Iterative Multiple Symbol Differential Detection for Turbo Coded Differential Unitary Space-Time Modulation

Pisit Vanichchanunt, Paramin Sangwongngam, Suvit Nakpeerayuth, and Lunchakorn Wuttisittikulkij

Abstract: **In this paper, an iterative multiple symbol differential detection for turbo coded differential unitary space-time modulation using *a posteriori* probability (APP) demodulator is investigated. Two approaches of different complexity based on linear prediction are presented to utilize the temporal correlation of fading for the APP demodulator. The first approach intends to take account of all possible previous symbols for linear prediction, thus requiring an increase of the number of trellis states of the APP demodulator. In contrast, the second approach applies Viterbi algorithm to assist the APP demodulator in estimating the previous symbols, hence allowing much reduced decoding complexity. These two approaches are found to provide a trade-off between performance and complexity. It is shown through simulation that both approaches can offer significant BER performance improvement over the conventional differential detection under both correlated slow and fast Rayleigh flat-fading channels. In addition, when comparing the first approach to a modified bit-interleaved turbo coded differential space-time modulation counterpart of comparable decoding complexity, the proposed decoding structure can offer performance gain over 3 dB at BER of 10^{-5}.**

Index Terms: **Differential space-time modulation, iterative decoding, multiple symbol differential detection, space-time codes, turbo codes.**

I. INTRODUCTION

Recent advances in diversity techniques across multiple antennas [1] have been a subject of intense research, as they are effective means to mitigate the fading phenomena caused by multipath and movement in radio links allowing significant improvement in spectral efficiency and reliability of data transmission over wireless channels. Traditionally, antenna diversity is applied at the receiver side, known as receive diversity. However, recent development on diversity techniques has paid much attention to the transmitter side, referred to as transmit diversity as opposed to receive diversity. Transmit diversity techniques [2]–[4] offer many advantageous features, particularly for improving downlink transmission in cellular mobile environment where mobile terminals are limited in size and space. Deploying multiple transmit and receive antennas simultaneously can broaden the capacity of wireless systems [5], [6] to support higher demanded bandwidth. However, in order to fully exploit available diversity advantages and achieve the maximum theoretical capacity limit, the concept of joint design of coding, modulation, and diversity has been perceived as a promising approach and currently under explored. Space-time coding [7] is one such promising technique where both spatial and time diversities are considered simultaneously. So far a number of space-time coding schemes, including space-time trellis codes [7] and space-time block codes [8]–[11] have been proposed and these coding schemes have shown to offer diversity gain and coding advantage, providing as an effective means to combat fading. Much further performance improvement can be achieved by concatenating them with powerful forward error-correcting codes, such as turbo codes [12]–[14].

Many works of the space-time codes in the literature assume that fading coefficients are perfectly known at the receiver, which can in practice be obtained through some forms of channel estimation such as using pilot signal insertion. However, under fast fading condition there is growing concern on the difficulty in obtaining accurate estimation and the inefficiency in channel bandwidth utilization due to an increase in the number of pilot symbols required especially for multiple input and multiple output (MIMO) channels. As a result, there have been remarkable efforts [15]–[18] to establish classes of space-time coding schemes which allow differential detection to be applied at the receiver side without using fading coefficients, i.e., noncoherent detection and are now commonly referred to as differential space-time modulation (DSTM). Various different DSTM schemes have been proposed and they can be classified based on theirs constellations as group [15], [16] or non-group [17], [18], see more details in [21] and [25]. Among them, the unitary space-time codes proposed by [15] and [16] are particularly suitable for low-complexity differential modulation and decoding [19] and hence opted for our system development; this will be referred to as differential unitary space-time modulation (DUSTM). DSTM with conventional differential detection (DD) is known to be effective only when the channel state changes slowly, i.e., channel coefficients are assumed fixed for two consecutive DSTM symbols. If the fading changes more rapidly, the performance of DSTM with conventional DD deteriorates severely [20]. To mitigate this problem, a well-known technique called multiple symbol differential detection (MSDD) widely used in single-transmit single-receive antenna systems has been extended and developed for DUSTM in [20] and [21] by extending the observation length of consecutively received matrix symbols more than two. Resultant performance of the detection confirms that significant improvement closer to that of coherent detection is possible as the number of the observed symbols increases.

To further improve the performance of differential space-time modulation, DUSTM is combined with forward error-correcting codes, especially those that use iterative decoding, see references [22]–[25]. In [22], a simple bit-interleaved coded mod-

Manuscript received May 4, 2006; approved for publication by Jinho Choi, Division II Editor, October 19, 2007.

The authors are with the Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand, email: pisit.v@student.chula.ac.th, sparamin@hotmail.com, {nsuvit, wlunchak}@chula.ac.th.

Financial support from the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0015/2545) is acknowledged.

ulation (BICM) comprising a convolutional encoder and a bitwise interleaver concatenated serially with DSTM to form a low-complexity yet power efficient transmission system that achieves both diversity and coding gains is introduced. At the receiver, simplified MSDD metrics and hard-decision Viterbi decoding are applied for iterative decision-feedback differential demodulation (DF-DM). It is shown by mathematical analysis and computer simulation that such a relatively low-complexity receiver structure can exploit both space and time diversity and results in performance improvement for transmission over spatially correlated Ricean flat-fading channels. In [23], a serial concatenation of turbo codes and DUSTM is investigated. Both conventional DD and MSDD metrics are considered and applied in the analysis for high performance joint iterative decoding of the turbo code and the modulation code. This is clearly a more sophisticated encoding and decoding structure than that of [22], as it is intended to achieve reliable communication at SNR reasonably close to the capacity limit [23]. Note that the MSDD metrics used in the analysis is based on the assumption that fading coefficients are constant over each observation interval. This can be in fact viewed as a specific case of more general metrics described in [20] and [21]. In [24], three coding schemes of different complexity for combining DSTM and MSDD, namely multilevel coding (MLC), BICM, and hybrid coded modulation (HCM), are investigated. In the paper, channel capacity is considered an appropriate measure for their performance study. However, BER performances of these schemes are also evaluated using computer simulation. It is shown that with respect to channel capacity, MSDD is useful in enhancing power efficiency when applied with MLC and HCM, but less effective with BICM. This suggests that the advantage of MSDD can be effectively exploited only with an appropriate combination of coding and modulation structure. In [25], serial concatenation of simple error control codes, i.e., convolutional or block codes with DUSTM is introduced. At the receiver, iterative decoding/demodulation is employed by exchanging *a posteriori* probability (APP) values between the demodulator of the inner space-time code and the decoder of outer error-correcting codes. It is interesting to see that outstanding BER performance can be achieved with the requirement of long interleaver length and many iterations. This achievement is mainly due to the use of APP demodulator and the introduction of pilot insertion and integrated channel estimation. Note that the paper does not consider MSDD for noncoherent detection.

Motivated by the advantage of MSDD in combating fast fading channels [20], [21] and the powerful APP demodulator for single and multi-antenna systems [25]–[27], in this paper, we develop a serial concatenation scheme of a turbo code and DUSTM for reliable signal transmission over both slow and fast Rayleigh flat-fading channels. At the receiver, a decoding structure that allows effective exchange of soft information between the turbo decoder and the APP demodulator is proposed. In the APP demodulator, soft outputs are computed based on the trellis diagram of DUSTM by using the BCJR algorithm [28]. We present two different approaches, which exploit the MSDD metrics in the APP demodulator in different ways. In the first approach, the classic APP demodulator is modified so that it becomes capable of exploiting the correlated nature of fading in the form of linear prediction. Since the actual code matrices are not known in advance, this approach increases the number of the APP detector's trellis states in order to consider all possible state-transitions affected by the Z previous code matrices used in the linear prediction. In the second approach, which aims for reducing the complexity of the first approach, the APP detector is designed to cooperate with the Viterbi algorithm (VA) for ensuring that the number of trellis states is not increased. More specifically, the task of the VA is to find the previous code matrices associated with the survivors of the APP demodulator's trellis states for each iteration of the iterative decoding/detection. These two approaches can be perceived as a trade-off between performance and complexity. In this study, the performance of our scheme is compared to the turbo coded DUSTM scheme described in [23] with some modifications at the receiver: the modulation decoder (in Fig. 3 of [23]) is replaced by a metric calculation unit and a more powerful APP demodulator for fair comparison purposes, i.e., both schemes possess comparable complexity. It is shown later that the proposed scheme can provide significant BER performance improvement over the modified scheme in [23].

The rest of this paper is organized as follows. A correlated Rayleigh flat-fading channel model and a turbo coded DUSTM encoder are explained in Section II. Iterative decoding with MSDD is detailed in Section III. Computer simulation results are shown and discussed in Section IV. Finally, some conclusions are given in Section V.

