

Reliability Distributions of Truncated Max-Log-MAP (MLM) Detectors Applied to Binary ISI Channels

Fabian Lim and Aleksandar Kavčić

Abstract—The max-log-MAP (MLM) receiver is an approximated version of the well-known Bahl–Cocke–Jelinek–Raviv algorithm. The MLM algorithm is attractive due to its implementation simplicity. In practice, sliding-window implementations are preferred, whereby truncated signaling neighborhoods (around each transmission time instant) are considered. In this paper, we consider binary signaling sliding-window MLM receivers, where the MLM detector is truncated to a length- m signaling neighborhood. Here, truncation is used to ease the burden of analysis. For any number n of chosen times instants, we derive exact expressions for both 1) the joint distribution of the MLM symbol reliabilities, and 2) the joint probability of the erroneous MLM symbol detections. We show that the obtained expressions can be efficiently evaluated using Monte–Carlo techniques. The most computationally expensive operation (in each Monte–Carlo trial) is an eigenvalue decomposition of a size $2mn \times 2mn$ matrix. The proposed method handles various scenarios such as correlated noise distributions, modulation coding, etc.

Index Terms—Detection, intersymbol interference, probability, reliability, Viterbi algorithm.

I. INTRODUCTION

THE intersymbol interference (ISI) channel has been widely studied. Optimal detection schemes for the ISI channel consider input–output sequences, rather than individual symbols [1]. Sequence detectors such as the Viterbi detector only compute hard decisions [2]. However, modern coding techniques often benefit from detection schemes that also compute *symbol reliabilities* also known as *soft-outputs*, *log-likelihood ratios* [3]–[5]. Some well-known detectors that perform this task include the soft-output Viterbi algorithm [6], the Bahl–Cocke–Jelinek–Raviv algorithm [7], and the *max-log-MAP* (MLM) detector [8]. There has been recent interest in the analysis of the MLM detector. The marginal symbol error probability has been derived for a two-state convolutional code in [9]; this was further extended for convolutional codes with constraint length two in [10]. Approximations for the MLM reliability distributions have been obtained in [11] and [12].

Manuscript received June 30, 2010; revised July 02, 2010; accepted August 04, 2011. Date of publication January 30, 2013; date of current version March 13, 2013. This work was supported by the National Science Foundation under Grant CCF-1018984. This paper was presented in part at the 2011 IEEE International Symposium on Information Theory.

F. Lim was with the University of Hawaii, Honolulu, HI 96822 USA. He is now with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: flim@mit.edu).

A. Kavčić is with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA (e-mail: alek@hawaii.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2013.2244277

In this paper, we consider the use of an MLM receiver for binary signaling over an ISI channel. In particular, we consider its *sliding-window* implementation. An 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 m -truncation is used here to facilitate the analysis of the MLM receiver. For the m -truncated MLM receiver, considering any number n of chosen time instants, we derive *exact, closed-form* expressions for both 1) the joint distribution of the symbol reliabilities, and 2) 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 follows from a simple observation.

Notation

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 uppercase italics (e.g., denoted A) and vectors denoted using uppercase bold italics (e.g., denoted \mathbf{A}). Note that we do not reserve specific notation for random matrices. Throughout the paper, both t and τ are used to denote time indices. Sets are denoted using curly braces, e.g., $\{a_1, a_2, a_3, \dots\}$. Also, both α and β are used for auxiliary notation as needed. Finally, the maximization over the components of the size- n vector $\mathbf{a} = [a_1, a_2, \dots, a_n]^T$ may be written either explicitly as $\max_{i \in \{1, 2, \dots, n\}} a_i$, 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 F is reserved to denote probability cumulative distribution functions, i.e., $F_A(a) = \Pr \{A \leq a\}$. The expectation of A is denoted as $\mathbb{E}\{A\}$.

II. MLM ALGORITHM

Let a random sequence of symbols drawn from the set $\{-1, 1\}$, denoted as $\dots, A_{-2}, A_{-1}, A_0, A_1, A_2, \dots$, be transmitted over an ISI channel with memory ℓ characterized by channel coefficients h_0, h_1, \dots, h_ℓ . The binary signaling ISI channel output sequence, denoted $\dots, Z_{-2}, Z_{-1}, Z_0, Z_1, Z_2, \dots$, satisfies the following input–output relationship:

$$Z_t = \sum_{i=0}^{\ell} h_i A_{t-i} - W_t \quad (1)$$

and the channel noise samples $\dots, W_{-2}, W_{-1}, W_0, W_1, W_2, \dots$ are assumed to be zero-mean and jointly Gaussian distributed (we do not assume they are independent). Note that the Gaussian noise sample W_t in (1) is subtracted (as opposed to being usually

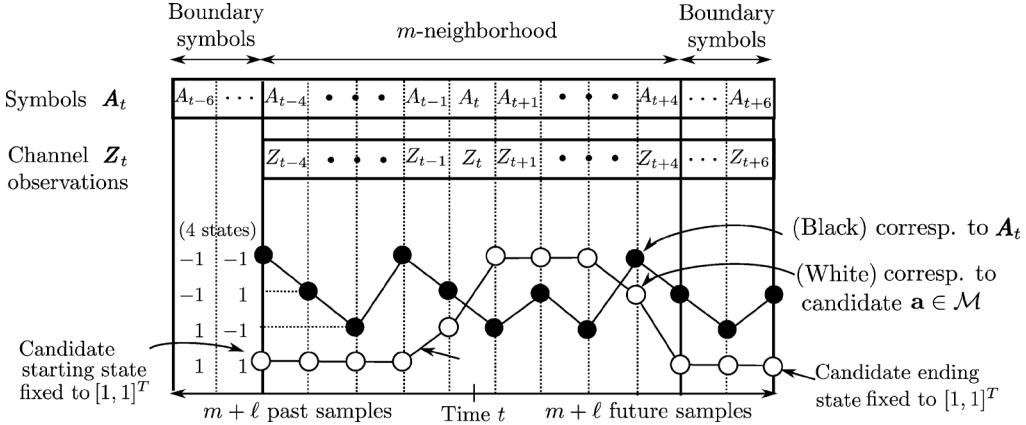


Fig. 1. Time evolution of the channel states of an m -truncated MLM detector for binary signaling over an ISI channel with memory length ℓ , where $m = 6$ and $\ell = 2$.

added in the literature) for purposes of obtaining neater expressions in the sequel. Clearly, subtraction incurs no loss in generality, as the Gaussian distribution is symmetric about its mean. The ISI *channel state* at time t equals the (length- ℓ) vector of input symbols $[A_{t-\ell+1}, A_{t-\ell+2}, \dots, A_t]^T$. The total number of possible states is 2^ℓ , exponential in the memory length ℓ .

At time instant t , the m -truncated MLM detector considers the neighborhood of $2m + \ell + 1$ channel outputs $\mathbf{Z}_t \triangleq [Z_{t-m}, Z_{t-m+1}, \dots, Z_{t+m+\ell}]^T$. Let \mathbf{A}_t denote the symbol neighborhood that contains the following $2(m + \ell) + 1$ input symbols:

$$\mathbf{A}_t \triangleq [A_{t-m-\ell}, A_{t-m-\ell+1}, \dots, A_{t+m+\ell}]^T. \quad (2)$$

Both \mathbf{A}_t and \mathbf{Z}_t are depicted in Fig. 1. Let $\Phi_{i,j}$ denote a matrix of size $i \times j$, whereby every entry of $\Phi_{i,j}$ equals zero. Let \mathbf{h}_i denote the following length- $(2m + \ell + 1)$ vector:

$$\mathbf{h}_i \triangleq [\Phi_{1,m+i}, h_0, h_1, \dots, h_\ell, \Phi_{1,m-i}]^T \quad (3)$$

where i can take values $|i| \leq m$, note that $\Phi_{1,i}$ is a row vector of length i containing only zeros. Let both \mathbf{H} and \mathbf{T} denote the size $2m + \ell + 1$ by $2(m + \ell) + 1$ matrices given as

$$\begin{aligned} \mathbf{H} &\triangleq [\Phi_{2m+\ell+1,\ell}, \mathbf{h}_{-m}, \mathbf{h}_{-m+1}, \dots, \mathbf{h}_m, \Phi_{2m+\ell+1,\ell}] \\ \mathbf{T} &\triangleq \left[\begin{array}{ccc} \mathbf{T}_1 & \Phi_{2m+\ell+1,2m+1} & \Phi_{2m+1,\ell} \\ \Phi_{2m+1,\ell} & \mathbf{T}_2 & \end{array} \right] \end{aligned} \quad (4)$$

where the two $\ell \times \ell$ submatrices \mathbf{T}_1 and \mathbf{T}_2 equal

$$\mathbf{T}_1 = \begin{bmatrix} h_\ell & h_{\ell-1} & \cdots & h_1 \\ h_\ell & & \vdots & \\ & \ddots & \ddots & \\ & & \vdots & \end{bmatrix}, \quad \mathbf{T}_2 = \begin{bmatrix} h_0 & & & \\ \vdots & \ddots & & \\ h_{\ell-2} & \cdots & h_0 & \\ h_{\ell-1} & \cdots & h_1 & h_0 \end{bmatrix}.$$

Using (4), rewrite $\mathbf{Z}_t \triangleq [Z_{t-m}, Z_{t-m+1}, \dots, Z_{t+m+\ell}]^T$ using (1) into the following form:

$$\mathbf{Z}_t = (\mathbf{H} + \mathbf{T}) \mathbf{A}_t - \mathbf{W}_t \quad (5)$$

where \mathbf{W}_t denotes the neighborhood of noise samples

$$\mathbf{W}_t \triangleq [W_{t-m}, W_{t-m+1}, \dots, W_{t+m+\ell}]^T. \quad (6)$$

Let $\mathbb{1}_\ell$ denote a vector of length ℓ , with all its entries equal to 1. Let \mathcal{M} denote the set of m -truncated MLM candidate sequences, defined as

$$\mathcal{M} \triangleq \left\{ \mathbf{a} \in \{-1, 1\}^{2(m+\ell)+1} : a_i = 1 \text{ for all } |i| > m \right\}. \quad (7)$$

Each candidate $\mathbf{a} \in \mathcal{M}$ has boundary symbols equal to 1, i.e., each \mathbf{a} has the form

$$\mathbf{a} = [\mathbb{1}_\ell^T, a_{-m}, a_{-m+1}, \dots, a_m, \mathbb{1}_\ell^T]^T.$$

Alternatively, the boundary symbols can be specified to be any sequence of choice in the set $\{-1, 1\}^\ell$; here, we choose $\mathbb{1}_\ell$ for the boundary sequence to simplify exposition. An example of a candidate $\mathbf{a} \in \mathcal{M}$ is illustrated in Fig. 1. The reason for fixing the boundary symbols of the candidates $\mathbf{a} \in \mathcal{M}$ *a priori* (to some chosen sequence) is so as to initialize them to some values (as the boundary symbols of the transmitted sequence \mathbf{A}_t are *unknown* to the detector). The start/end states of \mathbf{A}_t (colored black) are shown (see Fig. 1) to be different from the start/end states of the candidate $\mathbf{a} \in \mathcal{M}$ (colored white).

Let the sequence $\dots, B_{-2}, B_{-1}, B_0, B_1, B_2, \dots$ denote *symbol decisions* on the channel inputs $\dots, A_{-2}, A_{-1}, A_0, A_1, A_2, \dots$. Let $\mathbb{1}$ denotes an all-ones vector of nonspecified length. In the following, let $|\mathbf{a}|$ denote the Euclidean norm of the vector \mathbf{a} . For each time instant t , define the sequence $\mathbf{B}^{[t]}$ as

$$\begin{aligned} \mathbf{B}^{[t]} &\triangleq \arg \min_{\mathbf{a} \in \mathcal{M}} |\mathbf{Z}_t - (\mathbf{H} + \mathbf{T})\mathbf{a}|^2 \\ &= \arg \min_{\mathbf{a} \in \mathcal{M}} |\mathbf{Z}_t - \mathbf{T}\mathbb{1} - \mathbf{H}\mathbf{a}|^2. \end{aligned} \quad (8)$$

The symbol decision B_t on channel input A_t is obtained from $\mathbf{B}^{[t]}$ by setting $B_t \triangleq B_0^{[t]}$, where $B_0^{[t]}$ denotes the zeroth element of the candidate $\mathbf{B}^{[t]} \in \mathcal{M}$. Clearly, $\mathbf{B}^{[t]}$ only has length $2(m + \ell) + 1$ and, therefore, does not equal the MLM bit detection sequence $\dots, B_{-2}, B_{-1}, B_0, B_1, B_2, \dots$; however, note that each symbol B_t is obtained from each $\mathbf{B}^{[t]}$. Each sequence $\mathbf{B}^{[t]}$ is obtained by comparing the squared Euclidean distances of each candidate $\mathbf{H}\mathbf{a}$ from the received neighborhood $\mathbf{Z}_t - \mathbf{T}\mathbb{1}$, see (8).

