Research Article

Detector Design and Complexity Analysis for Cooperative Communications in Intersymbol Interference Fading Channels

Yanjie Peng, Andrew G. Klein, and Xinming Huang

Department of Electrical & Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Correspondence should be addressed to Yanjie Peng, yjpeng@wpi.edu

The area of cooperative communications has recently attracted lots of research interest because of the potential benefit of increased spatial diversity. In this paper, we study the detector design for cooperative communications in intersymbol interference (ISI) channels. A novel system framework employing nonorthogonal amplify-and-forward half-duplex relays through ISI channels is introduced. We focus on detector design and first consider an optimal detector that consists of a whitening filter and a maximum-likelihood sequence estimator (MLSE). However, such an optimal detector embedded with Viterbi algorithm has practical issues of high complexity if the relay period is long. Consequently, we adopt a multitrellis Viterbi algorithm (MVA) that reduces the complexity significantly but still achieves near-optimal performance. Simulation results demonstrate the performance of both the optimal and near-optimal detector designs.

1. Introduction

Multipath fading is one of the major impairments to the performance of wideband wireless communication systems [1]. Historically, such fading has been combated by using time and frequency diversity techniques. In the last fifteen years, results in information theory have shown that spatial diversity can yield significant gains in the spectral efficiency and power efficiency of point-to-point multiple-antenna communication (MIMO) systems [2, 3]. The transition from theory to practice has largely taken place with MIMO technology, as many modern consumer wireless standards exploit MIMO technology. To realize such gains, however, it is necessary that each of the paths between transmit and receive antennas is uncorrelated. For such an assumption to be valid, it is typically required that the antenna elements are spaced at least a half carrier wavelength apart [4], and perhaps even more in environments with minimal scattering. In many scenarios, however, it may not be practical for size-constrained nodes to have even two antennas with sufficient spacing between them. Furthermore, for each antenna that is added to the node, a complete RF front end must be added. There may be cost and power constraints that preclude

the inclusion of multiple antennas, as the RF portion of a communications system often accounts for the majority of the cost and power.

More recently, cooperative diversity [5, 6] and relay networks [7] have attracted a lot of attention for their ability to exploit increased spatial diversity available at distributed antennas on other nodes in the system. By intelligent cooperation among nodes in the network which may only have a single antenna, a virtual multiple antenna system can be formed. Indeed, information theoretic results demonstrate that some of the loss associated with using only a single antenna can be recuperated by using intelligent cooperation among distributed nodes [8, 9].

While communication via cooperative relays has seen a lot of active research interest in recent years, most of the existing work has largely come from the information theory and coding communities. While there are few exceptions, for example [10, 11], little research has yet been conducted into the implementation issues of relaying and cooperation. As such, the majority of works in the field of cooperative diversity assume that receivers employ optimal detectors.

In this paper, we set out to investigate detector design for half-duplex relays in frequency-selective fading channels
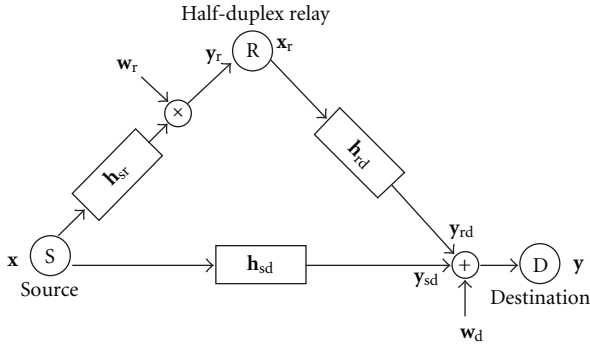
Figure 1: System model with one half-duplex relay.

encountered in practice. While a variety of forwarding protocols have been previously proposed, we will consider amplify-and-forward (AF) for its simplicity and reduced implementation costs. Frequency selective fading channels are an inevitable impairment in wideband communication systems [1], and such channels cause the receiver to observe the superposition of multiple delayed reflections of the transmitted signal, resulting in intersymbol interference (ISI). Even in channels which do not exhibit significant time dispersion, the nonorthogonal AF relay itself effectively introduces ISI since the destination observes a superposition of the source and relayed signals. As our focus is on the complexity of the detector itself, we do not treat the performance gains possible with relays as this has been demonstrated elsewhere [12]. Similarly, with our focus on the complexity of detector implementation, we do not address the problem of channel estimation and thus consider the optimistic case where the detector has perfect channel knowledge. After developing a system model for the case of AF relays in ISI channels, we will present an optimal detector realization based on the Viterbi algorithm (VA), and we will address its implementation complexity. Then we will show that the delay introduced by the half-duplex AF relay causes the effective channel impulse response to become sparse when the relay period is large. To exploit the inherent sparsity in the effective channel, a multitrellis Viterbi algorithm (MVA) [13] is adopted which results in much lower complexity but negligible performance loss. Finally, we conclude with numerical simulations demonstrating performance of the detector in slow standardized fading channels.

2. Cooperative Communication System Model

A basic model of a three-node relay system model is shown in Figure 1. Both source and relay can be considered as mobile users, and each has only one antenna. The source is attempting to send a message to the destination. Due to the broadcast nature of wireless communications, however, the relay receives transmissions from the source that are intended for the destination; therefore, the relay can assist by forwarding additional copies of these transmissions to the destination. Since the channels from source and relay to destination are statistically independent, the three-node cooperative communication scheme effectively forms

spatial diversity. We consider a system where a source transmits a continuous stream of data to a destination, and a simplistic AF relay assists the source by amplifying and forwarding the data to the destination. We do not assume the relay has performed any synchronization with the destination, and so the relay forwards information to the destination in an open-loop fashion.

Relays have mainly two types: full-duplex relays that can transmit and receive simultaneously and half-duplex relays that can either transmit or receive in any time slot. Since full duplex relays are difficult to implement due to self-interference which occurs when both transmit and receive operations are in the same band, half-duplex is considered more practical for cooperative communication systems. In our system model, the half-duplex relay period T is a parameter which defines the frame structure where the relay receives for T symbol periods and then transmits for T symbol periods. The relay repeats these two tasks alternately. The source and relay are assumed to transmit on the same channel, employing the so-called nonorthogonal amplify-forward protocol (NAF) [12].

