On Sufficient Conditions for Testing Optimality of Codewords in ISI Channels

Fabian Lim[1], Aleksandar Kavčić[1], and Marc Fossorier[2]
[1]EE Department, University of Hawaii at Manoa, Honolulu, HI 96822
[2]ETIS ENSEA/UCP/CNRS UMR-8051, 6 avenue du Ponceau, 95014, Cergy Pontoise, France
flim@hawaii.edu, alek@hawaii.edu, mfossorier@ieee.org

***Abstract*—For the memoryless AWGN channel, there exists low complexity methods to test the optimality of any chosen candidate codeword (i.e., whether the codeword in question equals the most-likely codeword or not). Such optimality tests find application in practical decoders that perform heuristic searches for the most-likely codeword. If some located codeword passes the optimality test, then the search may be terminated and computations saved. In this paper, we generalize techniques for determining if a codeword is optimal, for intersymbol interference (ISI) channels.**

I. Introduction

The maximum-likelihood decoder (MLD) minimizes the word-error rate, and thus serves as a benchmark for the performance of practical decoders. Practical list decoding schemes search a subset of the codebook for the most-likely codeword (the codeword with the highest likelihood of reception amongst all candidate codewords). Terminating the search when the most-likely codeword is found allows one to avoid unnecessary computations and makes the overall decoding algorithm more efficient. However, is it possible to identify the most-likely codeword without comparing its likelihood to that of every possible candidate codeword?

To answer this question we typically utilize *optimality tests*. Whenever a *sufficient condition for optimality* is satisfied by a codeword, then that particular codeword must equal the most-likely codeword. For memoryless channels, optimality tests include [1], [2], [3]. However, similar techniques have never been developed for intersymbol interference (ISI) channels, where (channel) memory is present. The techniques for the memoryless channels in [1], [2], [3] do not generalize trivially to ISI channels. Here, we develop such a generalization.

In Section II, we develop the relevant background. In Section III, we proceed to generalize the optimality tests to ISI channels from first principles. We show that the techniques in [1], [2], [3], are connected to our techniques for ISI channels by the *Moore-Penrose pseudoinverse* of the channel transfer matrix. We explicitly show that the optimality tests for ISI channels consist of two separate subproblems corresponding to the channel and code, respectively. We show that the optimality tests for ISI channels have comparable complexities to their memoryless channel counterparts. In Section IV, we evaluate the effectiveness of the proposed optimality tests via computer simulation. We conclude in Section V.

This work was supported by the NSF under grant number ECCS07-25649.

II. Background

Let $\mathbb{R}$ denote the set of real numbers. Let $\mathbb{B} \overset{\triangle}{=} \{-1, 1\}$ denote the *bipolar* set, and let $\mathcal{C}_{\mathbb{B}} \subseteq \mathbb{B}^N$ denote a binary code of length N in bipolar form. Vectors and matrices are denoted in boldface font. Let $\mathbf{a} \in \mathbb{B}^N$ denote a length-N bipolar sequence. The Hamming distance between any two bipolar sequences $\mathbf{a}, \mathbf{a}' \in \mathbb{B}^N$ is denoted as $d_{\mathrm{H}}(\mathbf{a}, \mathbf{a}')$. The minimum distance of $\mathcal{C}_{\mathbb{B}}$ is denoted as $d_{\min}$.

Let $\mathbf{r}$ and $\mathbf{w}$ respectively denote the observed received sequence, and i.i.d zero-mean Gaussian noise realization. Let $\mathbf{h} = [h_0, h_1, \cdots, h_I]^T$ denote the channel impulse response, where I is the ISI length. If $\mathbf{a} = [a_1, a_2 \cdots, a_N]^T \in \mathbb{B}^N$ is the channel input, the ISI channel input-output relationship is

$$r_t = \sum_{i=0}^{I} h_i a_{t-i} + w_t \quad \text{for, } 1 \leq t \leq N + I \tag{1}$$

where we implicitly assumed[1] in (1) that $a_t = 0$ for all $t < 1$ and $t > N$. We rewrite (1) as $\mathbf{r} = \mathbf{Ha} + \mathbf{w}$, where $\mathbf{H}$ is a $(N + I) \times N$ *banded* Toeplitz matrix

$$\mathbf{H} = \begin{bmatrix} h_0 & & & \\ \vdots & \ddots & & \\ h_I & & \ddots & \\ & \ddots & & h_0 \\ & & \ddots & \vdots \\ & & & h_I \end{bmatrix}, \tag{2}$$

For all $\mathbf{a} \in \mathbb{B}^N$, the quantity $|\mathbf{r} - \mathbf{Ha}|^2$ is the squared Euclidean distance between the received vector $\mathbf{r}$, and the (noiseless) channel output $\mathbf{Ha}$. For the code $\mathcal{C}_{\mathbb{B}}$ and channel observation $\mathbf{r}$, the *most-likely codeword* $\mathbf{c}_{\mathrm{ml}}(\mathbf{r})$ is obtained as

$$\mathbf{c}_{\mathrm{ml}}(\mathbf{r}) \overset{\triangle}{=} \arg\min_{\mathbf{a} \in \mathcal{C}_{\mathbb{B}}} |\mathbf{r} - \mathbf{Ha}|^2. \tag{3}$$

Note that for large codebooks, determining $\mathbf{c}_{\mathrm{ml}}(\mathbf{r})$ is impractical. In this paper, we consider a related problem (as considered in [1], [2], [3] for memoryless channels); given a codeword $\mathbf{c}$, we ask whether or not there exists another codeword with a higher likelihood than $\mathbf{c}$. If we may conclude that the answer to the previous question is no, then we may conclude that $\mathbf{c}$ is optimal (i.e. that $\mathbf{c} = \mathbf{c}_{\mathrm{ml}}(\mathbf{r})$).