In addition to computing *hard*, i.e., $\{-1, 1\}$, symbol decisions B_t , the m -truncated MLM also computes a symbol *reliability* sequence $\dots, R_{-2}, R_{-1}, R_0, R_1, R_2, \dots$. Consider the following log-likelihood approximation (see [8]):

$$\begin{aligned} \log \frac{\Pr\{A_t = B_t | \mathbf{Z}_t\}}{\Pr\{A_t \neq B_t | \mathbf{Z}_t\}} &= \log \frac{\sum_{\mathbf{a} \in \mathcal{M}: a_0 = B_t} \Pr\{\mathbf{Z}_t | A_t = \mathbf{a}\}}{\sum_{\mathbf{a} \in \mathcal{M}: a_0 \neq B_t} \Pr\{\mathbf{Z}_t | A_t = \mathbf{a}\}} \\ &\approx \min_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq B_t}} \frac{1}{2\sigma^2} |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\mathbf{a}|^2 \\ &\quad - \min_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 = B_t}} \frac{1}{2\sigma^2} |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\mathbf{a}|^2 \end{aligned} \quad (9)$$

where the first equality assumes¹ uniform signal priors, i.e., $\Pr\{\mathbf{A}_t = \mathbf{a}\} = 2^{-2(m+\ell)-1}$; see (2). Denote σ^2 to be the worst case noise variance

$$\sigma^2 \triangleq \sup_{t \in \mathbb{Z}} \mathbb{E}\{W_t^2\} \quad (10)$$

and assume that σ^2 is bounded, i.e., $\sigma^2 < \infty$. If W_t is stationary, then $\sigma^2 = \mathbb{E}\{W_t^2\}$. We want to set the (m -truncated MLM) reliability R_t to equal the log-likelihood approximation (9), written in the following form. Denote the difference in the obtained squared Euclidean distances

$$\begin{aligned} \Delta(\mathbf{a}, \bar{\mathbf{a}}) &= \Delta(\mathbf{a}, \bar{\mathbf{a}}; \mathbf{Z}_t) \\ &\triangleq |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\mathbf{a}|^2 - |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\bar{\mathbf{a}}|^2 \end{aligned} \quad (11)$$

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

Definition 1: The nonnegative m -truncated MLM reliability R_t is defined as

$$R_t \triangleq \min_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq B_t}} \frac{1}{2\sigma^2} \Delta(\mathbf{a}, \mathbf{B}^{[t]}) \quad (12)$$

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

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

III. KEY OBSERVATION AND MAIN RESULT

In Section III-A, we describe an important *key observation* of which the derivation of the main result is based on. In Section III-B, the main result provides closed-form expressions for the 1) joint reliability distribution $F_{R_{t_1}, R_{t_2}, \dots, R_{t_n}}(r_1, r_2, \dots, r_n)$, and 2) joint symbol error probability $\Pr\{\bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\}\}$, for n time instants t_i . A Monte-Carlo procedure is given to evaluate these closed forms. In Section III-C, an efficient method to run the Monte-Carlo is discussed.

¹The relaxation of this assumption is discussed in the latter-half of the upcoming Section III-C, where we allow some of the probabilities $\Pr\{\mathbf{A}_t = \mathbf{a}\}$ to equal zero, i.e., in the case of modulation coding.

A. Key Observation

For all t , define X_t and Y_t as

$$\begin{aligned} X_t &\triangleq \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq A_t}} \frac{1}{4} \Delta(\mathbf{A}_t, \mathbf{a}) \\ Y_t &\triangleq \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 = A_t}} \frac{1}{4} \Delta(\mathbf{A}_t, \mathbf{a}) \geq 0 \end{aligned} \quad (13)$$

where $\Delta(\mathbf{A}_t, \mathbf{a})$ [see (11)] equals the difference of the squared Euclidean distances corresponding to \mathbf{A}_t , and a candidate $\mathbf{a} \in \mathcal{M}$, respectively. Note that $Y_t \geq 0$ because there must exist a candidate $\mathbf{a} \in \mathcal{M}$ that satisfies $\Delta(\mathbf{A}_t, \mathbf{a}) = 0$; see (11); this particular candidate $\mathbf{a} \in \mathcal{M}$ 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). \square

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

$$\begin{aligned} \frac{\sigma^2}{2} \cdot R_t &= \min_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq B_t}} \frac{\Delta(\mathbf{a}, \mathbf{A}_t)}{4} + \frac{\Delta(\mathbf{A}_t, \mathbf{B}^{[t]})}{4} \\ &= \left(- \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq B_t}} \frac{\Delta(\mathbf{A}_t, \mathbf{a})}{4} \right) + \frac{\Delta(\mathbf{A}_t, \mathbf{B}^{[t]})}{4}. \end{aligned} \quad (15)$$

To obtain the last equality in (15), we used the relationship $\Delta(\mathbf{A}_t, \mathbf{a}) = -\Delta(\mathbf{a}, \mathbf{A}_t)$ [see (11)]. Recall the symbol decision $B_t \triangleq B_0^{[t]}$, where $\mathbf{B}^{[t]}$ is defined in (8). 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

$$\begin{aligned} \frac{\sigma^2}{2} \cdot R_t &= \left(- \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 = A_t}} \frac{\Delta(\mathbf{A}_t, \mathbf{a})}{4} \right) + \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq A_t}} \frac{\Delta(\mathbf{A}_t, \mathbf{a})}{4} \\ &= -Y_t + X_t = |X_t - Y_t| \end{aligned}$$

where the second equality follows from (13), and the third from the fact $R_t \geq 0$ (see Definition 1). 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 a similar manner. \blacksquare

Expression (14) is developed for purposes of analysis and cannot be used to compute R_t . In practice, the quantities X_t and Y_t cannot never be computed, as they require knowledge of the transmitted sequence \mathbf{A}_t [see (13)]. Such knowledge is never available at the detector, because the detector is in fact trying to estimate \mathbf{A}_t . The simple Proposition 1, which seems completely overlooked in past literature, enables the derivation of the main result.

B. Expressions for Joint Reliability Distribution and Symbol Error Probability: Main Result

For any n number of arbitrarily chosen time instants t_1, t_2, \dots, t_n , we wish to obtain the distribution of the vector $\mathbf{R}_{t_1^n}$, containing the following reliabilities:

$$\mathbf{R}_{t_1^n} \triangleq [R_{t_1}, R_{t_2}, \dots, R_{t_n}]^T. \quad (16)$$

Recall $\mathbb{0}_{1,i}$ denotes a length i vector with all entries equal to 0. Define a *binary vector* \mathbf{e}_i of length $2(m + \ell) + 1$ as

$$\mathbf{e}_i \triangleq [\mathbb{0}_{1,m+\ell+i}, 1, \mathbb{0}_{1,m+\ell-i}]^T \quad (17)$$

where i can take values $|i| \leq m + \ell$. Further define the matrix \mathbf{E} of size $2(m + \ell) + 1$ by $2m$ as

$$\mathbf{E} \triangleq [\mathbf{e}_{-m}, \mathbf{e}_{-m+1}, \dots, \mathbf{e}_{-1}, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m]. \quad (18)$$

Let $\text{diag}(\mathbf{A}_t)$ denote the *diagonal matrix*, whose diagonal equals the vector \mathbf{A}_t . Define the following $2m + \ell + 1$ by 2^{2m} matrix:

$$\mathbf{G}(\mathbf{A}_t) \triangleq \mathbf{H} \text{diag}(\mathbf{A}_t) \mathbf{E} \quad (19)$$

where the noise neighborhood \mathbf{W}_t is given by (6). Let $\mathbf{W}_{\mathbf{t}_1^n}$ denote the concatenation

$$\mathbf{W}_{\mathbf{t}_1^n} \triangleq \begin{bmatrix} \mathbf{W}_{t_1} \\ \mathbf{W}_{t_2} \\ \vdots \\ \mathbf{W}_{t_n} \end{bmatrix}. \quad (20)$$

Define the noise covariance matrix

$$\mathbf{K}_W \triangleq \begin{bmatrix} \mathbb{E}\{\mathbf{W}_{t_1} \mathbf{W}_{t_1}^T\} & \cdots & \mathbb{E}\{\mathbf{W}_{t_1} \mathbf{W}_{t_n}^T\} \\ \vdots & \ddots & \vdots \\ \mathbb{E}\{\mathbf{W}_{t_n} \mathbf{W}_{t_1}^T\} & \cdots & \mathbb{E}\{\mathbf{W}_{t_n} \mathbf{W}_{t_n}^T\} \end{bmatrix} = \mathbb{E}\{\mathbf{W}_{\mathbf{t}_1^n} \mathbf{W}_{\mathbf{t}_1^n}^T\} \quad (21)$$

where note that \mathbf{K}_W is generally not Toeplitz even if \mathbf{W}_t is stationary. As in (20), let $\mathbf{A}_{\mathbf{t}_1^n}$ denote the concatenation

$$\mathbf{A}_{\mathbf{t}_1^n} \triangleq \begin{bmatrix} \mathbf{A}_{t_1} \\ \mathbf{A}_{t_2} \\ \vdots \\ \mathbf{A}_{t_n} \end{bmatrix}. \quad (22)$$

Let \mathbf{I} denote the identity matrix; in particular, \mathbf{I}_{2m} has size $2m \times 2m$. Define the matrix \mathbf{S} of size $2m \times 2^{2m}$ as

$$\mathbf{S} \triangleq [\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{2^{2m}-1}] \quad (23)$$

where the columns $\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{2^{2m}-1}$ make up all 2^{2m} possible, length-($2m$) binary vectors, i.e., $\{\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{2^{2m}-1}\} = \{0, 1\}^{2m}$. The matrix \mathbf{SS}^T has the following simple expression:

$$\mathbf{SS}^T = \sum_{k=0}^{2^{2m}-1} \mathbf{s}_k \mathbf{s}_k^T = 2^{2(m-1)} \cdot [\mathbf{I}_{2m} + \mathbf{1}\mathbf{1}^T] \quad (24)$$

where the vector $\mathbf{1}$ has all entries equal to 1. Denote the matrix *Kronecker product* using the operation \otimes . Let $\text{diag}(\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n}))$ denote a block-diagonal matrix, whose block-diagonal entries equal $\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n})$.

Definition 2: Let the square matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{\mathbf{t}_1^n})$ of size $2mn \times 2mn$ satisfy the following two conditions:

- i) the matrix \mathbf{Q} decomposes

$$\begin{aligned} \mathbf{Q}\Lambda^2\mathbf{Q}^T &= \text{diag}(\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n}))^T \mathbf{K}_W \\ &\quad \cdot \text{diag}(\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n})) \end{aligned} \quad (25)$$

where $\Lambda = \Lambda(\mathbf{A}_{\mathbf{t}_1^n})$ is a diagonal matrix. The number of positive diagonal elements in the matrix Λ , equals the rank of the matrix (25).

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

$$\mathbf{Q}^T (\mathbf{I}_n \otimes \mathbf{SS}^T) \mathbf{Q} = \mathbf{I} \quad (26)$$

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

Appendix A describes the computation of $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{\mathbf{t}_1^n})$, and $\Lambda = \Lambda(\mathbf{A}_{\mathbf{t}_1^n})$ in (25). We partition matrix \mathbf{Q} into n partitions of equal size $2m \times 2mn$, i.e.,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \vdots \\ \mathbf{Q}_n \end{bmatrix}. \quad (27)$$

Let $\text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n})$ denote the diagonal matrix, whose diagonal equals $[A_{t_1}, A_{t_2}, \dots, A_{t_n}]^T$. Define the $n \times 2mn$ matrix $\mathbf{F}(\mathbf{A}_{\mathbf{t}_1^n})$ as²

$$\begin{aligned} \mathbf{F}(\mathbf{A}_{\mathbf{t}_1^n}) &\triangleq \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes \mathbf{h}_0^T \mathbf{K}_W \\ &\quad \cdot \begin{bmatrix} \mathbf{G}(\mathbf{A}_{t_1}) & & & \mathbf{SS}^T \mathbf{Q}_1 \\ & \mathbf{G}(\mathbf{A}_{t_2}) & & \mathbf{SS}^T \mathbf{Q}_2 \\ & & \ddots & \vdots \\ & & & \mathbf{G}(\mathbf{A}_{t_n}) \end{bmatrix} \begin{bmatrix} \mathbf{SS}^T \mathbf{Q}_1 \\ \mathbf{SS}^T \mathbf{Q}_2 \\ \vdots \\ \mathbf{SS}^T \mathbf{Q}_n \end{bmatrix} \Lambda^\dagger \end{aligned} \quad (28)$$

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

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{A}_t) &= [\mu_1, \mu_2, \dots, \mu_{2^{2m}-1}]^T \\ &\triangleq [\mathbf{G}(\mathbf{A}_t) \mathbf{S}]^T \cdot \mathbf{T}(\mathbf{1} - \mathbf{A}_t) \\ &\quad - [| \mathbf{G}(\mathbf{A}_t) \mathbf{s}_0 |^2, | \mathbf{G}(\mathbf{A}_t) \mathbf{s}_1 |^2, \dots, | \mathbf{G}(\mathbf{A}_t) \mathbf{s}_{2^{2m}-1} |^2]^T \end{aligned} \quad (29)$$

$$\begin{aligned} \boldsymbol{\nu}(\mathbf{A}_t) &= [\nu_1, \nu_2, \dots, \nu_{2^{2m}-1}]^T \\ &\triangleq \boldsymbol{\mu}(\mathbf{A}_t) - 2\mathbf{A}_t \cdot \mathbf{h}_0^T \mathbf{G}(\mathbf{A}_t) \mathbf{S} \end{aligned} \quad (30)$$

where $\mu_k = \mu_k(\mathbf{A}_t)$ and $\nu_k = \nu_k(\mathbf{A}_t)$ denote the k th components of $\boldsymbol{\mu}_k(\mathbf{A}_t)$ and $\boldsymbol{\nu}_k(\mathbf{A}_t)$, 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:

$$\begin{aligned} \mathbf{X}_{\mathbf{t}_1^n} &\triangleq [X_{t_1}, X_{t_2}, \dots, X_{t_n}]^T \\ \mathbf{Y}_{\mathbf{t}_1^n} &\triangleq [Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}]^T \end{aligned} \quad (31)$$

where both X_{t_i} and Y_{t_i} are given in (13). Let \mathbb{R} denote the set of real numbers.

Theorem 1: Define the following quantities:

- 1) let \mathbf{U} denote a standard zero-mean identity-covariance Gaussian random vector of length-($2mn$).

