Reliability Distributions of Truncated Max-log-map (MLM) Detectors Applied to Binary ISI Channels

Fabian Lim and Aleksandar Kavčič

Abstract—The max-log-map (MLM) receiver is an approximated version of the well-known, Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. The MLM algorithm is attractive due to its implementation simplicity. In practice, sliding-window implementations are preferred; these practical implementations consider truncated signaling neighborhoods around each transmission time instant. In this paper, we consider the binary signaling case. We consider sliding-window MLM receivers, where for any time instant, the MLM algorithm is attractive due to its estimated version of the well-known, Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm. The MLM algorithm is attractive due to its implementation simplicity. In practice, sliding-window implementations are preferred; these practical implementations consider truncated signaling neighborhoods around each transmission time instant. In this paper, we consider the binary signaling case. We consider sliding-window MLM receivers, where for any integer \( m \), the MLM detector is truncated to a length-\( m \) signaling neighborhood. For any number \( n \) of chosen times, we derive exact expressions for both i) the joint distribution of the MLM symbol reliabilities, and ii) the joint probability of the erroneous MLM symbol detections.

We show that the obtained expressions can be efficiently evaluated using Monte-Carlo techniques. Our proposed method is efficient; the most computationally expensive operation (in each Monte-Carlo trial) is an eigenvalue decomposition of a size \( 2mn \) by \( 2mn \) matrix. Finally, our proposed method handles various scenarios such as correlated noise distributions, modulation coding, etc.

Index Terms—detection, intersymbol interference, max-log-map, probability distribution, reliability

I. INTRODUCTION

The intersymbol interference (ISI) channel has been widely studied in communication theory. In optimal detection schemes for the ISI channel, input-output sequences, rather than individual symbols, have to be considered [1]. Sequence detectors such as the Viterbi detector, only compute hard decisions [2]. On the other hand, modern coding techniques require detection schemes that also compute symbol reliabilities (also known as soft-outputs, log-likelihood ratios, etc.) [3], [4], [5]. Some commonly cited detectors that perform this task include the soft-output Viterbi algorithm (SOVA) [6], the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [7], and the max-log-map (MLM) detector [8]. These detectors have been in use for some time, however there is scarce literature on their analysis.

That being said, it appears there has been recent interest in the analysis of the MLM detector. The marginal symbol error probability has been derived for a 2-state convolutional code in [9]; this has been further extended for convolutional codes with constraint length two in [10]. Also, approximations for the MLM reliability distributions are obtained in [11], [12].

In this paper, we consider the MLM receiver applied, using binary signaling, to an intersymbol interference (ISI) channel.

In particular we consider its sliding-window implementation. A MLM receiver is termed to be \( m \)-truncated, if it only considers a signaling window of length \( m \) around the time instant of interest. The analysis of \( m \)-truncated MLM receivers is shown to be tractable, in which for any number \( n \) of chosen time instants, we derive exact, closed-form expressions for both i) the joint distribution of the symbol reliabilities, and ii) the joint probability that the detected symbols are in error. While past work considered only marginal distributions, we provide analytic expressions for joint MLM receiver statistics. Our derivation is simple; and follows from a simple observation.

Notation: Deterministic quantities are denoted as follows. Bold fonts are used to distinguish both vectors and matrices (e.g. denoted \( \mathbf{a} \) and \( \mathbf{A} \), respectively) from scalar quantities (e.g. denoted \( a \)). Next, random quantities are denoted as follows. Scalars are denoted using upper-case italics (e.g. denoted \( A \)) and vectors denoted using upper-case bold italics (e.g. denoted \( \mathbf{A} \)). Note that we do not reserve specific notation for random matrices. Throughout the paper both \( t \) and \( \tau \) are used to denote time indices. Sets are denoted using curly braces, e.g. \( \{a_1, a_2, a_3, \cdots \} \). Also, both \( \alpha \) and \( \beta \) are used for auxiliary notation as needed. Finally, the maximization over the components of the size-\( n \) vector \( \mathbf{a} = [a_1, a_2, a_3, \cdots a_n]^T \), may be written either explicitly as \( \max_{a_1, \cdots a_n} a_i \), or concisely as \( \max \mathbf{a} \). Events are denoted in curly brackets, e.g. \( \{A \leq a\} \) is the event where \( A \) is at most \( a \). The probability of the event \( \{A \leq a\} \) is denoted \( \Pr \{A \leq a\} \). The letter \( P \) is reserved to denote probability cumulative distribution functions, i.e. \( F_A(a) = \Pr \{A \leq a\} \). The expectation of \( A \) is denoted as \( E \{A\} \).

II. THE MLM ALGORITHM

A random sequence of symbols drawn from the set \( \{-1, 1\} \), denoted as \( \cdots, A_{-2}, A_{-1}, A_0, A_1, A_2, \cdots \), is transmitted across the ISI channel. Let the following random sequence denoted as \( \cdots, Z_{-2}, Z_{-1}, Z_0, Z_1, Z_2, \cdots \) be the ISI channel output sequence. Let \( h_0, h_1, \cdots, h_{\ell} \) denote the ISI channel coefficients, here \( \ell \) is a non-negative integer. The
input-output relationship of the ISI channel is given by the following equation
\[ Z_t = \sum_{i=0}^{\ell} h_i A_{t-i} - W_t, \]  
(1)
and we assume that the noise \( W_1, W_2, \ldots \) are zero-mean and jointly Gaussian distributed (note that we do not assume they are independent).

**Definition 1.** The ISI channel state at time \( t \) equals the (length-\( \ell \)) vector of input symbols \([A_{t-\ell+1}, A_{t-\ell+2}, \ldots, A_t]^T\). The constant \( \ell \) in (1) is termed the ISI channel memory length.

Figure 1 depicts the time evolution of the ISI channel states. The total number of possible states is clearly \( 2^\ell \), which is exponential in the memory length \( \ell \).

**A. The m-truncated max-log-map (MLM) detector**

We proceed to describe the sliding-window MLM receiver. At time instant \( t \), the \( m \)-truncated MLM detector considers the neighborhood of \( 2m+\ell+1 \) channel outputs \( Z_t = [Z_{t-m}, Z_{t-m+1}, \ldots, Z_{t+m+\ell}]^T \). Define the symbol neighborhood \( A_t \) containing the following \( 2(m+\ell)+1 \) input symbols
\[ A_t \triangleq [A_{t-\ell}, A_{t-\ell+1}, \ldots, A_{t+m+\ell}]^T. \]
(2)
Both \( A_t \) and \( Z_t \) are depicted in Figure 2. Let \( h_t \) denote the following length-(2\( m+\ell \)+1) vector
\[ h_t \triangleq [0, 0, \ldots, 0, h_0, h_1, \ldots, h_\ell, 0, 0, \ldots, 0]^T, \]
(3)
where \( i \) can take values \( |i| \leq m \). Let 0 denote an all-zeros vector \( 0 \triangleq [0, 0, \ldots, 0]^T \). Let both \( H \) and \( T \) denote the size \( 2m+\ell+1 \) by \( 2(m+\ell)+1 \) matrices given as
\[ H \triangleq [0, 0, \ldots, 0, h_{-m}, h_{-m+1}, \ldots, h_m, 0, 0, \ldots, 0]^\ell, \]
\[ T \triangleq [T_1, 0, 0, \ldots, 0, T_2], \]
(4)
where the two submatrices \( T_1 \) and \( T_2 \) equal
\[
\begin{bmatrix}
  h_\ell & h_{\ell-1} & \cdots & h_1 \\
  h_\ell & \ddots & \ddots & \vdots \\
  \vdots & \ddots & \ddots & \vdots \\
  h_\ell & \cdots & h_0 & 0
\end{bmatrix}
\]
\[
\begin{bmatrix}
  h_0 \\
  \vdots \\
  h_{\ell-2} & h_0 \\
  h_{\ell-1} & h_0
\end{bmatrix}
\]
Using (4), rewrite \( Z_t \triangleq [Z_{t-m}, Z_{t-m+1}, \ldots, Z_{t+m+\ell}]^T \) using (1) into the following form
\[ Z_t = (H+T)A_t - W_t, \]
(5)
where here \( W_t \) denotes the neighborhood of noise samples
\[ W_t \triangleq [W_{t-m}, W_{t-m+1}, \ldots, W_{t+m+\ell}]^T. \]
(6)

**Definition 2.** Denote the set \( M \) that contains the \( m \)-truncated MLM candidate sequences
\[ M \triangleq \{ a \in \{-1, 1\}^{2(m+\ell)+1} : a_i = 1 \text{ for all } |i| > m \}. \]
(7)
Each candidate \( a \in M \) has the following form
\[ a = [1, 1, \ldots, 1, a_{-m}, a_{-m+1}, \ldots, a_m, 1, \ldots, 1]^T, \]
i.e. candidates \( a \in M \) have boundary symbols equal to 1.

An example of a candidate sequence in the set \( M \) is illustrated in Figure 2. The boundary symbols of the candidates \( a \in M \) are fixed, because the boundary symbols of the transmitted sequence \( A_t \) are unknown to the detector. The start/end states of \( A_t \) (colored black), is shown (see Figure 2) to be different from the start/end states of the candidate \( a \in M \) (colored white).

Let the following sequence \( \cdots, B_{-2}, B_{-1}, B_0, B_1, B_2, \cdots \) denote symbol decisions on the channel inputs \( \cdots, A_{-2}, A_{-1}, A_0, A_1, A_2, \cdots \). Let \( \mathcal{I} \) denote the all-ones vector \( \mathcal{I} \triangleq [1, 1, \ldots, 1]^T \). In the following let \( |a| \) denote the Euclidean norm of the vector \( a \).

**Definition 3.** The symbol decision \( B_t \) on channel input \( A_t \), is obtained by i) computing the sequence \( B^{[t]} \) that achieves the following minimum
\[ B^{[t]} \triangleq \arg \min_{a \in M} |Z_t - (H+T)a|^2, \]
\[ = \arg \min_{a \in M} |Z_t - T \mathcal{I} - Ha|^2, \]
(8)
and ii) setting the symbol decision \( B_t \) to the 0-th component of \( B^{[t]} \) in (8), i.e. set \( B_t \triangleq B_{-t}^{[0]} \) where the sequence \( B^{[t]} = [1, B_{-m}^{[t]}, B_{-m+1}^{[t]}, \ldots, B_{-1}^{[t]}, B_0^{[t]}, B_1^{[t]}, \ldots, B_m^{[t]}, \mathcal{I}]^T \).

