Universal Decoding for Frequency-Selective Fading Channels

Ofer Shayevitz, *Student Member, IEEE,* and Meir Feder, *Fellow, IEEE*

***Abstract*—We address the problem of universal decoding in unknown frequency-selective fading channels, using an orthogonal frequency-division multiplexing (OFDM) signaling scheme. A block-fading model is adopted, where the bands' fading coefficients are unknown yet assumed constant throughout the block. Given a codebook, we seek a decoder independent of the channel parameters whose worst case performance relative to a maximum-likelihood (ML) decoder that knows the channel is optimal. Specifically, the decoder is selected from a family of quadratic decoders, and the optimal decoder is referred to as a quadratic minimax (QMM) decoder for that family. As the QMM decoder is generally difficult to find, a suboptimal QMM decoder is derived instead. Despite its suboptimality, the proposed decoder is shown to outperform the generalized likelihood ratio test (GLRT), which is commonly used when the channel is unknown, while maintaining a comparable complexity. The QMM decoder is also derived for the practical case where the fading coefficients are not entirely independent but rather satisfy some general constraints. Simulations verify the superiority of the proposed QMM decoder over the GLRT and over the practically used training sequence approach.**

***Index Terms*—Decoding, fading channels, generalized likelihood ratio test (GLRT), maximum-likelihood (ML) decoding, minimax methods, orthogonal frequency-division multiplexing (OFDM), quadratic minimax (QMM) decoders, universal decoding.**

I. Introduction

In this work, we consider the long-standing problem of digital communication over an unknown frequency-selective fading channel. In many situations, neither the transmitter nor the receiver are familiar with the specific channel over which communication takes place, thus both the codebook and the decoder must be selected without knowledge of the law governing the channel. An important example for such a situation is found in mobile wireless communication, where variations of the transmitter location in a dense urban environment leads to constantly changing scattering scenarios, which in turn results in a varying channel law.

In this paper, we assume that the transmission scheme is given, and focus on receiver design. We consider slow frequency-selective fading channels [3], and we further adopt the simplified block-fading model, which assumes that the frequency response of the channel remains unchanged throughout a block of symbols, and only changes from block to block. The size of the block is usually selected to be proportional to the *coherence time* of the channel.

Within a block, the channel will be assumed to belong to a parametric family of channels. For instance, in the mobile wireless communication scenario, this family may include multipath channels with some limits on the delay spread and on the paths fading. Formally, if $\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$ are the input and the output of the channel, respectively, then the channel transition probability density function is assumed to belong to a parametric family

$$\mathcal{P} = \{f_\theta(\boldsymbol{y}|\boldsymbol{x}), \theta \in \Theta\}$$

where Θ is some index set. Such a family of channels is sometimes referred to as a *compound channel* [14]. Had the channel law $f_\theta(\boldsymbol{y}|\boldsymbol{x})$ been known in advance for each block, *maximum-likelihood* (ML) decoding rule could have been applied at the receiver to minimize the average probability of error. However, since the ML decoding rule typically varies with the channel parameter θ, it cannot be used in the compound channel setting, and thus, decoding turns into a composite hypothesis testing problem, where different hypotheses correspond to different channels in the family $\mathcal{P}$.

Several heuristic approaches to this problem have been suggested, where probably the most common is the use of a training sequence. The basic idea is for the transmitter to send a known sequence of symbols over the channel, allowing the receiver to use its knowledge of this sequence in order to estimate the specific channel law. Once the channel law is estimated, the receiver typically decodes the rest of the transmission by performing ML decoding with respect to the estimated channel. The training sequence approach has several drawbacks. First, since the channel estimation is imperfect and the decoding is performed using an incorrect likelihood function, there is a mismatched decoding penalty, which results in an increase in error rates [15] and a decrease in capacity [18]. Second, there is a throughput penalty, since the training sequence carries no information. This penalty is worse the longer the training sequence is compared to the length of the data sequence. The throughput penalty may be especially acute in wireless communications, where the channel parameters vary over time, thus requiring frequent retransmission of the training sequence in order to cope with the channel variations. We thus see that increasing the relative length of the training sequence results in throughput penalty, while decreasing it results in a more severe mismatch penalty. Due to delay constraints, this tradeoff cannot always be balanced.

Another commonly used decoder for unknown parametric channels is the *generalized likelihood ratio test* (GLRT), which is a heuristic generalization of the ML decoder. The GLRT finds

Manuscript received September 12, 2004. This work was supported in part by an Intel Research Grant. The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, March 2005.