²The matrix appearing in (28), with elements $\mathbf{G}(\mathbf{A}_{t_i})$, can also be written as $\text{diag}(\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n}))$.

2) let $\delta(\mathbf{U}, \mathbf{A}_{t_1^n}) = [\delta_1, \delta_2, \dots, \delta_n]^T$ denote a length- n vector in \mathbb{R}^n , which satisfies

$$\delta_i = \delta_i(\mathbf{U}, \mathbf{A}_{t_1^n}) \triangleq \max(\mathbf{S}^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} + \boldsymbol{\mu}(\mathbf{A}_{t_i})) - \max(\mathbf{S}^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} + \boldsymbol{\nu}(\mathbf{A}_{t_i})). \quad (32)$$

3) let $\boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n}) = [\eta_1, \eta_2, \dots, \eta_n]^T$ denote a length- n vector in \mathbb{R}^n , which satisfies

$$\begin{aligned} \boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n}) &\triangleq \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \mathbf{T} \\ &\cdot \left(\mathbf{1} \cdot \mathbf{1}^T - [A_{t_1}, A_{t_2}, \dots, A_{t_n}] \right)^T \mathbf{h}_0 \\ &- |\mathbf{h}_0|^2 \cdot \mathbf{1} + \mathbf{F}(\mathbf{A}_{t_1^n}) \mathbf{U}. \end{aligned} \quad (33)$$

4) let $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ denote an $n \times n$ matrix as follows:

$$\begin{aligned} \mathbf{K}_V(\mathbf{A}_{t_1^n}) &\triangleq \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes \mathbf{h}_0^T \mathbf{K}_W \\ &\cdot \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes \mathbf{h}_0 \\ &- \mathbf{F}(\mathbf{A}_{t_1^n}) \mathbf{F}(\mathbf{A}_{t_1^n})^T. \end{aligned} \quad (34)$$

Then, the distribution of $X_{t_1^n} - Y_{t_1^n}$ is given as

$$F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r}) = \mathbb{E} \left\{ \Phi_{\mathbf{K}_V(\mathbf{A}_{t_1^n})} (\mathbf{r} + \boldsymbol{\delta}(\mathbf{U}, \mathbf{A}_{t_1^n}) - \boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n})) \right\} \quad (35)$$

for all $\mathbf{r} \in \mathbb{R}^n$.

The proof of Theorem 1 is given in Section IV. Both i) the joint distribution of the reliabilities $\mathbf{R}_{t_1^n} \triangleq [R_{t_1}, R_{t_2}, \dots, R_{t_n}]^T$ in (16) and ii) the joint error probability $\Pr \{ \bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\} \}$ follow as corollaries from Theorem 1. In the following, we denote an index subset $\{\tau_1, \tau_2, \dots, \tau_j\} \subseteq \{t_1, t_2, \dots, t_n\}$ of size j , written compactly in vector form as $\boldsymbol{\tau}_1^j = [\tau_1, \tau_2, \dots, \tau_j]^T$.

Corollary 1: The distribution of $\mathbf{R}_{t_1^n} = 2/\sigma^2 \cdot |X_{t_1^n} - Y_{t_1^n}|$ (see Proposition 1) is given by

$$\begin{aligned} F_{\mathbf{R}_{t_1^n}}(\mathbf{r}) &= F_{|X_{t_1^n} - Y_{t_1^n}|}(\sigma^2/2 \cdot \mathbf{r}) \\ &= \sum_{j=0}^n \sum_{\substack{\{\tau_1, \tau_2, \dots, \tau_j\} \subseteq \\ \{t_1, t_2, \dots, t_n\}}} (-1)^j \cdot F_{X_{t_1^n} - Y_{t_1^n}} \left(\frac{\sigma^2}{2} \cdot \boldsymbol{\alpha}(\boldsymbol{\tau}_1^j, \mathbf{r}) \right) \end{aligned}$$

where the length- n vector $\boldsymbol{\alpha}(\boldsymbol{\tau}_1^j, \mathbf{r}) = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$ satisfies

$$\alpha_i = \alpha_i(\boldsymbol{\tau}_1^j, r_i) = \begin{cases} -r_i & \text{if } t_i \in \{\tau_1, \tau_2, \dots, \tau_j\} \\ r_i & \text{otherwise} \end{cases}$$

and $F_{X_{t_1^n} - Y_{t_1^n}} \left(\frac{\sigma^2}{2} \cdot \boldsymbol{\alpha}(\boldsymbol{\tau}_1^j, \mathbf{r}) \right)$ has the similar closed form as in Theorem 1. \square

Corollary 1 can be verified using recursion; for the n th case, we express

$$\begin{aligned} F_{|X_{t_1^n} - Y_{t_1^n}|}(\mathbf{r}) &= F_{|X_{t_1^{n-1}} - Y_{t_1^{n-1}}|, |X_{t_n} - Y_{t_n}|}(\mathbf{r}_1^{n-1}, r_n) \\ &- F_{|X_{t_1^{n-1}} - Y_{t_1^{n-1}}|, |X_{t_n} - Y_{t_n}|}(\mathbf{r}_1^{n-1}, -r_n). \end{aligned}$$

Observe that we still may apply Corollary 1 to each of the two terms on the r.h.s.; we apply Corollary 1 only to the variables $|X_{t_1^{n-1}} - Y_{t_1^{n-1}}|$, at the same time accounting for the (respective) joint events $\{X_{t_n} - Y_{t_n} \leq r_n\}$ and $\{X_{t_n} - Y_{t_n} \geq -r_n\}$.

The desired expression will be obtained after using some algebraic manipulations.

Procedure 1: Evaluate Joint Distribution $F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r})$

Initialize: Set $F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r}) := 0$ for all $\mathbf{r} \in \mathbb{R}^n$;

while $F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r})$ not converged **do**

1 Sample $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$ using $\Pr \{ \mathbf{A}_{t_1^n} = \mathbf{a}_1^n \}$. Sample the length- n , standard zero-mean identity-covariance Gaussian vector $\mathbf{U} = \mathbf{u}$;

2 Using the sampled realization $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$, obtain the matrices $\mathbf{Q} = \mathbf{Q}(\mathbf{a}_1^n)$ and $\mathbf{\Lambda} = \mathbf{\Lambda}(\mathbf{a}_1^n)$ satisfying Definition 2, see **Appendix A**;

3 Compute $\delta_i = \delta_i(\mathbf{u}, \mathbf{a}_1^n)$ for all $i \in \{1, 2, \dots, n\}$. For δ_i compute

$$\max_{k \in \{0, 1, \dots, 2^{2m}-1\}} \mathbf{s}_k^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{u} + \mu_k(\mathbf{a}),$$

$$\max_{k \in \{0, 1, \dots, 2^{2m}-1\}} \mathbf{s}_k^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{u} + \nu_k(\mathbf{a}),$$

see (32). Here \mathbf{a} is the sampled realization $\mathbf{A}_{t_1} = \mathbf{a}$, and both $\mu_k(\mathbf{a})$ and $\nu_k(\mathbf{a})$ are the k th components of $\boldsymbol{\mu}(\mathbf{a})$ and $\boldsymbol{\nu}(\mathbf{a})$, see (29) and (30);

4 Compute $\mathbf{F}(\mathbf{A}_{t_1^n})$ in (28); Also compute $\boldsymbol{\eta}(\mathbf{u}, \mathbf{a}_1^n)$ in (33) and $\mathbf{K}_V(\mathbf{a}_1^n)$ in (34);

5 Update

$$\begin{aligned} F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r}) \\ := F_{X_{t_1^n} - Y_{t_1^n}}(\mathbf{r}) + \Phi_{\mathbf{K}_V(\mathbf{a}_1^n)} (\mathbf{r} + \boldsymbol{\delta}(\mathbf{u}, \mathbf{a}_1^n) - \boldsymbol{\eta}(\mathbf{u}, \mathbf{a}_1^n)) \end{aligned}$$

for all $\mathbf{r} \in \mathbb{R}^n$;

end

Corollary 2: The probability $\Pr \{ \bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\} \}$ that all symbol decisions $B_{t_1}, B_{t_2}, \dots, B_{t_n}$ are in error equals

$$\begin{aligned} \Pr \left\{ \bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\} \right\} &= \Pr \{ X_{t_1^n} \geq Y_{t_1^n} \} \\ &= 1 + \sum_{j=1}^n \sum_{\substack{\{\tau_1, \tau_2, \dots, \tau_j\} \subseteq \\ \{t_1, t_2, \dots, t_n\}}} (-1)^j \cdot F_{X_{\boldsymbol{\tau}_1^j} - Y_{\boldsymbol{\tau}_1^j}}(\mathbf{0}) \end{aligned}$$

where the probability

$$F_{X_{\boldsymbol{\tau}_1^j} - Y_{\boldsymbol{\tau}_1^j}}(\mathbf{0}) = \Pr \left\{ \bigcap_{\tau \in \{\tau_1, \tau_2, \dots, \tau_j\}} \{X_\tau - Y_\tau \leq 0\} \right\}$$

has a similar closed form as in Theorem 1. \square

Proof: From (13), we clearly see that the event $\{X_t \geq Y_t\}$ indicates that the sequence $\mathbf{B}^{[t]}$ in (8) will have its zeroth component $B_0^{[t]} \neq A_t$. Because the symbol decision B_t is set to $B_t = B_0^{[t]}$ [where $\mathbf{B}^{[t]}$ is defined in (8)], the event $\{X_t \geq Y_t\}$ indicates that $B_t \neq A_t$, which is exactly a symbol decision error occurring at time t . \blacksquare

Denote the realizations of $\mathbf{A}_{t_1^n}$, \mathbf{A}_t and \mathbf{U} , as $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$, and $\mathbf{A}_t = \mathbf{a}$, and $\mathbf{U} = \mathbf{u}$. The Monte-Carlo procedure used to evaluate the closed form of $F_{\mathbf{X}_{t_1^n}-\mathbf{Y}_{t_1^n}}(\mathbf{r})$ in Theorem 1 is given in Procedure 1. We may reduce the number of computations used to obtain matrices $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{t_1^n})$ and $\Lambda = \Lambda(\mathbf{A}_{t_1^n})$ in Line 3, by sampling $\mathbf{U} = \mathbf{u}$ multiple times for a fixed $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$.

Remark 1: The matrix $\mathbf{K}_V(\mathbf{a}_1^n)$ computed in Line 5 [see also (34)] may not have full rank. Hence, when evaluating the Gaussian distribution function $\Phi_{\mathbf{K}_V(\mathbf{a}_1^n)}(\mathbf{r})$ with covariance matrix $\mathbf{K}_V(\mathbf{a}_1^n)$ in Line 6, we may require techniques designed for rank deficient covariances (see, for example, [13]).

Our proposed method requires no assumptions on the noise covariance matrix \mathbf{K}_W in (21) and can be applied even when the noise W_t is correlated and/or nonstationary. However, there is an implicit assumption that $\mathbf{A}_{t_1^n}$ is equally likely among all realizations $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$ that have nonzero probability. Further modifications will be required to extend our method to the general case of nonuniform priors $\Pr\{\mathbf{A}_{t_1^n} = \mathbf{a}_1^n\}$ (the first equality of (9) is not valid for such cases).

Remark 2: Because $\Phi_{\mathbf{K}}(\mathbf{r})$ is a probability distribution function, therefore

$$0 \leq \Phi_{\mathbf{K}_V(\mathbf{A}_{t_1^n})}(\mathbf{r} + \boldsymbol{\delta}(\mathbf{U}, \mathbf{A}_{t_1^n}) - \boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n})) \leq 1.$$

The well-known Hoeffding probability inequalities can be applied to obtain convergence guarantees (see [16]).

The main thrust of Section III-C is to address Line 4 of Procedure 1. It appears that to execute Line 4 of Procedure 1, we require an exhaustive search over 2^{2m} terms to perform the two maximizations. However, we point out in the next section that these maximizations can be performed more efficiently by utilizing dynamic programming optimization techniques. Also, in the next section, we address the computation of $F_{\mathbf{X}_{t_1^n}-\mathbf{Y}_{t_1^n}}(\mathbf{r})$, in instances where one wishes to only consider a subset $\mathcal{M} \subset \mathcal{M}$ [see (7)].

C. On Efficient Computation of the Closed-Form Expressions

To compute δ_i in (32) while executing Line 4 of Procedure 1, we need to perform the following two maximizations:

$$\begin{aligned} & \max_{\mathbf{s} \in \{0,1\}^{2m}} \mathbf{s}^T \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + [\mathbf{G}(\mathbf{a}) \mathbf{s}]^T \cdot \mathbf{T}(\mathbf{1} - \mathbf{a}) - |\mathbf{G}(\mathbf{a}) \mathbf{s}|^2 \\ & \max_{\mathbf{s} \in \{0,1\}^{2m}} \mathbf{s}^T \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + [\mathbf{G}(\mathbf{a}) \mathbf{s}]^T \cdot [\mathbf{T}(\mathbf{1} - \mathbf{a}) - 2a_0 \cdot \mathbf{h}_0] \\ & \quad - |\mathbf{G}(\mathbf{a}) \mathbf{s}|^2 \end{aligned} \quad (36)$$

where both \mathbf{a} and \mathbf{u} are realizations $\mathbf{A}_{t_i} = \mathbf{a}$ and $\mathbf{U} = \mathbf{u}$. Note that we obtain (36) from (32) by substituting for both $\boldsymbol{\mu}(\mathbf{a})$ and $\boldsymbol{\nu}(\mathbf{a})$ using (29) and (30), respectively. Index the realization $\mathbf{A}_{t_i} = \mathbf{a}$ as

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

Let $\text{diag}(\mathbf{a})$ denote the diagonal matrix, with diagonal \mathbf{a} . Let \mathbf{g}_τ denote the length $2(m + \ell) + 1$ vector

$$\mathbf{g}_\tau \triangleq [\mathbf{0}_{1,m+\tau}^T, h_\ell a_{\tau-\ell}, h_{\ell-1} a_{\tau-(\ell-1)}, \dots, h_0 a_\tau, \mathbf{0}_{1,m+\ell-\tau}^T]^T$$

where τ can take values $\tau \in \{-m, -, m + 1, \dots, m + \ell\}$. We rewrite $\mathbf{G}(\mathbf{a})$ as

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

Recall the definition of $\mathbf{G}(\mathbf{a})$ from (19). From the observed structure of \mathbf{g}_τ , it can be clearly seen from (37) that $\mathbf{G}(\mathbf{a})$ is a *sparse matrix* with many zero entries. The matrix $\mathbf{G}(\mathbf{a})$ 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].