II. CHANNEL MODEL AND TURBO CODED DUSTM ENCODER

A. Channel Model

Consider a wireless communication link comprising T transmit antennas and R receive antennas that operates in a Rayleigh flat-fading environment. Let $X_n = \{x_{jk}(n)\}$ denote a $T \times L$ matrix of transmitted DUSTM signals at symbol time n, where $x_{jk}(n)$ is the signal transmitted from antenna j at time slot k, $1 \leq k \leq L$, and L is the number of time slots per DUSTM symbol. Let $Y_n = \{y_{ik}(n)\}$ be an $R \times L$ matrix of received signals at symbol time n, where $y_{ik}(n)$ is the signal that arrives at receive antenna i at time slot k. At symbol time n, the signal $y_{ik}(n)$ at each receive antenna is a superposition of the T fading transmitted signals plus noise. Under this assumption, the channel can be described as the following baseband discrete-time equivalent model:

$$y_{ik}(n) = \sqrt{\rho/T} \sum_{j=1}^{T} h_{ij}^{k}(n) x_{jk}(n) + \eta_{ik}(n) \tag{1}$$

where $h_{ij}^{k}(n)$ is the fading coefficient from transmit antenna j to receive antenna i at time slot k of symbol time n, $\eta_{ik}(n)$ is the complex additive white Gaussian noise (AWGN) with zero-mean and unit-variance at receive antenna i, and ρ is the signal-to-noise ratio (SNR) per receive antenna. The fading coefficient $h_{ij}^{k}(n)$ is modeled as zero-mean, unit-variance, complex Gaussian random variable with autocorrelation given by

$$\begin{aligned} \phi_h((m-n)L + l - k) &= E\{h_{ij}^{k}(n) h_{ij}^{l*}(m)\} \\ &= J_0(2\pi f_d T_d((m-n)L + l - k)) \end{aligned} \tag{2}$$

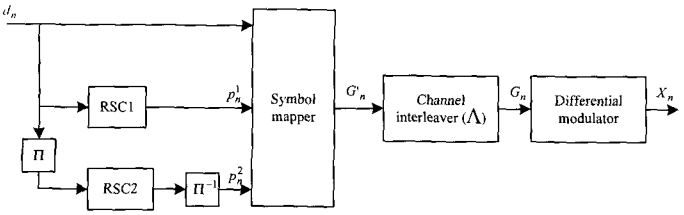


Fig. 1. Block diagram of the turbo coded DUSTM encoder.

where $E\{\cdot\}$ denotes expectation, $(\cdot)^*$ stands for complex conjugate, $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, f_d is the maximum Doppler frequency, T_d represents the duration of each time slot and f_dT_d is referred to as the *normalized maximum Doppler frequency*. If the fading coefficients are assumed to be constant over the DUSTM symbol duration, i.e., $h_{ij}^k(n) = h_{ij}(n)$ for $1 \le k \le L$, then (1) can be expressed in the matrix form as

$$Y_n = \sqrt{\rho_T} H_n X_n + N_n \tag{3}$$

where $H_n = \{h_{ij}(n)\}$ is an $R \times T$ fading matrix, $\rho_T = \rho/T$, and $N_n = \{\eta_{ik}(n)\}$ is an $R \times L$ noise matrix.

B. Turbo Coded DUSTM Encoder

Fig. 1 shows a DUSTM encoding system comprising a turbo encoder of code rate 1/3, a symbol mapper, a channel interleaver Λ and a differential modulator. A block of binary data sequence of length N_b denoted as $\underline{d}_1^{N_b} = d_1, d_2, \ldots, d_{N_b}$ is fed into the turbo encoder which consists of two identical recursive systematic convolutional encoders (RSC1 and RSC2) connected in a parallel way by an interleaver (Π). A simile odd-even helical interleaver is employed to guarantee that the trellis of each recursive systematic convolutional code can be terminated to the zero state with a common tail bit sequence [29]. The output of RSC2 is deinterleaved according to the inverse operation of the interleaver (Π^{-1}). This ensures that the parity bit p_n^2 transmitted with the data bit d_n is the one produced when d_n is encoded. The output sequence of the turbo encoder is passed through the symbol mapper which maps the data bit d_n and its associated parity bits p_n^1 and p_n^2 into a matrix G'_n of size $L \times L$ by using the Gray mapping. Following the symbol mapper, the code matrix sequence is reordered by the channel interleaver in order to break up the fading process correlation. Then the interleaved code matrix G_n is differentially modulated as follows:

$$X_n = X_{n-1} G_n \tag{4}$$

where X_0 is the initial transmitted matrix satisfying $X_0 X_0^H = LI_T$, and I_T is the $T \times T$ identity matrix. Since we consider the code matrix G_n that belongs to an algebraic group of $L \times L$ unitary matrices ($G_n G_n^H = I_L$), we obtain $X_n X_n^H = LI_T$ for $0 \le n \le N$, where N is the length of the transmitted code matrix sequence and $(\cdot)^H$ denotes Hermitian (complex conjugate) transpose. Eq.(4) can also be expressed as

$$X_n = X_0 D_n, \quad n \ge 0 \tag{5}$$

where

$$D_n = \begin{cases} G_1 G_2 \cdots G_n, & n > 0 \\ I_L, & n = 0. \end{cases} \tag{6}$$

III. ITERATIVE DECODING WITH MSDD

The developed decoding system shown in Fig. 2 comprises an APP demodulator, a metric calculation unit (MCU), a channel interleaver Λ, a channel deinterleaver Λ^{-1}, and two decoding units. The MCU calculates the reduced-complexity version of the maximum likelihood sequence estimation (MLSE) metrics for correlated Rayleigh flat-fading channels. These metrics will be referred to as the MSDD metrics. The task of the APP demodulator is to receive the MSDD metrics from the MCU and calculate the probabilities $\Gamma_n(G_n)$ of code symbols G_n for two constituent decoders. Each constituent decoder (CD) in the two decoding units is used to compute the joint probabilities $W_n(d_n, p_n^m)$, $m = 1, 2$, of data bits d_n and their associated parity bits p_n^m of the same trellis branches based on the trellis structure of the corresponding RSC encoder in Fig. 1. Then these probabilities are transferred to the other CD and the APP demodulator. Before these probabilities can be utilized by the latter CD, they are multiplied by the probabilities $\Gamma_n(G'_n)$ of the corresponding code symbols G'_n (fed by the APP demodulator). Then the obtained products are marginalized over the parity bits p_n^m that do not belong to the corresponding RSC encoder, and these marginal probabilities are used as the branch metrics in the CD. In the iterative decoding/detection process, the information is calculated in such a way that the extrinsic information is exchanged among the two CD's and the APP demodulator. After the iterative process is completed, the data bits are decoded by using the *a posteriori* probabilities $\Pr\{d_n|\underline{Y}_0^N\}$ of the data bits d_n from one of the two CD's. This proposed decoding scheme is quite different from that we have presented in [30], i.e., the joint probabilities $W_n(d_n, p_n^m)$ without the *sole* probabilities $\Pr\{d_n\}$ of data bits d_n are exchanged between the two CD's during the iterative process. In our investigation, the decoding scheme of Fig. 2 offers a better performance gain in the range of 0.5–1 dB over that presented in [30] (the performance comparison is not shown here).

A. MSDD Metrics for DUSTM

In this section we will extend the MSDD metrics of a single antenna system reported in [31] for DUSTM. The MSDD metrics for DUSTM are defined as the conditional probabilities of received matrices Y_n given by the sequences of current and previous transmitted code matrices of length Z denoted as $\underline{G}_{n-Z+1}^n$ and by the sequences of all previous received matrices $\underline{Y}_0^{n-1}$:

$$\begin{aligned} M_n(\underline{G}_{n-Z+1}^n) &= \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\} \\ &= \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{Y}_{n-Z}^{n-1}\} \qquad (7) \\ &= \frac{1}{\pi^{RL}\sigma_Z^{2RL}} \exp\{-\frac{1}{\sigma_Z^2}\|Y_n - \sum_{z=1}^{Z} P_z B_{n,z}\|^2\} \qquad (8) \end{aligned}$$

where

$$B_{n,z} = Y_{n-z} G_{n-z+1} G_{n-z+2} \cdots G_n = Y_{n-z} \prod_{l=1}^{z} G_{n-z+l}, \tag{9}$$

the operation $\|\cdot\|$ is the Frobenius norm, the notation P_z denotes the zth $R \times R$ linear prediction coefficient matrices, $1 \le z \le Z$,