The authors are with the Department of Electrical Engineering Systems, Tel-Aviv University, Ramat-Aviv 69978, Tel-Aviv, Israel (e-mail: ofersha@eng.tau.ac.il; meir@eng.tau.ac.il).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.851741

0018-9448/$20.00 © 2005 IEEE

the ML estimate of the channel parameters for each codeword, substitutes it into the likelihood function for that codeword, and favors the codeword with the maximum likelihood. In general, the GLRT does not claim optimality in any sense, and in some cases there exists a decoder with better performance for all channels in the family [7], which implies the GLRT is strictly suboptimal.

A more systematic approach to the general problem of decoding in unknown channels is the *universal decoding* approach [9], [14], [19]. Loosely speaking, a universal decoder for a parametric family of channels is a decoder independent of the specific channel in use, that nevertheless performs asymptotically as well as the ML decoder tuned to that channel. There are several different definitions of such universality. Universality with respect to (w.r.t.) the random coding exponent was introduced in [9], where a sequence of decoders was termed *(strong) random-coding universal* if the exponential decay rate of its error probability with increasing block length over a random selection of codebooks converged to the random coding exponent uniformly over the parameter set. The more stringent notion of *(strong) deterministic-coding universality* refers to the existence of a sequence of specific codebooks with increasing block length, so that the exponential decay rate of the error probability for the sequence of decoders using these specific codebooks converges to the random coding exponent uniformly over the parameter set.

Many families of channels admit universal decoding. Such families include *discrete memoryless channels* (DMCs) [6], finite-state channels [17], [27], and Gaussian intersymbol interference (ISI) channels [9]. An example showing that the training sequence approach is generally not universal was given in [9], and a corresponding example for the GLRT was given in [14], [16].

The aforementioned definitions of universality are asymptotic in their nature, and so the resulting universal decoders may not be suited for use in communications systems with stringent delay constraints. A somewhat different approach was presented in [10], where optimal decoders Ω_n were sought for any block length n, in the competitive minimax sense

$$\Omega_n \Leftarrow \inf \sup_{\theta\in\Theta} \frac{P_e(\theta,\Omega_n)}{P_e^*(\theta)}) \triangleq K_n \tag{1}$$

where the nominator represents the error probability of the decoder Ω_n over the channel indexed by θ, and the denominator represents the error probability of the ML decoder tuned to that channel, both for either specific or randomly selected codebooks. The error probability ratio represents the relative loss in performance incurred by employing a decoder ignorant of the channel in use, and therefore the proposed criterion seeks a decoder whose worst case relative loss is minimal.

In the case where $\frac{1}{n}\log K_n$ can be made small for a large enough n, the proposed decoder Ω_n is universal in the sense of attaining a probability of error approaching that of the ML decoder on an exponential scale. However, demanding competitive minimaxity w.r.t. the ML is sometimes too ambitious. A less demanding criterion also suggested in [10] is

$$\Omega_n \Leftarrow \inf \sup_{\theta\in\Theta} \frac{P_e(\theta,\Omega_n)}{[P_e^*(\theta)]^\xi} \triangleq K_n^\xi \tag{2}$$

for some $0 < \xi < 1$. Now, in the case where $\frac{1}{n}\log K_n^\xi$ can be made small for a large enough n, the proposed decoder Ω_n is universal in the sense of attaining an error probability approaching a fraction ξ, on an exponential scale, of the ML error probability. Naturally, the maximal value of ξ is sought, for which this still holds.

In this work, we seek practical universal decoders for unknown frequency-selective fading channels with additive independent and identically distributed (i.i.d.) Gaussian noise. For simplification, we use the orthogonal frequency-division multiplexing (OFDM) signaling scheme, which has gained much attention as an effective multicarrier technique for wireless transmissions over such channels [12]. By using the fast Fourier transform (FFT) and its inverse (IFFT) and adding a cyclic prefix to each data block, OFDM converts a frequency-selective fading channel with additive i.i.d. Gaussian noise into parallel independent subchannels (bands) with additive i.i.d. Gaussian noise [25]. Indeed, using OFDM signaling results in a rate penalty that is proportional to the ratio between the length of the cyclic prefix and the size of the data block. Despite that, since the length of the cyclic prefix is determined by the delay spread of the channel, and the data block size is practically limited by the coherence time of the channel, then if the delay spread is small relative to the coherence time, this penalty may be negligible.

Using OFDM greatly simplifies the equalization stage, when the fading coefficient are known to the receiver. When the fading coefficients are unknown, they are traditionally estimated before or while decoding. There has been extensive work on OFDM channel estimation, both training based [1], [5], [23] and blind [11], [20]. GLRT based joint channel estimation and decoding was considered as well [4].