The source sends the symbols $\mathbf{x} = [x[0], x[1], \ldots, x[N-1]]^T \in \mathbb{C}^N$ at a symbol rate of f, where N is the number of transmitted symbols. We assume that a squared root raised cosine (SRRC) filter with the impulse response $h_{Tx}(t)$ is applied as the transmitter filter. The received signal is processed by a matched filter with the impulse response $h_{Rx}(t) = h_{Tx}(-t)$ and then sampled at the symbol rate of f. The equivalent discrete-time channel impluse responses [14] which include the effect of pulse shaping are denoted by $\mathbf{h}_{sd}$, $\mathbf{h}_{sr}$, and $\mathbf{h}_{rd}$ for the source-destination, source-relay, and relay-destination channels, respectively, and they have corresponding channel lengths L_{sd}, L_{sr} and L_{rd} (e.g., $\mathbf{h}_{sd} = [h_{sd}[0], h_{sd}[1], \ldots, h_{sd}[L_{sd}-1]]^T$). The signals $\mathbf{w}_r$ and $\mathbf{w}_d$ are complex additive white Gaussian noise (AWGN) at the relay and the destination with variances σ_r^2 and σ_d^2, respectively.

The destination receives the superposition of the two signals from the source and the relay, and the received signal can be expressed as

$$\mathbf{y} = \mathbf{y}_{sd} + \mathbf{y}_{rd} + \mathbf{w}_d, \tag{1}$$

where $\mathbf{y}_{sd} \in \mathbb{C}^{N+L_{sd}-1}$ is the contribution from the source and $\mathbf{y}_{rd} \in \mathbb{C}^{N+L_{sr}+L_{rd}-2}$ is the contribution from the relay.

We first consider the source-destination link. The contribution from source to destination is written as

$$\mathbf{y}_{sd} = \mathbf{H}_{sd}\mathbf{x}, \tag{2}$$

where $\mathbf{H}_{sd} \in \mathbb{C}^{(N+L_{sd}-1)\times N}$ is the complex Tœplitz channel convolution matrix whose entries are defined by

$$[\mathbf{H}_{sd}]_{i,j} = \begin{cases} h_{sd}[i-j], & 0 \le i-j \le L_{sd}-1, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where $1 \le i \le N+L_{sd}-1$ and $1 \le j \le N$, that is,

$$\mathbf{H}_{sd} = \begin{bmatrix} h_{sd}[0] & 0 & 0 & 0 & 0 & 0 \\ h_{sd}[1] & h_{sd}[0] & 0 & 0 & 0 & 0 \\ h_{sd}[2] & h_{sd}[1] & h_{sd}[0] & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ h_{sd}[L_{sd}-1] & \cdots & & h_{sd}[1] & h_{sd}[0] & 0 \\ 0 & h_{sd}[L_{sd}-1] & \cdots & & h_{sd}[1] & h_{sd}[0] & \cdots \\ \vdots & \vdots & & & \vdots & \ddots \end{bmatrix}. \tag{4}$$

For the source-relay-destination link, the corresponding contribution is given by

$$\begin{aligned} \mathbf{y}_{rd} &= \mathbf{H}_{rd}\mathbf{x}_r \\ &= \mathbf{H}_{rd}\boldsymbol{\Gamma}\mathbf{y}_r \\ &= \mathbf{H}_{rd}\boldsymbol{\Gamma}(\mathbf{H}_{sr}\mathbf{x} + \mathbf{w}_r). \end{aligned} \tag{5}$$

The Tœplitz channel matrices $\mathbf{H}_{rd} \in \mathbb{C}^{(N+L_{sr}+L_{rd}-2)\times(N+L_{sr}-1)}$ and $\mathbf{H}_{sr} \in \mathbb{C}^{(N+L_{sr}-1)\times N}$ are defined in the same way as $\mathbf{H}_{sd}$, $\mathbf{y}_r \in \mathbb{C}^{N+L_{sr}-1}$ is the signal received by the relay, $\mathbf{x}_r \in \mathbb{C}^{N+L_{sr}-1}$ is the signal transmitted from the relay, and $\boldsymbol{\Gamma} \in \mathbb{C}^{(N+L_{sr}-1)\times(N+L_{sr}-1)}$ is a fixed matrix described below. Note that for the matrix dimensions to be compatible, we require that $L_{sd} = L_{sr} + L_{rd} - 1$; if this is not satisfied, we can append zeros to the appropriate matrix without loss of generality. The function of $\boldsymbol{\Gamma}$ is to impose the half-duplex constraint by selecting groups of T symbols from $\mathbf{y}_r$ (receiving), scaling these symbols by a factor β (amplifying), and then delaying the scaled symbols of $\mathbf{y}_r$ for transmission in the next T symbol block (forwarding). The value of β is typically chosen to satisfy an average power constraint at the relay by choosing $\beta = \sqrt{P_r/(\|\mathbf{h}_{sr}\|^2 P_s + \sigma_r^2)}$, where P_s and P_r are the source power and relay power, respectively. The constant matrix $\boldsymbol{\Gamma}$ is given by

$$\boldsymbol{\Gamma} \triangleq \beta\mathbf{I}_{((N+L_{sr}-1)/2T)} \otimes \begin{bmatrix} \mathbf{0}_{T\times T} & \mathbf{0}_{T\times T} \\ \mathbf{I}_{T\times T} & \mathbf{0}_{T\times T} \end{bmatrix}, \tag{6}$$

where $\otimes$ denotes Kronecker product. Here we implicitly require that $N + L_{sr} - 1$ be divisible by $2T$. As an example, when $T = 2$, the signals received and transmitted by the relay are shown in Figure 2, where the first eight time periods are considered, and the shadow indicates time period where the relay cannot receive because it is transmitting, or vice versa.

