CR with diversity reception can achieve significant performance improvement. However, when few receiving branches are available, the implementation of the DR or the VR is necessary for achieving better performance.

REFERENCES


Fig. 1. The comparison between the TD- and the FD-OFDM.

Fig. 2. The discrete-time baseband equivalent system model for the differential OFDM.

Tab. 1. Parameter settings for different channel conditions.

<table>
<thead>
<tr>
<th></th>
<th>f₀ (Hz)</th>
<th>ν (km/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>HT</td>
<td>200</td>
<td>120</td>
</tr>
<tr>
<td>(\tau_{rms}) (\mu s)</td>
<td>(d)</td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>1.20</td>
<td>1</td>
</tr>
<tr>
<td>HF</td>
<td>4.27</td>
<td>20</td>
</tr>
</tbody>
</table>
Fig. 3. BER comparisons for LTLF ($|\rho_1| = 0.9993$; $|\rho_2| = 0.9988$) and $Q = 1, 2, 4$.

Fig. 4. BER comparisons for LTHF ($|\rho_1| = 0.9993$; $|\rho_2| = 0.9859$) and $Q = 1, 2, 4$. 
Fig. 5. BER comparisons for HTLF ($|\rho| = 0.9826; \ |\rho| = 0.9972$) and $Q = 1, 2, 4$.

Fig. 6. BER comparisons for HTHF ($|\rho| = 0.9826; \ |\rho| = 0.9843$) and $Q = 1, 2, 4$. 