As these methods suffer from several disadvantages mentioned earlier, we pursuit another direction, and consider the problem within the framework of universal decoding. Focusing on receiver design, we assume that the codebook is given, and so it is only natural to adopt the competitive minimax criterion given in (2). However, since we are interested in a fixed codebook, we take a somewhat different approach then [10], and consider the asymptotic behavior of the minimax solution in the limit of high signal-to-noise ratio (SNR), rather than for an increasing block size. That is, we seek the maximal value of ξ (and the corresponding decoder) for which the minimax solution converges when the SNR is taken to infinity. We note, however, that the performance of the decoders typically coincides with their asymptotic behavior even for moderate SNR levels, and therefore our analysis is not limited to the asymptotic regime but is valid in the practical regime as well.

There is another important point to be made regarding the applicability of the above minimax approach in a practical context. A minimax solution is pessimistic by nature, since it is derived by considering a worst case scenario. However, note that our minimax criterion is essentially a *regret* criterion, since it seeks a decoder whose worst case loss w.r.t. the ML is minimal, thus minimizing the regret for not knowing the channel parameters. Consequently, the worse the channel the less we expect from our decoder, hence a situation where our performance is dominated by the worst channel is avoided. In fact, the performance is actually dominated by the typically good channels for which the decoder performs most poorly compared to the best that could be attained had the channel been known. It should also be pointed out that the GLRT is asymptotically minimax in a sense of minimizing the worst case error probability itself (which corresponds to $\xi = 0$) rather than the regret [10], thereby classifying it as a more pessimistic decision rule.

The outline of the paper is as follows. Section II presents the OFDM system model, introduces the notions of pairwise decoding and quadratic decoders, and describes the ML and GLRT decoders for the OFDM setting. In Section III, the SNR-asymptotic minimax criterion for universal decoding in frequency-selective fading channels is introduced. This criterion seeks optimal decoders in a minimax sense out of a given family of decoders, and for a family of quadratic decoders, the optimal decoder is termed a quadratic minimax (QMM) decoder. The main results of the paper are presented in Section IV, where a specific design of the QMM decoder is described in detail. It turns out that in order to end up with a practical decoding scheme, suboptimal design assumptions have to be taken. The resulting suboptimal QMM decoder is intuitively appealing, as demonstrated by some specific implementation examples given in Section V. In Section VI, the proposed decoder is shown to outperform the GLRT in the minimax sense. Section VII provides simulation results for the QMM decoder and compares its performance to that of the GLRT and the training sequence approach. A summary and discussion of future research is given in Section VIII.

II. OFDM: System Model and Decoders

We consider an OFDM signaling scheme over an unknown frequency-selective fading channel. It is assumed that there are L frequency bands, where each band suffers an unknown complex fading and an additive i.i.d. Gaussian noise. The FFT/IFFT and the cyclic prefix associated with the OFDM signaling scheme will be disregarded here, being a constant part of the encoding/decoding procedure. Adopting a block-fading model, we assume that the unknown fading coefficients are constant throughout a block of K consecutive time points. We also assume that a given codebook of M codewords is used, where codewords are selected with equal probability. Each codeword occupies a single block, and can be therefore represented by an $L \times K$ matrix. It should be noted that this model is not limited merely to the OFDM setting, but rather fits any system that can be converted into L parallel channels with unknown gains. For instance, a narrowband system with an L-block-fading unknown multiplicative gain fits this model well by considering codewords that span K consecutive blocks.

Specifically, for a codebook of M codewords $\{X^{(0)}, \dots, X^{(M-1)}\}$, the output of the channel when transmitting the ith codeword is

$$Y = AX^{(i)} + Z$$
$$A = \operatorname{diag}\{\boldsymbol{a}\}, \quad X^{(i)}, Y, Z \in \mathbb{C}^{L \times K}, \quad Z_{\ell k} \sim CN(0,1)$$

where the diagonal elements $\boldsymbol{a}$ are the L unknown complex fading coefficients, $X_{\ell k}^{(i)}$ is the complex-valued component of the ith codeword transmitted on band ℓ at time point k, and the elements of the matrix Z are i.i.d. complex normal random variables. For simplicity of exposition, we will consider for the rest of the paper a real OFDM setting, which corresponds to a real-valued codebook, real fading coefficients, and normal distributed noise. However, all the results derived herein are also valid for complex OFDM channels with minor modifications; see [22] for details.

Denote by $\boldsymbol{x}_\ell^{(i)}$ the transpose of the ℓth row of the ith codeword. We define the (transmitted) power of the ith codeword in the ℓth band to be

$$P_\ell^{(i)} = \left\| \boldsymbol{x}_\ell^{(i)} \right\|^2$$

and the correlation coefficient between codewords i and j on the ℓth band to be

$$\rho_\ell^{(i,j)} = \frac{\langle \boldsymbol{x}_\ell^{(i)}, \boldsymbol{x}_\ell^{(j)} \rangle}{\sqrt{P_l^{(i)} P_\ell^{(j)}}}.$$

Since the fading coefficients are unknown, a reasonable requirement for the codebook is that no pair of codewords is colinear in any of the bands. Otherwise, the codewords are indistinguishable in that band, since even without the additive noise observations may stem from any of the two codewords under different fading values. Therefore, we will assume throughout this work that the correlation coefficients satisfy

$$\left| \rho_\ell^{(i,j)} \right| < 1 \tag{3}$$

which is equivalent to the requirement above.