The sequence \( B^{[t]} \) in (8), and therefore the symbol decision \( B_t \), is obtained by considering the candidate sequences in the set \( M \), recall Definition 2 and refer to Figure 2. Note that \( B^{[t]} \) does not equal the MLM bit detection sequence \( \cdots, B_{-2}, B_{-1}, B_0, B_1, B_2, \cdots \); only the \( t \)-th symbol \( B_t \) is obtained from \( B^{[t]} \). To obtain \( B^{[t]} \), we compare the squared Euclidean distances of each candidate \( Ha \) from the received neighborhood \( Z_t - T \mathcal{I} \).

In addition to computing hard, i.e., \( \{-1, 1\} \), symbol decisions \( \cdots, B_{-2}, B_{-1}, B_0, B_1, B_2, \cdots \), the \( m \)-truncated MLM also computes the symbol reliability sequence, to be denoted as \( \cdots, R_{-2}, R_{-1}, R_0, R_1, R_2, \cdots \). Consider the following

---

1To obtain neater expressions in the sequel, the Gaussian noise sample \( W_1 \) in (1) is subtracted. This differs from convention where \( W_1 \) is typically added [1]. Note there is no loss in generality when subtracting, because the Gaussian distribution is symmetric about its mean.

2Alternatively, the boundary symbols can be specified to be any sequence of choice in the set \( \{-1, 1\} \); here we choose the boundary sequence \([1, 1, \ldots, 1] = \mathcal{I} \) simply for clearer exposition.
log-likelihood approximation (see [8])

\[
\log \frac{Pr \{A_t = B_t | Z_t\}}{Pr \{A_t \neq B_t | Z_t\}} = \log \frac{\sum_{a \in M: a_0 = B_t} Pr \{Z_t | A_t = a\}}{\sum_{a \in M: a_0 \neq B_t} Pr \{Z_t | A_t = a\}} \\
\approx \min_{a \in M} \frac{1}{2\sigma^2} |Z_t - T^a - Ha|^2 \\
- \min_{a \in M} \frac{1}{2\sigma^2} |Z_t - T^a - Ha|^2,
\]

where the first equality assumes uniform signal priors, i.e. \(Pr \{A_t = a\} = 2^{-2(2m+\ell)} - 1\), see (2). We also denote \(\sigma^2\) as the worst-case noise variance

\[
\sigma^2 \triangleq \sup_{a \in M} E\{W_t^2\}. \tag{10}
\]

We assume that \(\sigma^2\) is bounded, i.e. \(\sigma^2 < \infty\). We want to set the \((m,\ell)-truncated MLM\) reliability \(R_t\), to equal the log-likelihood approximation (9); before formally stating the expression for \(R_t\), we first make another definition. Denote the difference in the obtained squared Euclidean distances

\[
\Delta(a, \bar{a}) \triangleq \Delta(a, \bar{a}; Z_t) \\
\triangleq |Z_t - T^a - Ha|^2 - |Z_t - T^a - Ha|^2, \tag{11}
\]

where both \(a\) and \(\bar{a}\) are arbitrary sequences in \([-1,1]^{2(2m+\ell)+1}\). Recalling (8), we write \(R_t\) as follows.

**Definition 4.** The non-negative \((m,\ell)-truncated MLM\) reliability \(R_t\) is defined as

\[
R_t \triangleq \min_{a \in M} \frac{1}{2\sigma^2} \Delta(a, B^{[i]}), \tag{12}
\]

where \(\Delta(a, B^{[i]}) \geq 0\), is the difference in the obtained squared Euclidean distances corresponding to candidates \(a, B^{[i]} \in M\), and \(\sigma^2\) is the noise variance (10).

The relaxation of this assumption is discussed in the latter-half of the upcoming Subsection III-C, where we allow some of the probabilities \(Pr \{A_t = a\}\) to equal zero, i.e. in the case of modulation coding. We also comment on non-uniform signal priors in the upcoming Remark 4.

Note that \(\Delta(a, B^{[i]}) \geq 0\) for all \(a \in M\), simply because \(B^{[i]}\) achieves the minimum squared Euclidean distance amongst all candidates in \(M\), see (8).

### III. Key Observation and Statement of Main Result

This section contains three subsections. In the first subsection, we describe an important key observation; the main result of this paper is derived based on this observation. In the second subsection, we state the main result and give closed-form expressions for i) the joint reliability distribution \(F_{R_{t_1}, R_{t_2}, \ldots, R_{t_n}}(r_1, r_2, \ldots, r_n)\), and ii) the joint symbol error probability \(Pr \{\bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\}\}\). The result holds for any number \(n\) of arbitrarily chosen time instants \(t_1, t_2, \ldots, t_n\). Also, in the second subsection, a Monte-Carlo based procedure that evaluates these closed-form expressions is also given. In the third subsection, we address two important points regarding the given Monte-Carlo procedure, namely i) how to efficiently implement this procedure, and ii) how this procedure may be modified when one wishes to only consider a subset \(\mathcal{M} \subset M\) of the candidates \(M\) (recall Definition 2).

#### A. Key observation

For all times \(t\), define the following two random variables \(X_t\) and \(Y_t\) as

\[
X_t \triangleq \max_{a \in M} \frac{1}{4} \Delta(A_t, a), \\
Y_t \triangleq \max_{a \in M} \frac{1}{4} \Delta(A_t, a) \geq 0, \tag{13}
\]

where \(\Delta(A_t, a)\) is the difference in obtained squared Euclidean distances, corresponding to the transmitted sequence \(A_t\) and a candidate \(a \in M\), see (11). Note that the random variable \(Y_t\) satisfies \(Y_t \geq 0\), because there must exist a candidate \(a \in M\) that satisfies \(\Delta(A_t, a) = 0\), see (11); this particular candidate \(a \in M\) satisfies \(a_i = A_{t+i}\) for all values of \(i\) satisfying \(|i| \leq m\).
Proposition 1 (Key Observation). The \( m \)-truncated MLM reliability \( R_t \) in (12) satisfies

\[
R_t = \frac{2}{\sigma^2} |X_t - Y_t|, \quad (14)
\]

where both random variables \( X_t \) and \( Y_t \) are given in (13).

Proof: Scale (12) by \( \sigma^2/2 \) and write

\[
\frac{\sigma^2}{2} : R_t = \min_{a \in M : a_0 \neq B_t} \frac{\Delta(a, A_t)}{4} + \frac{\Delta(A_t, B^t)}{4},
\]

\[
= \left( -\max_{a \in M : a_0 = A_t} \frac{\Delta(a, A_t)}{4} \right) + \frac{\Delta(A_t, B^t)}{4}. \quad (15)
\]

To obtain the last equality in (15), we used the relationship \( \Delta(A_t, a) = -\Delta(a, A_t) \), see (11). Recall Definition 3 which states the symbol decision \( B_t \). Because \( B_t \) is either \( -1 \) or \( 1 \), we have either \( B_t \neq A_t \) or \( B_t = A_t \). Consider the former case \( B_t \neq A_t \), in which (15) reduces to

\[
\frac{\sigma^2}{2} : R_t = \left( -\max_{a \in M : a_0 = A_t} \frac{\Delta(a, A_t)}{4} \right) + \frac{\Delta(A_t, B^t)}{4},
\]

where the second equality follows from (13), and the third from the fact \( R_t \geq 0 \), see Definition 4. We have thus shown (14) for the case \( B_t \neq A_t \). The same conclusion follows for the other case \( B_t = A_t \) in similar manner.

Note that the expression (14) for \( R_t \) in Proposition 1, cannot be computed in practice; it is derived purely for analysis purposes. This is (14) relies on the ability to compute \( X_t \) and \( Y_t \), which in turn requires knowledge of the transmitted sequence \( A_t \). Clearly, it is absurd to assume that the detector knows \( A_t \).

Remark 1. From past literature (e.g. [11]), there seems to be a misconception that the reliability \( R_t \) must be expressed in terms of \( B^t \) (as in (12)). However as shown in Proposition 1, this is not true. The reliability \( R_t \) can be simply written as \( R_t = 2/\sigma^2 \cdot |X_t - Y_t| \), where we see from (13) that both \( X_t \) and \( Y_t \) depend only on the transmitted sequence \( A_t \). In other words, the reliability \( R_t \) can be alternatively computed using (14), which does not require any knowledge of \( B^t \).

As mentioned before, the key observation Proposition 1 will be used to prove the main result. However before going into detailed derivations, we would like to first state the main result. This will be done in the next subsection; we believe that by doing so this will better motivate the significance of this work.

B. Statement of main result

For any \( n \) number of arbitrarily chosen time instants \( t_1, t_2, \ldots, t_n \), we wish to obtain the distribution of the vector \( R_{t_1}^n \), containing the following reliabilities

\[
R_{t_1}^n \triangleq [R_{t_1}, R_{t_2}, \ldots, R_{t_n}]^T. \quad (16)
\]
Definition 8. Let the square matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_t^T)$ of size $2mn$ by $2mn$ satisfy the following two conditions:

i) the matrix $\mathbf{Q}$ decomposes the following size $2mn$ matrix

$$Q\Lambda^2Q^T = \text{diag} \left( \mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \ldots, \mathbf{G}(\mathbf{A}_{t_n}) \right)^T \mathbf{K}_W \cdot \text{diag} \left( \mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \ldots, \mathbf{G}(\mathbf{A}_{t_n}) \right),$$

where $\Lambda = \Lambda(\mathbf{A}_t^T)$ on the l.h.s. of (25) is a diagonal matrix. The number of positive diagonal elements in the matrix $\Lambda$, equals the rank of the matrix on the r.h.s. of (25).

ii) the matrix $\mathbf{Q}$ diagonalizes the matrix $\mathbf{I}_n \otimes \mathbf{S}^T$, i.e. the matrix $\mathbf{Q}$ satisfies

$$Q^T(\mathbf{I}_n \otimes \mathbf{S}^T)Q = \mathbf{I},$$

noting that the matrix $\mathbf{S}^T$ is square of size $2m$.