Procedure 2: Solve $\max_{\mathbf{s} \in \{0,1\}^{2m}} \mathbf{s}^T \mathbf{C} - |\mathbf{G}(\mathbf{a}) \mathbf{s}|^2$ using Dynamic Programming

Convention: Set $\mathcal{C}_0 := -\infty$ and also set values $\mathcal{C}_j := 0$ for all $|j| > m$;

Denote the length- ℓ binary vector by $\bar{\mathbf{s}} \triangleq [\bar{s}_{\ell-1}, \bar{s}_{\ell-2}, \dots, \bar{s}_0]^T$;

Input: Matrix $\mathbf{G}(\mathbf{a})$; Vector of constants $\mathbf{C} = [\mathcal{C}_{-m}, \mathcal{C}_{-m+1}, \dots, \mathcal{C}_{-1}, \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m]^T$;

Output: Value stored in $\beta_{m+\ell}(\bar{\mathbf{s}}) = \beta_{m+\ell}(\mathbf{0})$;

Initialize: For all $\bar{\mathbf{s}} \in \{0,1\}^\ell$, set the values

$$\beta_{-m-1}(\bar{\mathbf{s}}) := \begin{cases} 0 & \text{if } \bar{\mathbf{s}} = \mathbf{0}, \\ -\infty & \text{otherwise.} \end{cases}$$

forall the $\tau \in \{-m, -m + 1, \dots, m + \ell\}$ **do**

forall the $\bar{\mathbf{s}} \in \{0,1\}^\ell$ **do**

1 Set the value $\alpha = \alpha(\bar{\mathbf{s}}) := \sum_{j=0}^{\ell-1} h_j a_{\tau-j} \bar{s}_j$. Set the states \bar{s}_0 and \bar{s}_1 as

$$\begin{aligned} \bar{s}_0 &:= [0, \bar{s}_{\ell-1}, \dots, \bar{s}_2, \bar{s}_1]^T, \\ \bar{s}_1 &:= [1, \bar{s}_{\ell-1}, \dots, \bar{s}_2, \bar{s}_1]^T; \end{aligned}$$

2 Compute $\beta_\tau(\bar{\mathbf{s}}) := \max\{-\alpha^2 + \beta_{\tau-1}(\bar{s}_0), \mathcal{C}_{\tau-\ell} - [h_\ell a_{\tau-\ell} + \alpha]^2 + \beta_{\tau-1}(\bar{s}_1)\}$;

end

end

It is clear that the inner product $\mathbf{g}_\tau^T \mathbf{e}_j$ extracts the j th component of the vector \mathbf{g}_τ^T , i.e.,

$$\mathbf{g}_\tau^T \mathbf{e}_{\tau-j} = \begin{cases} h_j \cdot a_{\tau-j} & \text{if } 0 \leq j \leq \ell \\ 0 & \text{otherwise} \end{cases} \quad (38)$$

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

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

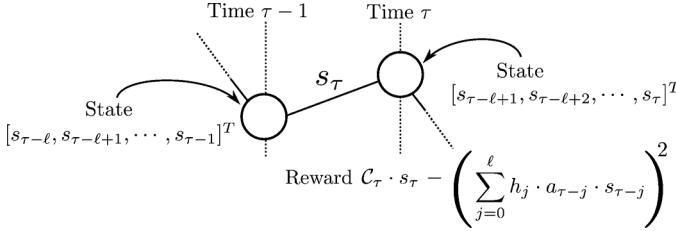


Fig. 2. Time evolution of the dynamic programming states.

It is clear that by using (38), the following is true for all vectors $\mathbf{g}_\tau^T \mathbf{E}\mathbf{s}$:

$$\begin{aligned} \mathbf{g}_\tau^T \mathbf{E}\mathbf{s} &= \sum_{j=-m-\ell}^{m+\ell} (\mathbf{g}_\tau^T \mathbf{e}_j) \cdot s_j \\ &= \sum_{j=0}^{\ell} h_j \cdot a_{\tau-j} \cdot s_{\tau-j} \end{aligned} \quad (39)$$

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

Define the length-($2m$) vector $\mathcal{C} \triangleq [\mathcal{C}_{-m}, \mathcal{C}_{-m+1}, \dots, \mathcal{C}_{-1}, \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m]^T$. Set $\mathcal{C}_0 := -\infty$ and $\mathcal{C}_\tau := 0$ for all $|\tau| > m$. By setting

$$\mathcal{C} := \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + [\mathbf{G}(\mathbf{a})]^T \cdot \mathbf{T}(\mathbb{1} - \mathbf{a})$$

and

$$\mathcal{C} := \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + [\mathbf{G}(\mathbf{a})]^T \cdot [\mathbf{T}(\mathbb{1} - \mathbf{a}) - 2a_0 \cdot \mathbf{h}_0]$$

respectively, we can solve both problems (36) as

$$\begin{aligned} \max_{\mathbf{s} \in \{0,1\}^{2m}} \mathbf{s}^T \mathcal{C} - |\mathbf{G}(\mathbf{a}) \mathbf{s}|^2 \\ = \max_{\mathbf{s} \in \{0,1\}^{2m}} \sum_{\tau=-m}^{m+\ell} \mathcal{C}_\tau \cdot s_\tau - (\mathbf{g}_\tau^T \mathbf{E}\mathbf{s})^2 \end{aligned} \quad (40)$$

where the τ th term is $\mathbf{g}_\tau^T \mathbf{E}\mathbf{s} = \sum_{j=0}^{\ell} h_j a_{\tau-j} s_{\tau-j}$. For the sake of completeness, we provide the dynamic programming procedure that solves (40). The dynamic programming state at time τ equals the length- ℓ vector of binary symbols $[s_{\tau-\ell+1}, s_{\tau-\ell+2}, \dots, s_\tau]^T \in \{0,1\}^\ell$. For the benefit of readers knowledgeable in dynamic programming techniques, we illustrate the time evolution of the dynamic programming states in Fig. 2. Dynamic programs can be solved with a complexity that is *linear* in the state size [2]; in our case, we have 2^ℓ states. The dynamic programming procedure optimizing (40) is given in Procedure 2.

The second part of this subsection addresses the following separate issue. Consider the case where some of the probabilities $\Pr\{\mathbf{A}_t = \mathbf{a}\}$ equal 0; one example of such a case is where a modulation code is present in the system [14], [15]. In these cases, we consider the subset $\bar{\mathcal{M}} \subset \mathcal{M}$, explicitly written as

$$\bar{\mathcal{M}} = \bar{\mathcal{M}}_t \triangleq \left\{ \mathbf{a} \in \mathcal{M} : \Pr \left\{ \bigcap_{j=-m}^m \{A_{t+j} = a_j\} \right\} = 0 \right\} \quad (41)$$

for each time instant t .

If we consider the subsets $\bar{\mathcal{M}} \subset \mathcal{M}$, 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. Note that 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.

Procedure 3: Evaluating $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r})$, for candidate subsets $\bar{\mathcal{M}} \subset \mathcal{M}$, see (41)

Initialize: Set $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r}) := 0$ for all $\mathbf{r} \in \mathbb{R}^n$;

while $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r})$ not converged **do**

1 Perform Lines 2–3 of Procedure 1;

2 Compute $\delta_i = \delta_i(\mathbf{u}, \mathbf{a}_1^n)$ for all $i \in \{1, 2, \dots, n\}$ by computing

$$\begin{aligned} \max_{k: \mathbf{a}(\mathbf{E}\mathbf{s}_k, \mathbf{a}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + \mu_k(\mathbf{a}), \\ \max_{k: \mathbf{a}(\mathbf{E}\mathbf{s}_k + \mathbf{e}_0, \mathbf{a}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T \mathbf{Q}_i \boldsymbol{\Lambda} \mathbf{u} + \nu_k(\mathbf{a}), \end{aligned}$$

see (32), where $\mu_k(\mathbf{a})$ and $\nu_k(\mathbf{a})$ denote the k th components of $\boldsymbol{\mu}(\mathbf{a})$ and $\boldsymbol{\nu}(\mathbf{a})$, see (29) and (30). Both \mathbf{E} and \mathbf{e}_0 are given in (17) and (18). Also, the vector $\mathbf{a}(\mathbf{e}, \mathbf{a}) = [\alpha_{-m-\ell}, \alpha_{-m-(\ell-1)}, \dots, \alpha_{m+\ell}]^T$ satisfies

$$\alpha_j = \alpha_j(e_j, a_j) = \begin{cases} -a_j & \text{if } e_j = 1, \\ a_j & \text{if } e_j = 0. \end{cases}$$

3 Perform Lines 5–6 of Procedure 1;

end

IV. PROOF OF THEOREM 1 AND SOME COMMENTS

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

$$V_t \triangleq A_t \cdot \mathbf{h}_0^T \mathbf{W}_t. \quad (42)$$

It is easy to verify that V_t is Gaussian: recall that $\mathbf{W}_t \triangleq [W_{t-M}, W_{t-M+1}, \dots, W_{t+M+I}]^T$ is the neighborhood of (Gaussian) noise samples. To improve clarity, we shall introduce the following new notation, both used only in this section:

$$\begin{aligned} \theta(\mathbf{A}_t) &\triangleq A_t \cdot [\mathbf{T}(\mathbb{1} - \mathbf{A}_t)]^T \mathbf{h}_0 - |\mathbf{h}_0|^2 \\ \mathbf{F}(\mathbf{A}_{t_1^n}) &\triangleq \text{diag}(\mathbf{G}(\mathbf{A}_{t_1}), \mathbf{G}(\mathbf{A}_{t_2}), \dots, \mathbf{G}(\mathbf{A}_{t_n})). \end{aligned} \quad (43)$$

Using (43), we may now more compactly write

$$\begin{aligned} \mathbf{Q} \boldsymbol{\Lambda}^2 \mathbf{Q}^T &= \mathbf{F}^T \mathbf{K}_W \mathbf{F}, \\ \mathbf{F}(\mathbf{A}_{t_1^n}) &= \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes \mathbf{h}_0^T \mathbf{K}_W \mathbf{F} \\ &\quad \cdot [\mathbf{I}_n \otimes \mathbf{S} \mathbf{S}^T] \cdot \mathbf{Q} \boldsymbol{\Lambda}^\dagger, \\ \boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n}) &= [\theta(\mathbf{A}_{t_1}), \theta(\mathbf{A}_{t_2}), \dots, \theta(\mathbf{A}_{t_n})]^T + \mathbf{F}(\mathbf{A}_{t_1^n}) \mathbf{U} \end{aligned} \quad (44)$$

where $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{t_1^n})$ and $\Lambda = \Lambda(\mathbf{A}_{t_1^n})$ are given in Definition 2, matrix $\mathbf{F}(\mathbf{A}_{t_1^n})$ in (28), and $\boldsymbol{\eta}(\mathbf{U}, \mathbf{A}_{t_1^n})$ in (33).