Another useful representation of this channel is derived by stacking the rows of each matrix into a column vector. If we row-stack the rows of $X^{(i)}$ into a column, we get a $KL \times 1$ column vector, which is denoted by $\boldsymbol{x}^{(i)}$. Let $\boldsymbol{y}$ and $\boldsymbol{z}$ be constructed by a similar row stacking of the noise matrix Z and the channel output matrix Y, respectively. We get

$$\boldsymbol{y} = \bar{A}\boldsymbol{x}^{(i)} + \boldsymbol{z}, \qquad \bar{A} = A \otimes I_K$$

where I_K is the $K \times K$ identity matrix, and $\otimes$ stands for the Kronecker matrix product.

A. Pairwise Decoding and Quadratic Decoders

Adopting the row-stacked representation, a general *decoder* for the OFDM setting above is a mapping

$$\Omega_{\boldsymbol{a}} : \mathbb{R}^{KL} \mapsto \{0, \dots, M-1\}.$$

In general, the decoder may or may not be dependent on the value of the fading coefficients. We will naturally be interested in decoders that do not depend on the fading values, since in our setting these values are assumed unknown.

Associated with a decoder are the *decision regions* $R_i(\Omega)$ for each codeword i, defined as

$$R_i(\Omega) = \{\boldsymbol{y} \in \mathbb{R}^{KL} | \Omega(\boldsymbol{y}) = i\}.$$

We will further assume here that each decision region contains its own codeword for all fading values, i.e.,

$$\bar{A}\boldsymbol{x}^{(i)} \in R_i(\Omega), \qquad \forall\, \bar{A}, \; \forall\, i. \tag{4}$$

This is a *detectability* property which guarantees a correct decision when no noise is present. There is no reason to consider decoders that do not satisfy that property, since their probability of error is bounded from below for some of the channels, and can never approach zero for an increasing SNR.

The probability of error associated with a decoder Ω for a specific fading coefficients vector $\boldsymbol{a}$ is denoted by $P_e(\boldsymbol{a},\Omega)$. Since a precise analysis of $P_e(\boldsymbol{a},\Omega)$ for general decoders is usually hard, we will resort to high SNR approximations. We define the decoder's *power error exponent* as the asymptotic slope of the error probability as a function of the SNR on a logarithmic scale

$$E^{\Omega}(\boldsymbol{a}) = \lim_{r\to\infty} -\frac{1}{r}\log P_e(\sqrt{r}\boldsymbol{a},\Omega)$$

whenever the limit exists. As this quantity is easier to determine, we will later use it instead of the error probability. Note that the error probability typically decreases exponentially for even moderate SNR levels, and therefore using the power error exponent in lieu of the error probability does not limit our discussion to the asymptotic SNR regime, but rather renders it valid in the practical SNR regime as well.

Every decoder Ω can be decomposed (though not uniquely) into *pairwise decoders* Ω^{ij} that decide only between codewords i and j

$$\Omega^{ij}: \mathbb{R}^{KL} \mapsto \{i,j\}.$$

The essence of this decomposition is that instead of applying the decision rule Ω and directly decode one codeword out of M, decoding is performed in pairs, and the decoded codeword is the one favored by all its pairwise decoders. If no such codeword exists, then decision is made according to some inconsistency resolving rule. Practically, one can first decide between the first and second codewords, then take the "winner" and decide between it and the third codeword, and so on, until only one codeword survives. That way, a minimal number of decisions is made, and inconsistencies are inherently resolved. This implementation of pairwise decoding will be used throughout. Decoders that uniquely assign a metric to different codewords (such as the ML and the GLRT) can be decomposed into pairwise decoders simply by comparing pairwise metrics. Inversely, a decoder Ω can also be defined by stating its pairwise components Ω^{ij} (and possibly an inconsistency resolving rule).