[1]Note there is no loss of generality by assuming $a_t = 0$ for all $t < 1$ and $t > N$ in (1). If $a_t \neq 0$ for any $t < 1$ and/or $t > N$, we can always write $\mathbf{r} = \mathbf{Ha} + \boldsymbol{\ell} + \mathbf{w}$ for some $\boldsymbol{\ell} \in \mathbb{R}^N$. Our exposition holds for any $\boldsymbol{\ell} \in \mathbb{R}^N$ if we substitute $\mathbf{r} - \boldsymbol{\ell}$ for $\mathbf{r}$ in appropriate places in the text.

III. Optimality Criteria for ISI channels

In the sequel let us fix the channel observation $\mathbf{r}$ and let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$ denote a fixed codeword. Any bipolar sequence $\mathbf{a} \in \mathbb{B}^N$ can be written with respect to the codeword $\mathbf{c}$ as

$$\mathbf{a} = \mathbf{a}(\mathbf{c}, \mathbf{e}) = \mathbf{c} - 2\mathbf{C}\mathbf{e} \tag{4}$$

where $\mathbf{e} \in \{0,1\}^N$, and $\mathbf{C}$ is a diagonal matrix whose diagonal equals $\mathbf{c}$. In other words for any fixed $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$, equation (4) establishes a one-to-one correspondence between some $\mathbf{a} \in \mathbb{B}^N$ and some $\mathbf{e} \in \{0,1\}^N$ (i.e. $\mathbf{e}$ represents the classical "error vector" of $\mathbf{a}$ from $\mathbf{c}$).

Definition 1. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$. For a fixed channel observation $\mathbf{r} \in \mathbb{R}^N$, define the set $\mathcal{M}(\mathbf{c}) \subseteq \{0,1\}^N$ as*

$$\mathcal{M}(\mathbf{c}) \stackrel{\triangle}{=} \left\{ \mathbf{e} \in \{0,1\}^N : \; |\mathbf{r} - \mathbf{H}\left(\mathbf{c} - 2\mathbf{C}\mathbf{e}\right)|^2 \leq |\mathbf{r} - \mathbf{H}\mathbf{c}|^2 \right\}.$$

That is, the set $\mathcal{M}(\mathbf{c})$ contains all sequences $\mathbf{e}$, that correspond (via (4)) to bipolar sequences $\mathbf{a} = \mathbf{c} - 2\mathbf{C}\mathbf{e} \in \mathbb{B}^N$ with smaller squared Euclidean distance to $\mathbf{r}$ than the codeword $\mathbf{c}$.

Definition 2. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$. Define the set $\mathcal{C}_{0,1}(\mathbf{c}) \subseteq \{0,1\}^N$ as*

$$\mathcal{C}_{0,1}(\mathbf{c}) \stackrel{\triangle}{=} \{\mathbf{e} \in \{0,1\}^N : \; \mathbf{c} - 2\mathbf{C}\mathbf{e} \in \mathcal{C}_{\mathbb{B}}\}.$$

That is, the set $\mathcal{C}_{0,1}(\mathbf{c})$ contains all sequences $\mathbf{e}$, that correspond (via (4)) to (bipolar) codewords $\mathbf{c} - 2\mathbf{C}\mathbf{e} \in \mathcal{C}_{\mathbb{B}}$.

Consider the following. If

$$\mathcal{M}(\mathbf{c}) \cap \mathcal{C}_{0,1}(\mathbf{c}) = \{\mathbf{0}\}, \tag{5}$$

then it follows by definition of $\mathbf{c}_{\mathrm{ml}}(\mathbf{r})$ in (3) that we must have $\mathbf{c} = \mathbf{c}_{\mathrm{ml}}(\mathbf{r})$. Generally, both sets $\mathcal{M}(\mathbf{c})$ and $\mathcal{C}_{0,1}(\mathbf{c})$ are (generally) extremely large and difficult to obtain. Our aim is to demonstrate computationally efficient ways to test condition (5). Note the following observation.

Theorem 1. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$. Let the set $\mathcal{S} \subset \mathcal{C}_{0,1}(\mathbf{c})$ satisfy $\mathcal{M}(\mathbf{c}) \cap \mathcal{S} = \{\mathbf{0}\}$. If there exists $\boldsymbol{\mu} \in \mathbb{R}^N$ such that*

$$\max_{\mathbf{e} \in \mathcal{M}(\mathbf{c})} |\mathbf{e} - \boldsymbol{\mu}|^2 < \min_{\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}} |\mathbf{e} - \boldsymbol{\mu}|^2, \tag{6}$$

is satisfied, then we must have $\mathbf{c} = \mathbf{c}_{\mathrm{ml}}(\mathbf{r})$.

Proof: Let us first prove

$$\mathcal{M}(\mathbf{c}) \cap (\mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}) = \emptyset. \tag{7}$$

Assume the contrary to (7). Then there must exist some $\mathbf{e}' \in \mathcal{M}(\mathbf{c}) \cap (\mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S})$, which implies the contradiction

$$|\mathbf{e}' - \boldsymbol{\mu}|^2 \leq \max_{\mathbf{e} \in \mathcal{M}(\mathbf{c})} |\mathbf{e} - \boldsymbol{\mu}|^2 < \min_{\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}} |\mathbf{e} - \boldsymbol{\mu}|^2 \leq |\mathbf{e}' - \boldsymbol{\mu}|^2$$

(see Figure 1 for a depiction of this argument in Euclidean space). Therefore (7) must hold. Recall that the set $\mathcal{S}$ satisfies $\mathcal{M}(\mathbf{c}) \cap \mathcal{S} = \{\mathbf{0}\}$, which when combined with (7) implies that (5) must hold. Consequently $\mathbf{c} = \mathbf{c}_{\mathrm{ml}}(\mathbf{r})$. ■

Condition (6) is still hard to evaluate. We are interested in sufficient conditions for optimality of the following form.