For this example,

$$\boldsymbol{\Gamma} = \beta\mathbf{I}_2 \otimes \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \beta\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{7}$$

such that $\mathbf{x}_r = \boldsymbol{\Gamma}\mathbf{y}_r$ as in Figure 2.

From (1), (2), and (5), the received signal at the destination is expressed as

$$\begin{aligned} \mathbf{y} &= \overbrace{\mathbf{H}_{sd}\mathbf{x}}^{\text{from source}} + \overbrace{\mathbf{H}_{rd}\boldsymbol{\Gamma}(\mathbf{H}_{sr}\mathbf{x} + \mathbf{w}_r)}^{\text{from relay}} + \mathbf{w}_d \\ &= \underbrace{(\mathbf{H}_{sd} + \mathbf{H}_{rd}\boldsymbol{\Gamma}\mathbf{H}_{sr})}_{\triangleq \tilde{\mathbf{H}}}\mathbf{x} + \underbrace{(\mathbf{H}_{rd}\boldsymbol{\Gamma}\mathbf{w}_r + \mathbf{w}_d)}_{\triangleq \tilde{\mathbf{w}}}, \end{aligned} \tag{8}$$

where $\tilde{\mathbf{w}} \sim \mathcal{CN}(0, \sigma_d^2\mathbf{I} + \sigma_r^2\mathbf{H}_{rd}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^H\mathbf{H}_{rd}^H)$. Note that $\tilde{\mathbf{w}}$ is colored, not white, since the AWGN on the source-relay link is amplified-and-forwarded over the relay-destination ISI channel which colors the noise. Additionally, from (6) and (7), we see that the relay matrix $\boldsymbol{\Gamma}$ has a repetitive structure with a period of $2T$. Accordingly, the channel matrix $\tilde{\mathbf{H}}$ shows the same structure as $\boldsymbol{\Gamma}$. Consequently, $\tilde{\mathbf{H}}$ can be interpreted as a periodically time-varying FIR channel which consists of $2T$ sets of different channel coefficients. In summary, (8) allows us to describe the input-output behavior of the system with a linear equation. While the constituent channels themselves are not time-varying, the effective impulse response of the overall system is indeed time-varying due to the on/off behavior of the relay.

3. Maximum-Likelihood Detector

Assuming that receivers can acquire perfect channel knowledge, maximum likelihood sequence estimation (MLSE)

Relay received signal $\mathbf{y}_r$ | $y_r[0]$ | $y_r[1]$ | ░░ | ░░ | $y_r[4]$ | $y_r[5]$ | ░░ | ░░ |

Γ

Relay transmitted signal $\mathbf{x}_r$ | ░░ | ░░ | $\beta y_r[0]$ | $\beta y_r[1]$ | ░░ | ░░ | $\beta y_r[4]$ | $\beta y_r[5]$ |
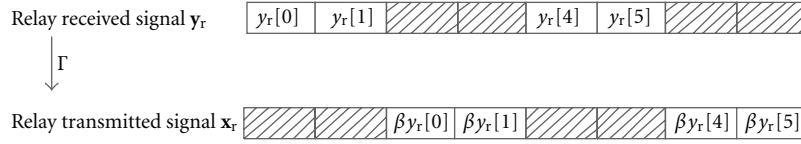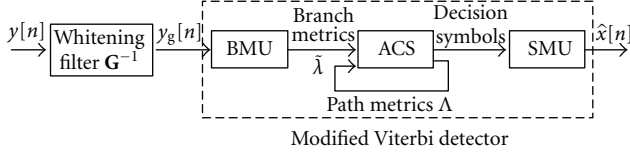
FIGURE 2: Signals received and transmitted by the relay when $T = 2$.

FIGURE 3: Block diagram of the proposed detector.

can be employed to combat the ISI by searching for the minimum Euclidean distance between observed signal and any given transmitted signals [15]. The Viterbi algorithm [16] is an efficient technique for solving the minimum distance problem, and its implementation has been investigated extensively [17–21]. The traditional Viterbi algorithm as proposed in [16] is directly applicable only to time-invariant channels. A modified Viterbi detector is proposed to address the periodically time-varying *effective* channel induced by the half-duplex relay. Furthermore, since the minimum Euclidean distance is not optimal in the presence of colored Gaussian noise, we employ a whitening filter before detection, which is optimal as shown in [15]. The block diagram of our design is given in Figure 3.

To whiten the noise, spectral factorization of the composite noise covariance must be performed. We factor the noise covariance matrix as.

$$\mathbf{G}\mathbf{G}^H = \sigma_d^2 \mathbf{I} + \sigma_r^2 \mathbf{H}_{rd} \mathbf{\Gamma} \mathbf{\Gamma}^H \mathbf{H}_{rd}^H \tag{9}$$

which can be accomplished by taking $\mathbf{G}$ to be the Cholesky factorization of the covariance. We note that the Cholesky

factorization is not the only such factorization $\mathbf{G}$, as the factorization in (9) is not unique. By filtering the received signal with $\mathbf{G}^{-1}$ (i.e., by forming $\mathbf{G}^{-1}\mathbf{y}$), the noise becomes whitened since the covariance of the filtered noise $\mathbf{G}^{-1}\widetilde{\mathbf{w}}$ is given by

$$E\left[(\mathbf{G}^{-1}\widetilde{\mathbf{w}})(\mathbf{G}^{-1}\widetilde{\mathbf{w}})^H \right] = (\mathbf{G}^{-1})E\left[\widetilde{\mathbf{w}}\widetilde{\mathbf{w}}^H \right](\mathbf{G}^{-1})^H$$

$$= (\mathbf{G}^{-1}\mathbf{G})(\mathbf{G}^{-1}\mathbf{G})^H \tag{10}$$

$$= \mathbf{I}.$$

We note that noise covariance matrix in (9) is positive definite, so the inverse of $\mathbf{G}$ always exists. Ignoring end effects (or, equivalently, taking the block length $N \rightarrow \infty$), $\mathbf{G}^{-1}$ follows the same repetitive structure as $\mathbf{\Gamma}$ and thus also exhibits the periodically time-varying property.