A decoder Ω will be called a *quadratic decoder* if there exists a set of symmetric matrices H_{ij} so that Ω can be decomposed into pairwise decoders Ω^{ij}

$$\Omega^{ij}(\boldsymbol{y}) = \begin{cases} i, & \boldsymbol{y}^T H_{ij}\boldsymbol{y} > 0 \\ j, & \boldsymbol{y}^T H_{ij}\boldsymbol{y} < 0 \\ \text{arbitrary}, & \text{o.w.} \end{cases}$$

Notice that there is an implicitly assumed dependence between the matrices H_{ij} and H_{ji}, needed to ensure that the decoders Ω^{ij} and Ω^{ji} describe the same decoding rule. Specifically, it is assumed that for every i,j there is some positive constant α so that $H_{ij} = -\alpha H_{ji}$. The family of quadratic decoders, defined via the matrices H_{ij} of their pairwise components, will be of a special interest in the following sections.

For a given decoder Ω, a decomposition Ω^{ij}, a specific value of the fading coefficients vector $\boldsymbol{a}$, and two codewords i,j define the *pairwise error probability* as

$$P_e^{i\to j}(\boldsymbol{a},\Omega^{ij}) = P_r\{\boldsymbol{x}^{(j)} \text{ decoded } | \ \boldsymbol{x}^{(i)} \text{ transmitted}; \Omega^{ij},\boldsymbol{a}\}.$$

Similarly to the decoder's power error exponent, we define the *pairwise power error exponent* as

$$E_{ij}^{\Omega}(\boldsymbol{a}) = \lim_{r\to\infty} -\frac{1}{r}\log P_e^{i\to j}\left(\sqrt{r}\boldsymbol{a},\Omega^{ij}\right)$$

whenever the limit exists. Notice that generally $E_{ij}^{\Omega}(\boldsymbol{a}) \neq E_{ji}^{\Omega}(\boldsymbol{a})$.

For a constant number of codewords M, the decoder's error probability is dominated by the worst pair, and so the power error exponent can easily be shown to be equal to the minimal pairwise exponent

$$E^{\Omega}(\boldsymbol{a}) = \min_{i\neq j} E_{ij}^{\Omega}(\boldsymbol{a}). \tag{5}$$

Associated with any pairwise decomposition are the *pairwise separating surfaces* defined as

$$S_{ij} = \overline{R_i(\Omega^{ij})} \cap \overline{R_j(\Omega^{ij})}$$

where the bar stands for the closure operator. For instance, a pairwise separating surface of a quadratic decoder is given by

$$S_{ij} = \{\boldsymbol{y} | \boldsymbol{y}^T H_{ij}\boldsymbol{y} = 0\}.$$

For a specific value of the fading coefficients $\boldsymbol{a}$, the *pairwise minimal distance* for codewords i,j is defined as

$$d_{ij}^{\Omega}(\boldsymbol{a}) = \inf_{\boldsymbol{y}\in S_{ij}} \|\boldsymbol{y} - \bar{A}\boldsymbol{x}^{(i)}\|$$

and the *decoder's minimal distance* for the channel $\boldsymbol{a}$ is defined to be

$$d^{\Omega}(\boldsymbol{a}) = \min_{i\neq j} d_{ij}^{\Omega}(\boldsymbol{a}).$$

Notice that generally $d_{ij}(\boldsymbol{a}) \neq d_{ji}(\boldsymbol{a})$, since in general one codeword may be closer to the separation surface than the other. The ML decoder, whose decision rule depends on the knowledge of the specific channel realization, is the only decoder for which these pairwise minimal distances coincide for all fading values.

Under some general conditions on the separating surface, and since the noise is assumed to be Gaussian i.i.d., it is easy to verify that there exist a simple relation between the pairwise power error exponent and the pairwise minimal distance, given by

$$E_{ij}^{\Omega}(\boldsymbol{a}) = \frac{1}{2}\left|d_{ij}^{\Omega}(\boldsymbol{a})\right|^2 \tag{6}$$

see [22] for details. Specifically, relation (6) holds for quadratic decoders and for the ML decoder. Using (5), the decoder's power error exponent is similarly related to the decoder's minimal distance

$$E^{\Omega}(\boldsymbol{a}) = \frac{1}{2}|d^{\Omega}(\boldsymbol{a})|^2. \tag{7}$$

B. The ML and GLRT Decoders

The ML decoder for the OFDM setting is defined by

$$\Omega^*(\boldsymbol{y};\boldsymbol{a}) = \arg\max_i f_{\boldsymbol{a}}(\boldsymbol{y} \,|\, \boldsymbol{x}^{(i)})$$

where ties are broken arbitrarily. Since the noise is additive Gaussian and i.i.d., the ML decoder can be stated as a minimum distance decoder

$$\Omega^*(\boldsymbol{y};\boldsymbol{a}) = \arg\min_i \|\boldsymbol{y} - \bar{A}\boldsymbol{x}^{(i)}\|^2.$$