Corollary 1. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$ and let the set $\mathcal{S} \subset \mathcal{C}_{0,1}(\mathbf{c})$ satisfy $\mathcal{M}(\mathbf{c}) \cap \mathcal{S} = \{\mathbf{0}\}$. If there exist nonnegative numbers $\mathcal{U}(\mathbf{c})$ and $\mathcal{L}(\mathcal{S})$ and if there exist vectors $\hat{\mathbf{r}} \in \mathbb{R}^N$ and $\boldsymbol{\mu} \in \mathbb{R}^N$ that simultaneously satisfy the following 3 inequalities*

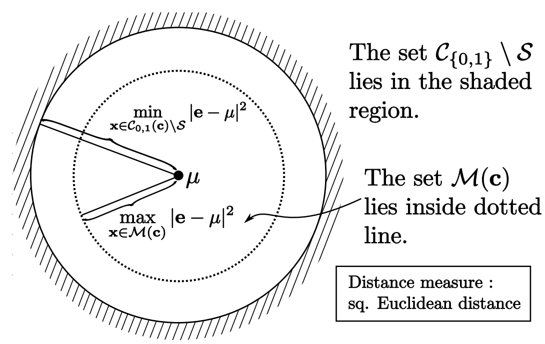

Fig. 1. Inequality (6) implies (7) because (6) implies that $\mathcal{M}(\mathbf{c})$ and $\mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}$ are disjoint.

$$\mathcal{U}(\mathbf{c}) \geq \max_{\mathbf{e} \in \mathcal{M}(\mathbf{c})} |\mathbf{e} - \boldsymbol{\mu}|^2, \tag{8}$$

$$\mathcal{L}(\mathcal{S}) \leq \min_{\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}} |\mathbf{e} - \boldsymbol{\mu}|^2 + \left(\sum_{t:\; c_t \hat{r}_t < 0} |\hat{r}_t| \right) - |\boldsymbol{\mu}|^2, \tag{9}$$

$$\mathcal{L}(\mathcal{S}) > \left(\sum_{t:\; c_t \hat{r}_t < 0} |\hat{r}_t| \right) + \left(\mathcal{U}(\mathbf{c}) - |\boldsymbol{\mu}|^2 \right) \tag{10}$$

then $\mathbf{c} = \mathbf{c}_{\mathrm{ml}}(\mathbf{r})$.

To ensure the effectiveness of the sufficient condition in Corollary 1, the bounds $\mathcal{U}(\mathbf{c})$ and $\mathcal{L}(\mathcal{S})$ should be chosen such that both inequalities (8) and (9) are satisfied reasonably tightly. We also need to make judicious choices of $\boldsymbol{\mu}$ and $\hat{\mathbf{r}}$. In this paper we will assume that the channel transfer matrix $\mathbf{H}$ satisfies $\mathrm{rank}(\mathbf{H}) = N$. We will fix our choice for $\boldsymbol{\mu}$ to be

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{c}) \quad \stackrel{\triangle}{=} \quad \frac{1}{2}\left(\mathbb{1} - \mathbf{C}\hat{\mathbf{r}}\right) \tag{11}$$

where $\mathbb{1} \stackrel{\triangle}{=} [1, 1, \cdots, 1]^T$ and the vector $\hat{\mathbf{r}}$ satisfies

$$\hat{\mathbf{r}} \stackrel{\triangle}{=} (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{r}. \tag{12}$$

In Subsection III-A we will show that the set $\mathcal{M}(\mathbf{c})$ is contained in an ellipsoid centered at $\boldsymbol{\mu} = (\mathbb{1} - \mathbf{C}\hat{\mathbf{r}})/2$. Using this fact, we will show that a reasonable bound $\mathcal{U}(\mathbf{c})$ (see (8)) may be efficiently computed. Also because we choose $\boldsymbol{\mu}$ as in (11), we are able to show in Subsection III-B that the problem of computing the bound $\mathcal{L}(\mathcal{S})$ reduces to problems considered before in [1], [2], [3] for memoryless AWGN channels.

Remark 1. *Note that when $\mathrm{rank}(\mathbf{H}) = N$, the matrix $(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ used in (12) is known as the Moore-Penrose psuedoinverse of $\mathbf{H}$. Since $\mathbf{H}^T\mathbf{H}$ is an I-banded matrix (see (2)), the complexity of obtaining $\hat{\mathbf{r}}$ in (12) is linear in N because we need not invert $\mathbf{H}^T\mathbf{H}$; simply solve $(\mathbf{H}^T\mathbf{H})\hat{\mathbf{r}} = \mathbf{H}^T\mathbf{r}$ for $\hat{\mathbf{r}}$ using Cholesky factorization methods [4].*

When we set $\mathbf{H} = \mathbf{I}$ where $\mathbf{I}$ is the $N \times N$ identity matrix, Theorem 1 collapses[2] down to optimality criteria for memoryless AWGN channels [1], [2], [3]. Also, note that

[2]When $\mathbf{H} = \mathbf{I}$ then we may set $\mathcal{U}(\mathbf{c}) = |\boldsymbol{\mu}|^2$ as shown in Subsection III-A

$\mathcal{U}(\mathbf{c}) - |\boldsymbol{\mu}|^2 \geq 0$. This follows easily from (8) and the fact that $\mathbf{0} \in \mathcal{M}(\mathbf{c})$. It is clear from the form of (10) that if the term $\mathcal{U}(\mathbf{c}) - |\boldsymbol{\mu}|^2$ is large then Cor. 1 is rendered less effective.