It is shown in Appendix A how to compute such a matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_t^T)$, and also obtain the diagonal matrix $\Lambda = \Lambda(\mathbf{A}_t^T)$ in (25). We partition the matrix $\mathbf{Q}$ into $n$ partitions of equal size $2m$ by $2mn$, i.e.,

$$\mathbf{Q} = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_n \end{bmatrix}.$$  

(27)

Let $\text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n})$ denote the diagonal matrix, whose diagonal equals $[\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}]^T$. Define the size $n$ by $2mn$ matrix $\mathbf{F}(\mathbf{A}_t^T)$ as

$$\mathbf{F}(\mathbf{A}_t^T) \triangleq \text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}) \otimes \mathbf{h}_0^T \mathbf{K}_W$$

where $\mathbf{h}_0$ is given in (3), and $\Lambda^\dagger$ is formed by reciprocating only the non-zero diagonal elements of $\Lambda$. Define the following length-$2^{2m}$ vectors $\bm{\mu}(\mathbf{A}_t)$ and $\bm{\nu}(\mathbf{A}_t)$ as

$$\bm{\mu}(\mathbf{A}_t) = \left[ \mu_1, \mu_2, \ldots, \mu_{2^{2m-1}} \right]^T$$

$$= \left[ \mathbf{G}(\mathbf{A}_t)\mathbf{S}^T \cdot \mathbf{T} \cdot (\mathbf{I} - \mathbf{A}_t) \right. - \left[ |\mathbf{G}(\mathbf{A}_t)\mathbf{s}_0|^2, |\mathbf{G}(\mathbf{A}_t)\mathbf{s}_1|^2, \ldots, |\mathbf{G}(\mathbf{A}_t)\mathbf{s}_{2^{2m-1}}|^2 \right] \mathbf{T}^T,$$

$$\bm{\nu}(\mathbf{A}_t) = \left[ \nu_1, \nu_2, \ldots, \nu_{2^{2m-1}} \right]^T$$

$$\triangleq \bm{\mu}(\mathbf{A}_t) - 2\mathbf{A}_t \cdot \mathbf{h}_0^T \mathbf{G}(\mathbf{A}_t)\mathbf{S},$$

where $\mu_k = \mu_k(\mathbf{A}_t)$ and $\nu_k = \nu_k(\mathbf{A}_t)$ denote the $k$-th components of $\bm{\mu}(\mathbf{A}_t)$ and $\bm{\nu}(\mathbf{A}_t)$ respectively, and $\mathbf{T}$ is given in (4). Let $\Phi_\mathbf{K}(\mathbf{r})$ denote the distribution function of a zero-mean Gaussian random vector with covariance matrix $\mathbf{K}$. Finally define the following length-$n$ random vectors

$$\mathbf{X}_{t_1} \triangleq [X_{t_1}, X_{t_2}, \ldots, X_{t_n}]^T,$$

$$\mathbf{Y}_{t_1} \triangleq [Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}]^T,$$

where both $X_{t_i}$ and $Y_{t_i}$ are given in (13). Let $\mathbb{R}$ denote the set of real numbers. We are now ready to state the main result.

**Theorem 1.** The distribution of $\mathbf{X}_{t_1} - \mathbf{Y}_{t_1}$ equals

$$\mathbf{F}\mathbf{X}_{t_1} - \mathbf{F}\mathbf{Y}_{t_1}(\mathbf{r}) = \mathbb{E} \left\{ \Phi_{\mathbf{K}_W(\mathbf{A}_t^T)} \left( \mathbf{r} + \delta(\mathbf{U}, \mathbf{A}_t^T) - \eta(\mathbf{U}, \mathbf{A}_t^T) \right) \right\}$$

(32)

for all $\mathbf{r} \in \mathbb{R}^n$, where the following random vectors and matrices appear in (32)

- $\mathbf{U}$ is a standard zero-mean identity-covariance Gaussian random vector of length-$2mn$.
- $\delta(\mathbf{U}, \mathbf{A}_t^T) = [\delta_1, \delta_2, \ldots, \delta_n]^T$ is a length-$n$ vector in $\mathbb{R}^n$, where

$$\delta_i = \delta_i(\mathbf{U}, \mathbf{A}_t^T) \triangleq \max(S^T\mathbf{Q}\mathbf{A}_t^T + \mathbf{u}(\mathbf{A}_t)) - \max(S^T\mathbf{Q}\mathbf{A}_t^T + \mathbf{v}(\mathbf{A}_t)).$$

(33)

- $\eta(\mathbf{U}, \mathbf{A}_t^T) = [\eta_1, \eta_2, \ldots, \eta_n]^T$ is a length-$n$ vector in $\mathbb{R}^n$, where

$$\eta(\mathbf{U}, \mathbf{A}_t^T) \triangleq \text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}) \cdot \left( \mathbf{1} \cdot \mathbf{1}^T - [\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}]^T \mathbf{h}_0 \right)$$

$$- \left| \mathbf{h}_0 \right|^2 \mathbf{1} + \mathbf{F}(\mathbf{A}_t^T)\mathbf{U}.$$  

(34)

- $\mathbf{K}_W(\mathbf{A}_t^T)$ is the $n$ by $n$ matrix

$$\mathbf{K}_W(\mathbf{A}_t^T) = \text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}) \otimes \mathbf{h}_0^T \mathbf{K}_W$$

$$- \text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}) \otimes \mathbf{h}_0$$

$$- \mathbf{F}(\mathbf{A}_t^T)\mathbf{F}(\mathbf{A}_t^T)^T.$$  

(35)

A., Refer to (3), (19), (25), (27), (28), (29) and (30) for clarifications of the notation used above.

The proof of Theorem 1 is given in Subsection IV-A. Both i) the joint distribution of the reliabilities $\mathbf{R}_{t_1} \triangleq [R_{t_1}, R_{t_1}, \ldots, R_{t_n}]^T$ in (16), and ii) the joint error probability $\Pr \left\{ \bigcap_{i=1}^n \{ B_{t_i} \neq A_{t_i} \} \right\}$, follow as corollaries from our main result Theorem 1. In the following we denote an index subset $\{ \tau_1, \tau_2, \ldots, \tau_j \} \subseteq \{ 1, 2, \ldots, n \}$ of size $J$, written compactly in vector form as $\mathbf{r}_j^T = [\tau_1, \tau_2, \ldots, \tau_j]^T$.

**Corollary 1.** The distribution of $\mathbf{R}_{t_1} \triangleq 2/\alpha^2 \cdot |\mathbf{X}_{t_1} - \mathbf{Y}_{t_1}|$, see Proposition 1, is given as

$$\mathbb{F}(\mathbf{r}) = \mathbb{F}\mathbf{X}_{t_1} - \mathbf{Y}_{t_1}(\mathbf{r})$$

$$\triangleq \sum_{j=0}^n \left( -1 \right)^j \cdot \mathbb{F}\mathbf{X}_{t_1} - \mathbf{Y}_{t_1}(\mathbf{r})$$

$$\left( \alpha^2/2 \cdot \bm{\alpha}(\mathbf{r}, \mathbf{r}) \right),$$

where the length-$n$ vector $\bm{\alpha}(\mathbf{r}, \mathbf{r}) = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$ satisfies

$$\alpha_i = \alpha_i(\mathbf{r}, \mathbf{r}) \equiv \left\{ \begin{array}{ll} -r_i & \text{if } i \in \{ \tau_1, \tau_2, \ldots, \tau_j \}, \\ r_i & \text{otherwise,} \end{array} \right.$$
Corollary 2. The probability \( \Pr \{ \bigcap_{i=1}^{n} \{ B_i \neq A_i \} \} \) that all symbol decisions \( B_1, B_2, \ldots, B_n \) are in error, equals
\[
\Pr \left\{ \bigcap_{i=1}^{n} \{ B_i \neq A_i \} \right\} = \Pr \{ X_{t_i} \geq Y_{t_i} \} = 1 + \sum_{j=1}^{n} \sum_{\{t_1, t_2, \ldots, t_j\} \subseteq \{t_1, t_2, \ldots, t_n\}} (-1)^j \cdot F_{X_{t_j} - Y_{t_j}}(0),
\]
where the probability
\[
F_{X_{t_j} - Y_{t_j}}(0) = \Pr \left\{ \bigcap_{\tau \in \{t_1, t_2, \ldots, t_j\}} \{ X_{\tau} - Y_{\tau} \leq 0 \} \right\}
\]
has the similar closed form as in Theorem 1.

Proof: From (13) we clearly see that the event \( \{ X_t \geq Y_t \} \) indicates that the sequence \( B[0] \) in (8) will have its 0-th component \( B[0] \neq A_1 \). Because the symbol decision \( B_1 \) is set to \( B_1 = B[0] \), see Definition 3, the event \( \{ X_t \geq Y_t \} \) indicates that \( B_1 \neq A_1 \), which is exactly a symbol decision error occurring at time \( t \).

Denote the realizations of \( A_{t_i}, \ A_t \), and \( U \), as \( A_{t_i} = a_{t_i} \) and \( A_t = a_t \), and \( U = u \). The Monte-Carlo procedure used to evaluate the closed-form of \( F_{X_{t_i} - Y_{t_i}}(r) \) in Theorem 1, is given in Procedure 1. The following Remarks 2-5 pertain to Procedure 1.

Remark 2. We may reduce the number of computations used to obtain matrices \( Q = Q(A_{t_i}) \) and \( \Lambda = \Lambda(A_{t_i}) \) in Line 3, by selecting \( U = u \) multiple times for a fixed \( A_{t_i} = a_{t_i} \).

Remark 3. The matrix \( K_{W}(a_{t_i}) \) computed in Line 5 (also see (35)) may not have full rank. Hence when evaluating the Gaussian distribution function \( \Phi_{K_{W}(a_{t_i})}(r) \) with covariance matrix \( K_{W}(a_{t_i}) \) in Line 6, we may require techniques designed for rank deficient covariances, see for example [13].

Remark 4. Our proposed method requires no assumptions on the noise covariance matrix \( K_W \) in (22), and can be applied even when the noise \( W_t \) is correlated and/or non-stationary. Also at the end of this subsection, we present a modification of the previous Procedure 1, which addresses certain cases where we do not want to consider all candidates in \( M \) (see Definition 2), i.e. \( \Pr \{ A_{t_i} = a_{t_i} \} = 0 \) for some \( a_{t_i} \). This particular situation arises, for example, when we have a modulation code (see [14, 15]) present in the system.

Here we always assume that \( A_{t_i} \) is equally-likely amongst all its realizations \( A_{t_i} = a_{t_i} \). Further modifications will be required to extend our method to the general case of non-uniform priors \( \Pr \{ A_{t_i} = a_{t_i} \} \) (the first equality of (9) is not valid for such cases).

Remark 5. Because we have that
\[
0 \leq \Phi_{K_{W}(A_{t_i})}(r + \delta(U, A_{t_i}) - \eta(U, A_{t_i})) \leq 1,
\]
the well-known Hoeffding probability inequalities can be applied to obtain convergence guarantees, see [16].