A straightforward decomposition of the ML decoder to pairwise decoders Ω^{*ij} is achieved simply by comparing distances of pairs. The separating surface S_{ij} for this decomposition is therefore the locus of all points with equal Euclidian distance to codewords i and j, which is a hyperplane. Consequently, the pairwise error exponent for the ML decoder is given by

$$\begin{aligned} E^*_{ij}(\boldsymbol{a}) &= \frac{1}{8}\|\bar{A}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})\|^2 = \frac{1}{8}\sum_{\ell=0}^{L-1}\left\|\boldsymbol{x}_\ell^{(i)} - \boldsymbol{x}_\ell^{(j)}\right\|^2 a_\ell^2 \\ &= \frac{1}{8}\sum_{\ell=0}^{L-1}\left(P_\ell^{(i)} + P_\ell^{(j)} - 2\rho_\ell^{(ij)}\sqrt{P_\ell^{(i)}P_\ell^{(j)}}\right)a_\ell^2 \end{aligned}$$

and the power error exponent for the ML decoder is

$$E^*(\boldsymbol{a}) = \frac{1}{8}\min_{i\neq j}\left\{\sum_{\ell=0}^{L-1}\left(P_\ell^{(i)} + P_\ell^{(j)} - 2\rho_\ell^{(ij)}\sqrt{P_\ell^{(i)}P_\ell^{(j)}}\right)a_\ell^2\right\}. \tag{8}$$

The corresponding probability of error for the ML decoder will be denoted by $P_e^*(\boldsymbol{a})$.

Since the ML decoder is tuned to a specific channel, it cannot be used for decoding when the channel parameters are unknown, and have no statistical model. In that case, a reasonable candidate for a decoder may be the GLRT, which is a heuristic generalization of the ML. The idea is to replace the parameters with their respective ML estimates under each codeword, and then perform ML decoding. Therefore, the decoding rule of the GLRT is

$$\Omega^{\mathrm{GLRT}}(\boldsymbol{y}) = \arg\max_i\left\{\max_{\boldsymbol{a}} f_{\boldsymbol{a}}(\boldsymbol{y} \,|\, \boldsymbol{x}^{(i)})\right\}$$

where ties are broken arbitrarily. In the OFDM setting, a straightforward calculation shows that the GLRT has the following compact and comprehensible decision rule [22]:

$$\Omega^{\mathrm{GLRT}}(\boldsymbol{y}) = \arg\max_i \sum_{\ell=0}^{L-1}\frac{|\langle \boldsymbol{x}_\ell^{(i)}, \boldsymbol{y}_\ell\rangle|^2}{\|\boldsymbol{x}_\ell^{(i)}\|^2}. \tag{9}$$

That is, one can calculate a GLRT metric for every codeword by projecting the observation onto the direction of that codeword in each band separately, and summing the squares of the distances. The decoded codeword would be the one with the largest sum. Again, a natural decomposition to pairwise decoders is available by comparing pair metrics. Notice that the GLRT does not take into account the power of the codeword in each band, and considers only its direction. This is an inherent weakness of the GLRT as we shall later see.

It is easily verified from (9) that the GLRT is a quadratic decoder under the pairwise metrics decomposition. The power error exponent for the GLRT is therefore related to the minimal distance according to (7), and it is discussed in Section VI.

III. Universal Minimax Decoders for Fading Channels

In this section, universal decoding based on the competitive minimax approach will be suggested for unknown frequency-selective fading channels, assuming the OFDM block-fading model. This approach is based on the work presented in [10], with two main differences: First, a fixed codebook is assumed and the optimal decoder is considered in the limit of high SNR rather than in the limit of increasing block size. This is done by using the power error exponents in lieu of the error probability in the minimax criterion, and allowing unlimited fading power. Second, unlike [10] where the decoder's selection is unrestricted and therefore the optimal decoder is hard to find, only candidate decoders belonging to some given family are considered. While this may lead to some loss in performance, the family is chosen so that determining the optimal decoder becomes more tractable. Specifically, given a family $\mathcal{F}$ of decoders, we seek an optimal decoder $\Omega \in \mathcal{F}$ in the competitive minimax sense

$$\inf_{\Omega\in\mathcal{F}}\sup_{\boldsymbol{a}}\frac{P_e(\Omega,\boldsymbol{a})}{P_e^*(\boldsymbol{a})}. \tag{10}$$