For soon to be apparent reasons, the problems of computing tight bounds $\mathcal{U}(\mathbf{c})$ and $\mathcal{L}(\mathcal{S})$ are termed the *channel*, and *code* subproblems, respectively. This is because our proposed technique for computing $\mathcal{U}(\mathbf{c})$ only depends on the channel $\mathbf{H}$ and not the code $\mathcal{C}_{\mathbb{B}}$ (and vice-versa for $\mathcal{L}(\mathcal{S})$).

A. Channel Subproblem: Obtaining $\mathcal{U}(\mathrm{c})$

In this subsection, we address the problem of obtaining a reasonably tight bound $\mathcal{U}(\mathbf{c})$ (see (8)). Our bound $\mathcal{U}(\mathbf{c})$ is obtained by considering an ellipsoid centered at $\boldsymbol{\mu}(\mathbf{c})$ (see (11)) that contains the set $\mathcal{M}(\mathbf{c})$.

Definition 3. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$. Let $\mathcal{E}(\mathbf{c}) \subset \mathbb{R}^N$ denote an ellipsoid in N-dimensional Euclidean space*

$$\mathcal{E}(\mathbf{c}) \quad \triangleq \quad \left\{\mathbf{x} \in \mathbb{R}^N : (\mathbf{x}-\boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x}-\boldsymbol{\mu}) \leq \boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\mu}\right\} \tag{13}$$

where the $N \times N$ matrix $\mathbf{Q} = \mathbf{Q}(\mathbf{c})$ is given as

$$\mathbf{Q}(\mathbf{c}) \quad \triangleq \quad \mathbf{C}\mathbf{H}^T\mathbf{H}\mathbf{C} \tag{14}$$

and $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{c}) = (\mathbb{1} - \mathbf{C}\hat{\mathbf{r}})/2$ is the center of the ellipsoid.

We assumed that $\operatorname{rank}(\mathbf{H}) = N$ so $\mathbf{Q}$ is positive definite. Note that $\boldsymbol{\mu}$ implicitly depends on $\mathbf{r}$ because of (12).

Proposition 1. *Let $\mathbf{c} \in \mathcal{C}_{\mathbb{B}}$. The set $\mathcal{M}(\mathbf{c})$ (see Definition 1) is contained in the ellipsoid $\mathcal{E}(\mathbf{c})$, i.e. $\mathcal{M}(\mathbf{c}) \subset \mathcal{E}(\mathbf{c})$.*

Proof: First, note that we may rewrite

$$\begin{aligned} &|\mathbf{r} - \mathbf{H}(\mathbf{c} - 2\mathbf{C}\mathbf{e})|^2 - |\mathbf{r} - \mathbf{H}\mathbf{c}|^2 \\ &\qquad = 4|\mathbf{H}\mathbf{C}\mathbf{e}|^2 + 4(\mathbf{r} - \mathbf{H}\mathbf{c})^T \mathbf{H}\mathbf{C}\mathbf{e}. \end{aligned} \tag{15}$$

If $\mathbf{e} \in \mathcal{M}(\mathbf{c})$ then $\mathbf{e}$ satisfies $|\mathbf{r} - \mathbf{H}(\mathbf{c} - 2\mathbf{C}\mathbf{e})|^2 \leq |\mathbf{r} - \mathbf{H}\mathbf{c}|^2$ (see Definition 1), then from (15) we may conclude that such an $\mathbf{e}$ also satisfies

$$\mathbf{e}^T\mathbf{C}\mathbf{H}^T\mathbf{H}\mathbf{C}\mathbf{e} + (\mathbf{r} - \mathbf{H}\mathbf{c})^T \mathbf{H}\mathbf{C}\mathbf{e} \leq 0. \tag{16}$$

Next, note that

$$(\mathbf{r} - \mathbf{H}\mathbf{c})^T \mathbf{H}\mathbf{C} = -2\boldsymbol{\mu}^T\mathbf{Q} \tag{17}$$

follows from (11), (12) and (14) (note the identity $\mathbf{C}^2 = \mathbf{I}$ when verifying (17)). Substituting (14) and (17) into (16) gives

$$(\mathbf{e}-\boldsymbol{\mu})^T\mathbf{Q}(\mathbf{e}-\boldsymbol{\mu}) \leq \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu} \tag{18}$$

which implies $\mathbf{e} \in \mathcal{E}(\mathbf{c})$. Finally (18) holds for any $\mathbf{e} \in \mathcal{M}(\mathbf{c})$ therefore $\mathcal{M}(\mathbf{c}) \subset \mathcal{E}(\mathbf{c})$. ∎

Proposition 2. *Let $\operatorname{rank}(\mathbf{H}) = N$ and let $\mathcal{E}(\mathbf{c})$ be the ellipsoid in Definition 3. Then*

$$\max_{\mathbf{x} \in \mathcal{E}(\mathbf{c})} |\mathbf{x} - \boldsymbol{\mu}|^2 = \frac{1}{\lambda_{\min}} \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu} \tag{19}$$

where $\lambda_{\min} > 0$ is the minimum eigenvalue of $\mathbf{H}^T\mathbf{H}$.