After applying the whitening filter to the received signal (8), the whitening filter output becomes

$$\mathbf{y}_g = \overbrace{\mathbf{G}^{-1}\widetilde{\mathbf{H}}}^{\mathbf{H}_{eff}} \mathbf{x} + \overbrace{\mathbf{G}^{-1}\widetilde{\mathbf{w}}}^{\mathbf{w}_{eff}} \tag{11}$$

$$= \mathbf{H}_{eff}\mathbf{x} + \mathbf{w}_{eff}$$

where $\mathbf{w}_{eff}$ is now *white* Gaussian noise. Note that the effective whitened channel matrix $\mathbf{H}_{eff}$ maintains the periodically time-varying property due to the similar structures of $\mathbf{G}^{-1}$ and $\widetilde{\mathbf{H}}$. The structure of the effective channel matrix $\mathbf{H}_{eff}$ is given as

$$\begin{bmatrix} \ddots & & \vdots & & & & & & & \\ \cdots & h_0[L-1] & \cdots & & h_0[0] & & & & & \\ & & h_1[L-1] & \cdots & & h_1[0] & & & & \\ & & & \ddots & & & & & & \\ & & & & h_{2T-1}[L-1] & \cdots & h_{2T-1}[0] & & & \\ & & & & & h_0[L-1] & \cdots & & h_0[0] & \\ & & & & & & h_1[L-1] & \cdots & & h_1[0] \\ & & & & & & & \ddots & & \\ & & & & & & & h_{2T-1}[L-1] & \cdots & h_{2T-1}[0] & \cdots \\ & & & & & & & & \vdots & \ddots \end{bmatrix}, \tag{12}$$

where we see that the matrix has a block Tœplitz structure with rows repeating every multiple of $2T$. It defines $2T$ sets of effective channel coefficients as $\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{2T-1}$, where $\mathbf{h}_0 = [h_0[0], h_0[1], \ldots, h_0[L-1]]^T$, $\mathbf{h}_1 = [h_1[0], h_1[1], \ldots, h_1[L-1]]^T, \ldots$, and L is the effective channel length. The effective length L may be significantly extended by the delay introduced by the relay, as well as the group delay introduced by the whitening filter. We can determine the lower bound of the effective channel length in terms of the constituent channel lengths and the relay period as

$$L \geq \max(L_{sd}, L_{sr} + L_{rd} + T - 1). \tag{13}$$

At time n, the corresponding coefficients of the periodically time-varying effective channel are

$$\mathbf{h}_n = \mathbf{h}_{\bmod(n,2T)} \in \{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{2T-1}\}, \quad n = 0, 1, \ldots, \tag{14}$$

where $\bmod(\cdot)$ is the modulus. The ideal output of the whitening filter at time n is then given by

$$s[n] = \sum_{i=0}^{L-1} h_n[i]x[n-i], \quad n = 0, 1, \ldots, \tag{15}$$

which is simply the convolution of the source symbols with the periodically time-varying channel coefficients. Thus, the system model for relay-aided transmission through ISI channels reduces to a classical MLSE problem, with the additional twist that the effective channel is periodically time-varying. As the true output is of course corrupted by AWGN, the maximum-likelihood detector for estimating the source symbols $\mathbf{x}$ from (11) can be accomplished most efficiently with the Viterbi algorithm.

For the branch metric unit (BMU) in Figure 3, the branch metrics along the trellis path are not only related to state transitions but also the current time instant. The branch metric calculation is modified as

$$\tilde{\lambda}[n] = \left| y_g[n] - \tilde{s}[n] \right|^2, \quad n = 0, 1, \ldots, \tag{16}$$

where $y_g[n]$ is the signal from the whitening filter at the time instant n, $\tilde{s}[n]$ and $\tilde{\lambda}[n]$ are the estimated output signal and the corresponding branch metric the time instant n, respectively.

The add-compare-select (ACS) unit in Figure 3 recursively computes path metrics and decision bits,

$$\Lambda^{(j)}[n] = \min_i \left(\Lambda^{(i)}[n-1] + \tilde{\lambda}^{(i,j)}[n] \right), \quad n = 0, 1, \ldots, \tag{17}$$

where $\Lambda^{(j)}[n]$ denotes the path metric at state j at the time instant n, and $\tilde{\lambda}^{(i,j)}[n]$ is the branch metric from state i to state j at instant n. The path metrics for each state are updated for the next iteration, and the decision indicating the survivor path for state j is recorded and retrieved from the survivor-path memory unit (SMU) in order to estimate the transmitted symbols along the final survivor path.

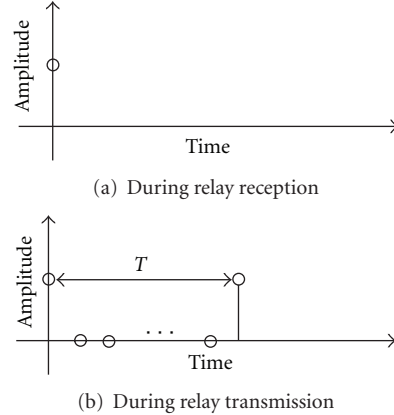(a) During relay reception

(b) During relay transmission

Figure 4: Effective channel impulse response in non-ISI channels.

Similar to the traditional MLSE, the implementation cost of the ML detector for relay networks increases exponentially with respect to the effective channel length. Furthermore, the overhead of the proposed detector when compared with the traditional MLSE comprises the whitening filter, the multiplexers in the BMU, and additional control logic to account for the periodically time-varying effective channel [22]. As indicated in (13), the effective channel length increases with the relay period T. In cooperative relay systems, the relay period T is likely chosen to be very long, possibly spanning hundreds of symbols, so that the relay is not required to switch frequently between transmit and receive modes. When the relay period T is large, however, an implementation of the Viterbi algorithm-based optimal detector becomes not practical. This problem will be addressed in Section 4.