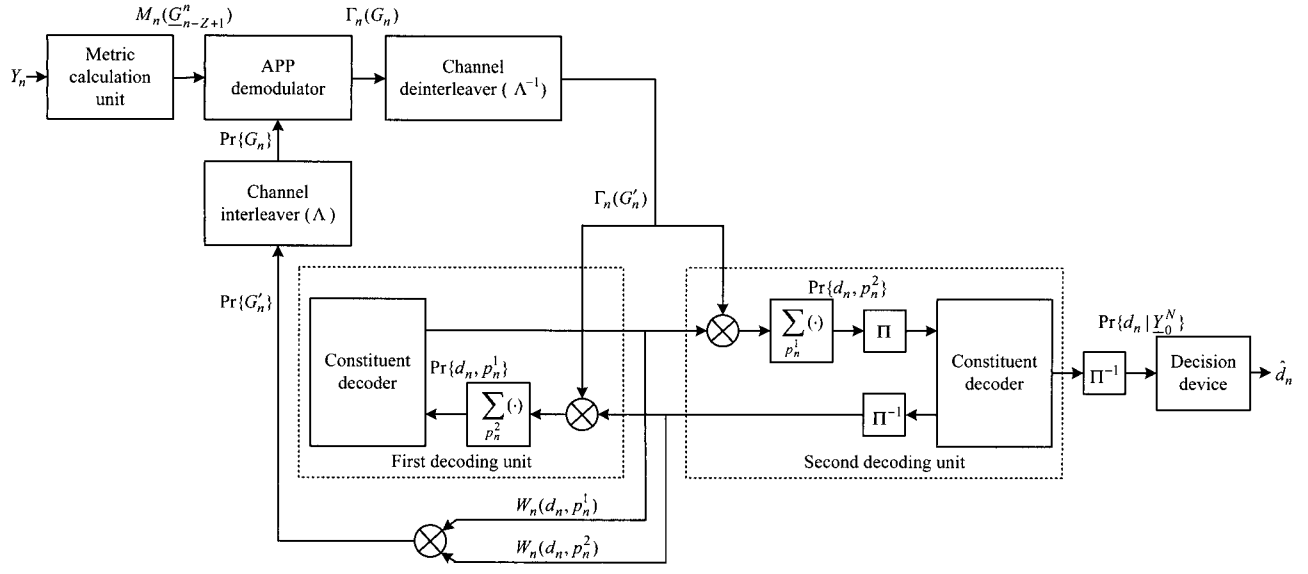


Fig. 2. Block diagram of the iterative decoding with MSDD.

and the variable σ_Z^2 is the minimum mean-squared prediction error. The term $\prod_{l=1}^{z} G_{n-z+l}$ in the right-hand side of (9) can be calculated by using the look-up table of group multiplication. Comparing (8) to the coherent detection metric

$$\Pr\{Y_n|H_n, X_n\} = \frac{1}{\pi^{RL}} \exp\{-\|Y_n - \sqrt{\rho_T} H_n X_n\|^2\}, \quad (10)$$

it can be seen that $\prod_{z=1}^{Z} P_z B_{n,z}$ in (8) is the predicted matrix of $\sqrt{\rho_T} H_n X_n$ in (10) using the linear prediction order of Z. For the case where all fading processes are spatially uncorrelated, the matrices P_z, $1 \le z \le Z$, are diagonal matrices (see Appendix A)

$$P_z = p_z I_R \quad (11)$$

where the variables p_z, $1 \le z \le Z$, represent scalar values which are obtained by solving the following linear system:

$$\begin{bmatrix} \phi_h(0)+\lambda & \phi_h(L) & \cdots & \phi_h((Z-1)L) \\ \phi_h(L) & \phi_h(0)+\lambda & \cdots & \phi_h((Z-2)L) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_h((Z-1)L) & \phi_h((Z-2)L) & \cdots & \phi_h(0)+\lambda \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_Z \end{bmatrix} = \begin{bmatrix} \phi_h(L) \\ \phi_h(2L) \\ \vdots \\ \phi_h(ZL) \end{bmatrix} \quad (12)$$

where

$$\lambda = \frac{1}{\rho}. \quad (13)$$

The minimum mean-squared prediction error σ_Z^2 is given by

$$\sigma_Z^2 = 1 + \rho \left(\phi_h(0) - \sum_{z=1}^{Z} p_z \phi_h(zL) \right). \quad (14)$$

It should be noted that if we have $Z = 1$, there are only two consecutive received symbols Y_{n-1} and Y_n employed in the MSDD metrics. So, we will refer to the detection of this case as the conventional differential detection (CDD).

B. APP Demodulator

In Fig. 2, the APP demodulator receives both the MSDD metric sequence $\underline{M}_1^N$ from the MCU and the *a priori* probabilities $\Pr\{G_n\}$ of code matrices G_n from the two CD's. Then it calculates the extrinsic information $\Gamma_n(G_n)$ of code matrices G_n for the two CD's, based on the structure of the differential modulation. The differential modulation in (4)–(6) can be represented by a trellis diagram in which D_n represents the states of the diagram at time n and code matrices G_n are assigned as the labels of branches (D_{n-1}, D_n). Hence, the BCJR algorithm in [28] can be applied to the APP demodulator. Before calculating the extrinsic information $\Gamma_n(G_n)$, the APP demodulator has to recursively calculate the forward probabilities α_n and the backward probabilities β_n. In this section, the summary of the analysis of the APP demodulator will be presented for the two approaches.

B.1 APP Demodulator Using an Increased Number of Trellis States (Approach 1)

In the first approach, the BCJR algorithm is revisited so that the APP demodulator can utilize the MSDD metrics. From the analysis provided in Appendix B, the recursive formulae for calculating the forward probabilities, the backward probabilities, and the extrinsic information can respectively be expressed as follows:

$$\alpha_n(D_n, \underline{G}^n_{n-Z+2}) = \sum_{D_{n-1}} \sum_{G_{n-Z+1}} \alpha_{n-1}(D_{n-1}, \underline{G}^{n-1}_{n-Z+1}) \times M_n(\underline{G}^n_{n-Z+1}) \Pr\{G_n\} \times \Pr\{D_n|D_{n-1}, G_n\}, \quad (15)$$

$$\beta_n(D_n, \underline{G}_{n-Z+2}^n) = \sum_{D_{n+1}} \sum_{G_{n+1}} \beta_{n+1}(D_{n+1}, \underline{G}_{n-Z+3}^{n+1}) \times M_{n+1}(\underline{G}_{n-Z+2}^{n+1}) \Pr\{G_{n+1}\} \times \Pr\{D_{n+1}|D_n, G_{n+1}\}, \quad (16)$$

and

$$\Gamma_n(G_n) = \frac{\sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \alpha_{n-1}(D_{n-1}, \underline{G}_{n-Z+1}^{n-1})}{\sum_{G_n} \sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \alpha_{n-1}(D_{n-1}, \underline{G}_{n-Z+1}^{n-1})} \times \frac{M_n(\underline{G}_{n-Z+1}^n)\beta_n(D_n, \underline{G}_{n-Z+2}^n)}{M_n(\underline{G}_{n-Z+1}^n)\beta_n(D_n, \underline{G}_{n-Z+2}^n)} \times \frac{\Pr\{D_n|D_{n-1}, G_n\}}{\Pr\{D_n|D_{n-1}, G_n\}} \quad (17)$$

where $\Pr\{D_n|D_{n-1}, G_n\}$ represents the state transition probabilities of the trellis given by the associated code matrices G_n,

$$\Pr\{D_n|D_{n-1}, G_n\} = \begin{cases} 1, & \text{possible event} \\ 0, & \text{otherwise,} \end{cases} \quad (18)$$

and $\Pr\{G_n\}$ denotes the *a priori* probabilities of code matrices G_n which can be obtained by using the product of the corresponding extrinsic information fed from the two CD's (see Section III-D and Fig. 2). The denominator in the right-hand side of (17) is a normalization factor which renders the summation of $\Gamma_n(G_n)$ over all possible G_n unity. The tuple $(D_n, \underline{G}_{n-Z+2}^n)$ in the left-hand side of (15) may be viewed as the extended state at time n. For the CDD ($Z = 1$), substitute $\alpha_n(D_n)$ and $\beta_n(D_n)$ into $\alpha_n(D_n, \underline{G}_{n-Z+2}^n)$ and $\beta_n(D_n, \underline{G}_{n-Z+2}^n)$, respectively. We will refer to the decoding scheme using $Z = 1$ as the iterative CDD scheme.