Proof: It follows from (13) that the l.h.s. of (19) equals $\max\{|\mathbf{x}|^2 : \mathbf{x}^T\mathbf{Q}\mathbf{x} \leq \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu}\}$. Since $\operatorname{rank}(\mathbf{H}) = N$, the matrix $\mathbf{Q}$ is positive definite, thus by applying an orthonormal transformation we can show that the l.h.s. of (19) equals $\max\{|\mathbf{x}|^2 : \sum_{i=1}^N \lambda_i x_i^2 \leq \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu}\}$, where $\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ are the eigenvalues of $\mathbf{Q} = \mathbf{C}\mathbf{H}^T\mathbf{H}\mathbf{C}$. Therefore result (19) follows. It remains to show that $\lambda_{\min}$ is the minimum eigenvalue of $\mathbf{H}^T\mathbf{H}$. This follows from the fact that both $\mathbf{Q}$ and $\mathbf{H}^T\mathbf{H}$ share the same eigenvalues (verified by utilizing the property $\mathbf{C}^2 = \mathbf{I}$). ∎

It follows from Proposition 2 and the fact that $\mathcal{M}(\mathbf{c}) \subset \mathcal{E}(\mathbf{c})$ that we may then choose

$$\mathcal{U}(\mathbf{c}) \triangleq \frac{1}{\lambda_{\min}} \boldsymbol{\mu}^T\mathbf{Q}\boldsymbol{\mu} \tag{20}$$

and (8) is satisfied. Also note from Propsition 2 that $\lambda_{\min}$ does not depend on the codeword $\mathbf{c}$. Hence the minimum eigenvalue $\lambda_{\min}$ of any $\mathbf{Q}$ can be computed *offline*.

We end this subsection with a quick discussion for the memoryless AWGN channel (i.e., when $\mathbf{H} = \mathbf{I}$). We verify from (20) that $\mathcal{U}(\mathbf{c}) = |\boldsymbol{\mu}|^2$. Also, from the fact that $\mathcal{U}(\mathbf{c}) \geq |\boldsymbol{\mu}|^2$ (follows from (8) and the fact that $\mathbf{0} \in \mathcal{M}(\mathbf{c})$), we conclude that the bound $\mathcal{U}(\mathbf{c}) = |\boldsymbol{\mu}|^2$ is tight for the case $\mathbf{H} = \mathbf{I}$.

B. Code Subproblem: Obtaining $\mathcal{L}(\mathcal{S})$

Next, we address the problem of obtaining a reasonably tight bound $\mathcal{L}(\mathcal{S})$ (see (9)). We will show that the problem of computing $\mathcal{L}(\mathcal{S})$, is equivalent to a set of integer programming problems considered in [1], [2], [3] for optimality criteria for memoryless AWGN channels. We will establish that Corollary 1 is indeed a generalization of the work in [1], [2], [3].

To show the relationship with [1], [2], [3], we first need to define a set of binary sequences similar to $\mathcal{C}_{0,1}(\mathbf{c})$ in Definition 2. First, define a bipolar sequence $\mathbf{b} \in \mathbb{B}^N$ as $b_t \triangleq \operatorname{sign} \hat{r}_t$ for all t (see (12) for definition of $\hat{\mathbf{r}}$). Define $\mathbf{B}$ as the diagonal matrix whose diagonal is $\mathbf{b}$. Define the set of binary sequences

$$\mathcal{C}^\dagger_{0,1}(\mathbf{b}) \triangleq \{\boldsymbol{\varepsilon} \in \{0,1\}^N : \ \mathbf{b} - 2\mathbf{B}\boldsymbol{\varepsilon} \in \mathcal{C}_{\mathbb{B}}\}, \tag{21}$$

For a fixed pair $(\mathbf{c}, \mathbf{b})$ (or equivalently a fixed $(\mathbf{c}, \mathbf{r})$) it is clear that the equivalence

$$\mathbf{c} - 2\mathbf{C}\mathbf{e} = \mathbf{b} - 2\mathbf{B}\boldsymbol{\varepsilon} \tag{22}$$

induces a one-to-one correspondence between the elements $\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c})$ and $\boldsymbol{\varepsilon} \in \mathcal{C}^\dagger_{0,1}(\mathbf{b})$. We write $\mathcal{S}^\dagger \subset \mathcal{C}^\dagger_{0,1}(\mathbf{b})$ for the set of elements of $\mathcal{S} \subset \mathcal{C}_{0,1}(\mathbf{c})$ under this correspondence, i.e.,

$$\mathcal{S}^\dagger = \{\boldsymbol{\varepsilon} : \mathbf{b} - 2\mathbf{B}\boldsymbol{\varepsilon} = \mathbf{c} - 2\mathbf{C}\mathbf{e} \text{ where } \mathbf{e} \in \mathcal{S}\}. \tag{23}$$

Proposition 3. *For $\boldsymbol{\mu}$ given in (11), we have*

$$\min_{\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}} |\mathbf{e} - \boldsymbol{\mu}|^2 = |\boldsymbol{\mu}|^2 - \sum_{t:\ c_t\hat{r}_t < 0} |\hat{r}_t| + \min_{\boldsymbol{\varepsilon} \in \mathcal{C}^\dagger_{0,1}(\mathbf{b}) \setminus \mathcal{S}^\dagger} \sum_t |\hat{r}_t| \varepsilon_t$$