The main thrust of the next subsection is to address Line 4 of Procedure 1. It appears that to execute Line 4 of Procedure 1, we require an exhaustive search over an exponential \( 2^{2m_n} \) number of terms, in order to perform the two maximizations. However, we point out in the next subsection, that these maximizations can be performed more efficiently by utilizing dynamic programming optimization techniques. Also in the next subsection, we address the computation of \( F_{X_{t_i} - Y_{t_i}}(r) \), in instances where one wishes to only consider a subset \( \mathcal{M} \subset \mathcal{M} \) of the candidates \( \mathcal{M} \) (see Definition 2).
C. On computing the closed-form of $FX_n^v - Y_{1,v}^r(r)$ using Procedure 1

To compute $\delta_t$ in (33) while executing Line 4 of Procedure 1, we need to perform the following two maximizations

$$
\max_{s \in \{0,1\}^{2m}} s^T Q_\tau A u + [G(a)s]^T \cdot [T(1-a) - |G(a)s|^2, \quad (36)

$$

where both $a$ and $u$ are realizations $A_{t_i} = a$ and $U = u$. Note that we obtain (36) from (33), by substituting for both $\mu(a)$ and $\nu(a)$ using (29) and (30) respectively. Index the realization $A_{t_i}$ as $a$ similarly as in Definition 2

$$
a \triangleq [a_{-m-\ell}, a_{-m-\ell+1}, \ldots, a_{m+\ell}]^T.
$$

Let $\text{diag}(a)$ denote the diagonal matrix, with diagonal $a$.

The matrix $G(a)$ appearing in both maximization problems (36), has a distinctive structure. We now proceed to clarify by exploiting this structure.

**Definition 9.** Let $g_\tau$ denote the $2(m+\ell)+1$ vector

$$
g_\tau \triangleq m+\ell\begin{bmatrix} 0, 0, \ldots, 0, h_{\ell}a_{\tau-\ell}, h_{\ell-1}a_{\tau-(\ell-1)}, \ldots, h_0a_\tau, 0, 0, \ldots, 0 \end{bmatrix}^T,
$$

where $\tau$ can take values $\tau \in \{-m, -m+1, \ldots, m+\ell\}$.

Using the $2m+\ell+1$ vectors $g_\tau$, we rewrite $G(a)$ as

$$
G(a) \triangleq H \text{diag}(a)E = \begin{bmatrix} g_{-m}^T \\ g_{-m+1}^T \\ \vdots \\ g_{m+\ell}^T \end{bmatrix} E, \quad (37)
$$

recall the definition of $G(a)$ from (20). From the observed structure of $g_\tau$, it can be clearly seen from (37) that $G(a)$ is a sparse matrix with many zero entries. The matrix $G(a)$ is an $(\ell + 1)$-banded matrix, see [17], p. 16. As it is well-known in the literature on ISI channels, it is efficient to employ dynamic programming techniques to solve both problems (36), by exploiting this $(\ell + 1)$-banded sparsity [2].

It is clear that the inner product $g_\tau^T e_j$ extracts the $j$-th component of the vector $g_\tau^T$, i.e.

$$
g_\tau^T e_j = \begin{cases} h_j \cdot a_{\tau-j} & \text{if } 0 \leq j \leq \ell, \\ 0 & \text{otherwise}, \end{cases}
$$

where $j$ satisfies $|j| \leq m+\ell$. Both problems (36) are optimized over all $s \in \{0,1\}^{2m}$, we index

$$
s \triangleq [s_{-m}, s_{-m+1}, \ldots, s_{-1}, s_1, s_2, \ldots, s_m]^T.
$$

It is clear that by using (38), the following is true for all vectors $g_{\tau}^T$ given in Definition 9

$$
g_{\tau}^T E s = \sum_{j=0}^{m+\ell} (g_{\tau}^T e_j) \cdot s_j = \sum_{j=0}^{m+\ell} h_j \cdot a_{\tau-j} \cdot s_{\tau-j}, \quad (39)
$$

if we set $s_0 = 0$ and $s_\tau = 0$ for all $|\tau| > m$.

Define the length-$(2m)$ vector $C \triangleq [C_{-m}, C_{-m+1}, \ldots, C_{-1}, C_1, C_2, \ldots, C_m]^T$. Set $C_0 := -\infty$ and $C_{\tau} := 0$ for all $|\tau| > m$. By setting

$$
C := Q_\tau A u + [G(a)]^T \cdot [T(1-a) - 2a_0 \cdot h_0],
$$

and

$$
C := Q_\tau A u + [G(a)]^T \cdot [T(1-a) - 2a_0 \cdot h_0],
$$

respectively, we can solve both problems (36) as

$$
\max_{s \in \{0,1\}^{2m}} s^T C - |G(a)s|^2
$$

$$
= \max_{s \in \{0,1\}^{2m}} \sum_{\tau=-m}^{m+\ell} C_{\tau} \cdot s_{\tau} - (g_{\tau}^T E s)^2, \quad (40)
$$

where the $\tau$-th term $g_{\tau}^T E s = \sum_{j=0}^{m+\ell} h_j a_{\tau-j} s_{\tau-j}$. For the sake of completeness, we shall state the dynamic programming procedure that solves (40).
Definition 10. The dynamic programming state at time $\tau$ equals the length-$\ell$ vector of binary symbols $[s_{\tau-\ell+1}, s_{\tau-\ell+2}, \cdots, s_{\tau}]^T \in \{0, 1\}^\ell$.

For the benefit of readers knowledgeable in dynamic programming techniques, we illustrate the time evolution of the dynamic programming states in Figure 3. Dynamic programs can be solved with complexity that is linear in the state size [2]; in our case we have $2^{2t}$ states. The dynamic programming procedure optimizing (40) is given in Procedure 2.

The second part of this subsection addresses the following separate issue. Recall from Remark 4 that Theorem 1 requires no assumptions on the distribution $\Pr \{A_{t_{\tau}} = a_{t_{\tau}}\}$. In other words, the distribution $\Pr \{A_t = a\}$ for each time $t$ can be arbitrary specified. One may particularly want to consider certain cases, where some of the probabilities $\Pr \{A_t = a\}$ are equal 0; one example of such a case is where a modulation code is present in the system [14], [15]. In these cases we would not want to consider candidates in the set $M$ (see Definition 2) that have zero probability of occurrence. We would consider the subset $\tilde{M} \subset M$, explicitly written as

$$\tilde{M} = \tilde{M}_t \triangleq \left\{ a \in M : \Pr \left\{ \bigcap_{j=-m}^{m} \{A_{t_{j}} = a_{j}\} \right\} = 0 \right\}$$

for each time instant $t$.

If we consider the subsets $\tilde{M} \subset M$ then Procedure 1 has to be modified. The modification of Procedure 1 is given as Procedure 3; this modification will be justified in the upcoming Section IV.

Remark 6. Line 4 of Procedure 3 may also be efficiently solved using dynamic programming techniques.

Thus far, we have completed the statement of our main result Theorem 1 and the two main Corollaries 1 and 2. We have given Procedures 1-3 (also see Appendix A), used to efficiently evaluate the given closed-form expressions. The rest of this paper is organized as follows. In the following Section IV, we shall prove the correctness of both Theorem 1, and also Procedure 3. A simple upper bound on the rank of $K_W(A_{t_{\tau}})$ in (35) will also be given. In Section V, numerical computations will be presented for various commonly-cited ISI channels in magnetic recording literature [18]. The computations are performed for various scenarios, so that we may demonstrate a range of applications of our results. We conclude in Section VI.

IV. DISTRIBUTION OF $X_{t_{\tau}} - Y_{t_{\tau}}$ AND RELIABILITY

$R_{t_{\tau}} = 2/\sigma^2 \cdot |X_{t_{\tau}} - Y_{t_{\tau}}|$

A. Proof of Theorem 1

We begin by showing the correctness of Theorem 1, which was stated in the previous section. Define the random variable

$$V_t \triangleq A_t \cdot h_0^T W_t.$$

(42)

It is easy to verify that $V_t$ is Gaussian: recall that $W_t \triangleq [W_{t-M}, W_{t-M+1}, \cdots, W_{t+M+1}]^T$ is the neighborhood of

(Gaussian) noise samples. To improve clarity, we shall introduce the following new notation, both used only in this section

$$\theta(A_t) \triangleq A_t \cdot [T(1 - A_t)]^T h_0 - |h_0|^2,$$

$$\Gamma = \Gamma(A_{t_{\tau}}) \triangleq \text{diag}(G(A_{t_{1}}), G(A_{t_{2}}), \cdots, G(A_{t_{n}})).$$

(43)

Recall that $I_n$ denotes a size $n$ identity matrix, and that $\otimes$ denotes the matrix Kronecker product. Using (43), we may now more compactly write

$$QA^2Q^T = \Gamma^T K_W \Gamma,$$

$$F(A_{t_{\tau}}) = \text{diag}(A_{t_{1}}, A_{t_{2}}, \cdots, A_{t_{n}}) \otimes h_0^T K_W \Gamma$$

$$[I_n \otimes SS^T] \cdot QA^1,$$

$$\eta(U, A_{t_{\tau}}) = [\theta(A_{t_{1}}), \theta(A_{t_{2}}), \cdots, \theta(A_{t_{n}})]^T + F(A_{t_{\tau}}) U,$$

(44)

where (recall that) matrices $Q = Q(A_{t_{\tau}})$ and $A = A(A_{t_{\tau}})$ are given in Definition 8, matrix $F(A_{t_{\tau}})$ in (28), and $\eta(U, A_{t_{\tau}})$ in (34).

Proposition 2. The random variables $X_t$ and $Y_t$ in (13) can be written as

$$X_t = \max \left( [G(A_t)^T W_t + \nu(A_t)] \cdot \bar{1} \right),$$

$$Y_t = \max \left( [G(A_t)^T W_t + \mu(A_t)] \right),$$

where $\theta(A_t) \triangleq A_t \cdot [T(1 - A_t)]^T h_0 - |h_0|^2$ as given in (43).