B.2 APP Demodulator Using the VA (Approach 2)

As opposed to Approach 1, Approach 2 considers only the sequences that are associated with the survivors so that the number of trellis states of the APP demodulator does not increase in relation to the prediction order. To accomplish this, the VA is first modified to process the *a priori* information $\Pr\{G_n\}$ from the two CDs. Next, in each iteration of the decoding/detection, the path metric of each trellis state D_n

$$A_n(D_n) = \max_{D_{n-1}} \{A_{n-1}(D_{n-1}) + \log \Pr\{G_n\} + \log M_n(G_n, \underline{\hat{G}}_{n-Z+1}^{n-1}(D_{n-1}))\} \quad (19)$$

is recursively calculated and used to determine the survivors of trellis states D_n where $\underline{\hat{G}}_{n-Z+1}^{n-1}(D_{n-1})$ represents the code matrix sequences of the survivors associated with states D_{n-1}. For compact notation, we would like to omit the explicit dependence on D_{n-1} and use only $\underline{\hat{G}}_{n-Z+1}^{n-1}$. After the VA has been finished in each iteration of the decoding/detection, the sequences $\underline{\hat{G}}_{n-Z+2}^n$ of length $Z-1$ corresponding to the survivors associated with each trellis states D_n at all times are known. Each sequence is retained and used for the forward/backward recursions:

$$\alpha_n(D_n) = \sum_{D_{n-1}} \sum_{G_n} M_n(G_n, \underline{\hat{G}}_{n-Z+1}^{n-1}) \alpha_{n-1}(D_{n-1}) \times \Pr\{G_n\} \Pr\{D_n|D_{n-1}, G_n\} \quad (20)$$

and

$$\beta_n(D_n) = \sum_{D_{n+1}} \sum_{G_{n+1}} M_{n+1}(G_{n+1}, \underline{\hat{G}}_{n-Z+2}^n) \beta_{n+1}(D_{n+1}) \times \Pr\{G_{n+1}\} \Pr\{D_{n+1}|D_n, G_{n+1}\}. \quad (21)$$

Then the extrinsic information of a code matrix G_n can be calculated as follows:

$$\Gamma_n(G_n) = \frac{\sum_{\underline{D}_{n-1}^n} \alpha_{n-1}(D_{n-1}) M_n(G_n, \underline{\hat{G}}_{n-Z+1}^{n-1})}{\sum_{G_n} \sum_{\underline{D}_{n-1}^n} \alpha_{n-1}(D_{n-1}) M_n(G_n, \underline{\hat{G}}_{n-Z+1}^{n-1})} \times \frac{\beta_n(D_n) \Pr\{D_n|D_{n-1}, G_n\}}{\beta_n(D_n) \Pr\{D_n|D_{n-1}, G_n\}}. \quad (22)$$

For the case of $Z = 1$, the VA of (19) is no longer needed for the APP demodulator. This leads to the case of Approach 1 using $Z = 1$ outlined in Section III-B.1.

C. Analysis of Constituent Decoders

There are two constituent decoders (CD), each is in the corresponding decoding unit. The two CD's operate based on the BCJR algorithm. In this section, the summary of the analysis of both constituent decoders is presented. For simplicity of expression, only the first CD is considered. The second CD can be analyzed in a similar way by taking the effect of the interleaver Π into account. The APP of a data bit is given by

$$\Pr\{d_n|\underline{Y}_0^N\} = \sum_{(S_{n-1}, S_n):d_n} \Pr\{S_{n-1}, S_n|\underline{Y}_0^N\} \quad (23)$$

$$= \frac{\sum_{(S_{n-1}, S_n):d_n} \alpha_{n-1}(S_{n-1}) \gamma_n(S_{n-1}, S_n) \beta_n(S_n)}{\sum_{d_n} \sum_{(S_{n-1}, S_n):d_n} \alpha_{n-1}(S_{n-1}) \gamma_n(S_{n-1}, S_n) \beta_n(S_n)} \quad (24)$$

where branches $(S_{n-1}, S_n) : d_n$ are all possible state transitions from S_{n-1} to S_n whose data bits driving the constituent encoder are d_n. The branch metrics for the first CD are formulated as

$$\gamma_n(S_{n-1}, S_n) = \Pr\{d_n, p_n^1\} = \sum_{p_n^2} \Gamma_n(G_n') W_n(d_n, p_n^2) = \sum_{p_n^2} \Gamma_n(d_n, p_n^1, p_n^2) W_n(d_n, p_n^2) \quad (25)$$

where $\Pr\{d_n, p_n^1\}$ are the *a priori* joint probabilities of data bits d_n and their associated parity bits p_n^1, and $W_n(d_n, p_n^2)$ are the joint probabilities of data bits d_n and their associated parity bits p_n^2, which are supplied by the second CD (see Section III-D).

D. The Extrinsic Information of Constituent Decoders

The joint probabilities $W_n(d_n, p_n^1)$ of data bits d_n and their associated parity bits p_n^1 of RSC1 can be calculated by the first CD as follows:

$$W_n(d_n, p_n^1) = \frac{\sum_{(S_{n-1}, S_n):(d_n, p_n^1)} \alpha_{n-1}(S_{n-1}) \beta_n(S_n)}{\sum_{d_n} \sum_{p_n^1} \sum_{(S_{n-1}, S_n):(d_n, p_n^1)} \alpha_{n-1}(S_{n-1}) \beta_n(S_n)} \quad (26)$$

where branches $(S_{n-1}, S_n) : (d_n, p_n^1)$ are all possible state transitions from S_{n-1} to S_n whose data bits driving the constituent encoder are d_n and the corresponding parity bits are p_n^1. Similarly, the joint probabilities $W_n(d_n, p_n^2)$ of data bits d_n and their associated parity bits p_n^2 of RSC2 can be calculated by the second CD as follows:

$$W_n(d_n, p_n^2) = \frac{\sum_{(S_{n-1},S_n):(d_n,p_n^2)} \alpha_{n-1}(S_{n-1})\beta_n(S_n)}{\sum_{d_n}\sum_{p_n^2}\sum_{(S_{n-1},S_n):(d_n,p_n^2)} \alpha_{n-1}(S_{n-1})\beta_n(S_n)}. \tag{27}$$

The *a priori* probability $\Pr\{G_n'\}$ of a code matrix G_n' for the APP demodulator can be computed by using the product of the corresponding probabilities $W_n(d_n, p_n^1)$ and $W_n(d_n, p_n^2)$ as follows:

$$\begin{aligned}\Pr\{G_n'\} &= \Pr\{d_n, p_n^1, p_n^2\} \\ &= \frac{W_n(d_n, p_n^1)W_n(d_n, p_n^2)}{\sum_{d_n}\sum_{p_n^1}\sum_{p_n^2} W_n(d_n, p_n^1)W_n(d_n, p_n^2)}.\end{aligned} \tag{28}$$

E. Computational Complexity Analysis

In this section, we will compare the computational complexity of Approach 1 and Approach 2. For ease of comparison, we assume that in both approaches, the metric calculation unit calculates the metrics $M_n(\underline{G}_{n-Z+1}^n)$ for all possible $\underline{G}_{n-Z+1}^n$ even though Approach 2 may use some of them. So, at a given linear prediction order Z, the metric calculation unit requires the same number of additions and multiplications for both approaches. Then the complexity comparison can be performed by evaluating the amount of calculation needed for the APP demodulator. Table 1 shows the number of multiplications, additions, and (greater/less) comparisons required by the APP demodulator per trellis section per iteration. Note that when $Z = 1$, Approach 2 becomes Approach 1 and VA is not needed (and there is no (greater/less) comparison also). In such a case, the number of multiplications and additions of Approach 2 is the same as that of Approach 1 with $Z = 1$. As we can see from Table 1, for Approach 1, the number of multiplications and addition increases exponentially with the linear prediction order Z, whereas for Approach 2, the number of multiplications, additions and comparisons for the APP demodulator is not a function of the linear prediction order.

Table 1. The computational complexity of APP demodulator.

Approach	Complexity of APP demodulator per trellis section per iteration		
	Multiplications	Additions	Comparisons
Approach 1	$6N_S^{Z+1} + 2N_S^Z + N_S$	$2N_S^Z(N_S - 1) + 2(N_S^Z - 1) + N_S(N_S^Z - 1) + (N_S - 1)$	-
Approach 2 ($Z > 1$)	$8N_S^2 + 3N_S$	$3(N_S + 1)(N_S - 1)$	$N_S(N_S - 1)$

Note: N_S is the number of trellis states D_n.