4. Near-Optimal Detector for Long Relay Periods

Consider the practical effect of the relay period T in a simplified relay network with only AWGN channels and no ISI so that $L_{\{sd, sr, rd\}} = 1$. When the relay is in the receiving period, the destination only receives the signal from the source. The channel impulse response at the destination is shown in Figure 4(a). After T time instants, the relay begins to forward the copy of the signal to the destination. The effective channel impulse response as seen at the destination during relay transmission is depicted in Figure 4(b). We observe that the relay actually introduces ISI even in the non-ISI channels, and the effective channel impulse response alternates between that shown in Figures 4(a) and 4(b). Also, large values of T will increase the number of zero coefficients in the channel impulse response, which make the effective channels become sparse.

Viterbi algorithms for sparse channels have been investigated by several independent researchers. The parallel trellis Viterbi algorithm (PTVA) proposed in [23] reformulated the original single trellis into a set of independent trellises. These independent trellises operate in parallel and have less overall complexity than a single trellis. The PTVA requires that

the sparse channel have equispaced coefficients, however, which usually cannot be satisfied in practice. Although a generalized PTVA is given to deal with general sparse channels, its performance loss is remarkable if the channel taps are not well-approximately by an equispaced structure.

A multitrellis Viterbi algorithm (MVA) was proposed in [13] for near-optimal detection in sparse ISI channels. For the MVA, the complexity does not depend on the channel impulse response length but only on the number of nonzero coefficients. In order to process the sparse time-varying channels for relay networks, the MVA is modified and incorporated in our MVA-based ML detector.

To illustrate the operation of the MVA, we begin by considering an example. Assume $\mathbf{h}_n \in \{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{2T-1}\}$ has only a few nonzero coefficients, for example, $h_n[i] \neq 0$ for $i = 0, K, L-1$, $(0 < K < L-1)$. The ideal (noiseless) output signal at the time n is given by

$$s[n] = h_n[0]x[n] + h_n[K]x[n-K] + h_n[L-1]x[n-L+1]. \tag{18}$$

When $x[0]$ is to be estimated, we can see that $x[0]$ is needed in $s[0], s[K], s[L-1]$. Next, in $s[K]$, a new symbol $x[K]$ appears, and it is also needed in $s[L-1]$ and $s[2K]$. In this way, we record all the output signals and symbols related with $x[0]$ in Table 1, and we use the notation of $f(\cdot)$ to indicate the dependency of outputs on input symbols and channels. Note that some output signals and symbols are not needed when $x[0]$ is under detection; for example, if $K \neq 1$, there is no need to record $s[1]$, since $x[1], x[1-K], x[2-L]$ in $s[1]$ do not affect the estimation of $x[0]$. With the traceback length $L_{tb} = 3(L-1)$, for example, the estimation of $x[0]$ depends on $x[K], x[L-1], x[2K], x[K+L-1], x[2L-2]$ by a noninstantaneous relationship, assuming that $x[n], n < 0$, are known.

We note that some related symbols appear only once in the first column, for example, $x[L-1-K]$, and do not need to be recorded. Its value can be determined by an instant decision given by

$$\hat{x}[L-1-K] = \arg\min_{\tilde{x}[L-1-K]} \left| y_g[L-1] - \tilde{s}[L-1] \right|, \tag{19}$$

where $\tilde{s}[L-1] = f(\mathbf{h}_{L-1}, x[L-1], \tilde{x}[L-1-K], x[0])$, and $x[L-1]$ and $x[0]$ are known for a given state.

When two or more symbols are determined by the instant decision, for example, $x[3K]$ and $x[3K-L+1]$ in $s[3K]$, the estimation is given by

$$\{\hat{x}[3K], \hat{x}[3K-L+1]\} = \arg\min_{\tilde{x}[3K], \tilde{x}[3K-L+1]} \left| y_g[3K] - \tilde{s}[3K] \right|, \tag{20}$$

where $\tilde{s}[3K] = f(\mathbf{h}_{3K}, \tilde{x}[3K], x[2K], \tilde{x}[3K-L+1])$, and $x[2K]$ is known for a given state.

The definition of state depends only on the related time instant. By the list of related symbols, the state definition is derived and given in Table 2. Note that the state definition excludes the symbols assumed to be known, that is, $x[-K]$, $x[-L+1]$, $x[K-L+1]$, and the symbols that can be determined by the instant decision.

From Tables 1 and 2, it is observed that the corresponding trellis shrinks in two dimensions, which leads to a significant reduction in computational complexity. Furthermore, the process of traceback is faster, since, for some instant given the current state, the previous state can be obtained immediately without survivor path decisions. There are two categories for these instant tracebacks. First, the previous state definition is a subset of the current state definition. For example, the state at instant $2K$ is defined as $[x[2K], x[L-1], x[K]]$, while the previous state $[x[L-1], x[K]]$ at the instant $L-1$ can be obtained from the current state without the help of the survivor path decision. Second, the state definition is the same for the current and previous state. For example, considering the instants $K+2L-1$, $2K+L-1$, $3K$, and $2L-2$, we can bypass the traceback from $K+2L-2$ to $2L-2$. Once the start state at the instant $K+2L-2$ is available, we can begin to traceback from the instant $2L-2$ at the same state.

When subsequent symbols are under detection, the structure of the trellis remains the same, except that the branch metric calculation for the first several instants are slightly different, since the initial symbols (e.g., $x[-K]$, $x[-L+1]$, $x[K-L+1]$, for $x[0]$) have been estimated. The available estimated symbols will be used in the calculation of output signal $s[n]$ when needed.