Proof: We expand $\Delta(A_i, a)$ in (11) by substituting for
$\mathbf{Z}_t$ using (5) to get

$$
\Delta(\mathbf{A}_t, a) = |\mathbf{Z}_t - \mathbf{T} - \mathbf{H} \mathbf{A}_t|^2 - |\mathbf{Z}_t - \mathbf{T} - \mathbf{H} a|^2
$$

$$
= - \mathbf{W}_t + \mathbf{T}(\mathbf{A}_t - \mathbf{I})^2 + |\mathbf{Z}_t - \mathbf{H} (\mathbf{A}_t - a)|^2
$$

$$
= -2 |\mathbf{W}_t + \mathbf{T}(\mathbf{A}_t - \mathbf{I})^T \mathbf{H} (\mathbf{A}_t - a) - |\mathbf{H} (\mathbf{A}_t - a)|^2.
$$

We substitute (45) into the definition of $X_t$ and $Y_t$ in (13) to obtain

$$
X_t = \max_{a \in \mathcal{A}_t, a_0 \neq A_t} \left[ |\mathbf{W}_t + \mathbf{T}(\mathbf{I} - \mathbf{A}_t)\mathbf{H} (\mathbf{A}_t - a)|^2 - \frac{1}{2} |\mathbf{H} (\mathbf{A}_t - a)|^2 \right].
$$

Using (17) and Definitions 2, 5 and 6, we establish the following equality of sets

$$
\left\{ \frac{1}{2} (A_t - a) : a \in \mathcal{M}, a_0 \neq A_t \right\}
$$

$$
= \left\{ \text{diag}(\mathbf{A}_t) \mathbf{E}_{j} + A_t \cdot \mathbf{E}_{j} : 0 \leq j \leq 2^{m_2} - 1 \right\},
$$

$$
\left\{ \frac{1}{2} (A_t - a) : a \in \mathcal{M}, a_0 = A_t \right\}
$$

$$
= \left\{ \text{diag}(\mathbf{A}_t) \mathbf{E}_{j} : 0 \leq j \leq 2^{m_2} - 1 \right\}.
$$

Next, we utilize both (46) and (20) to rewrite (45) as

$$
X_t = \max_{j \in \{0, 1, \ldots, 2^{m_2} - 1\}} \left[ |\mathbf{W}_t + \mathbf{T}(\mathbf{I} - \mathbf{A}_t)\mathbf{H} (\mathbf{A}_t - a)|^2 - |\mathbf{G}(\mathbf{A}_t) \mathbf{s}_{j} + A_t \mathbf{h}_0|^2 \right],
$$

$$
Y_t = \max_{j \in \{0, 1, \ldots, 2^{m_2} - 1\}} \left[ |\mathbf{W}_t + \mathbf{T}(\mathbf{I} - \mathbf{A}_t)\mathbf{H} (\mathbf{A}_t - a)|^2 - |\mathbf{G}(\mathbf{A}_t) \mathbf{s}_{j}|^2 \right].
$$

By the definition of $\mathbf{\mu}(\mathbf{A}_t)$ in (29) and $\mathbf{S}$ in Definition 6, the expression for $Y_t$ in the proposition statement follows from (48). For $X_t$, we continue to expand (48) to get

$$
X_t = \max_{\mathbf{\mu}(\mathbf{A}_t)} \left( \mathbf{G}(\mathbf{A}_t) \mathbf{S}^T \mathbf{W}_t + \mathbf{\mu}(\mathbf{A}_t) - 2 \mathbf{\lambda} \cdot \mathbf{h}_0^T \mathbf{G}(\mathbf{A}_t) \mathbf{S} \right)
$$

$$
+ \mathbf{\lambda}^T \mathbf{h}_0^T \mathbf{W}_t \mathbf{I} + \left\{ \mathbf{\lambda} \cdot |\mathbf{W}_t + \mathbf{T}(\mathbf{I} - \mathbf{A}_t)\mathbf{H} (\mathbf{A}_t - a)|^2 \mathbf{h}_0^T - \mathbf{h}_0^2 \mathbf{I} \right\},
$$

in the same form as in the proposition statement, where $\mathbf{\nu}(\mathbf{A}_t)$ is defined in (30), and $V_t$ in (42), and $\theta(\mathbf{A}_t)$ in (43).

We now are ready to prove Theorem 1. The proof is split up into the following two separate cases:

1. rank $|\mathbf{G}(\mathbf{A}_t)|^T \mathbf{K}_\mathbf{W} \mathbf{G}(\mathbf{A}_t)| = 2mn$, and
2. rank $|\mathbf{G}(\mathbf{A}_t)|^T \mathbf{K}_\mathbf{W} \mathbf{G}(\mathbf{A}_t)| < 2mn$ for some realization $\mathbf{A}_t = \mathbf{a}_t^\dagger$.

We begin with the first case.

**Proof of Theorem 1 when rank $|\mathbf{G}(\mathbf{A}_t)|^T \mathbf{K}_\mathbf{W} \mathbf{G}(\mathbf{A}_t)| = 2mn$**

- We first derive the following equalities

$$
(\mathbf{A}^T \mathbf{Q}) (\mathbf{I}_n \otimes \mathbf{S}^T) \mathbf{G}(\mathbf{A}_t)|^T \mathbf{W}_t = \mathbf{A}^T \mathbf{Q} \mathbf{A} \mathbf{U} = \mathbf{A}^T \mathbf{A} \mathbf{U} = \mathbf{U}.
$$

The first two equalities follow by respectively applying properties i) and ii) of the matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_t^\dagger)$. The last equality holds because of the assumption rank $|\mathbf{G}(\mathbf{A}_t)|^T \mathbf{K}_\mathbf{W} \mathbf{G}(\mathbf{A}_t)| = 2mn$, in which case $\mathbf{A}^T$ is strictly an inverse of $\mathbf{A}$. Recall both $V_t \in \mathbf{A}_t$, $\mathbf{h}_0^T \mathbf{W}_t$, and $V_t^T = [V_{t_1}, V_{t_2}, \ldots, V_{t_n}]^T$. Taking (51) together with (42), we have the following transformation

$$
V_{t_1} \quad \mathbf{U} = \left[ \text{diag}(\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_n}) \otimes \mathbf{h}_0^T \right] \left( \mathbf{A}^T \mathbf{Q} \mathbf{A} \mathbf{U} \right) \mathbf{W}_t.
$$

Consider the conditional event

$$
\{ X_{t_1} - Y_{t_1} \leq r | \mathbf{A}_t, \mathbf{U} \}.
$$
where \( r = [r_1, r_2, \ldots, r_n]^T \in \mathbb{R}^n \). It is clear from both Proposition 2 and (52), that after conditioning on both \( A_{t_i}^2 \) and \( U \) in (53), the only quantity that remains random in (53) is the Gaussian vector \( V_{t_i} \). Using Lemma 1, we have the transformation

\[
S^T Q_i(A_{t_i}) A(A_{t_i}) U = [G(A_{t_i})S]^T W_{t_i},
\]

therefore we may rewrite both \( X_{t_i} \) and \( Y_{t_i} \) from Proposition 2 as

\[
\begin{align*}
X_{t_i} &= \max (S^T Q_i A U + \nu(A_{t_i})) + V_{t_i} + \theta(A_{t_i}), \\
Y_{t_i} &= \max (S^T Q_i A U + \mu(A_{t_i})). \tag{54}
\end{align*}
\]

The event (53) can then be written as

\[
\{ X_{t_i} - Y_{t_i} \leq r | A_{t_i}^2, U \} = \bigcap_{1 \leq i \leq n} \{ X_{t_i} \leq r_i + Y_{t_i} | A_{t_i}^2, U \}
\]

\[
= \bigcap_{1 \leq i \leq n} \left\{ \max \left( \frac{S^T Q_i A U + \nu(A_{t_i})}{+V_{t_i} + \theta(A_{t_i})} \right) \leq r_i + Y_{t_i} \right\} \bigcap_{1 \leq i \leq n} A_{t_i}^2, U \bigcap_{1 \leq i \leq n} V_{t_i} + \theta(A_{t_i}) \leq r_i + \max \left( \frac{S^T Q_i A U + \mu(A_{t_i})}{-\max S^T Q_i A U + \nu(A_{t_i})} \right) \bigcap_{1 \leq i \leq n} A_{t_i}^2, U.
\]

Continuing from (55), we utilize (33) to rewrite

\[
\{ X_{t_i} - Y_{t_i} \leq r | A_{t_i}^2, U \} = \bigcap_{1 \leq i \leq n} \left\{ V_{t_i} + \theta(A_{t_i}) \leq r_i + \delta(U, A_{t_i}) | A_{t_i}^2, U \right\} \tag{56}
\]

We now determine both the mean and variance of \( V_{t_i} \), after conditioning on both \( A_{t_i}^2 \) and \( U \). From (52), we derive the formula

\[
E\{V_{t_i} U^T | A_{t_i}^2\} = \text{diag}(A_{t_i}, A_{t_2}, \ldots, A_{t_n}) \otimes h_0^T K_W \cdot \Gamma(A_{t_i}^2)(I_n \otimes SS^T)QA^T \triangleq F(A_{t_i}^2), \tag{57}
\]

where \( F(A_{t_i}^2) \) is given in (28). Next, we compute the conditional mean

\[
E\{V_{t_i} | A_{t_i}^2, U\} = E\{V_{t_i} U^T | A_{t_i}^2\} U = F(A_{t_i}^2) U, \tag{58}
\]

where the second equality follows from \( E\{V_{t_i} | A_{t_i}^2\} = 0 \) (because \( W_{t_i} \) has zero mean, see (42)), and substituting (57). The conditional covariance matrix \( \text{Cov}\{V_{t_i} | A_{t_i}^2, U\} \) is obtained as follows

\[
\text{Cov}\{V_{t_i} | A_{t_i}^2, U\} = E\{V_{t_i} V_{t_i}^T | A_{t_i}^2\} - E\{V_{t_i} U^T | A_{t_i}^2\} \cdot E\{U U^T | A_{t_i}^2\} U
\]

\[
= \text{diag}(A_{t_1}, A_{t_2}, \ldots, A_{t_n}) \otimes h_0^T K_W \cdot \text{diag}(A_{t_1}, A_{t_2}, \ldots, A_{t_n}) \otimes h_0 - F(A_{t_i}^2) F(A_{t_i}^2)^T \triangleq K_V(A_{t_i}^2), \tag{59}
\]

where \( K_V(A_{t_i}^2) \) is given in (35). The expression for \( F_{X_{t_i} - Y_{t_i} | A_{t_i}^2, U} \) in Theorem 1 now follows easily from (56)

\[
\begin{align*}
\{ X_{t_i} - Y_{t_i} \leq r | A_{t_i}^2, U \} &= \{ V_{t_i} + [\theta(A_{t_1}), \theta(A_{t_2}), \ldots, \theta(A_{t_n})]^T \leq r + \delta(U, A_{t_i}) | A_{t_i}^2, U \}
\end{align*}
\]

and noticing that the random vector

\[
V_{t_i} + [\theta(A_{t_1}), \theta(A_{t_2}), \ldots, \theta(A_{t_n})]^T \tag{60}
\]

is (conditionally on \( A_{t_i}^2 \) and \( U \)) Gaussian distributed with distribution function

\[
\Phi_K(A_{t_i}^2)(r - \eta(U, A_{t_i}^2)),
\]

where both the conditional mean and covariance \( \eta(U, A_{t_i}^2) \) and \( K_V(A_{t_i}^2) \), are given respectively in (58) and (59).