IV. SIMULATION RESULTS

The performance of the decoding system is evaluated by using computer simulation. Two RSC encoders are identical with the forward polynomial $1 + \mathrm{D}^4$ and the backward polynomial $1 + \mathrm{D} + \mathrm{D}^2 + \mathrm{D}^3 + \mathrm{D}^4$ where D is a unit delay. The data block size N_b is 930. The channel interleaver is the block interleaver of size 41×23 as described in [31]. The simile odd-even helical interleaver [29] is selected for the interleaver Π so that both RSC encoders terminate to the zero state with the same input tail bits. The performance of the decoding system is measured in terms of the bit error rate (BER) of decoded data bits. The BER is measured for the bit energy over the noise spectral density E_b/N_0 which is defined as [23]

$$\frac{E_b}{N_0} = \frac{\rho}{R_s} \tag{29}$$

where R_s is the system rate given in uncoded information bits per channel use. In the simulation, two transmit antennas are considered. The number of time slots per DUSTM symbol (channel use per DUSTM symbol) or L is also two so that transmitted matrices are square. Since we use a code rate of 1/3 and a mapping of 3 bits per DUSTM symbol, the system rate is $R_s = 1/3 \times 3 \times 1/2 = 1/2$ bits per channel use (tail bits are ignored). Although, in the analysis of the linear prediction, the fading processes are assumed to be constant during each transmitted matrix X_n, the channel model of (1) used in the simulation allows the fading processes to vary within the transmitted matrices in order to see the effects of Doppler shift on the decoding performance. All fading processes are assumed to be spatially uncorrelated. The set of unitary group matrices used for the symbol mapper is

$$\left\{\pm\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \pm\begin{bmatrix}j & 0\\ 0 & -j\end{bmatrix}, \pm\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}, \pm\begin{bmatrix}0 & j\\ j & 0\end{bmatrix}\right\}.$$

For signal transmission, we use the reference DUSTM symbol

$$X_0 = \begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}. \tag{30}$$

A. Performance for Each Iteration

The BER versus E_b/N_0 results for the two proposed decoding approaches: Approach 1 and Approach 2, using different number of decoding iterations are respectively illustrated in Figs. 3 and 4 in order to show how fast these two approaches converge under two very different fading channel conditions with normalized maximum Doppler frequencies of 0.01 and 0.1. These results are based on systems with two receive antennas and fixed linear prediction order of two. As we can see, the BER performance of both decoding approaches improves rapidly in the first few iterations as one may expect from an iterative decoding behavior. Later iterations offer less and less additional improvement and finally the BER performance converges, say roughly at the 15th iteration. Notice also that convergence behavior remains similar even the fading of the channel changes from slow to very fast as much as 10 times, i.e., from $f_dT_d = 0.01$ to $f_dT_d = 0.1$ as depicted in Figs. 3(a) and 3(b) for Approach 1 and in Figs. 4(a) and 4(b) for Approach 2. Therefore, in the following sections, all BER performance results presented here are

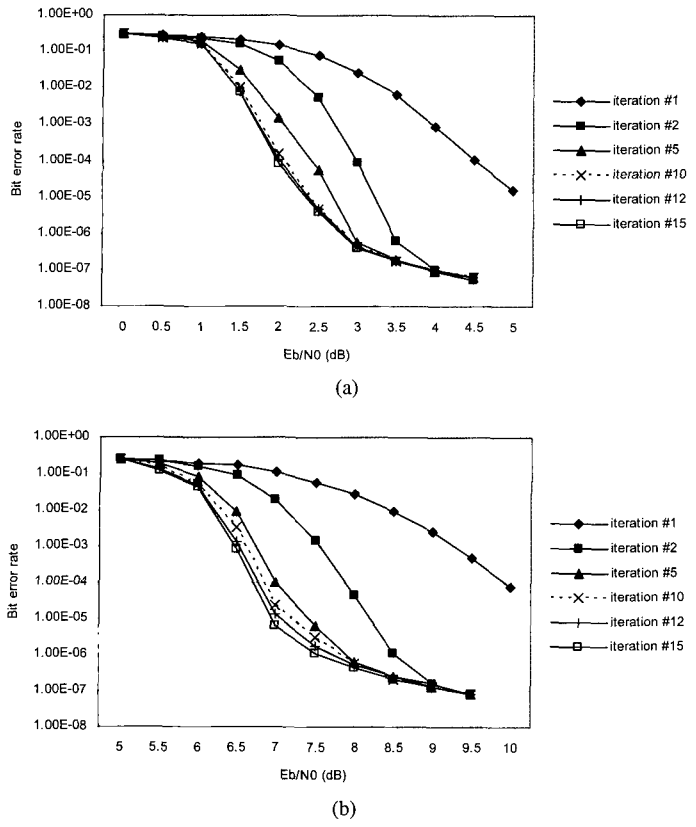


Fig. 3. The BER performance of Approach 1 using two transmit and two receive antennas, and setting the linear prediction order to $Z = 2$: (a) $f_dT_d = 0.01$ and (b) $f_dT_d = 0.1$.

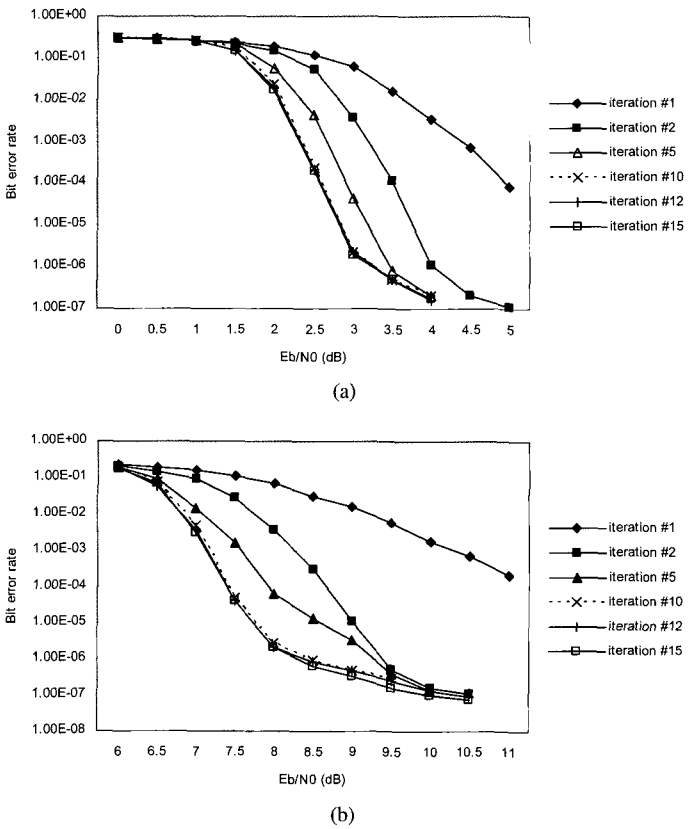


Fig. 4. The BER performance of Approach 2 using two transmit and two receive antennas, and setting the linear prediction order to $Z = 2$: (a) $f_dT_d = 0.01$ and (b) $f_dT_d = 0.1$.

obtained at the 15th iteration in order to ensure the decoding convergence.

B. Performance Comparison

In this section, we shall compare the BER performance of the two proposed approaches to that of the iterative CDD scheme ($Z = 1$) at two normalized maximum Doppler frequencies of 0.01 and 0.1; this is shown in Figs. 5(a) and 5(b), respectively. Consider the case of slow fading ($f_dT_d = 0.01$) as in Fig. 5(a), it is apparent that both proposed approaches offer performance improvement over the iterative CDD scheme as anticipated. For instance, at the BER of 10^{-5}, performance gains of 1.1, 1.7, and 1.8 dB are achieved with Approach 1 using $Z = 2, 3$, and 4, respectively, whereas Approach 2, for the same system settings, provides less gains. For the case of fast fading ($f_dT_d = 0.1$) as shown in Fig. 5(b), the performance gains of both approaches are even much greater. Approach 1 in particular provides significant improvement, namely gains of 3.1, 3.3, and 3.4 dB are attained by setting the linear prediction order to $Z = 2$, 3, and 4, respectively. These results clearly highlight the great benefit of introducing MSDD ($Z \geq 2$) in conjunction with APP demodulator as such combined feature allows fading to be more accurately estimated and hence capable of enhancing the robustness of the decoding system against fading variation when compared to the iterative CDD scheme. The results also show that increasing the linear predictor order can result in improved overall decoding performance. Such an increase in performance improvement is mainly because of the fact that higher order of prediction provides better fading estimates, thereby permitting more accurate detection. However, it must be emphasized that such performance gains come at a price of complexity. Notice also that the performance improvement is substantial only when increasing the linear prediction order from $Z = 1$ to 2. Further increases in the linear prediction order do not offer as much performance gains. This coincides with the decrease of the minimum mean-squared prediction error σ_Z^2, when Z is increased as the following. At $f_dT_d = 0.01$ and $E_b/N_0 = 2$ dB, $\sigma_Z^2 = 1.4449, 1.3155, 1.2526$, and 1.2194 for $Z = 1$, 2, 3, and 4, respectively, whereas at $f_dT_d = 0.10$ and $E_b/N_0 = 7$ dB, $\sigma_Z^2 = 2.7665, 2.4884, 2.4197$, and 2.3628, respectively. Notice that substantial reduction of σ_Z^2 is obtained when Z increases from 1 to 2. So using the linear prediction order as low as $Z = 2$ can be considered as highly desirable with respect to performance-complexity trade-offs, as significant performance improvement can be attained with moderate increase in the decoding complexity.