Due to the reduced-size trellis, the detector can be realized by utilizing L_{tb} trellises working in parallel to increase the throughput. The received signals $y_g[n]$ are filled in the L_{tb} trellises sequentially. At the instant $L_{tb} - 1$, the first trellis is full, and $x[0]$ is estimated. At the instant L_{tb}, the received signal $y_g[L_{tb}]$ is ready to fill in the first trellis, and, also, $x[1]$ is available from the second trellis. Notice that these trellis are similar in structure; however, the channel coefficients used in the branch metric calculation are not the same at different instants. For example, in the first trellis, $\mathbf{h}_0, \mathbf{h}_K, \mathbf{h}_{L-1}, \ldots, \mathbf{h}_{K+2L-2}$ are used in sequence for each step, while in the second trellis, $\mathbf{h}_1, \mathbf{h}_{K+1}, \mathbf{h}_L, \ldots, \mathbf{h}_{K+2L-1}$ are used for each step. The corresponding channel coefficients for the received signals $y_g[n]$ are $\mathbf{h}_{\bmod(n,2T)} \in \{\mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_{2T-1}\}$.

For general Viterbi detectors in M-ary modulation systems, complex multiplications dominate the computational cost. There are M^{L-1} states, and each state corresponds to M complex multiplications for branch metric calculation. Then the estimated total computational cost is M^L. For the proposed MVA detector in the sparse channel with 3 nonzero coefficients, there are M^3 states at most, and the corresponding computational cost for each trellis is M^4. Thus the estimated total computational cost for MVA detector is $L_{tb}M^4$. Furthermore, survivor path decisions are recorded in the SMU which requires a significant amount of memory. The memory cost in bits for the general Viterbi detector is $L_{tb}M^{L-1}\log_2 M$. For the proposed MVA detector, we do not need to record the survivor path decisions for all L_{tb} instants. Assuming that only $L'_{tb}(< L_{tb})$ instants are considered in each trellis, the memory size for each trellis is $L'_{tb}M^3\log_2 M$.

In general, the complexity of Viterbi detector is in $\mathcal{O}(M^L)$, where L is the channel length. In comparison, the complexity of the MVA detector does not depend on the

TABLE 1: The dependencies between $x[0]$ and related output signal.

Related output signal	Symbols to be recorded
$s[0] = f(\mathbf{h}_0, x[0], x[-K], x[-L+1])$	$x[0]$
$s[K] = f(\mathbf{h}_K, x[K], x[0], x[K-L+1])$	$x[0], x[K]$
$s[L-1] = f(\mathbf{h}_{L-1}, x[L-1], x[L-1-K], x[0])$	$x[0], x[K], x[L-1]$
$s[2K] = f(\mathbf{h}_{2K}, x[2K], x[K], x[2K-L+1])$	$x[K], x[L-1], x[2K]$
$s[K+L-1] = f(\mathbf{h}_{K+L-1}, x[K+L-1], x[L-1], x[K])$	$x[K], x[L-1], x[2K], x[K+L-1]$
$s[2L-2] = f(\mathbf{h}_{2L-2}, x[2L-2], x[2L-2-K], x[L-1])$	$x[L-1], x[2K], x[K+L-1], x[2L-2]$
$s[3K] = f(\mathbf{h}_{3K}, x[3K], x[2K], x[3K-L+1])$	$x[2K], x[K+L-1], x[2L-2]$
$s[2K+L-1] = f(\mathbf{h}_{2K+L-1}, x[2K+L-1], x[K+L-1], x[2K])$	$x[2K], x[K+L-1], x[2L-2]$
$s[K+2L-2] = f(\mathbf{h}_{K+2L-2}, x[K+2L-2], x[2L-2], x[K+L-1])$	$x[K+L-1], x[2L-2]$

TABLE 2: State definition ($x[0]$ under estimation).

Related time instant	State definition
0	$[x[0]]$
K	$[x[K], x[0]]$
$L-1$	$[x[L-1], x[K]]$
$2K$	$[x[2K], x[L-1], x[K]]$
$K+L-1$	$[x[K+L-1], x[2K], x[L-1]]$
$2L-2$	$[x[2L-2], x[L-1+K], x[2K]]$
$3K$	$[x[2L-2], x[L-1+K], x[2K]]$
$2K+L-1$	$[x[2L-2], x[L-1+K], x[2K]]$
$K+2L-2$	$[x[2L-2], x[L-1+K], x[2K]]$

FIGURE 5: Average BER performance for different channel length.

- ○ 2 independent paths
- □ 3 independent paths
- ∗ 6 independent paths

channel length but on the number of nonzero coefficients. When instantaneous decisions are made appropriately so that the symbol dependency table simplifies [13], the computing and memory resource required for the MVA detector is in $\mathcal{O}(M^{L'})$, where L' is the number of nonzero coefficients. Therefore, the MVA detector is a better solution for sparse channels.

5. Simulation Results

We simulate the bit error rate (BER) performance of the proposed ML detector with the following parameters: the transmitted BPSK signal consists of i.i.d unit-power symbols $x[n] = \{\pm 1\}$, the relay transmitting power is $P_r = 1$. The SNR values $E_b/N_0^{(d)} = 1/\sigma_d^2$ and $E_b/N_0^{(r)} = 1/\sigma_r^2$ denote bit energy-to-noise ratio for the destination and the relay, respectively. Unless otherwise specified, we assumed $E_b/N_0^{(r)} = E_b/N_0^{(d)} + 10$ dB, which represents a scenario where the source-relay link is better, on average, than the source-destination link.

We assume the SRRC pulse shaping filter is employed at both the transmitting and receiving ends. The SRRC filter is truncated to $[-2/f, 2/f]$ with a roll-off factor 0.5, where $f = 5$ MHz is the symbol rate. We simulate over 200 fading realizations in the ITU-R 3G indoor office test environment [24] with 6 independent channel paths. The time delay relative to the first path is $[0, 50, 110, 170, 290, 310]$ (ns), and the average power relative to the strongest path is $[0, -3, -10, -18, -26, -32]$ (dB). As a practical matter, it is
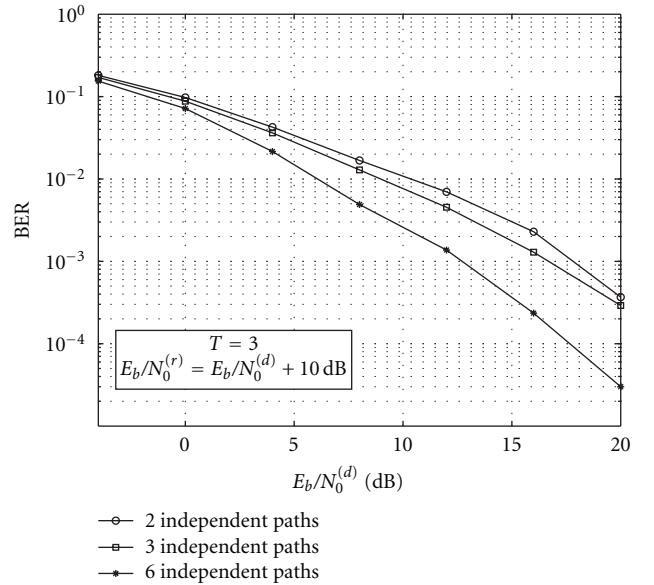
possible for the whitening filter to be quite long, depending on the noise covariance matrix. In these simulations, we truncate the whitening filter to have $L = \max(L_{sd}, L_{sr} + L_{rd} + T - 1)$ taps.