Next we consider the other case where the rank of \( \Gamma(A_{t_i}^2)^T K_W \Gamma(A_{t_i}^2) < 2mn \) for some value of \( A_{t_i}^2 = a_{t_i}^2 \). In this case, the arguments of the preceding proof fail in equation (51), where the final equality does not hold because then \( A^\dagger \) is strictly not the inverse of \( A \). However as we shall see, the expression for \( F_{X_{t_i} - Y_{t_i} | A_{t_i}^2, U} \) in Theorem 1 still holds for this case.

**Proof of Theorem 1 when rank(\( \Gamma(A_{t_i}^2)^T K_W \Gamma(A_{t_i}^2) \)) < 2mn for some \( A_{t_i}^2 = a_{t_i}^2 \)**

Recall that the matrix \( [\Lambda(A_{t_i}^2)]^T = A^\dagger \) is formed by only reciprocating the non-zero diagonal elements of \( \Lambda(A_{t_i}^2) = \Lambda \). For a particular realization \( A_{t_i}^2 = a_{t_i}^2 \), let the value \( j = \text{rank}(\Gamma(A_{t_i}^2)^T K_W \Gamma(A_{t_i}^2)) \equiv \text{rank}(\Gamma(A_{t_i}^2)^T K_W \Gamma(A_{t_i}^2)) \). Consider what happens if \( j < 2mn \).

Without loss of generality, assume that all non-zero diagonal elements of \( \Lambda(A_{t_i}^2) = \Lambda \), are located at the first \( j < 2mn \) diagonal elements of \( \Lambda \). Define the following size-\( j \) quantities

- the random vector \( U_{j}^T = [U_1, U_2, \ldots, U_j]^T \), a truncated version of \( U = [U_1, U_2, \ldots, U_{2mn}]^T \).
- the size \( 2mn \) by \( j \) matrix \( Q \), containing the first \( j \) columns of the \( Q \), see Definition 8.
- the size \( j \) diagonal square matrix \( \Lambda \), containing the \( j \) positive diagonal elements of \( \Lambda \), also see Definition 8.

If we substitute the new quantities \( U_{j}^T, Q \) and \( \Lambda \) for \( U \), \( Q \) and \( \Lambda \) in equation (51), it is clear that (51) holds true, i.e.,

\[
(\Lambda^\dagger Q^T)(I_n \otimes SS^T)\Gamma(A_{t_i}^2)^T W_{t_i} = (\Lambda^\dagger Q^T)(I_n \otimes SS^T)Q\Lambda^T U_{j}^T = \Lambda^\dagger \Lambda^T U_{j}^T = U_{j}^T, \tag{61}
\]

where note from Definition 8 that it must be true that \( Q^T(I_n \otimes SS^T)Q = I_j \), here \( I_j \) is the size \( j \) identity matrix. Hence, Theorem 1 clearly holds when we substitute \( U_{j}^T, Q \) and \( \Lambda \) for \( U \), \( Q \) and \( \Lambda \).

Further, we can verify the following facts:

- \( Q\Lambda U_{j}^T = Q\Lambda U \), and therefore \( \delta(U_{j}^T, A_{t_i}^2) = \delta(U, A_{t_i}^2) \). Also,
- \( F(A_{t_i}^2) \) remains unaltered whether we use \( Q, \Lambda \) or \( Q, \Lambda \), therefore \( \eta(U_{j}^T, A_{t_i}^2) = \eta(U, A_{t_i}^2) \). Also,
\* \( \text{K}_V(A_{t^\gamma}) \) remains unaltered whether we use \( Q_\Lambda \) or \( Q_\Lambda \).

Thus we conclude that
\[
\begin{align*}
E \left\{ \Phi_{ \text{K}_V(A_{t^\gamma})}(\delta(U_{t^1}, A_{t^\gamma}) - \eta(U_{t^1}, A_{t^\gamma})) \right\} A_{t^\gamma} \\
= E \left\{ \Phi_{ \text{K}_V(A_{t^\gamma})}(\delta(U, A_{t^\gamma}) - \eta(U, A_{t^\gamma})) \right\} A_{t^\gamma}
\end{align*}
\]

must hold, and thus Theorem 1 must be true even when
\[
\text{rank}(\Gamma^T \text{K}_V \Gamma(A_{t^\gamma})) < 2mn \text{ for certain values of } A_{t^\gamma}.
\]

We have thus far completed our proof of Theorem 1; we next show an upper bound for the rank of the matrix \( \text{K}_V(A_{t^\gamma}) \) in (59). We point out that \( \text{K}_V(A_{t^\gamma}) \) sometimes may even have rank 0, i.e. \( \text{K}_V(A_{t^\gamma}) \) equals the zero matrix.

**B. Other comments**

The following proposition states that the rank of \( \text{K}_V(A_{t^\gamma}) \) depends on both the chosen time instants \( \{t_1, t_2, \ldots, t_n\} \), and the MLM truncation length \( m \). The following proposition gives the upper bound on \( \text{rank}(\text{K}_V(A_{t^\gamma})) \).

**Proposition 3.** The rank of \( \text{K}_V(A_{t^\gamma}) \) equals at most the number of time instants \( t \in \{t_1, t_2, \ldots, t_n\} \), that satisfy \( |t - t'| > m \) for all \( t' \in \{t_1, t_2, \ldots, t_n\} \setminus \{t\} \).

Proposition 3 is proved using the following lemma.

**Lemma 2.** If two time instants \( t_1 \) and \( t_2 \) satisfy \( |t_1 - t_2| \leq m \), then observation of \( [G(A_{t_1})S]^T W_{t_1} \) uniquely determines \( V_{t_1} = A_{t_1} \cdot h_0 W_{t_1} \) (and vice versa observation of \( [G(A_{t_2})S]^T W_{t_2} \) uniquely determines \( V_{t_2} = A_{t_2} \cdot h_0 W_{t_2} \)).

**Proof:** Recall that \( V_{t_2} \) equals
\[
V_{t_2} = A_{t_2} \cdot h_0 W_{t_2} = A_{t_2} \cdot (h_0 W_{t_2} + \ldots + h_j W_{t_2+j}).
\]

If the condition \( |t_1 - t_2| \leq m \) is satisfied, then \( W_{t_2+t_1}, \ldots, W_{t_2+m} \) is a length-\((s+1)\) subsequence of \( W_{t_1} = [W_{t_1-m}, W_{t_1-m+1}, \ldots, W_{t_1+m}]^T \). From the definition of \( S \) (see Definition 6) and because \( |t_1 - t_2| \leq m \), then the matrix \( S \) must have a column \( s \) that satisfies \( s = e_{t_2-t_1} \), see Definition 5 for \( E \) and its columns \( e_i \). Then for this particular column \( s \) we have
\[
[G(A_{t_1})S]^T W_{t_1} = [H \text{ diag}(A_{t_1}) s]^T W_{t_1} = A_{t_2} \cdot [H e_{t_2-t_1}]^T W_{t_2} = A_{t_2} \cdot h_0 W_{t_2} = V_{t_2},
\]

where the second equality holds because \( s \) satisfies \( \text{diag}(A_{t_1}) s = \text{diag}(A_{t_1}) e_{t_2-t_1} = A_{t_2} \cdot e_{t_2-t_1} \), and also
\[
[He_{t_2-t_1}]^T W_{t_2} = [He_{t_2-t_1}]^T [W_{t_1-m}, \ldots, W_{t_1+m}] = h_0 W_{t_2} + h_1 W_{t_2+1} + \ldots + h_l W_{t_2+l}.
\]

By symmetry, the same argument holds for \( [G(A_{t_1})S]^T W_{t_2} \) and \( V_{t_2} = A_{t_1} \cdot h_0 W_{t_1} \).

**Proof of Proposition 3:** Recall from (59) that
\[
K_V(A_{t^\gamma}) = \text{Cov}(V_{t^\gamma} | A_{t^\gamma}, U) \text{ is the (conditional) covariance matrix of } V_{t^\gamma}.
\]
After conditioning on \( U \), the vector
\[
Q_\Lambda U = [G(A_{t_1})S]^T W_{t_1}
\]
is uniquely determined, see Lemma 1. Furthermore by Lemma 2, if \( Q_\Lambda U = [G(A_{t_1})S]^T W_{t_1} \) is uniquely determined then \( V_{t_1} \triangleq A_{t_1} \cdot h_0 W_{t_1} \) is determined whenever \( |t_1 - t_2| \leq m \). Thus we conclude that the only variables \( V_{t_1} \) that may contribute to the rank of \( \text{K}_V(A_{t^\gamma}) \), must be those with corresponding \( t_1 \) that are separated from all other \( \{t_1, t_2, \ldots, t_n\} \setminus \{t_1\} \) by greater than \( m \).

**Remark 7.** From the expression for \( F_{X_{t^1}, Y_{t^1}}(r) \) in Theorem 1, the distribution function \( F_{X_{t^1}, Y_{t^1}}(r) \) must be left-continuous [19], if the rank(\( \text{K}_V(A_{t^\gamma}) \)) = \( n \).

We conclude this section by verifying the correctness of Procedure 3, used to evaluate \( F_{X_{t^1}, Y_{t^1}}(r) \) when candidate subsets \( M \subset M \) (see (41)) are considered. The only difference between Procedures 1 and 3, is that Line 3 of Procedure 3 replaces Line 4 of Procedure 1. First verify that the following equality of sets is true
\[
\{a \in M_{t_1} : a_0 \neq A_{t_1} \} = \{a \in \text{Es}_k + e_0 A_{t_1} : 0 \leq k \leq 2^{2m-1} \},
\]

where here the function \( \alpha(\text{e}, A_{t_1}) \) is given in Line 3 of Procedure 3. Next perform the following verifications in the order presented:

- Replace \( M \) by \( M_{t_1} \) in the definitions of \( R_1 \) in (12). Replace \( M \) by \( M_{t_1} \) in both \( X_{t_1} \) and \( Y_{t_1} \) in (13). The validity of Proposition 1 remains unaffected.
- Replace \( M \) by \( M_{t_1} \) in the proof of Proposition 2. The change first affects the proof starting from (40), and (47) needs to be slightly modified using (62). The new Proposition 2 finally reads

\[
X_{t_1} = \max_{k : \alpha(\text{Es}_k + e_0 A_{t_1}) \in M_{t_1}} s_k \text{[G}(A_{t_1})\text{]}^T W_{t_1} + \nu_k(A_{t_1}) + \theta(A_{t_1}),
\]

where \( s_k \text{[G}(A_{t_1})\text{]}^T W_{t_1} + \mu_k(A_{t_1}) \).