Proof: For any $\mathbf{e} \in \mathcal{C}_{0,1}(\mathbf{c})$, let $\boldsymbol{\varepsilon}$ be the corresponding element in $\mathcal{C}^\dagger_{0,1}(\mathbf{b})$ that satisfies (22). Manipulate $|\mathbf{e} - \boldsymbol{\mu}|^2$ as

$$\begin{aligned} |\mathbf{e} - \boldsymbol{\mu}|^2 &= \sum_{t:\ e_t = 1} (1 - 2\mu_t) + |\boldsymbol{\mu}|^2 = \sum_{t=1}^N (c_t\hat{r}_t) \cdot e_t + |\boldsymbol{\mu}|^2 \\ &= -\sum_{t:\ c_t\hat{r}_t < 0} |\hat{r}_t| + \sum_{t=1}^N |\hat{r}_t| \cdot \varepsilon_t + |\boldsymbol{\mu}|^2. \end{aligned} \tag{24}$$

where the last equality follows because $e_t = \boldsymbol{\varepsilon}_t$ only when $c_t = b_t \stackrel{\triangle}{=} \mathrm{sign}(\hat{r}_t)$. Minimizing the l.h.s. of (24) over $\mathcal{C}_{0,1}(\mathbf{c}) \setminus \mathcal{S}$ is equivalent to minimizing the r.h.s. over $\mathcal{C}^{\dagger}_{0,1}(\mathbf{b}) \setminus \mathcal{S}^{\dagger}$. ∎

By Proposition 3, it follows from (9) that

$$\mathcal{L}(\mathcal{S}) \leq \min_{\boldsymbol{\varepsilon} \in \mathcal{C}^{\dagger}_{0,1}(\mathbf{b}) \setminus \mathcal{S}^{\dagger}} |\hat{r}_t| \varepsilon_t. \tag{25}$$

However, lower bounding the r.h.s. of (25) has already been considered in [1], [2], [3] (for $|\mathcal{S}|$ equals 1, 2, and ≥ 3, respectively). We will utilize these existing results.

Definition 4. *Let the set $\mathcal{S} \subset \mathcal{C}_{0,1}(\mathbf{c})$ satisfy $\mathcal{M}(\mathbf{c}) \cap \mathcal{S} = \{\mathbf{0}\}$ and let $\mathcal{S}^{\dagger}$ be obtained from $\mathcal{S}$ via equivalence (23). We define $\mathcal{L}^*(\mathcal{S})$ to be the optimal cost to the following integer program*

$$\mathcal{L}^*(\mathcal{S}) \stackrel{\triangle}{=} \min_{\boldsymbol{\varepsilon} \in \{0,1\}^N} \sum_{t=1}^{N} |\hat{r}_t| \cdot \varepsilon_t$$

under constraints $d_{\mathrm{H}}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}') \geq d_{\min}$ for all $\boldsymbol{\varepsilon}' \in \mathcal{S}^{\dagger}$.

Proposition 4 ([1], [2], [3])**.** *The optimal cost $\mathcal{L}^*(\mathcal{S})$ given in Definition 4 is at most the r.h.s. of (25), i.e.*

$$\mathcal{L}^*(\mathcal{S}) \leq \min_{\boldsymbol{\varepsilon} \in \mathcal{C}^{\dagger}_{0,1}(\mathbf{b}) \setminus \mathcal{S}^{\dagger}} |\hat{r}_t| \varepsilon_t$$

Proof: The claim follows from the fact that the mapping from $\mathcal{C}_{\mathbb{B}}$ to $\mathcal{C}^{\dagger}_{0,1}(\mathbf{b})$ preserves Hamming distance, therefore it follows that any element $\boldsymbol{\varepsilon} \in \mathcal{C}^{\dagger}_{0,1}(\mathbf{b}) \setminus \mathcal{S}^{\dagger}$ (see (25)) is feasible to the integer program in Definition 4. ∎

By Proposition 4 and (25), we set

$$\mathcal{L}(\mathcal{S}) \stackrel{\triangle}{=} \mathcal{L}^*(\mathcal{S})$$

in the sequel. In [1], [2], [3] it was shown how to efficiently compute $\mathcal{L}(\mathcal{S})$ when the size $|\mathcal{S}| \leq 4$. The motivation for considering larger set sizes $|\mathcal{S}|$ is as follows. First, note that if $\mathcal{S} \subseteq \mathcal{S}'$, then $\mathcal{L}(\mathcal{S}) \leq \mathcal{L}(\mathcal{S}')$ must hold. Therefore it may be possible that the sufficient condition (10) fails for a smaller $\mathcal{S}$ but passes when a larger $\mathcal{S}'$ is used. Secondly, by the constraint that $\mathcal{S}$ must contain $\mathbf{0}$, in the case $|\mathcal{S}| = 1$ it is clear that $\mathcal{L}(\mathcal{S}) = \mathcal{L}(\{\mathbf{0}\}) \leq \mathcal{L}(\mathcal{S}')$ for any $\mathcal{S}'$ satisfying $|\mathcal{S}'| \geq 2$. However, a larger size $|\mathcal{S}|$ results in a higher complexity in the computation of $\mathcal{L}(\mathcal{S})$ (see [1], [2], [3]).

Due to lack of space, we refer the reader to [1], [2], [3] for methods to compute $\mathcal{L}(\mathcal{S})$ for sizes $|\mathcal{S}| \geq 2$. We conclude this section by stating the results for computing $\mathcal{L}(\mathcal{S})$ in the simplest case $\mathcal{S} = \{\mathbf{0}\}$, i.e. $|\mathcal{S}| = 1$.

Proposition 5 ([1])**.** *For $b_t \stackrel{\triangle}{=} \mathrm{sign}(\hat{r}_t)$, the optimal cost $\mathcal{L}^*(\mathcal{S})$ to the integer program in Definition 4 for the case $\mathcal{S} = \{\mathbf{0}\}$, is given as the sum of the $(d_{\min} - |\{t : c_t \neq b_t\}|)$-smallest values in the set*

$$\{|\hat{r}_t| : c_t = b_t\}$$

whenever $d_{\min} > |\{t : c_t \neq b_t\}|$ and $\mathcal{L}^*(\mathcal{S}) = 0$ otherwise.