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

$$\begin{aligned} X_t &= \max ([\mathbf{G}(\mathbf{A}_t)\mathbf{S}]^T \mathbf{W}_t + \boldsymbol{\nu}(\mathbf{A}_t) + [V_t + \theta(\mathbf{A}_t)] \cdot \mathbf{1}) \\ Y_t &= \max ([\mathbf{G}(\mathbf{A}_t)\mathbf{S}]^T \mathbf{W}_t + \boldsymbol{\mu}(\mathbf{A}_t)) \end{aligned}$$

where $\theta(\mathbf{A}_t) \triangleq A_t \cdot [\mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T \mathbf{h}_0 - |\mathbf{h}_0|^2$ as given in (43). \square

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

$$\begin{aligned} \Delta(\mathbf{A}_t, \mathbf{a}) &= |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\mathbf{A}_t|^2 - |\mathbf{Z}_t - \mathbf{T}\mathbf{1} - \mathbf{H}\mathbf{a}|^2 \\ &= |-\mathbf{W}_t + \mathbf{T}(\mathbf{A}_t - \mathbf{1})|^2 \\ &\quad - |-\mathbf{W}_t + \mathbf{T}(\mathbf{A}_t - \mathbf{1}) + \mathbf{H}(\mathbf{A}_t - \mathbf{a})|^2 \\ &= -2[-\mathbf{W}_t + \mathbf{T}(\mathbf{A}_t - \mathbf{1})]^T \mathbf{H}(\mathbf{A}_t - \mathbf{a}) - |\mathbf{H}(\mathbf{A}_t - \mathbf{a})|^2. \quad (45) \end{aligned}$$

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

$$\begin{aligned} X_t &= \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 \neq A_t}} [\mathbf{W}_t + \mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T \left(\frac{1}{2} \cdot \mathbf{H}(\mathbf{A}_t - \mathbf{a}) \right) \\ &\quad - \left| \frac{1}{2} \cdot \mathbf{H}(\mathbf{A}_t - \mathbf{a}) \right|^2 \\ Y_t &= \max_{\substack{\mathbf{a} \in \mathcal{M} \\ a_0 = A_t}} [\mathbf{W}_t + \mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T \left(\frac{1}{2} \cdot \mathbf{H}(\mathbf{A}_t - \mathbf{a}) \right) \\ &\quad - \left| \frac{1}{2} \cdot \mathbf{H}(\mathbf{A}_t - \mathbf{a}) \right|^2. \quad (46) \end{aligned}$$

Using (17) and Definitions (7), (18), and (23), we establish the following equality of sets:

$$\begin{aligned} \left\{ \frac{1}{2}(A_t - \mathbf{a}) : \mathbf{a} \in \mathcal{M}, a_0 \neq A_t \right\} &= \{ \text{diag}(\mathbf{A}_t) \mathbf{E} \mathbf{s}_j + A_t \cdot \mathbf{e}_0 : 0 \leq j \leq 2^{2m} - 1 \} \\ \left\{ \frac{1}{2}(A_t - \mathbf{a}) : \mathbf{a} \in \mathcal{M}, a_0 = A_t \right\} &= \{ \text{diag}(\mathbf{A}_t) \mathbf{E} \mathbf{s}_j : 0 \leq j \leq 2^{2m} - 1 \}. \quad (47) \end{aligned}$$

Next, we utilize both (46) and (19) to rewrite (45) as

$$\begin{aligned} X_t &= \max_{j \in \{0, 1, \dots, 2^{2m}-1\}} [\mathbf{W}_t + \mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T [\mathbf{G}(\mathbf{A}_t) \mathbf{s}_j + A_t \mathbf{h}_0] \\ &\quad - |\mathbf{G}(\mathbf{A}_t) \mathbf{s}_j + A_t \mathbf{h}_0|^2 \\ Y_t &= \max_{j \in \{0, 1, \dots, 2^{2m}-1\}} [\mathbf{W}_t + \mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T [\mathbf{G}(\mathbf{A}_t) \mathbf{s}_j] \\ &\quad - |\mathbf{G}(\mathbf{A}_t) \mathbf{s}_j|^2. \quad (48) \end{aligned}$$

By the definition of $\boldsymbol{\mu}(\mathbf{A}_t)$ in (29) and \mathbf{S} in Definition (23), the expression for Y_t in the proposition statement follows from (48). For X_t , we continue to expand (48) to get

$$\begin{aligned} X_t &= \max \left([\mathbf{G}(\mathbf{A}_t)\mathbf{S}]^T \mathbf{W}_t + \overbrace{\boldsymbol{\mu}(\mathbf{A}_t) - 2A_t \cdot \mathbf{h}_0^T \mathbf{G}(\mathbf{A}_t)\mathbf{S}}^{\boldsymbol{\nu}(\mathbf{A}_t)} \right. \\ &\quad \left. + \underbrace{A_t \cdot \mathbf{h}_0^T \mathbf{W}_t \cdot \mathbf{1}}_{V_t} + \underbrace{\{A_t [\mathbf{T}(\mathbf{1} - \mathbf{A}_t)]^T \mathbf{h}_0 - |\mathbf{h}_0|^2\} \cdot \mathbf{1}}_{\theta(\mathbf{A}_t)} \right) \end{aligned}$$

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

Recall $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{t_1^n})$ and $\Lambda = \Lambda(\mathbf{A}_{t_1^n})$ from Definition 2. To prove Theorem 1, we require the following lemma.

Lemma 1: Let \mathbf{U} denote a standard zero-mean identity-covariance Gaussian random vector of length-($2mn$). Recall $\mathbf{W}_{t_1^n}$ in (20). The following transformation of random vectors holds:

$$\begin{aligned} &\begin{bmatrix} \mathbf{S}^T \mathbf{Q}_1(\mathbf{A}_{t_1^n}) \\ \mathbf{S}^T \mathbf{Q}_2(\mathbf{A}_{t_1^n}) \\ \vdots \\ \mathbf{S}^T \mathbf{Q}_n(\mathbf{A}_{t_1^n}) \end{bmatrix} \Lambda(\mathbf{A}_{t_1^n}) \mathbf{U} \\ &= \begin{bmatrix} \mathbf{G}(\mathbf{A}_{t_1})\mathbf{S} & & & \\ & \mathbf{G}(\mathbf{A}_{t_2})\mathbf{S} & & \\ & & \ddots & \\ & & & \mathbf{G}(\mathbf{A}_{t_n})\mathbf{S} \end{bmatrix}^T \begin{bmatrix} \mathbf{W}_{t_1} \\ \mathbf{W}_{t_2} \\ \vdots \\ \mathbf{W}_{t_n} \end{bmatrix} \quad (49) \end{aligned}$$

or more concisely, we equivalently write

$$\begin{aligned} &(\mathbf{I}_n \otimes \mathbf{S}^T) \cdot \mathbf{Q}(\mathbf{A}_{t_1^n}) \Lambda(\mathbf{A}_{t_1^n}) \mathbf{U} \\ &= (\mathbf{I}_n \otimes \mathbf{S}^T) \cdot \Gamma(\mathbf{A}_{t_1^n})^T \mathbf{W}_{t_1^n} \quad (50) \end{aligned}$$

using $\mathbf{Q}(\mathbf{A}_{t_1^n})$ in (27) and $\Gamma(\mathbf{A}_{t_1^n})$ in (43). \square

Proof: After conditioning on $\mathbf{A}_{t_1^n}$, both vectors that appear on either side of (50) are seen to be zero-mean Gaussian random vectors (recall that \mathbf{W}_t is zero mean). Therefore, to prove the lemma, we only need to verify that after conditioned on $\mathbf{A}_{t_1^n}$, both l.h.s. and r.h.s. of (50) have the same covariance matrix. This is easily done by using property i) of $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{t_1^n})$ in Definition 2, which yields

$$\begin{aligned} \mathbb{E} \{ \mathbf{Q} \Lambda \mathbf{U} \mathbf{U}^T \Lambda \mathbf{Q}^T \mid \mathbf{A}_{t_1^n} \} &= \mathbf{Q}(\mathbf{A}_{t_1^n}) \Lambda(\mathbf{A}_{t_1^n})^2 \mathbf{Q}(\mathbf{A}_{t_1^n})^T \\ &= \Gamma(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \Gamma(\mathbf{A}_{t_1^n}). \end{aligned}$$

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

- 1) $\text{rank}[\Gamma(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \Gamma(\mathbf{A}_{t_1^n})] = 2mn$, and
- 2) $\text{rank}[\Gamma(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \Gamma(\mathbf{A}_{t_1^n})] < 2mn$ for some realization $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$.

We begin with the first case.

Proof of Theorem 1 when $\text{rank}(\Gamma(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \Gamma(\mathbf{A}_{t_1^n})) = 2mn$: We first derive the following equalities:

$$\begin{aligned} &(\Lambda^\dagger \mathbf{Q}^T)(\mathbf{I}_n \otimes \mathbf{S}^T) \Gamma(\mathbf{A}_{t_1^n})^T \mathbf{W}_{t_1^n} \\ &= (\Lambda^\dagger \mathbf{Q}^T)(\mathbf{I}_n \otimes \mathbf{S}^T) \mathbf{Q} \Lambda \mathbf{U} \\ &= \Lambda^\dagger \Lambda \mathbf{U} = \mathbf{U}. \quad (51) \end{aligned}$$

The first two equalities follow by, respectively, applying properties i) and ii) of the matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{A}_{t_1^n})$. The last equality holds because by virtue of the assumption $\text{rank}(\Gamma(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \Gamma(\mathbf{A}_{t_1^n})) = 2mn$, in which Λ^\dagger is strictly an inverse of Λ . Recall both $V_{t_i} \triangleq A_{t_i} \cdot \mathbf{h}_0^T \mathbf{W}_{t_i}$ and $\mathbf{V}_{t_1^n} \triangleq [V_{t_1}, V_{t_2}, \dots, V_{t_n}]^T$. Taking (51) together with (42), we have the following transformation:

$$\begin{bmatrix} \mathbf{V}_{t_1^n} \\ \mathbf{U} \end{bmatrix} = \begin{bmatrix} \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes \mathbf{h}_0^T \\ (\Lambda^\dagger \mathbf{Q}^T)(\mathbf{I}_n \otimes \mathbf{S}^T) \Gamma(\mathbf{A}_{t_1^n})^T \end{bmatrix} \mathbf{W}_{t_1^n}. \quad (52)$$

Consider the conditional event

$$\{X_{t_1^n} - Y_{t_1^n} \leq r | A_{t_1^n}, U\} \quad (53)$$

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

$$S^T Q_i(A_{t_1^n}) \Lambda(A_{t_1^n}) U = [G(A_{t_i}) S]^T W_{t_i}.$$

Therefore, we may rewrite both X_{t_i} and Y_{t_i} from Proposition 2 as

$$\begin{aligned} X_{t_i} &= \max(S^T Q_i \Lambda U + \nu(A_{t_i})) + V_{t_i} + \theta(A_{t_i}) \\ Y_{t_i} &= \max(S^T Q_i \Lambda U + \mu(A_{t_i})). \end{aligned} \quad (54)$$

The event (53) can then be written as

$$\begin{aligned} \{X_{t_1^n} - Y_{t_1^n} \leq r | A_{t_1^n}, U\} &= \bigcap_{1 \leq i \leq n} \{X_{t_i} \leq r_i + Y_{t_i} | A_{t_1^n}, U\} \\ &= \bigcap_{1 \leq i \leq n} \left\{ \max \left(\begin{array}{l} [S^T Q_i \Lambda U + \nu(A_{t_i})] \\ + V_{t_i} + \theta(A_{t_i}) \end{array} \right) \leq r_i + Y_{t_i} | A_{t_1^n}, U \right\} \\ &= \bigcap_{1 \leq i \leq n} \left\{ V_{t_i} + \theta(A_{t_i}) \leq \left(\begin{array}{l} r_i + \max[S^T Q_i \Lambda U + \mu(A_{t_i})] \\ - \min[S^T Q_i \Lambda U + \nu(A_{t_i})] \end{array} \right) | A_{t_1^n}, U \right\}. \end{aligned} \quad (55)$$

Continuing from (55), we utilize (32) to rewrite

$$\begin{aligned} \{X_{t_1^n} - Y_{t_1^n} \leq r | A_{t_1^n}, U\} &= \bigcap_{1 \leq i \leq n} \{V_{t_i} + \theta(A_{t_i}) \leq r_i + \delta_i(U, A_{t_1^n}) | A_{t_1^n}, U\}. \end{aligned} \quad (56)$$

We now determine both the mean and variance of $V_{t_1^n}$, after conditioning on both $A_{t_1^n}$ and U . From (52), we derive the formula

$$\begin{aligned} \mathbb{E}\{V_{t_1^n} U^T | A_{t_1^n}\} &= \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes h_0^T K_W \\ &\quad \cdot \Gamma(A_{t_1^n})(I_n \otimes SS^T) Q \Lambda^\dagger \\ &\triangleq F(A_{t_1^n}) \end{aligned} \quad (57)$$

where $F(A_{t_1^n})$ is given in (28). Next, we compute the conditional mean

$$\begin{aligned} \mathbb{E}\{V_{t_1^n} | A_{t_1^n}, U\} &= \mathbb{E}\{V_{t_1^n} | A_{t_1^n}\} + \mathbb{E}\{V_{t_1^n} U^T | A_{t_1^n}\} U \\ &= F(A_{t_1^n}) U \end{aligned} \quad (58)$$

where the second equality follows from $\mathbb{E}\{V_{t_1^n} | A_{t_1^n}\} = 0$ [because $W_{t_1^n}$ has zero mean, see (42)], and substituting (57). The conditional covariance matrix $\text{Cov}\{V_{t_1^n} | A_{t_1^n}, U\}$ is obtained as follows:

$$\begin{aligned} \text{Cov}\{V_{t_1^n} | A_{t_1^n}, U\} &= \mathbb{E}\{V_{t_1^n} V_{t_1^n}^T | A_{t_1^n}\} - \mathbb{E}\{V_{t_1^n} U^T | A_{t_1^n}\} \cdot \mathbb{E}\{U V_{t_1^n}^T | A_{t_1^n}\} \\ &= \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes h_0^T K_W \\ &\quad \cdot \text{diag}(A_{t_1}, A_{t_2}, \dots, A_{t_n}) \otimes h_0 - F(A_{t_1^n}) F(A_{t_1^n})^T \\ &\triangleq K_V(A_{t_1^n}) \end{aligned} \quad (59)$$

where $K_V(A_{t_1^n})$ is given in (34). The expression for $F_{X_{t_1^n} - Y_{t_1^n}}(r)$ in Theorem 1 now follows easily from (56)

$$\begin{aligned} \{X_{t_1^n} - Y_{t_1^n} \leq r | A_{t_1^n}, U\} &= \left\{ V_{t_1^n} + [\theta(A_{t_1}), \theta(A_{t_2}), \dots, \theta(A_{t_n})]^T \right. \\ &\quad \left. \leq r + \delta(U, A_{t_1^n}) \middle| A_{t_1^n}, U \right\} \end{aligned}$$

and noticing that the random vector

$$V_{t_1^n} + [\theta(A_{t_1}), \theta(A_{t_2}), \dots, \theta(A_{t_n})]^T \quad (60)$$

is (conditionally on $A_{t_1^n}$ and U) Gaussian distributed with distribution function

$$\Phi_{K_V(A_{t_1^n})}(r - \eta(U, A_{t_1^n}))$$

where both the conditional mean and covariance $\eta(U, A_{t_1^n})$ and $K_V(A_{t_1^n})$ are given, respectively, in (58) and (59). ■

Next we consider the other case where the rank of $\Gamma(A_{t_1^n})^T K_W \Gamma(A_{t_1^n}) < 2mn$ for some value of $A_{t_1^n} = a_1^n$. In this case, the arguments of the preceding proof fail in (51), where the final equality does not hold because Λ^\dagger is strictly not the inverse of Λ . However, as we soon shall see, the expression for $F_{X_{t_1^n} - Y_{t_1^n}}(r)$ in Theorem 1 still holds for this case.