Now, in the limit of high SNR, the minimax criterion above may be well approximated by considering only the exponential behavior of the error probabilities. The minimax criterion can now be stated in terms of the power error exponents

$$\begin{aligned} \inf_{\Omega\in\mathcal{F}}\sup_{\boldsymbol{a}}\frac{P_e(\Omega,\boldsymbol{a})}{P_e^*(\boldsymbol{a})} &\approx \inf_{\Omega\in\mathcal{F}}\sup_{\|\boldsymbol{a}\|=1}\sup_{r\gg 0}\frac{\exp\{-rE^{\Omega}(\boldsymbol{a})\}}{\exp\{-rE^*(\boldsymbol{a})\}} \\ &\Rightarrow \inf_{\Omega\in\mathcal{F}}\sup_{\|\boldsymbol{a}\|=1}\sup_{r\gg 0}(E^*(\boldsymbol{a}) - E^{\Omega}(\boldsymbol{a}))r. \end{aligned} \tag{11}$$

Again, it should be emphasized that despite the asymptotic approach, our analysis is typically valid even for moderate SNR levels, where the behavior of the decoders usually coincides with their asymptotic behavior. Moreover, since our criterion minimizes the worst case regret rather than the worst case error exponent, it is more optimistic by nature and is not dominated by the worst channel.

Considering (11), it can easily be seen that the maximization will diverge to infinity for any decoder that has a power error exponent inferior to that of the ML even for a single value of the channel fading vector. Since the ML decoder is the decoder with the minimal probability of error, it is unlikely that a decoder ignorant of the fading will match its power error exponent for all values of $\boldsymbol{a}$, for a fixed-sized codebook. Consequently, we modify the minimax criterion (10), in the same manner suggested in [10], and demand optimal minimaxity relative to a fraction of the ML power error exponent

$$\inf_{\Omega\in\mathcal{F}}\sup_{\boldsymbol{a}}\frac{P_e(\Omega,\boldsymbol{a})}{(P_e^*(\boldsymbol{a}))^{\xi}} \tag{12}$$

for $0 \le \xi \le 1$. Again, the above can be approximated at high SNR by

$$\inf_{\Omega\in\mathcal{F}} \sup_{\|\boldsymbol{a}\|=1} \sup_{r\gg 0} \frac{\exp\{-rE^{\Omega}(\boldsymbol{a})\}}{\exp\{-r\xi E^{*}(\boldsymbol{a})\}} \Rightarrow \inf_{\Omega\in\mathcal{F}} \sup_{\|\boldsymbol{a}\|=1} \sup_{r\gg 0} (\xi E^{*}(\boldsymbol{a}) - E^{\Omega}(\boldsymbol{a}))r.$$

A decoder that is optimal by this criteria, must achieve a power error exponent at least as good as a fraction ξ of the ML exponent, uniformly over all fading values. For a high value of ξ, such a decoder may not exist at all, since the supremum will always diverge. For a low value of ξ, there may be many optional decoders. Therefore, we would be interested in the maximal value of ξ for which such a decoder exists

$$\xi^{*} \triangleq \sup\left\{\xi : \exists\Omega\in\mathcal{F}, \inf_{\|\boldsymbol{a}\|=1}\{E^{\Omega}(\boldsymbol{a}) - \xi E^{*}(\boldsymbol{a})\} \ge 0\right\} \tag{13}$$

and the decoder we seek is one attaining that maximal fraction ξ^{*} of the ML exponent. Another useful expression for ξ^{*} can be obtained by defining ξ^{Ω} as the guaranteed fraction of the ML exponent attained by a decoder Ω

$$\xi^{\Omega} \triangleq \inf_{\|\boldsymbol{a}\|=1} \frac{E^{\Omega}(\boldsymbol{a})}{E^{*}(\boldsymbol{a})}$$

and taking ξ^{*} as the supremum of ξ^{Ω} over the decoders $\Omega \in \mathcal{F}$

$$\xi^{*} = \sup_{\Omega\in\mathcal{F}} \xi^{\Omega} = \sup_{\Omega\in\mathcal{F}} \inf_{\|\boldsymbol{a}\|=1} \frac{E^{\Omega}(\boldsymbol{a})}{E^{*}(\boldsymbol{a})}. \tag{14}$$

In the rest of the paper we focus on the case where $\mathcal{F}$ is a family of quadratic decoders. The decoder attaining ξ^{*} is then termed the *quadratic minimax (QMM) decoder* w.r.t. $\mathcal{F}$, and it is discussed in detail in the next section.