- Utilize the new Proposition 2 in the proof of Theorem 1. The change first affects the proof starting from (54). Proceeding from (55)-(56) we arrive at the new formulas

\[
\delta_i = \delta_i(U, A_{t^\gamma}) = \max_{k : \alpha(\text{Es}_k + e_0 A_{t_1}) \in M_{t_1}} s_k Q_i A \lambda + \nu_k(A_{t_1}) - \max_{k : \alpha(\text{Es}_k + e_0 A_{t_1}) \in M_{t_1}} s_k Q_i A U + \mu_k(A_{t_1}).
\]

This is exactly the way \( \delta_i \) is computed in Procedure 3, Line 3.

| Channel | Coefficients | Memory Length t |
|---------|--------------|-----------------|
| PR1     | 1            | -1             |
| PR2     | 2            | 2              |
| PR3     | 1            | 0              |
| PR4     | 1            | -1            |

| Channel | Coefficients | Memory Length t |
|---------|--------------|-----------------|
| PR1     | 1            | -1             |
| PR2     | 2            | 2              |
| PR3     | 1            | 0              |
| PR4     | 1            | -1            |

TABLE 1

VARIOUS ISI CHANNELS IN MAGNETIC RECORDING [18]
Thus i.i.d. A. Marginal distribution

Define the signal-to-noise (SNR) ratio as ing literature [18], [15].

Therefore, both quantities i) and ii) will be displayed utilizing a single graphical plot of $F_{X_t^1-Y_t^1}(\sigma^2/2 \cdot r)$.

The chosen ISI channels for our tests are given in Table I; these are commonly-cited channels in the magnetic recording literature [18], [15]. Define the signal-to-noise (SNR) ratio as $10 \log_{10}(\sum_{i=0}^{\ell} h_i^2/\sigma^2)$. The input symbol distribution $Pr\{A_t = a\}$ will always be uniform, i.e. $Pr\{A_t = a\} = 2^{-2(m+\ell) - 1}$ see (2), unless stated otherwise.

A. Marginal distribution $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ when the noise is i.i.d.

First, consider the case where the noise samples $W_t$ are i.i.d, thus $\sigma^2 = E\{W_t^2\}$. Figure 4 shows the marginal distribution $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ computed for the PR1 channel (see Table I) with memory $\ell = 1$. The distribution is shown for various truncation lengths $m = 1$ to 5, and two different SNRs : 3 dB and 10 dB. At SNR 3 dB, we observe that with the exception of $m = 1$, all curves appear to be extremely close. At SNR 10 dB, a good choice for the truncation length $m$ appears to be $m = 2$; the computed distribution for $m = 2$ appears close to the simulated distribution. At SNR 10 dB, it appears that $m = 5$ is a good choice. The probability of symbol error $Pr\{B_t \neq A_t\} = Pr\{X_t \geq Y_t\} = 1 - F_{X_t-Y_t}(0)$ is observed to decrease as the truncation length $m$ increases; this is expected. At SNR 3 dB, the (error) probability $Pr\{X_t \geq Y_t\} = 1 - F_{X_t-Y_t}(0) \approx 1.4 \times 10^{-1}$ for truncation lengths $m > 1$. For SNR 10 dB, the (error) probability $Pr\{X_t \geq Y_t\}$ is seen to vary significantly for both truncation lengths $m = 1$ and 5; the probability $Pr\{X_t \geq Y_t\} \approx 1.1 \times 10^{-1}$ and $1 \times 10^{-2}$ for $m = 1$ and 5, respectively.

For the PR1 channel and a fixed truncation length $m = 4$, the marginal distributions $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ are compared across various SNRs in Figure 5. As SNR increases, the distributions $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ appear to concentrate more probability mass over negative values of $X_t - Y_t$. This is intuitively expected, because as the SNR increases, the symbol error probability $Pr\{B_t \neq A_t\} = Pr\{X_t \geq Y_t\} = 1 - F_{X_t-Y_t}(0)$ should decrease. From Figure 5, the (error) probabilities $Pr\{X_t \geq Y_t\}$ are found to be approximately $1.2 \times 10^{-1}, 8 \times 10^{-2}, 3 \times 10^{-2}$, and $1 \times 10^{-2}$, respectively for SNRs 3 to 10 dB.
B. Joint distribution $F_{X_{t^2}-Y_{t^2}}(\sigma^2/2 \cdot r)$, here $n = 2$, when the noise is i.i.d.

We consider again i.i.d noise $W_t$, and the PR1 and PR2 channels (see Table I). Here, we choose the SNR to be moderate at 5 dB. For the PR1 channel with memory length $\ell = 1$, the truncation length is fixed to be $m = 2$. For the PR2 channel with $\ell = 2$, we fix $m = 5$. Figure 6 compares the joint distributions $F_{X_{t^2}-Y_{t^2}}(\sigma^2/2 \cdot r)$, computed for both PR1 and PR2 channels and for both time lags $|t_1 - t_2| = 1$ (i.e. neighboring symbols) and $|t_1 - t_2| = 7$. The difference between the two cases $|t_1 - t_2| = 1$ and 7 is subtle (but nevertheless inherent) as observed from the differently labeled points in the figure. For the PR1 channel, the joint symbol error probability $\Pr \{B_{t_1} \neq A_{t_1}, B_{t_2} \neq A_{t_2}\} = \Pr \{X_{t^2} \geq Y_{t^2}\}$ is approximately $6 \times 10^{-2}$ and $2 \times 10^{-2}$ for both cases $|t_1 - t_2| = 1$ and 7, respectively. Similarly for the PR2, the (error) probability is approximately $3 \times 10^{-2}$ and $1 \times 10^{-2}$ for both respective cases $|t_1 - t_2| = 1$ and 7. Finally note that for the PR1 channel when $|t_1 - t_2| = 7$, both MLM reliability values $R_{t_1} = 2/\sigma^2 |X_{t_2} - Y_{t_1}|$ and $R_{t_2} = 2/\sigma^2 |X_{t_2} - Y_{t_2}|$ are independent; this is because then $|t_1 - t_2| = 7 > 2(m + \ell) = 6$, refer to Figure 2.

C. Marginal distribution $F_{X_t-Y_1}(\sigma^2/2 \cdot r)$ when the noise is correlated.

Consider the PR2 channel, and now consider the case where the noise samples $W_t$ are correlated. For simplicity of argument we consider single lag correlation, i.e. $\E \{W_t \cdot W_{t^2}\} = 0$ for all $|t - t^2| > 1$, and consider the following two cases:

1. **the correlation coefficient** $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = 0.5$, and
2. **the correlation coefficient** $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = -0.5$.

We consider a moderate SNR of 5 dB. Figure 7 shows the distributions $F_{X_t-Y_1}(\sigma^2/2 \cdot r)$ computed for both cases. Also in Figure 7, the power spectral densities of the correlated noise samples $W_t$ (see [19], p. 408) are shown for both cases. It is apparent that the truncated MLM detector performs better (i.e. smaller symbol error probability) when the correlation coefficient $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = -0.5$. This is explained intuitively as follows. The detector should be able to tolerate more noise in the signaling frequency region. Observe the PR2 frequency response [18], [15] displayed in Figure 7. When the correlation coefficient equals $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = -0.5$, the noise power is strongest amongst signaling frequencies, and the symbol error probability $\Pr \{B_{t_1} \neq \bar{A}_{t_1}\}$ is observed to be the lowest (approximately $8 \times 10^{-2}$). On the other hand when the correlation coefficient is $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = 0.5$, the noise is strongest at frequencies near the spectral null of the PR2 channel, and the (error) probability $\Pr \{X_t \geq Y_t\}$ is the highest (approximately $1.6 \times 10^{-1}$). Note that in the latter case $\E \{W_t \cdot W_{t^2+1}\}/\sigma^2 = -0.5$, the MLM performs even better than the i.i.d case, see Figure 7. In the i.i.d case, the error probability $\Pr \{X_t \geq Y_t\} \approx 1.3 \times 10^{-1}$.

**Remark 8.** One intuitively expects that similar observations will be made even for other (more complicated) choices for the noise covariance matrix $K_W$, recall (22). We stress that our results are general in the sense that we may arbitrarily specify $K_W$; even if the noise samples $W_t$ are non-stationary our methods still apply.
D. Marginal distribution $F_{X_i - Y_i}(\sigma^2/2 \cdot r)$ when the noise is i.i.d., and when run-length limited (RLL) codes are used.

We demonstrate Procedure 3 in Subsection III-C, used to compute the distribution $F_{X_i - Y_i}(\sigma^2/2 \cdot r)$ when a modulation code is present in the system. In particular, consider a run-length limited (RLL) code; we test the simple RLL code that prevents neighboring symbol transitions [14], [15]. This code improves transmission over ISI channels, that have spectral nulls near the Nyquist frequency [15]; one such channel is the PR4, see Table I. Figure 8 shows $F_{X_i - Y_i}(\sigma^2/2 \cdot r)$ computed for both the PR4, as well as the dicode channel, see Table I. The PR4 channel has a spectral null at Nyquist frequency (recall Subsection V-C), but the dicode channel does not.

It is clearly seen from Figure 8 that the RLL code improves the performance when used on the PR4 channel. For the PR4 channel, the distribution $F_{X_i - Y_i}(\sigma^2/2 \cdot r)$ appears to concentrate more probability mass over negative values of $X_i - Y_i$ (similar to the observations made in Figure 5 when there is a SNR increase). The error probability $\Pr \{ B_t \neq A_t \} = \Pr \{ X_t \geq Y_t \} = 1 - F_{X_i - Y_i}(0)$ decreases by a factor of 2, dropping from approximately $9.5 \times 10^{-2}$ to $4 \times 10^{-2}$. On the other hand, the RLL code worsens the performance when applied to the dicode channel. For the dicode channel, $F_{X_i - Y_i}(r)$ concentrates more probability mass over positive values of $X_i - Y_i$ (similar to the observations made in Figure 5 when there is an SNR decrease), and the (error) probability $\Pr \{ X_t \geq Y_t \}$ increases from approximately $8.8 \times 10^{-2}$ to $1.35 \times 10^{-1}$.