Proof of Theorem 1 when $\text{rank}(\Gamma(A_{t_1^n})^T K_W \Gamma(A_{t_1^n})) < 2mn$ for some $A_{t_1^n} = a_1^n$: Recall that the matrix $[\Lambda(A_{t_1^n})]^\dagger = \Lambda^\dagger$ is formed by only reciprocating the nonzero diagonal elements of $\Lambda(A_{t_1^n}) = \Lambda$. For a particular realization $A_{t_1^n} = a_1^n$, let the value $j = \text{rank}(\Gamma(A_{t_1^n})^T K_W \Gamma(A_{t_1^n}))$ equal the rank of the matrix $\Gamma(A_{t_1^n})^T K_W \Gamma(A_{t_1^n})$. Consider what happens if $j < 2mn$. Without loss of generality, assume that all nonzero diagonal elements of $\Lambda(A_{t_1^n}) = \Lambda$ are located at the first $j < 2mn$ diagonal elements of Λ . Define the following size- j quantities:

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

If we substitute the new quantities U_1^j , \bar{Q} , and $\bar{\Lambda}$ for U , Q , and Λ in (51), it is clear that (51) holds true, i.e.,

$$\begin{aligned} (\bar{\Lambda}^\dagger \bar{Q}^T)(I_n \otimes SS^T) \Gamma(A_{t_1^n})^T W_{t_1^n} &= (\bar{\Lambda}^\dagger \bar{Q}^T)(I_n \otimes SS^T) \bar{Q} \bar{\Lambda} U_1^j \\ &= \bar{\Lambda}^\dagger \bar{\Lambda} U_1^j = U_1^j \end{aligned} \quad (61)$$

where note from Definition 2 that it must be true that $\bar{Q}^T(I_n \otimes SS^T)\bar{Q} = I_j$, where I_j is the size j identity matrix. Hence, Theorem 1 clearly holds when we substitute U_1^j , \bar{Q} , and $\bar{\Lambda}$ for U , Q , and Λ .

Further, we can verify the following facts:

- 1) $\bar{Q}_i \bar{\Lambda} U_1^j = Q_i \Lambda U$, and therefore
- 2) $\delta(U_1^j, A_{t_1^n}) = \delta(U, A_{t_1^n})$. Also
- 3) $F(A_{t_1^n})$ remains unaltered whether we use Q, Λ or $\bar{Q}, \bar{\Lambda}$, therefore
- 4) $\eta(U_1^j, A_{t_1^n}) = \eta(U, A_{t_1^n})$. Also
- 5) $K_V(A_{t_1^n})$ remains unaltered whether we use Q, Λ or $\bar{Q}, \bar{\Lambda}$.

Thus, we conclude that

$$\begin{aligned} & \mathbb{E} \left\{ \Phi_{\mathbf{K}_V(\mathbf{A}_{t_1^n})}(\delta(\mathbf{U}_1^j, \mathbf{A}_{t_1^n}) - \eta(\mathbf{U}_1^j, \mathbf{A}_{t_1^n})) \middle| \mathbf{A}_{t_1^n} \right\} \\ &= \mathbb{E} \left\{ \Phi_{\mathbf{K}_V(\mathbf{A}_{t_1^n})}(\delta(\mathbf{U}, \mathbf{A}_{t_1^n}) - \eta(\mathbf{U}, \mathbf{A}_{t_1^n})) \middle| \mathbf{A}_{t_1^n} \right\} \end{aligned}$$

must hold, and thus, Theorem 1 must be true even when $\text{rank}[\mathbf{\Gamma}(\mathbf{A}_{t_1^n})^T \mathbf{K}_W \mathbf{\Gamma}(\mathbf{A}_{t_1^n})] < 2mn$ for certain values of $\mathbf{A}_{t_1^n} = \mathbf{a}_1^n$. ■

We have thus far completed our proof of Theorem 1; we next show an upper bound for the rank of the matrix $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ in (59). We point out that $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ sometimes may even have rank 0, i.e., $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ equals the zero matrix.

B. Other Comments

The following proposition states that the rank of $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ depends on both the chosen time instants $\{t_1, t_2, \dots, t_n\}$, and the MLM truncation length m . The following proposition gives the upper bound on $\text{rank}(\mathbf{K}_V(\mathbf{A}_{t_1^n}))$.

Proposition 3: The rank of $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ equals at most the number of time instants $t \in \{t_1, t_2, \dots, t_n\}$, which satisfy $|t - t'| > m$ for all $t' \in \{t_1, t_2, \dots, t_n\} \setminus \{t\}$. □

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 $[\mathbf{G}(\mathbf{A}_{t_1})\mathbf{S}]^T \mathbf{W}_{t_1}$ uniquely determines $V_{t_2} \triangleq A_{t_2} \cdot \mathbf{h}_0^T \mathbf{W}_{t_2}$ (and vice versa observation of $[\mathbf{G}(\mathbf{A}_{t_2})\mathbf{S}]^T \mathbf{W}_{t_2}$ uniquely determines $V_{t_1} \triangleq A_{t_1} \cdot \mathbf{h}_0^T \mathbf{W}_{t_1}$). □

Proof: Recall that V_{t_2} equals

$$V_{t_2} \triangleq A_{t_2} \cdot \mathbf{h}_0^T \mathbf{W}_{t_2} = A_{t_2} \cdot (h_0 W_{t_2} + \dots + h_I W_{t_2+I}).$$

If the condition $|t_1 - t_2| \leq m$ is satisfied, then $W_{t_2}, \dots, W_{t_2+I}$ is a length- $(I + 1)$ subsequence of $\mathbf{W}_{t_1} \triangleq [W_{t_1-m}, W_{t_1-m+1}, \dots, W_{t_1+m+I}]^T$. From the definition of \mathbf{S} [see (23)] and because $|t_1 - t_2| \leq m$, the matrix \mathbf{S} must have a column \mathbf{s} that satisfies $\mathbf{E}\mathbf{s} = \mathbf{e}_{t_2-t_1}$ [see (18)]. Then, for this particular column \mathbf{s} , we have

$$\begin{aligned} [\mathbf{G}(\mathbf{A}_{t_1})\mathbf{S}]^T \mathbf{W}_{t_1} &= [\mathbf{H} \text{diag}(\mathbf{A}_{t_1})\mathbf{E}\mathbf{s}]^T \mathbf{W}_{t_1} \\ &= A_{t_2} \cdot [\mathbf{H}\mathbf{e}_{t_2-t_1}]^T \mathbf{W}_{t_1} \\ &= A_{t_2} \cdot \mathbf{h}_0^T \mathbf{W}_{t_2} \triangleq V_{t_2} \end{aligned}$$

where the second equality holds because \mathbf{s} satisfies $\text{diag}(\mathbf{A}_{t_1})\mathbf{E}\mathbf{s} = \text{diag}(\mathbf{A}_{t_1})\mathbf{e}_{t_2-t_1} = A_{t_2} \cdot \mathbf{e}_{t_2-t_1}$, and also

$$\begin{aligned} & [\mathbf{H}\mathbf{e}_{t_2-t_1}]^T \mathbf{W}_{t_1} \\ &= [\mathbf{H}\mathbf{e}_{t_2-t_1}]^T [W_{t_1-m}, W_{t_1-m+1}, \dots, W_{t_1+m+I}] \\ &= h_0 W_{t_2} + h_1 W_{t_2+1} + \dots + h_I W_{t_2+I}. \end{aligned}$$

By symmetry, the same argument holds for $[\mathbf{G}(\mathbf{A}_{t_1})\mathbf{S}]^T \mathbf{W}_{t_2}$ and $V_{t_2} \triangleq A_{t_2} \cdot \mathbf{h}_0^T \mathbf{W}_{t_2}$. ■

Proof of Proposition 3: Recall from (59) that $\mathbf{K}_V(\mathbf{A}_{t_1^n}) \triangleq \text{Cov}\{\mathbf{V}_{t_1^n} | \mathbf{A}_{t_1^n}, \mathbf{U}\}$ is the (conditional) covariance matrix of $\mathbf{V}_{t_1^n}$. After conditioning on \mathbf{U} , the vector $\mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} = [\mathbf{G}(\mathbf{A}_{t_i})\mathbf{S}]^T \mathbf{W}_{t_i}$ is uniquely determined (see Lemma 1). Furthermore, by Lemma 2, if $\mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} = [\mathbf{G}(\mathbf{A}_{t_i})\mathbf{S}]^T \mathbf{W}_{t_i}$ is uniquely determined, then $V_{t_j} \triangleq A_{t_j} \cdot \mathbf{h}_0^T \mathbf{W}_{t_j}$ is determined whenever $|t_i - t_j| \leq m$. Thus, we conclude that the only

variables V_{t_i} that may contribute to the rank of $\mathbf{K}_V(\mathbf{A}_{t_1^n})$ must be those with corresponding t_i that are separated from all other $\{t_1, t_2, \dots, t_n\} \setminus \{t_i\}$ by greater than m . ■

Remark 3: From the expression for $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r})$ in Theorem 1, the distribution function $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r})$ must be left-continuous [19], if the $\text{rank}(\mathbf{K}_V(\mathbf{A}_{t_1^n})) = n$.

We conclude this section by verifying the correctness of Procedure 3 used to evaluate $F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\mathbf{r})$ when candidate subsets $\bar{\mathcal{M}} \subset \mathcal{M}$ [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:

$$\begin{aligned} & \{\mathbf{a} \in \bar{\mathcal{M}}_{t_i} : a_0 \neq A_{t_i}\} \\ &= \{\boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k + \mathbf{e}_0, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i} : 0 \leq k \leq 2^{2m} - 1\} \\ & \{\mathbf{a} \in \bar{\mathcal{M}}_{t_i} : a_0 = A_{t_i}\} \\ &= \{\boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i} : 0 \leq k \leq 2^{2m} - 1\} \end{aligned} \quad (62)$$

where the function $\boldsymbol{\alpha}(\mathbf{e}, \mathbf{A}_{t_i})$ is given in Line 3 of Procedure 3. Next perform the following verifications in the order presented:

- 1) Replace \mathcal{M} by $\bar{\mathcal{M}}_{t_i}$ in the definitions of R_{t_i} in (12). Replace \mathcal{M} by $\bar{\mathcal{M}}_{t_i}$ in both X_{t_i} and Y_{t_i} in (13). The validity of Proposition 1 remains unaffected.
- 2) Replace \mathcal{M} by $\bar{\mathcal{M}}_{t_i}$ in the proof of Proposition 2. The change first affects the proof starting from (46), and (47) needs to be slightly modified using (62). The new Proposition 2, finally, reads

$$\begin{aligned} X_{t_i} &= \max_{k: \boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k + \mathbf{e}_0, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T [\mathbf{G}(\mathbf{A}_{t_i})]^T \mathbf{W}_{t_i} \\ &\quad + \nu_k(\mathbf{A}_{t_i}) + V_{t_i} + \theta(\mathbf{A}_{t_i}) \\ Y_{t_i} &= \max_{k: \boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T [\mathbf{G}(\mathbf{A}_{t_i})]^T \mathbf{W}_{t_i} + \mu_k(\mathbf{A}_{t_i}). \end{aligned}$$

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

$$\begin{aligned} \delta_i &= \delta_i(\mathbf{U}, \mathbf{A}_{t_1^n}) \\ &= \max_{k: \boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k + \mathbf{e}_0, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} + \nu_k(\mathbf{A}_{t_i}) \\ &\quad - \max_{k: \boldsymbol{\alpha}(\mathbf{E}\mathbf{s}_k, \mathbf{A}_{t_i}) \in \bar{\mathcal{M}}_{t_i}} \mathbf{s}_k^T \mathbf{Q}_i \mathbf{\Lambda} \mathbf{U} + \mu_k(\mathbf{A}_{t_i}). \end{aligned}$$

This is exactly the way δ_i is computed in Procedure 3, Line 3.

This concludes our verification of Procedure 3.

V. NUMERICAL COMPUTATIONS

We now present numerical computations performed for various ISI channels. To demonstrate the generality of our results, various cases will be considered. Both i) the reliability distribution $F_{\mathbf{R}_{t_1^n}}(\mathbf{r})$ and ii) the symbol error probability $\Pr\{\bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\}\}$ will be graphically displayed in the following manner. Recall from Corollaries 1 and 2 that we have $F_{\mathbf{R}_{t_1^n}}(\mathbf{r}) = F_{\mathbf{X}_{t_1^n} - \mathbf{Y}_{t_1^n}}(\sigma^2/2 \cdot \mathbf{r})$ [here, σ^2 denotes the noise variance in (10)] and $\Pr\{\bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\}\} = \Pr\{\mathbf{X}_{t_1^n} \geq \mathbf{Y}_{t_1^n}\}$. Therefore, both quantities i) and ii)

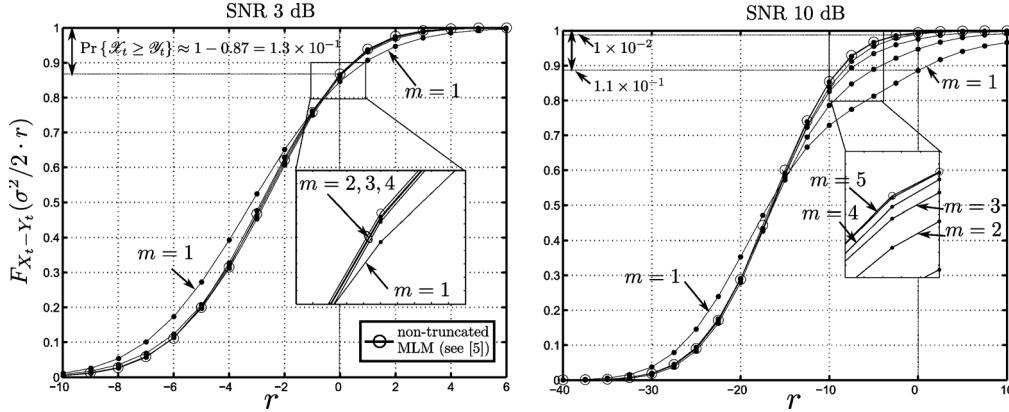


Fig. 3. Marginal reliability distribution $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$ computed for the PR1 channel (see Table I). Truncation lengths m are varied from 1 to 5.