When comparing between the performance of Approach 1 and of Approach 2, it is clear that Approach 1 always offers better performance than that of Approach 2 (the case of the iterative CDD scheme is excluded in this comparison) since Approach 1 uses more trellis complexity in the detection than Approach 2. More specifically, in order to obtain the possibility of performance enhancement, Approach 1 tries to utilize all useful information regarding the MSDD metrics based linear prediction by increasing the number of trellis states so that the for-

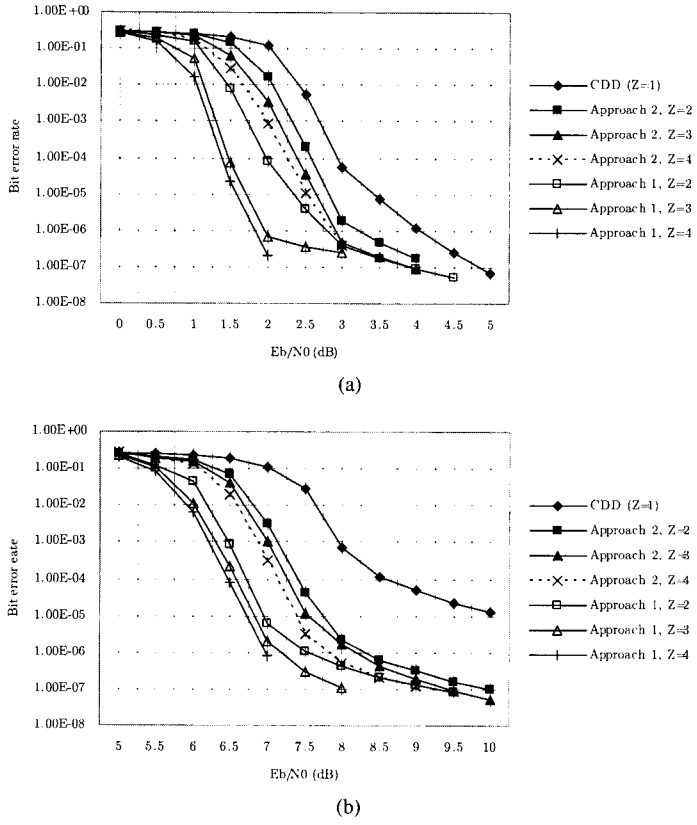


Fig. 5. The performance comparison between Approach 1 and Approach 2 for two transmit and two receive antennas and using the linear prediction order as Z = 1, 2, 3, and 4: (a) $f_dT_d = 0.01$ and (b) $f_dT_d = 0.1$.

ward/backward probabilities accumulate the MSDD metrics of all possible code matrix sequences $\underline{G}_{n-Z+1}^{n}$. In contrast, due to complexity reduction by using VA as early explained, the forward/backward probabilities of Approach 2 accumulate only the MSDD metrics of code matrix sequences associated with survivors, inevitably causing the information loss and resulting in the performance degradation. Therefore, when performance is the main priority, Approach 1 is recommended but if fast decoding is of prime importance then Approach 2 will be preferred.

C. Performance Comparison between the Modified Turbo Coded DUSTM Scheme of [23] and the Proposed Scheme

In this section, we will compare the performance of our proposed scheme to a modified version of bit-interleaved turbo coded differential space-time modulation scheme described in [23]. For fair comparison, the modulation decoder in Fig. 3 of [23] is substituted by a metric calculation unit and an APP demodulator. With these modifications, both schemes will have similar complexity. For this comparison, both schemes are tested under the same parameter settings as used in the previous section. Fig. 6 illustrates the BER performance of both schemes at two normalized maximum Doppler frequencies of 0.01 and 0.1. As we can see, at the BER of 10^{-5} and the normalized maximum Doppler frequency of 0.01, the proposed scheme offers performance gains of 3.1 and 3.4 dB over that of the modified scheme of [23] when using $Z = 1$ and 2, respectively. When fading changes more rapidly as $f_dT_d = 0.1$, these gains further increase to 4 and 3.5 dB, respectively. These results indicate that transmission of each data bit and its associated parity bits by grouping them together as a single code matrix symbol can provide improvement to the decoding performance as opposed to transmitting them in different symbols. However, bit-interleaved turbo coded DUSTM exhibits lower error floors as compared to the symbol interleaving technique used in our scheme.

D. Effects of the Number of Receive Antennas

The effects of the number of receive antennas on the decoding performance of both decoding approaches will be investigated in this section. The number of transmit antennas and the linear prediction order are fixed at two while the number of receive antennas is varied as $R = 1$, 2, 3, and 4. The BER performance curves of Approach 1 and Approach 2 are shown in Fig. 7(a) and Fig. 7(b), respectively, both at two normalized maximum Doppler frequencies of 0.01 and 0.10. In Fig. 7(a), at the BER of 10^{-5}, the performance gains of Approach 1 using $R = 2$, 3, and 4 are respectively 2.5, 4.2, and 5.2 dB over that using $R = 1$ for $f_dT_d = 0.01$. When fading changes rapidly as $f_dT_d = 0.10$, these gains are significantly increased as 5.5, 7.5, and 8.5 dB, respectively. The performance results of Approach 2 in Fig. 7(b) also have the same trend as Approach 1. As one may expect, the use of multiple receive antennas can always lead performance improvement. Notice also that the improvement is particularly substantial for fast fading conditions.

V. CONCLUSIONS

We have demonstrated that the combination of MSDD and APP demodulator can be applied to improve the performance of turbo coded DUSTM systems over both slow and fast Rayleigh flat-fading channels without using pilot insertion. In order to utilize the MSDD metrics in the APP demodulator, we have presented two different and effective approaches which provide a trade-off between the performance and complexity. As illustrated in the simulation results, both approaches using MSDD can offer significant gains over that using CDD, especially when the fading changes rapidly. However, the substantial BER improvement is observed only when increasing the linear prediction order from $Z = 1$ to 2. When applying higher prediction order, additional gains are less significant. When comparing the BER performance of both approaches, as expected, Approach 1 which uses more number of trellis states is superior to Approach 2 for all cases. Next, we have also shown that the concatenation structure of turbo codes and DUSTM plays an important role in enabling effective decoding at the receiver. To demonstrate this, we have compared the BER performance of the modified bit-interleaved turbo coded DUSTM scheme of [23] to that of the proposed scheme which employs a symbol-wise channel interleaver in conjunction with sending each data bit and its associated parity bits into the same code symbol. The results have shown that the proposed scheme provides significant gains in the range of 3–4 dB over the modified scheme of [23] for fading varied from $f_dT_d = 0.01$ to $f_dT_d = 0.1$ and at the BER of 10^{-5}. Finally, the results have revealed that increasing the number of receive antennas offers great benefit to cope

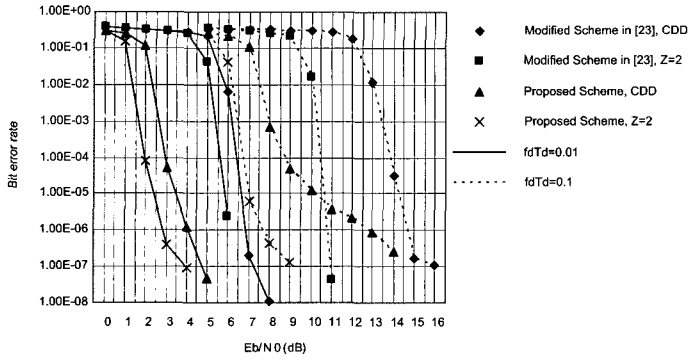
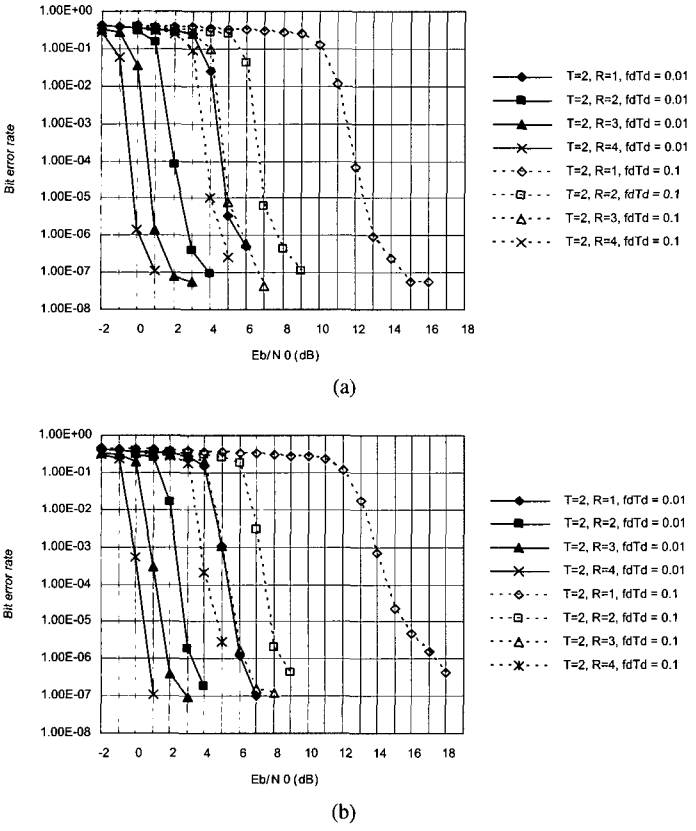


Fig. 6. The performance comparison between the modified scheme of [23] and the proposed scheme for two transmit and two receive antennas; all schemes use Approach 1 and the linear prediction order as Z = 1 and 2.