E. Marginal distribution of $2/\sigma^2 \cdot (X_i - Y_i)$, when conditioning on neighboring error events $\{ B_{t-1} \neq A_{t-1} \}$ and $\{ B_{t+1} \neq A_{t+1} \}$

Here we consider three neighboring symbol reliabilities, i.e. we consider $R_{t+1} = [R_{t-1}, R_t, R_{t+1}]^T$. We consider the following two conditional distributions:

\begin{enumerate}
  \item \( \Pr \{ X_t - Y_t \leq r | X_{t-1} < Y_{t-1}, X_{t+1} < Y_{t+1} \} = \frac{1}{C_1} \cdot F_{X_{t-1} - Y_{t-1}}(0, r, 0) \), and
  \item \( \Pr \{ X_t - Y_t \leq r | X_{t-1} \geq Y_{t-1}, X_{t+1} \geq Y_{t+1} \} = \frac{1}{C_2} \left( F_{X_{t-1} - Y_{t-1}}(r) - F_{X_{t+1} - Y_{t+1}}(r, 0) \right) \)
\end{enumerate}

where the normalization constants $C_1$ and $C_2$ equal the probabilities of the (respective) events that were conditioned on. Dis-
both distributions (a) and (b) always (practically) equal the
unconditioned distribution

\[ \{ B_{t-1} = A_{t-1}, B_{t+1} = A_{t+1} \} \]
and \{ \{ B_{t-1} = A_{t-1}, B_{t+1} = A_{t+1} \} \}. These two events correspond to error
(or non-error) events at neighboring time instants \( t-1 \) and \( t+1 \). The solid
black line represents the unconditioned marginal distribution of \( X_t - Y_t \).

In this paper, we derived closed-form expressions for both
i) the reliability distributions \( F_{X_1} - Y_1 \) \((\sigma^2/2 \cdot r)\), and ii) the
symbol error probabilities \( \Pr \{ \bigcap_{n=1}^{\infty} \{ B_{t_i} \neq A_{t_i} \} \} \), for the
m-truncated MLM detector. Our results hold jointly for any
number \( n \) of arbitrarily chosen time instants \( t_1, t_2, \ldots, t_n \).

The general applicability of our result has been demonstrated
for a variety of scenarios. Efficient Monte-Carlo procedures
that utilize dynamic programming simplifications have been
given, that can be used to numerically evaluate the closed-
form expressions.

It would be interesting to further generalize the exposition
to consider infinite impulse response (IIR) filters, such as in
convolutional codes.

APPENDIX

A. Computing the matrix \( Q = Q(A_{t_1}) \) in Definition 8

In this appendix, we show that the size 2mn square
matrix \( Q \) with both properties i) and ii) as stated in
Definition 8, can be easily found. We begin by noting
from (24) that rank(\( SS^T \)) = 2m, therefore the matrix
\( I_n \otimes SS^T \) has rank 2mn and is positive definite. Recall
diag(\( G(A_{t_1}), G(A_{t_2}), \ldots, G(A_{t_n}) \)) is block diagonal with
entries (20).

Lemma 3. Let \( S \) be given as in Definition 6. Let the size 2mn
by 2mn square matrix \( \alpha \) diagonalize

\[ \alpha^T (I_n \otimes SS^T) \alpha = I. \]  

Let \( \beta \) be the size 2mn by 2mn eigenvector matrix \( \beta \) in the
following decomposition

\[ \alpha^T (I_n \otimes SS^T) \text{diag}(G(A_{t_1}), G(A_{t_2}), \ldots, G(A_{t_n}))^T KW \]
\[ \cdot \text{diag}(G(A_{t_1}), G(A_{t_2}), \ldots, G(A_{t_n}))(I_n \otimes SS^T)\alpha \]
\[ = \beta \Lambda^2 \beta^T, \]  

and \( \Lambda \) is the eigenvalue matrix of (64), therefore \( \Lambda^2 \) in (64)
is diagonal of size 2mn. Then

\[ Q = \alpha \beta \]  

satisfies both properties i) and ii) stated in Definition 8.

Proof: Because \( \alpha \) diagonalizes \( I_n \otimes SS^T \) to an identity
matrix \( I \), it follows that \( \alpha \) must have full rank, and thus have an
inverse \( \alpha^{-1} \). It follows from (63) that \( \alpha^{-1} = \alpha^T (I_n \otimes SS^T) \). Replacing \( \alpha^T (I_n \otimes SS^T) = \alpha^{-1} \) in (64), we see that \( \beta \) satisfies

\[ \alpha^{-1} \text{diag}(G(A_{t_1}), G(A_{t_2}), \ldots, G(A_{t_n}))^T KW \]
\[ \cdot \text{diag}(G(A_{t_1}), G(A_{t_2}), \ldots, G(A_{t_n})) \alpha^{-T} = \beta \Lambda^2 \beta^T. \]  

Consider the matrix \( Q = \alpha \beta \). It follows from (66) that
\( Q = \alpha \beta \) satisfies property i) in Definition 8, as seen after
multiplying (the matrices satisfying (66) on the left and right
by \( \alpha \) and \( \alpha^{-T} \), respectively. It also follows that \( Q = \alpha \beta \)
satisfies property ii) in Definition 8, this is because

\[ Q^T (I_n \otimes SS^T) Q = \beta^T \alpha^T (I_n \otimes SS^T) \alpha \beta = \beta^T \beta = I, \]
where the last equality follows because \( \beta \) is unitary (i.e. \( \beta^{-1} = \beta^T \)) by virtue of the fact that it is an eigenvector matrix [17], p. 311.
To summarize Lemma 3, the matrix $Q = Q(A_{1t})$ in Definition 8, is obtained by first computing two size $2mn$ matrices $\alpha$ and $\beta$ respectively satisfying (63) and (64), and then setting $Q = \alpha \beta$. The matrix $\beta$ is obtained from an eigenvalue decomposition of the $2mn$ matrix (64), and clearly $\beta$ depends on the symbols $A_{1t}$. The matrix $\alpha$ however, is simpler to obtain. This is due to the simple form of $SS^T$ in (24), and we may even obtain closed form expressions for $\alpha$, see the next remark.

**Remark 9.** It can be verified that the following are eigenvectors of the matrix $SS^T$ in (24). The first $2n - 1$ eigenvectors are

$$(i + 1)^{-\frac{1}{2}} \cdot \begin{bmatrix} 1, & 1, & \cdots, & 1, & -i, & 0, & 0, & \cdots, & 0 \end{bmatrix}^T$$

where $i$ can take values $1 \leq i < 2m$, and the last eigenvector is simply $1/\|\alpha\| = 1/(2m)$. 

**REFERENCES**

[1] G. D. Forney, Jr., “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” IEEE Trans. on Inform. Theory, vol. 18, no. 3, pp. 363–378, May 1972.

[2] ——, “The Viterbi algorithm,” Proceedings of the IEEE, vol. 61, no. 3, pp. 268 – 278, Mar. 1973.

[3] C. Douillard, A. Picart, P. Didier, M. Jezequel, C. Berrou, and A. Glavieux, “Iterative correction of intersymbol interference: turboequalization,” European Trans. on Telecomms., vol. 6, no. 5, pp. 507–512, Sep-Oct. 1995.

[4] A. Kavcic, X. Ma, and M. Mitzenmacher, “Binary intersymbol interference channels: Gallager codes, density evolution and code performance bounds,” IEEE Trans. on Inform. Theory, vol. 49, no. 7, pp. 1636–1652, Jul. 2003.

[5] F. Lim, A. Kavcic, and M. Fossorier, “List decoding techniques for intersymbol interference channels using ordered statistics,” IEEE Journal on Selected Areas Commun., vol. 28, no. 2, pp. 241–251, Jan. 2010.

[6] J. Hagenauer and P. Hoeher, “A Viterbi algorithm with soft-decision outputs and its applications,” in IEEE Global Telecomm. Conference (GLOBECOM ’89), Dallas, TX, Nov. 1989, pp. 1680–1686.

[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” IEEE Trans. on Inform. Theory, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[8] M. P. C. Fossorier, F. Burkert, S. Lin, and J. Hagenauer, “On the equivalence between SOVA and max-log-MAP decodings,” Communications Letters, IEEE, vol. 2, no. 5, pp. 137 – 139, May 1998.

[9] H. Yoshikawa, “Theoretical analysis of bit error probability for maximum a posteriori probability decoding,” in Proc. IEEE International Symposium on Inform. Theory (ISIT’03), Yokohama, Japan, 2003, p. 276.

[10] M. Lentmaier, D. V. Truhachev, and K. S. Zigangirov, “Analytic expressions for the bit error probabilities of rate-$1/2$ memory $2$ convolutional encoders,” IEEE Trans. on Inform. Theory, vol. 50, no. 6, pp. 1303–1311, 2004.

[11] L. Reggiani and G. Tartara, “Probability density functions of soft information,” IEEE Commun. Letters, vol. 6, no. 2, pp. 52–54, Feb. 2002.

[12] A. Avudainayagam, J. M. Shea, and A. Roongta, “On approximating the density function of reliabilities of the max-log-map decoder,” in Fourth IASTED International Multi-Conference on Wireless and Optical Communications (CSA ’04), Banff, Canada, 2004, pp. 358–363.

[13] P. N. Somerville, “Numerical computation of multivariate normal and multivariate-$t$ probabilities over convex regions,” Journal of Computational and Graphical Statistics, vol. 7, no. 4, pp. 529–544, 1998.

[14] D. T. Tang and L. R. Bahl, “Block codes for a class of constrained noiseless channels,” Information and Control, vol. 17, no. 5, pp. 436–461, Dec. 1970.

[15] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, “Codes for digital recorders,” IEEE Trans. on Inform. Theory, vol. 44, no. 6, pp. 2260 – 2299, Oct. 1998.

[16] H. Hofelding, “Probability inequalities for sums of bounded random variables,” Journal of the American Statistical Association, vol. 58, no. 301, pp. 13–30, Mar. 1963.

[17] G. H. Golub and C. F. Van Loan, Matrix computations. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[18] P. Kabal and S. Pasupathy, “Partial response signaling,” IEEE Trans. on Commun., vol. 23, no. 9, pp. 921 – 934, Sep. 1975.

[19] A. Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4th ed. McGraw Hill, 2002.