TABLE I
VARIOUS ISI CHANNELS IN MAGNETIC RECORDING [18]

Channel	Coefficients	Memory Length ℓ	
	h_0	h_1	h_2
PR1	1	1	-
Dicode	1	-1	-
PR2	1	2	1
PR4	1	0	-1

will be displayed utilizing a *single* graphical plot of $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$.

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

A. Marginal Distribution When the Noise is i.i.d

First, consider the case where the noise samples W_t are i.i.d; thus, $\sigma^2 = \mathbb{E}\{W_t^2\}$. Fig. 3 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 and 10 dB. At an SNR of 3 dB, we observe that with the exception of $m = 1$, all curves appear to be extremely close. At the SNR of 3 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 the SNR of 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 the SNR of 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 the SNR of 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 1×10^{-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 Fig. 4. As SNR increases, the distributions $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$ appear to concentrate more probability mass

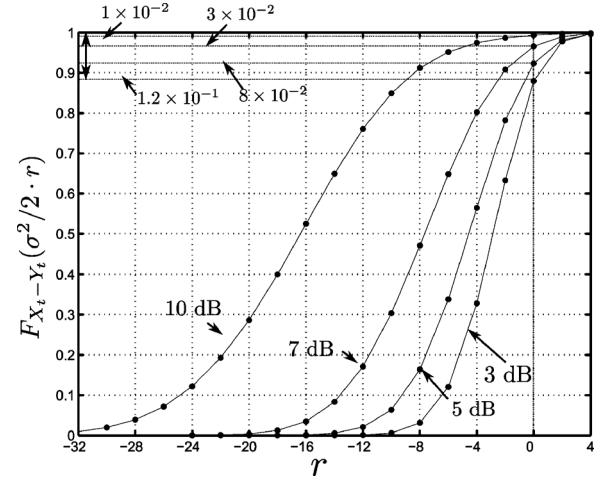


Fig. 4. Comparing the distributions $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$ across different SNRs for a fixed truncation length $m = 5$. The channel is the PR1 channel (see Table I).

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 Fig. 4, the (error) probabilities $\Pr\{X_t \geq Y_t\}$ are found to be approximately 1.2×10^{-1} , 8×10^{-2} , 3×10^{-2} , and 1×10^{-2} , respectively, for SNRs 3 to 10 dB.

B. Joint Distribution for $n = 2$ Case, 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 $m = 5$. Fig. 5 compares the joint distributions $F_{X_{t_1} - Y_{t_1}}(\sigma^2/2 \cdot r)$, 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 $\Pr\{B_{t_1} \neq A_{t_1}, B_{t_2} \neq A_{t_2}\} = \Pr\{X_{t_1} \geq Y_{t_1}, X_{t_2} \geq Y_{t_2}\}$ is approximately 6×10^{-2} and 2×10^{-2} for both cases $|t_1 - t_2| =$

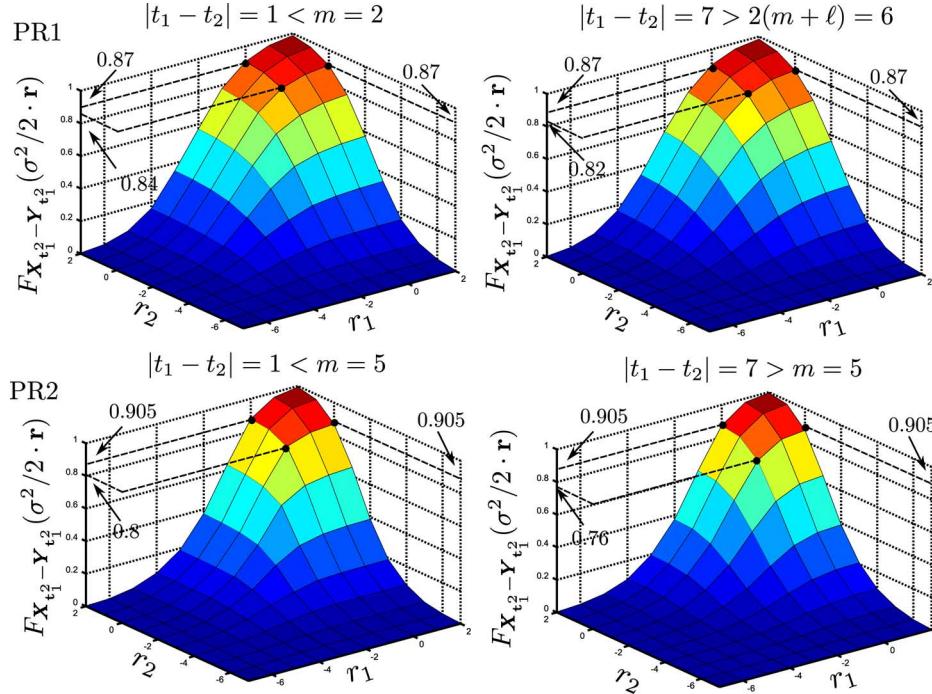


Fig. 5. Joint reliability distribution $F_{X_{t_1^2}-Y_{t_1^2}}(\sigma^2/2 \cdot r)$ computed for both the PR1 and PR2 channels, with chosen truncation lengths $m = 2$ and 5 .

1 and 7, respectively. Similarly, for the PR2, the (error) probability is approximately 3×10^{-2} and 1×10^{-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 $R_{t_1} = 2/\sigma^2 \cdot |X_{t_1} - Y_{t_1}|$ and $R_{t_2} = 2/\sigma^2 \cdot |X_{t_2} - Y_{t_2}|$ are *independent*; this is because $|t_1 - t_2| = 7 > 2(m + \ell) = 6$ (see Fig. 1).

C. Marginal Distribution 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., $\mathbb{E}\{W_t \cdot W_{\bar{t}}\} = 0$ for all $|t - \bar{t}| > 1$, and consider the following two cases:

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

We consider a moderate SNR of 5 dB. Fig. 6 shows the distributions $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ computed for both cases. Also in Fig. 6, 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 $\mathbb{E}\{W_t \cdot W_{t+1}\}/\sigma^2 = -0.5$. 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* [15], [18] displayed in Fig. 6. When the correlation coefficient equals $\mathbb{E}\{W_t \cdot W_{t+1}\}/\sigma^2 = -0.5$, the noise power is strongest among signaling frequencies, and the symbol error probability $\Pr\{B_t \neq A_t\} = \Pr\{X_t \geq Y_t\}$ is observed to be the lowest (approximately 8×10^{-2}). On the other hand, when the correlation coefficient is $\mathbb{E}\{W_t \cdot W_{t+1}\}/\sigma^2 = 0.5$, 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 1.6×10^{-1}). Note that in the latter case

$\mathbb{E}\{W_t \cdot W_{t+1}\}/\sigma^2 = -0.5$, the MLM performs even better than the i.i.d case (see Fig. 6). In the i.i.d case, the error probability $\Pr\{X_t \geq Y_t\} \approx 1.3 \times 10^{-1}$.

Remark 4: One intuitively expects that similar observations will be made even for other (more complicated) choices for the noise covariance matrix \mathbf{K}_W ; recall (21). We stress that our results are general in the sense that we may arbitrarily specify \mathbf{K}_W ; even if the noise samples W_t are **nonstationary**, our methods still apply.

D. Marginal Distribution When the Noise is i.i.d., and When Run-Length Limited (RLL) Codes are Used

We demonstrate Procedure 3 in Section III-C, used to compute the distribution $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ when a modulation code is present in the system. In particular, consider an 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). Fig. 7 shows $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ 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 Section V-C), but the dicode channel does not.

It is clearly seen from Fig. 7 that the RLL code improves the performance when used for the PR4 channel. For the PR4 channel, the distribution $F_{X_t-Y_t}(\sigma^2/2 \cdot r)$ appears to concentrate more probability mass over negative values of $X_t - Y_t$ similar to the observations made in Fig. 4 when there is an increase in SNR. The error probability $\Pr\{B_t \neq A_t\} = \Pr\{X_t \geq Y_t\} = 1 - F_{X_t-Y_t}(0)$ decreases by a factor of 2, dropping from approximately 9.5×10^{-2} to 4×10^{-2} . On the other hand, the RLL code has a negative impact on the performance when applied to the dicode channel.

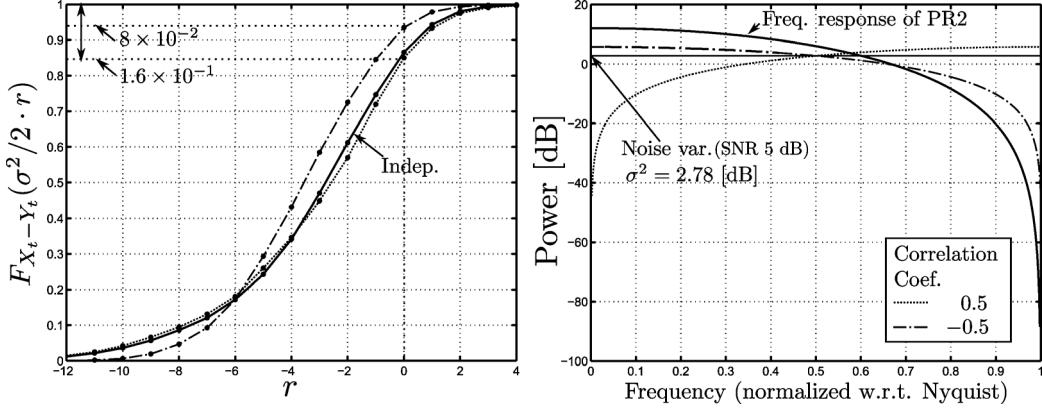


Fig. 6. Marginal distribution $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$ for correlated noises, for the PR2 channel, at the SNR of 5 dB. Truncation length $m = 5$.

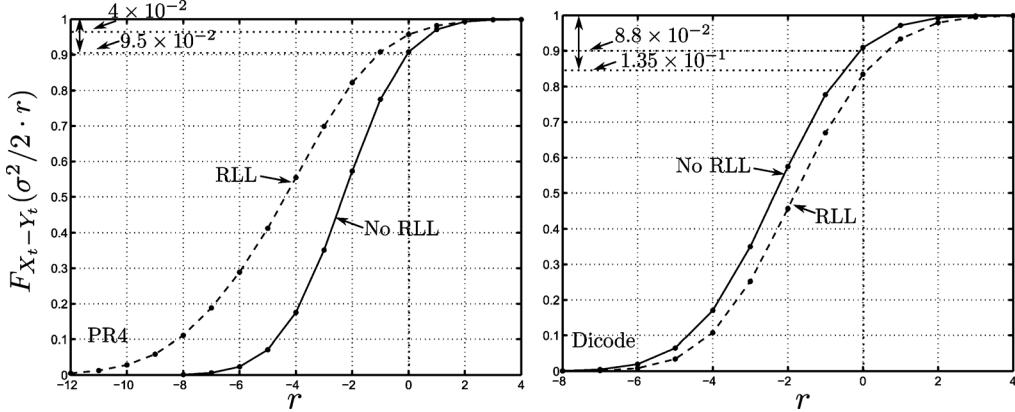


Fig. 7. Marginal distributions $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$ computed for cases when an RLL code is present. Here, we compare both the PR4 and dicode (see Table I) channels at the SNR of 5 dB.

For the dicode channel, $F_{X_t - Y_t}(r)$ concentrates more probability mass over positive values of $X_t - Y_t$ (similar to the observations made in Fig. 4 when there is an SNR decrease), and the (error) probability $\Pr\{X_t \geq Y_t\}$ increases from approximately 8.8×10^{-2} to 1.35×10^{-1} .