Proof: Let $\mathcal{S}^{\dagger} = \{\boldsymbol{\varepsilon}'\}$ correspond to $\mathcal{S} = \{\mathbf{0}\}$ under equivalence (23). Note that any feasible $\boldsymbol{\varepsilon}$ in the integer program in Definition 4 satisfies

$$\begin{aligned} d_{\mathrm{H}}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}') &= \sum_{t:\ \varepsilon'_t = 1} (1 - \varepsilon_t) + \sum_{t:\ \varepsilon'_t = 0} \varepsilon_t \\ &= -\sum_{t:\ \varepsilon'_t = 1} \varepsilon_t + \sum_{t:\ \varepsilon'_t = 0} \varepsilon_t + |\{t : \varepsilon'_t = 1\}| \geq d_{\min}. \end{aligned} \tag{26}$$

From the equivalence (22), we verify that $\varepsilon'_t = 1$ for all values of t for which $c_t \neq b_t$ (and $\varepsilon'_t = 0$ otherwise). Therefore continuing from (26) we get that

$$-\sum_{t:\ c_t \neq b_t} \varepsilon_t + \sum_{t:\ c_t = b_t} \varepsilon_t \geq d_{\min} - |\{t : c_t \neq b_t\}|. \tag{27}$$

Since all cost coefficients $|\hat{r}_t| > 0$ (with probability 1), the optimal solution ε^* must satisfy the following properties (or else there exists a feasible solution with lower cost)

- $\varepsilon^*_t = 0$ for all values of t where $c_t \neq b_t$.
- The inequality in (27) is strictly satisfied.

Thus $\varepsilon^*_t = 1$ only for values of t corresponding to the $(d_{\min} - |\{t : c_t \neq b_t\}|)$-smallest values in the set $\{|\hat{r}_t| : c_t = b_t\}$. ∎

IV. Computer Simulations

In this section, we present test results to show the effectivenesses of the sufficient conditions for 5 different ISI channels. The $[24, 12, 8]$ extended Golay, the $[32, 16, 8]$ and $[64, 42, 8]$ Reed-Muller (RM) codes were utilized for the tests. Table I shows the chosen ISI channels. Note from Table I that the channels are numbered in decreasing order of their minimum eigenvalue $\lambda_{\min}$. Also, note that all even numbered channels have 2 non-zero taps (channel memory $I = 1$), whereas all odd numbered channels have 3 non-zero taps ($I = 2$). All channels have unit energy $\sum_{i=0}^{I} |h_i|^2 = 1$.

No.	Channel $\mathbf{h}$	$\lambda_{\min}$ $(N = 24)$	$\lambda_{\min}$ $(N = 32)$	$\lambda_{\min}$ $(N = 64)$
1	$[0.3, -0.9, -0.316]^T$	0.7873	0.7848	0.7824
2	$[0.3, 0.953]^T$	0.4309	0.4290	0.4271
3	$[0.3, -0.2, -0.932]^T$	0.2052	0.1974	0.1895
4	$[0.5, 0.866]^T$	0.1408	0.1379	0.1350
5	$[0.8, -0.2, 0.565]^T$	0.0783	0.0683	0.0579

TABLE I
5 ISI channels with corresponding $\lambda_{\min}$ for different N.

The proposed sufficient conditions are utilized in an actual list decoder, namely the ordered statistics decoding (OSD) algorithm for ISI channels [5], which performs close to MLD for the chosen test cases. In each decoding instance, the codeword $\mathbf{c}$, and the set $\mathcal{S}$, are chosen from the OSD codeword list. The sizes of $\mathcal{S}$ range from $1 \leq |\mathcal{S}| \leq 3$.

Figure 2 shows the results for the $[24, 12, 8]$ extended Golay code. The results are displayed for all the 5 channels in Table I and shown for 3 different signal-to-noise ratios[3] (SNRs). For comparison, we also show the memoryless (i.e. $\mathbf{h} = [1]$) case, where $\lambda_{\min} = 1$. The improvements obtained when using the (nested) sets $\mathcal{S}$ of larger sizes (see discussion below Proposition 5) are shown as bar-graph increments in Figure 2.

[3]Our definition of SNR is $10 \log_{10} \frac{1}{\sigma^2}$, where σ^2 is the variance of the noise (whose realization is denoted w_t in (1)).