(a)

(b)

Fig. 7. The BER performance of Approach 1 and Approach 2 for two transmit and one to four receive antennas and setting the linear prediction order to Z = 2: (a) Approach 1 and (b) Approach 2.

with fast fading channels.

ACKNOWLEDGMENTS

The authors also would like to thank all anonymous reviewers and the editor for their valuable comments and suggestions in the revision of this paper.

APPENDIX A

Linear Prediction

In the following, we will analyze the prediction coefficient matrices P_z, $1 \le z \le Z$, used in the term $\sum_{z=1}^{Z} P_z B_{n,z}$ which is the predicted matrix of $\sqrt{\rho_T} H_n X_n$. Let error matrix ϵ_n defined as

$$\epsilon_n = \sqrt{\rho_T} H_n X_n - \sum_{z=1}^{Z} P_z B_{n,z}. \tag{A.1}$$

From the orthogonality principle, we have

$$E\{\epsilon_n B_{n,z}^H\} = \mathbf{0} \tag{A.2}$$

where $(\cdot)^H$ denotes Hermitian (complex conjugate) transpose and $\mathbf{0}$ is the $R \times R$ zero matrix. From (A.1) and (A.2), it can be proved by using the unitary property of the codes that

$$E\left\{H_n H_{n-m}^H\right\} = \sum_{z=1}^{Z} P_z \left[E\left\{H_{n-z} H_{n-m}^H\right\} + \lambda T \delta(z-m) I_R\right], \quad m = 1, 2, \ldots, Z \tag{A.3}$$

where $\lambda = 1/\rho$ and $\delta(n)$ is the Kronecker delta function. If all fading processes become spatially uncorrelated or

$$E\{h_{ij}(n) h_{i'j'}^*(n')\} = 0, \quad i \ne i', j \ne j'; n, n' \in I \tag{A.4}$$

then

$$E\{H_{n-z} H_{n-m}^H\} = T \phi_h((z-m)L) I_R, \tag{A.5}$$

and

$$P_z = p_z I_R, \qquad 1 \le z \le Z. \tag{A.6}$$

Hence, (A.3) can be reduced to the following equation

$$\phi_h(mL) = \sum_{z=1}^{Z} p_z \left[\phi_h((z-m)L) + \lambda \delta(z-m)\right], \quad m = 1, 2, \ldots, Z. \tag{A.7}$$

This relation can be written in the matrix form as (12). It should be noted that in the analysis of the linear prediction, the fading coefficients are assumed to be constant over the time duration of each space-time code. However, the fading coefficients in the simulation are not constant in the duration but varied according to the autocorrelation in (2).

APPENDIX B

APP Demodulator Using Approach 1

The *a posteriori* probability of a code matrix can be expressed as

$$\Pr(G_n | \underline{Y}_0^N) = \frac{\Pr(G_n, \underline{Y}_0^N)}{\Pr(\underline{Y}_0^N)} = \frac{\Pr(G_n, \underline{Y}_0^N)}{\sum_{G_n} \Pr(G_n, \underline{Y}_0^N)}, \tag{B.1}$$

where $\Pr(G_n|\underline{Y}_0^N)$ can be analyzed by applying Bayes' rule as follows.

$$\begin{aligned}
&\Pr(G_n|\underline{Y}_0^N) \\
&\quad = \sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \Pr\{\underline{G}_{n-Z+1}^n, \underline{D}_{n-1}^n, \underline{Y}_0^N\} \\
&\quad = \sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \Pr\{\underline{G}_{n-Z+1}^{n-1}, D_{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{G_n|\underline{G}_{n-Z+1}^{n-1}, D_{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{D_n|\underline{G}_{n-Z+1}^n, D_{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{D}_{n-1}^n, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{\underline{Y}_{n+1}^N|\underline{G}_{n-Z+1}^n, \underline{D}_{n-1}^n, \underline{Y}_0^n\}. \qquad \text{(B.2)}
\end{aligned}$$

Based on the MSDD metrics defined in (7) and (8), for different branches (D_{n-1}, D) which have the same code matrix G_n, the MSDD metrics are the same, that is

$$\Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{D}_{n-1}^n, \underline{Y}_0^{n-1}\} = \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\}. \qquad \text{(B.3)}$$

If the channel interleaver can fully shuffle the sequences of code matrices $\underline{G}_{n-Z+1}^{n-1}$, it may be assumed that the current code matrices G_n are independent of the sequences of the previous code matrices $\underline{G}_{n-Z+1}^{n-1}$, the previous states X_{n-1}, and the previous received sequence $\underline{Y}_0^{n-1}$. Hence,

$$\Pr\{G_n|\underline{G}_{n-Z+1}^{n-1}, D_{n-1}, \underline{Y}_0^{n-1}\} = \Pr\{G_n\}. \qquad \text{(B.4)}$$

When the previous states D_{n-1} and the current code matrices G_n are given, the current states D_n are exactly known. Hence, other conditions are no longer needed, or

$$\Pr\{D_n|\underline{G}_{n-Z+1}^n, D_{n-1}, \underline{Y}_0^{n-1}\} = \Pr\{D_n|G_n, D_{n-1}\}. \qquad \text{(B.5)}$$

Based on the limited order of the linear prediction used in the MSDD metrics, the linear prediction looks back only the Z previous symbols, also

$$\begin{aligned}
&\Pr\{\underline{Y}_{n+1}^N|\underline{G}_{n-Z+1}^n, \underline{D}_{n-1}^n, \underline{Y}_0^n\} \\
&\qquad \approx \Pr\{\underline{Y}_{n+1}^N|\underline{G}_{n-Z+2}^n, D_n, \underline{Y}_0^n\}. \qquad \text{(B.6)}
\end{aligned}$$

Use (B.3)–(B.6) for (B.2), it can be obtained that

$$\begin{aligned}
&\Pr\{G_n, \underline{Y}_0^N\} \\
&\quad = \sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \Pr\{\underline{G}_{n-Z+1}^{n-1}, D_{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{G_n\} \Pr\{D_n|G_n, D_{n-1}\} \\
&\qquad\qquad \times \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{\underline{Y}_{n+1}^N|\underline{G}_{n-Z+2}^n, D_n, \underline{Y}_0^n\} \\
&\quad = \sum_{\underline{D}_{n-1}^n} \sum_{\underline{G}_{n-Z+1}^{n-1}} \alpha_{n-1}(D_{n-1}, \underline{G}_{n-Z+1}^{n-1}) \\
&\qquad\qquad \times \Pr\{G_n\} \Pr\{D_n|G_n, D_{n-1}\} \\
&\qquad\qquad \times M_n(\underline{G}_{n-Z+1}^n)\beta_n(D_n, \underline{G}_{n-Z+2}^n), \qquad \text{(B.7)}
\end{aligned}$$

where $\alpha_n(D_n, \underline{G}_{n-Z+2}^n$ and $\beta_n(D_n, \underline{G}_{n-Z+2}^n)$ are the forward and backward probabilities of the APP demodulator using MSDD. It can be analyzed so that these probabilities can be computed recursively as follows:

$$\begin{aligned}
&\alpha_n(D_n, \underline{G}_{n-Z+2}^n) = \Pr\{\underline{G}_{n-Z+2}^n, D_n, \underline{Y}_0^n\} \\
&\quad = \sum_{D_{n-1}} \sum_{G_{n-Z+1}} \Pr\{D_{n-1}, D_n, \underline{G}_{n-Z+1}^n, \underline{Y}_0^n\} \\
&\quad = \sum_{D_{n-1}} \sum_{G_{n-Z+1}} \Pr\{D_{n-1}, \underline{G}_{n-Z+1}^{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{G_n|D_{n-1}, \underline{G}_{n-Z+1}^{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{D_n|D_{n-1}, \underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{Y_n|\underline{D}_{n-1}^n, \underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\} \\
&\quad = \sum_{D_{n-1}} \sum_{G_{n-Z+1}} \Pr\{D_{n-1}, \underline{G}_{n-Z+1}^{n-1}, \underline{Y}_0^{n-1}\} \\
&\qquad\qquad \times \Pr\{G_n\} \Pr\{D_n|G_n, D_{n-1}\} \\
&\qquad\qquad \times \Pr\{Y_n|\underline{G}_{n-Z+1}^n, \underline{Y}_0^{n-1}\} \\
&\quad = \sum_{D_{n-1}} \sum_{G_{n-Z+1}} \alpha_{n-1}(D_{n-1}, \underline{G}_{n-Z+1}^{n-1}) \Pr\{G_n\} \\
&\qquad\qquad \times \Pr\{D_n|G_n, D_{n-1}\} M_n(\underline{G}_{n-Z+1}^n), \qquad \text{(B.8)}
\end{aligned}$$

and

$$\begin{aligned}
&\beta_n(D_n, \underline{G}_{n-Z+2}^n) = \Pr\{\underline{Y}_{n+1}^N|\underline{G}_{n-Z+2}^n, D_n, \underline{Y}_0^n\} \\
&\quad = \sum_{G_{n+1}} \sum_{D_{n+1}} \Pr\{D_{n+1}, G_{n+1}, \underline{Y}_{n+1}^N|D_n, \underline{G}_{n-Z+2}^n, \underline{Y}_0^n\} \\
&\quad = \sum_{G_{n+1}} \sum_{D_{n+1}} \Pr\{G_{n+1}|D_n, \underline{G}_{n-Z+2}^n, \underline{Y}_0^n\} \\
&\qquad\qquad \times \Pr\{D_{n+1}|D_n, \underline{G}_{n-Z+2}^{n+1}, \underline{Y}_0^n\} \\
&\qquad\qquad \times \Pr\{Y_{n+1}|\underline{D}_n^{n+1}, \underline{G}_{n-Z+2}^{n+1}, \underline{Y}_0^n\} \\
&\qquad\qquad \times \Pr\{\underline{Y}_{n+2}^N|\underline{D}_n^{n+1}, \underline{G}_{n-Z+2}^{n+1}, \underline{Y}_0^{n+1}\} \\
&\quad = \sum_{G_{n+1}} \sum_{D_{n+1}} \Pr\{G_{n+1}\} \Pr\{D_{n+1}|D_n, G_{n+1}\} \\
&\qquad\qquad \times \Pr\{Y_{n+1}|\underline{G}_{n-Z+2}^{n+1}, \underline{Y}_0^n\} \\
&\qquad\qquad \times \Pr\{\underline{Y}_{n+2}^N|D_{n+1}, \underline{G}_{n-Z+3}^{n+1}, \underline{Y}_0^{n+1}\} \\
&\quad = \sum_{G_{n+1}} \sum_{D_{n+1}} \Pr\{G_{n+1}\} \Pr\{D_{n+1}|D_n, G_{n+1}\} \\
&\qquad\qquad \times M_{n+1}(\underline{G}_{n-Z+2}^{n+1})\beta_{n+1}(D_{n+1}, \underline{G}_{n-Z+3}^{n+1}). \qquad \text{(B.9)}
\end{aligned}$$

REFERENCES

[1] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Process. Mag.*, vol. 17, Issue 3, pp. 76–92, May 2000.

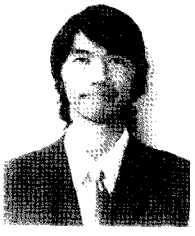
[2] A. Wittneben, "Base station modulation diversity for digital simulcast," in *Proc. IEEE VTC*, 1991, pp. 848–853.

[3] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," in *Proc. IEEE VTC*, 1993, pp. 508–511.

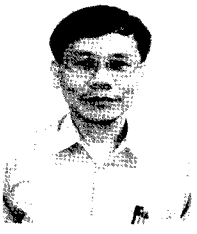
[4] J. Guey, M. P. Fitz, M. R. Bell, and W. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
[9] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
[10] H. Jafakhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.
[11] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
[12] H.-J. Su and E. Geraniotis, "Space-time turbo codes with full antenna diversity," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 47–57, Jan. 2001.
[13] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Full rate space-time turbo codes," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 969–980, May 2001.
[14] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 958–968, May 2001.
[15] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.
[16] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.
[17] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 1169–1174, July 2000.
[18] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2626–2631, Sept. 2001.
[19] H. Li and J. Li, "Differential and coherent decorrelating multiuser receivers for space-time-coded CDMA systems," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2529–2537, Oct. 2002.
[20] R. Schober and L. H.-J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 768–777, May 2002.
[21] C. Ling, K. H. Li, and A. C. Kot, "Noncoherent sequence detection of differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2727–2734, Oct. 2003.
[22] L. H.-J. Lampe and R. Schober, "Bit-interleaved coded differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1429–1439, Sept. 2002.
[23] A. Steiner, M. Peleg, and S. Shamai, "Iterative decoding ofspace-time differentially coded unitary matrix modulation," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2385–2395, Oct. 2002.
[24] L. H.-J. Lampe, R. Schober, and R. F. H. Fischer, "Coded differential space-time modulation for flat fading channels," *IEEE Trans. Wireless Commun.*, vol. 2 no. 3, pp. 582–590, May 2003.
[25] C. Schlegel and A. Grant, "Differential space-time turbo codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2298–2306, Sept. 2003.
[26] P. Hoeher and J. Lodge, "Turbo DPSK: Iterative differential PSK demodulation and channel decoding," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 837–843, June 1999.
[27] P. Vanichchanunt, C. Sritiapetch, S. Nakpeerayuth, and L. Wuttisittikulkij, "APP demodulator for turbo coded multiple symbol differential detection under correlated Rayleigh fading channels," in *Proc. IEEE GLOBECOM*, 2004, pp. 2578–2582.
[28] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, Mar. 1974.
[29] A. S. Barbulescu and S. S. Pietrobon, "Terminating the trellis of turbo-codes in the same state," *IEE Electron. Letters*, vol. 31, no. 1, pp. 22–23, Jan. 1995.
[30] P. Vanichchanunt, P. Sangwongngam, S. Nakpeerayuth, and L. Wuttisittikulkij, "APP demodulator for turbo coded differential unitary space-time modulation," in *Proc. IEEE ICC*, 2005, pp. 2906–2910.
[31] I. D. Marsland and P. T. Mathiopoulos, "Multiple differential detection of parallel concatenated convolutional (turbo) codes in correlated fast Rayleigh fading," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 265–275, Feb. 1998.

Pisit Vanichchanunt received the B.Eng. degree (with Honors) in Telecommunication Engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 1991. He received the M.Eng. degree in Electrical Engineering from Chulalongkorn University, Bangkok, Thailand, in 2001. He is currently pursuing a doctoral degree in Electrical Engineering at the Telecommunication System Research Laboratory (TSRL), Chulalongkorn University. His main research interests are iterative decoding/demodulation, space-time codes, and signal processing for communications.



Paramin Sangwongngam received the M.Eng. degree in Electrical Engineering from Chulalongkorn University in 2005. He is presently serving as an assistant researcher at Optical and Quantum Communications Laboratory, the National Electronics and Computer Technology Center (NECTEC), NSTDA, Thailand, where he is involved in research and development on error control coding for mobile communications and hard disk systems. His research focuses on iterative decoding, signal processing and coding for magnetic recording systems, and quantum communications.



Suvit Nakpeerayuth received B.Eng. and M.Eng. degrees in Electrical Engineering from Chulalongkorn University in 1980 and 1984, respectively. He is presently a lecturer at the Department of Electrical Engineering, Chulalongkorn University and also in charge of the DSP Basic Laboratory. His main research interests are signal processing and the hardware implementation of wireless transceivers.

Lunchakorn Wuttisittikulkij received the B.Eng. degree in electrical engineering from Chulalongkorn University Bangkok Thailand in 1990, the M.Sc. and Ph.D. degrees in telecommunications from the University of Essex, the United Kingdom in 1992 and 1997, respectively. In 1995, he was involved in a European Commission funded project R2028: Multiwavelength Transport Network (MWTN) responsible for network design and dimensioning. In 1997, he joined the faculty of engineering at Chulalongkorn University, where he is currently an associate professor at the Department of Electrical Engineering. His research interests are broadband wireless access networks. He is the author of thirteen books and over 50 research papers.