We consider the effect of the number of independent paths on system performance. While additional ISI is often viewed as an impairment to reliable communication, the additional fading paths result in increased diversity, and hence increased BER performance through the cooperative relay. This effect is observed in Figure 5, where we consider three cases. Specifically, we truncate the ITU-R 3G indoor office channel so that it only uses either the first two paths, the first three paths, or all the six paths. We note that the BER performance with respect to $E_b/N_0^{(d)}$ improves as the number of independent paths increases.

We next consider the effect of the relay cooperation period T. Recall that the relay receives T symbols and then amplifies and retransmits those T symbols. As shown in Figure 6 for an uncoded system, better performance is obtained when choosing a smaller relay period. If the relay period T is much larger than the channel delay spread, the
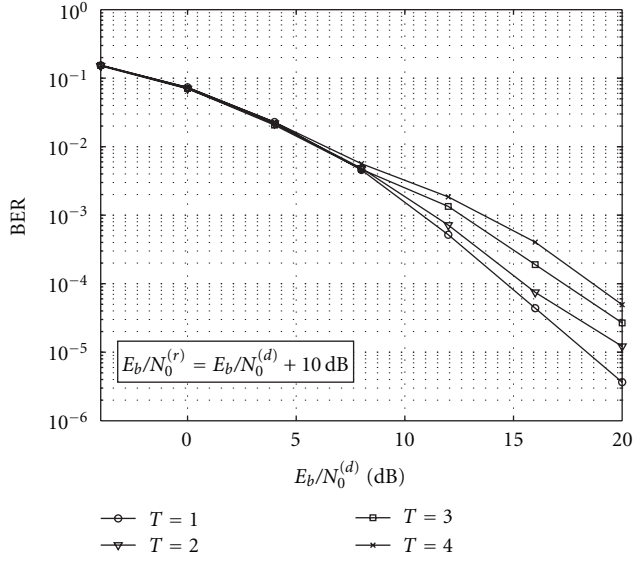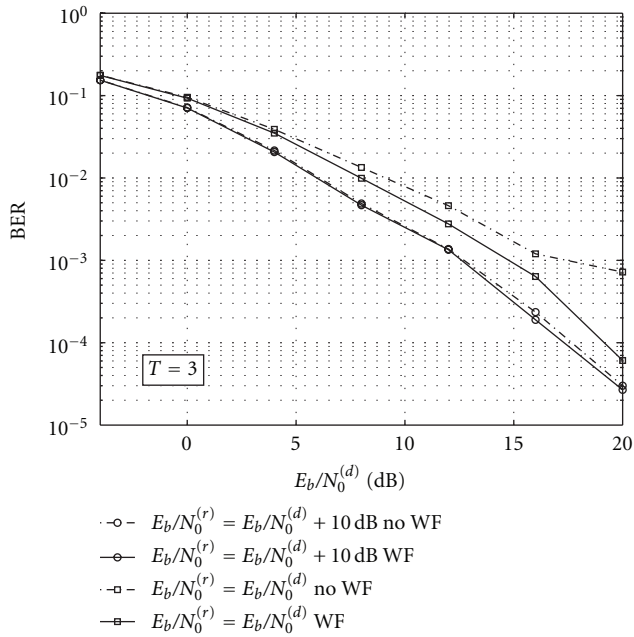
Figure 6: BER performance for different relay period T.

Figure 7: BER performance with and without whitening filter (WF).

(a) $T = 3$

(b) $T = 5$

Figure 8: BER performance of the proposed MVA-based detector.

diversity technique of using cooperative relays becomes less effective since there is almost no overlap between the symbols directly received from the source and forwarded by the relay.