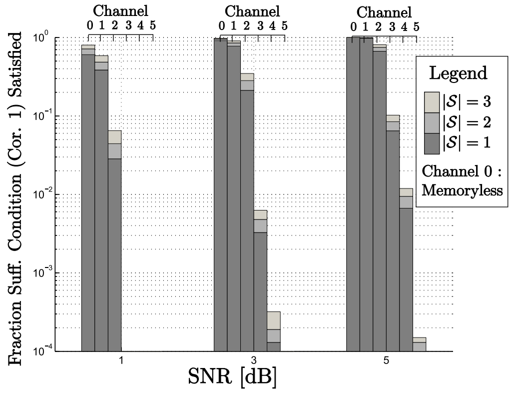

Fig. 2. $[24, 12, 8]$ Extended Golay Code

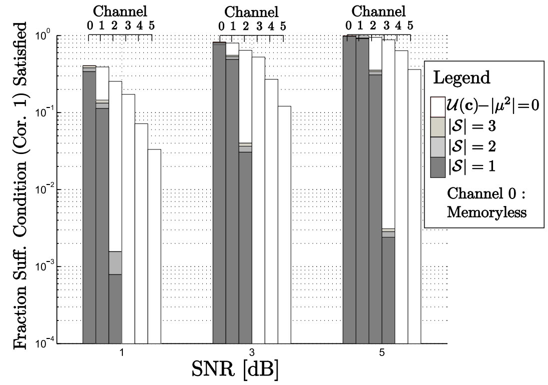

Fig. 3. $[32, 16, 8]$ Reed-Muller Code

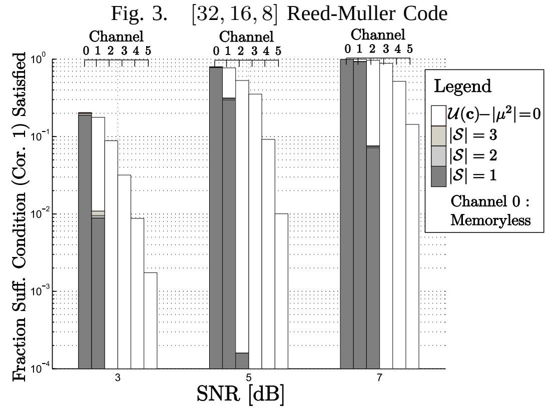

Fig. 4. $[64, 42, 8]$ Reed-Muller Code

We see from both Figure 2 and Table I that the effectiveness of the sufficient condition increase as both SNR and $\lambda_{\min}$ increase. In particular, we note that the effectiveness does not correlate with the channel memory I at all. The effectiveness vary widely amongst the 5 channels. Specifically at SNR 1 dB, in channel 1 the sufficient condition is satisfied approximately 0.4 fraction of the time, in contrast to channel 5 where it is never satisfied at all. For channel 5, the sufficient condition was observed to work at SNR 5 dB (for an insignificant 10^{-4} fraction of the time). On average, an increase in the size $|\mathcal{S}|$ by 1 is seen to improve the effectiveness by about $10 \sim 20$ %.

Figure 3 shows the results obtained for the $[32, 16, 8]$ Reed-Muller code. The results are displayed in exactly the same fashion as Figure 2, with the following difference. In Figure 3, we also show the fraction of times (10) is satisfied, when we allow $\mathcal{U}(\mathbf{c})$ to equal its lower theoretical limit $|\boldsymbol{\mu}|^2$ (trivially satisfied for the memoryless case, see Subsection III-A). This provides an optimistic estimate on the effectiveness of the sufficient condition (10) for all possible choices for $\mathcal{U}(\mathbf{c})$. We see that for channel 5 and SNR 1 dB, the sufficient condition is optimistically estimated to satisfy 3×10^{-2} fraction of the time. In contrast this fraction is seen to be lower than the case $|\mathcal{S}| = 1$ for channel 1 of the same SNR (without the optimistic estimate). This observation suggests that it is more challenging to apply sufficient conditions to channels with low $\lambda_{\min}$ (e.g. channel 5).

Finally Figure 4 shows the results obtained for the $[64, 42, 8]$ Reed-Muller Code. By comparing Figures 2 to 4, we notice that the effectiveness of the sufficient condition (10) tends to decrease as codelength N increases. We also notice diminishing improvement when using sets $\mathcal{S}$ with larger sizes $|\mathcal{S}| > 1$. For the $[64, 42, 8]$ Reed-Muller Code, the sufficient condition works on channel 1 for only a small 10^{-2} fraction of the time, at the lowest SNR. However, the optimistic estimates look reasonably high at SNR 7 dB, and it may be possible to further improve the sufficient condition for this SNR regime (which most practical schemes operate in).

V. Conclusion

We have generalized optimality conditions for testing whether a codeword is a maximum likelihood codeword to ISI channels. In our generalization, we showed that there exist 2 subproblems. The first subproblem relates to the channel characteristics, and a method was proposed to solve the first subproblem. The second subproblem is related to integer programming problems considered earlier in the literature for memoryless channels. Computer simulations show that for 3 tested codes, the proposed sufficient conditions are effective in the high SNR regime and for channels with reasonably high minimum eigenvalues $\lambda_{\min}$ of $\mathbf{H}^T\mathbf{H}$. Finally, we would like to comment that these sufficient conditions were derived irrespectively of the decoding scheme and therefore may be specialized for different coding schemes (see for example [1]), which may further improve on the results shown in this paper.

References

[1] D. J. Taipale and M. B. Pursley, "An improvement to generalized minimum distance decoding," *IEEE Trans. on Inform. Theory*, vol. 37, no. 1, pp. 167–172, Jan. 1991.

[2] T. Kaneko, T. Nishijima, H. Inazumi, and S. Hirasawa, "An efficient maximum liklihood decoding algorithm for linear block codes with algebraic decoder," *IEEE Trans. on Inform. Theory*, vol. 40, no. 2, pp. 320–327, Mar. 1994.

[3] H. Tang, T. Kasami, and T. Fujiwara, "An optimality testing algorithm for a decoded codeword of binary block codes and its computational complexity," in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, 13th International Symposium, AAECC-13, Honolulu, Hawaii, November 15-19, 1999, Proceedings*, M. Fossorier, H. Imai, S. Lin, and A. Poli, Eds., vol. 1719, 1999, pp. 201–210.

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

[5] F. Lim, A. Kavčić, and M. Fossorier, "Ordered statistics decoding of linear block codes over intersymbol interference channels," *IEEE Trans. on Magn.*, vol. 44, no. 11, pp. 3765–3768, Nov. 2008.