E. Marginal Distribution When Conditioning on Neighboring Error Events

Here, we consider three *neighboring* symbol reliabilities, i.e., we consider $\mathbf{R}_{t_1^3} = [R_{t-1}, R_t, R_{t+1}]^T$. We consider the following two conditional distributions:

- $\Pr\{X_t - Y_t \leq r | X_{t-1} < Y_{t-1}, X_{t+1} < Y_{t+1}\}$
 $= \frac{1}{C_1} \cdot F_{X_{t_1^3} - Y_{t_1^3}}(0, r, 0)$, and
- $\Pr\{X_t - Y_t \leq r | X_{t-1} \geq Y_{t-1}, X_{t+1} \geq Y_{t+1}\}$
 $= \frac{1}{C_2} \left(F_{X_t - Y_t}(r) - F_{X_{t_1^2} - Y_{t_1^2}}(r, 0) - F_{X_{t_1^1} - Y_{t_1^1}}(0, r) + F_{X_{t_1^3} - Y_{t_1^3}}(0, r, 0) \right)$

where the normalization constants C_1 and C_2 equal the probabilities of the (respective) events that were conditioned on. Distribution (a) is conditioned on the event that both neighboring symbols are *correct*, i.e., $\{B_{t-1} = A_{t-1}, B_{t+1} = A_{t+1}\}$. Distribution (b) is conditioned on the event that both neighboring symbols are *wrong*, i.e., $\{B_{t-1} \neq A_{t-1}, B_{t+1} \neq A_{t+1}\}$. For

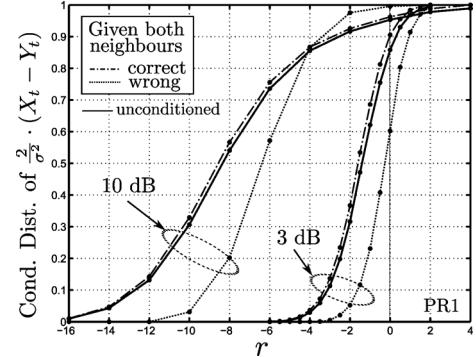


Fig. 8. Marginal distributions of $X_t - Y_t$ computed for the PR1 channel, obtained when conditioning on either events $\{B_{t-1} \neq A_{t-1}, B_{t+1} \neq A_{t+1}\}$ and $\{B_{t-1} = A_{t-1}, B_{t+1} = A_{t+1}\}$. These two events correspond to error (or nonerror) events at neighboring time instants $t - 1$ and $t + 1$. The solid black line represents the unconditioned marginal distribution of $X_t - Y_t$.

the PR1, PR2, and PR4 channels, both conditional distributions (a) and (b) are shown in Figs. 8 and 9. We compare two different SNRs of 3 and 10 dB. For comparison purposes, we also show the *unconditioned* distribution $F_{X_{t_1^n} - Y_{t_1^n}}(\sigma^2/2 \cdot r)$ in both Figs. 8 and 9. We make the following observations.

In all considered cases, distribution (a) is seen to be similar to the unconditioned distribution. However, distribution (b) is observed to vary for all the considered cases. Take, for example, the PR2 channel, we see from Fig. 9 that distribution (b) has probability mass concentrated to the right of the unconditioned

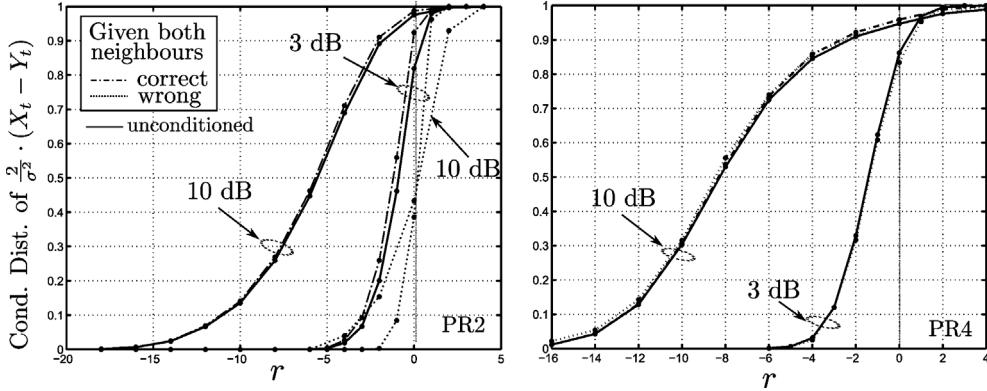


Fig. 9. Marginal distributions of $X_t - Y_t$ computed for both the PR2 and PR4 channels, obtained when conditioning on either events $\{B_{t-1} \neq A_{t-1}, B_{t+1} \neq A_{t+1}\}$ and $\{B_{t-1} = A_{t-1}, B_{t+1} = A_{t+1}\}$. These two events correspond to error (or nonerror) events at neighboring time instants $t-1$ and $t+1$. The solid black line represents the unconditioned marginal distribution of $2/\sigma^2 \cdot (X_t - Y_t)$.

$F_{X_t - Y_t}(\sigma^2/2 \cdot r)$. This is true for both SNRs of 3 and 10 dB. In contrast for the PR1, the MLM detector behaves differently at the two SNRs. We see from Fig. 8 that at the SNR of 10 dB, the distribution (b) has a lower symbol error probability than that of the unconditioned $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$. At SNR 3 dB, however, the opposite is observed, i.e., the symbol error probability is higher than that of the distribution $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$. This is because at SNR 10 dB, errors occur sparsely, interspersed by correct symbols; it is uncommon to encounter consecutive symbols in error. Hence, conditioned on adjacent symbols B_{t-1} and B_{t+1} being wrong, it is uncommon for B_t to be also wrong, as this is the event where we have three consecutive erroneous symbols. Finally, the observations made for the PR4 channel are again different. We notice that both distributions (a) and (b) practically equal the unconditioned distribution $F_{X_t - Y_t}(\sigma^2/2 \cdot r)$. This is because the even/odd output subsequences of the PR4 channel are independent of each other.

VI. CONCLUSION

In this paper, for the m -truncated MLM detector, we derived closed-form expressions for both 1) the reliability distributions $F_{X_{t_1^n} - Y_{t_1^n}}(\sigma^2/2 \cdot r)$, and 2) the symbol error probabilities $\Pr\{\bigcap_{i=1}^n \{B_{t_i} \neq A_{t_i}\}\}$. Our results hold jointly for any number n of arbitrarily chosen time instants t_1, t_2, \dots, 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 filters, such as in convolutional codes.

APPENDIX

A) Computing the Matrix $\mathbf{Q} = \mathbf{Q}(A_{t_1^n})$ in Definition 2: In this appendix, we show that the size $2mn$ square matrix \mathbf{Q} with both properties i) and ii) as stated in Definition 2 can be easily found. We begin by noting from (24) that $\text{rank}(\mathbf{SS}^T) = 2m$; therefore, the matrix $\mathbf{I}_n \otimes \mathbf{SS}^T$ has rank $2mn$ and is positive

definite. Recall $\text{diag}(\mathbf{G}(A_{t_1}), \mathbf{G}(A_{t_2}), \dots, \mathbf{G}(A_{t_n}))$ is block diagonal with entries (19).

Lemma 3: Let \mathbf{S} be given as in (23). Let the size $2mn \times 2mn$ square matrix $\boldsymbol{\alpha}$ diagonalize

$$\boldsymbol{\alpha}^T (\mathbf{I}_n \otimes \mathbf{SS}^T) \boldsymbol{\alpha} = \mathbf{I}. \quad (63)$$

Let $\boldsymbol{\beta}$ be a $2mn \times 2mn$ eigenvector matrix $\boldsymbol{\beta}$ in the following decomposition:

$$\begin{aligned} & \boldsymbol{\alpha}^T (\mathbf{I}_n \otimes \mathbf{SS}^T) \text{diag}(\mathbf{G}(A_{t_1}), \mathbf{G}(A_{t_2}), \dots, \mathbf{G}(A_{t_n}))^T \mathbf{K}_W \\ & \cdot \text{diag}(\mathbf{G}(A_{t_1}), \mathbf{G}(A_{t_2}), \dots, \mathbf{G}(A_{t_n})) (\mathbf{I}_n \otimes \mathbf{SS}^T) \boldsymbol{\alpha} \\ & = \boldsymbol{\beta} \boldsymbol{\Lambda}^2 \boldsymbol{\beta}^T \end{aligned} \quad (64)$$

where $\boldsymbol{\Lambda}^2$ is the eigenvalue matrix of (64); therefore, $\boldsymbol{\Lambda}^2$ in (64) is diagonal of size $2mn$. Then

$$\mathbf{Q} = \boldsymbol{\alpha} \boldsymbol{\beta} \quad (65)$$

satisfies both properties i) and ii) stated in Definition 2.

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

$$\begin{aligned} & \boldsymbol{\alpha}^{-1} \text{diag}(\mathbf{G}(A_{t_1}), \mathbf{G}(A_{t_2}), \dots, \mathbf{G}(A_{t_n}))^T \mathbf{K}_W \\ & \cdot \text{diag}(\mathbf{G}(A_{t_1}), \mathbf{G}(A_{t_2}), \dots, \mathbf{G}(A_{t_n})) \boldsymbol{\alpha}^{-T} = \boldsymbol{\beta} \boldsymbol{\Lambda}^2 \boldsymbol{\beta}^T. \end{aligned} \quad (66)$$

Consider the matrix $\mathbf{Q} = \boldsymbol{\alpha} \boldsymbol{\beta}$. It follows from (66) that $\mathbf{Q} = \boldsymbol{\alpha} \boldsymbol{\beta}$ satisfies property i) in Definition 2, as seen after multiplying (the matrices satisfying) (66) on the left and right by $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^T$, respectively. It also follows that $\mathbf{Q} = \boldsymbol{\alpha} \boldsymbol{\beta}$ satisfies property ii) in Definition 2, this is because

$$\mathbf{Q}^T (\mathbf{I}_n \otimes \mathbf{SS}^T) \mathbf{Q} = \boldsymbol{\beta}^T \boldsymbol{\alpha}^T (\mathbf{I}_n \otimes \mathbf{SS}^T) \boldsymbol{\alpha} \boldsymbol{\beta} = \boldsymbol{\beta}^T \boldsymbol{\beta} = \mathbf{I}$$

where the last equality follows because $\boldsymbol{\beta}$ is unitary (i.e., $\boldsymbol{\beta}^{-1} = \boldsymbol{\beta}^T$) by virtue of the fact that it is an eigenvector matrix [17, p. 311]. ■

To summarize Lemma 3, the matrix $\mathbf{Q} = \mathbf{Q}(A_{t_1^n})$ in Definition 2 is obtained by first computing two size $2mn$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, satisfying (63) and (64), and then setting

$\mathbf{Q} = \boldsymbol{\alpha}\boldsymbol{\beta}$. The matrix $\boldsymbol{\beta}$ is obtained from an eigenvalue decomposition of the $2mn$ matrix (64) and clearly $\boldsymbol{\beta}$ depends on the symbols $\mathbf{A}_{\mathbf{t}_1^n}$. The matrix $\boldsymbol{\alpha}$, however, is simpler to obtain. This is due to the simple form of \mathbf{SS}^T in (24), and we may even obtain closed-form expressions for $\boldsymbol{\alpha}$; see the following remark.

Remark 5: It can be verified that the following are eigenvectors of the matrix \mathbf{SS}^T in (24). The first $2m - 1$ eigenvectors are

$$(i + i^2)^{-\frac{1}{2}} \cdot [\underbrace{1, 1, \dots, 1}_i, \underbrace{-i, 0, 0, \dots, 0}_{2m-(i+1)}]^T$$

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

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor, and the anonymous reviewers, for their help in greatly improving the presentation of our initial draft.

REFERENCES

- [1] G. D. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.
- [2] G. D. Forney Jr., "The Viterbi algorithm," *Proc. 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: Turbo-equalization," *Eur. Trans. Telecommun.*, 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. Inf. 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 J. Sel. 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 *Proc. IEEE Global Telecommun. Conf.*, Dallas, TX, USA, 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. Inf. Theory*, vol. IT-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," *IEEE Commun. Lett.*, 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 Int. Symp. Inf. Theory*, 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. Inf. Theory*, vol. 50, no. 6, pp. 1303–1311, Jun. 2004.

- [11] L. Reggiani and G. Tartara, "Probability density functions of soft information," *IEEE Commun. Lett.*, 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 *Proc. 4th LASTED Int. Multi-Conf. Wireless Opt. Commun.*, Banff, AB, Canada, 2004, pp. 358–363.
- [13] P. N. Somerville, "Numerical computation of multivariate normal and multivariate-t probabilities over convex regions," *J. Comput. Graph. Statist.*, vol. 7, no. 4, pp. 529–544, Dec. 1998.
- [14] D. T. Tang and L. R. Bahl, "Block codes for a class of constrained noiseless channels," *Inf. 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. Inf. Theory*, vol. 44, no. 6, pp. 2260–2299, Oct. 1998.
- [16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, 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 Univ. Press, 1996.
- [18] P. Kabal and S. Pasupathy, "Partial response signaling," *IEEE Trans. Commun.*, vol. COM-23, no. 9, pp. 921–934, Sep. 1975.
- [19] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

Fabian Lim received the B.Eng and M.Eng degrees from the National University of Singapore in 2003 and 2006, respectively, and the Ph.D. degree from the University of Hawaii, Manoa in 2010, all in electrical engineering. Currently, he is a postdoctoral associate at the Massachusetts Institute of Technology.

Dr. Lim has held short-term visiting research positions at Harvard University in 2004 and 2005. From Oct 2005 to May 2006, he was a staff member in the Data Storage Institute, Singapore. From May 2008 to July 2008, he was an intern at Hitachi Global Storage Technologies, San Jose. In March 2009, he was a visitor at the Research Center for Information Security, Japan.

His research interests include error-control coding and signal processing.

Aleksandar Kavčić received the Dipl. Ing. degree in Electrical Engineering from Ruhr-University, Bochum, Germany in 1993, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, Pennsylvania in 1998.

Since 2007 he has been with the University of Hawaii, Honolulu where he is presently Professor of Electrical Engineering. Prior to 2007, he was in the Division of Engineering and Applied Sciences at Harvard University, as Assistant Professor and Associate Professor of Electrical Engineering. He also served as Visiting Associate Professor at the City University of Hong Kong in the Fall of 2005 and as Visiting Scholar at the Chinese University of Hong Kong in the Spring of 2006.

Prof. Kavčić received the IBM Partnership Award in 1999 and the NSF CAREER Award in 2000. He is a co-recipient, with X. Ma and N. Varnica, of the 2005 IEEE Best Paper Award in Signal Processing and Coding for Data Storage. He served on the Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY as Associate Editor for Detection and Estimation from 2001 to 2004, as Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE in 2003–2004, and as Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2008–2009. From 2005 until 2007, he was the Chair of the Data Storage Technical Committee of the IEEE Communications Society.