The effect of the whitening filter is illustrated in Figure 7, where $T = 3$. As we expect, the BER performance with the whitening filter is better than the performance without it. It is also observed that if the noise on the source-relay link is small relative to the noise on the source-destination link so that $\sigma_d \gg \sigma_r$, the whitening filter does little to help. The reason for this is that the noise looks approximately white when the
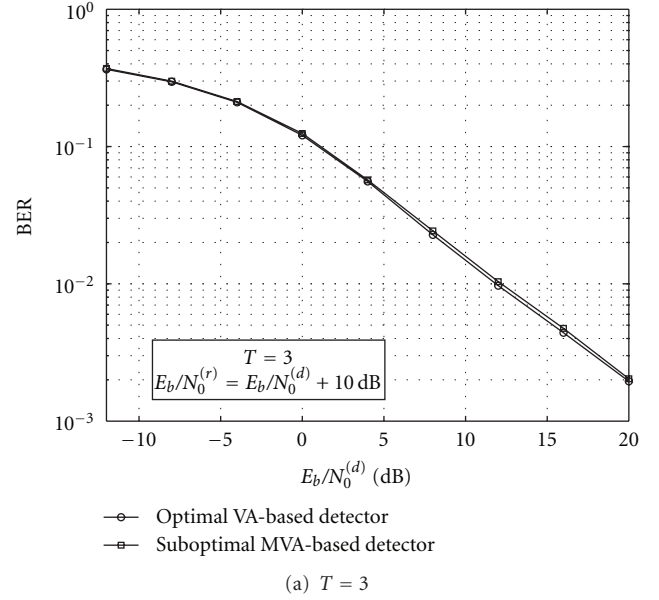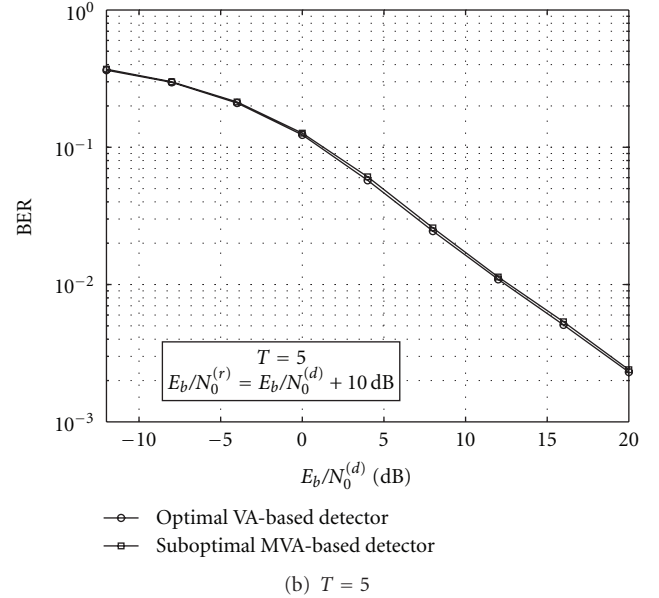
noise on the source-destination link dominates since $\sigma_d^2 \mathbf{I} + \sigma_r^2 \mathbf{H}_{rd} \mathbf{\Gamma} \mathbf{\Gamma}^H \mathbf{H}_{rd}^H \approx \sigma_d^2 \mathbf{I}$. Thus, in situations where the source-relay link is particularly good, it may be possible to reduce complexity without much performance penalty by removing the whitening filter.

To simulate the BER performance of the MVA-based detector, we set the channel length $L_{\{sd,rd\}} = 1$, $L_{sr} = 2$, so that each effective channel $\mathbf{h}_n$ has 3 nonzero coefficients, and Table 1 can be applied directly. Without loss of generality, we assume that each coefficient is i.i.d. as $\mathcal{CN}(0, 1)$. The performance is given in Figure 8(a) with $T = 3$ and Figure 8(b) with $T = 5$. It is shown that performance loss from the MVA-based detector is negligible, as was claimed previously in [13].

6. Conclusion

A system model is introduced for cooperative communication in ISI channels with amplify-and-forward relays. Based on the system model, we present a maximum-likelihood detector design based on the Viterbi algorithm. For long relay periods, the equivalent channel was shown to become sparse. Moreover, the channel length becomes much larger which makes the ML detector hardly practical. Therefore, we proposed use of the MVA algorithm in which the complexity is determined by the number of nonzero taps in the effective channel impulse response. The simulation results show that the MVA algorithm provides near-optimal BER performance but at significantly lower hardware complexity.

References

[1] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2000.

[2] I. E. Telatar, "Capacity of multi-antennagaussian channels," Tech. Rep., AT&T Bell Laboratories, October 1995, http://mars.bell-labs.com/papers/proof/.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[4] W. Jakes Jr., *Microwave Mobile Communications*, chapter 1 and 5, John Wiley & Sons, New York, NY, USA, 1974.

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity, parts I & II," *IEEE Transactions on Communications*, vol. 51, 2003.

[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[7] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[8] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '98)*, p. 156, August 1998.

[9] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '01)*, p. 294, June 2001.

[10] P. Murphy, A. Sabharwal, and B. Aazhang, "Building a cooperative communications system," *Computing Research Repository*, http://arxiv.org/abs/0707.2998.

[11] S. Wei, D. L. Goeckel, and M. C. Valenti, "Asynchronous cooperative diversity," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1547–1557, 2006.

[12] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4152–4172, 2005.

[13] N. Benvenuto and R. Marchesani, "The Viterbi algorithm for sparse channels," *IEEE Transactions on Communications*, vol. 44, no. 3, pp. 287–289, 1996.

[14] P. Hoeher, "A statistical discrete-time model for the wssus multipath channel," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 4, pp. 461–468, 1992.

[15] G. D. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, 1972.

[16] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[17] M. Kamuf, V. Owall, and J. B. Anderson, "Optimization and implementation of a Viterbi decoder under flexibility constraints," *IEEE Transactions on Circuits and Systems I*, vol. 55, no. 8, pp. 2411–2422, 2008.

[18] Y. Gang, A. T. Erdogan, and T. Arslan, "An efficient pre-traceback architecture for the Viterbi decoder targeting wireless communication applications," *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 9, pp. 1918–1927, 2006.

[19] G. Feygin and P. G. Gulak, "Architectural tradeoffs for survivor sequence memory management in Viterbi decoders," *IEEE Transactions on Communications*, vol. 41, no. 3, pp. 425–429, 1993.

[20] R. Cypher and C. B. Shung, "Generalized trace-back techniques for survivor memory management in the Viterbi algorithm," *Journal of VLSI Signal Processing*, vol. 5, no. 1, pp. 85–94, 1993.

[21] P. J. Black and T. H. Meng, "A 140-Mb/s, 32-state, radix-4 Viterbi decoder," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 12, pp. 1877–1885, 1992.

[22] Y. Peng, A. G. Klein, and X. Huang, "Design of a maximum-likelihood detector for cooperative communications in intersymbol interference channels," in *Proceedings of the 19th ACM Great Lakes Symposium on VLSI (GLSVLSI '09)*, pp. 429–432, ACM, May 2009.

[23] N. C. McGinty, R. A. Kennedy, and P. Hoeher, "Parallel trellis viterbi algorithm for sparse channels," *IEEE Communications Letters*, vol. 2, no. 5, pp. 143–145, 1998.

[24] *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, Recommendation ITU-R M.1225, 1997.