IV. The QMM Decoder

We are now ready to derive the main results of the paper, as we address the problem of QMM decoding with the goal of establishing a practical decoding scheme in mind. The outline of the section is as follows. In Section IV-A, we first show that it is sufficient to consider decoders whose pairwise decision rule takes into account only the projection of the observation vector onto the subspace spanned by the two corresponding codewords in each band. This is explained intuitively by asserting that anything orthogonal is noise. Such pairwise decoders are each dependent on the selection of a 2×2 symmetric matrix per band. To further simplify the analysis, at the cost of some possible performance degradation, we restrict our attention to a family denoted by $\mathcal{F}$, which includes the GLRT, and for which the 2×2 matrices above are diagonal. In Section IV-B, we derive a lower bound for the power error exponent of decoders in $\mathcal{F}$, since the exact exponent does not have an analytic expression. It is further shown, that for the sake of maximizing that lower bound, it is sufficient to consider a family $\mathcal{F}^{*} \subset \mathcal{F}$ of decoders Ω with pairwise components Ω^{ij} each dependent only on a single weight parameter λ_{ij}. The decoding rule for decoders in $\mathcal{F}^{*}$ is described in Section IV-C. The task of determining the optimal weights λ_{ij} is addressed in Section IV-D. A lower bound on the guaranteed fraction of the ML exponent ξ is derived using the power error exponent bound, and the procedure for selecting the weights so that this bound is maximized is described therein. Section IV-E concludes our discussion of QMM decoding by extending the decoding scheme to the practical case where the fading coefficients are known to be related, and provides two illustrative examples for that case.

A. Specifying the Family of Decoders

When making a pairwise decision for some pair of codewords, it is only sensible to take into account, in each band, only the projection of the observation onto the subspace spanned by those codewords in that band. This notion is now made precise.

Let $X^{(i)}$, $X^{(j)}$ be a pair of $L \times K$ codewords, and $\boldsymbol{x}_\ell^{(i)}$, $\boldsymbol{x}_\ell^{(j)}$ the transpose of their respective rows. As we have mentioned in Section II, we consider codebooks that are "suited" for universal decoding by satisfying condition (3), which means that $\boldsymbol{x}_\ell^{(i)}, \boldsymbol{x}_\ell^{(j)}$ are not colinear for any ℓ. Let $\{\boldsymbol{\psi}_{\ell,k}\}_{k=0}^{K-3}$ be a set of $K-2$ orthonormal vectors constituting a basis for the orthogonal complementary subspace of $\mathrm{span}\{\boldsymbol{x}_\ell^{(i)}, \boldsymbol{x}_\ell^{(j)}\}$. Now, define the $KL \times KL$ block-diagonal matrix

$$X^{(ij)} = \mathrm{diag}\left\{X_0^{(ij)}, X_1^{(ij)}, \ldots, X_{L-1}^{(ij)}\right\}$$

where the matrices $X_\ell^{(ij)}$ on the diagonal are

$$X_\ell^{(ij)} = \left[\boldsymbol{x}_\ell^{(i)}, \boldsymbol{x}_\ell^{(j)}, \boldsymbol{\psi}_{\ell,0}, \ldots, \boldsymbol{\psi}_{\ell,K-3}\right].$$

For the sake of brevity, we will omit the subscripts i, j throughout the rest of this section, and use X, H instead of $X^{(ij)}$, H_{ij}. Condition (3) guarantees that the matrix X is invertible, hence, the observation vector $\boldsymbol{y}$ can be represented w.r.t. X by

$$\boldsymbol{u} = X^{-1}\boldsymbol{y}. \tag{15}$$

Notice that the first two elements of $\boldsymbol{u}$ correspond to the projection of the observation onto the subspace spanned by the two codewords in the first band, followed by $K-2$ elements corresponding to the orthogonal (noise) subspace in that band. Then, the next two elements correspond to a similar projection in the second band, followed by $K-2$ elements of noise, and so forth. We now have the following result.

Theorem 1: For any pairwise decoder $\hat{\Omega}^{ij}$ represented by a symmetric matrix $\hat{H}$, there exists another pairwise decoder Ω^{ij} represented by a symmetric matrix

$$H = X^{-T} Q X^{-1} \tag{16}$$

where Q is a $KL \times KL$ block-diagonal matrix of the form

$$Q = \mathrm{diag}\{Q_0, Q_1, \ldots, Q_{L-1}\}$$

$$Q_\ell = \begin{bmatrix} \alpha_\ell & \beta_\ell & 0 & \cdots & 0 \\ \beta_\ell & \gamma_\ell & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & 0 \end{bmatrix}_{K\times K}$$

so that Ω^{ij} has a minimal distance equal or higher than that of $\hat{\Omega}^{ij}$ uniformly for all fading values.

Proof: See the Appendix. □

From Theorem 1 it follows that for the sake of finding the QMM decoder for the family of all quadratic decoders, it is sufficient to consider only the family of decoders whose pairwise components are of the form given in (16). Each pairwise decoder in that family is dependent on the selection of a 2×2 symmetric matrix per band. For simplification, we shall only consider decoders with pairwise components for which this matrix is diagonal, i.e., $\forall\, \ell,\, \beta_\ell = 0$, and we define $\mathcal{F}$ to be the family of all such decoders. In the case of flat fading (single band, $L = 1$), it can be shown that this restriction incurs in no loss of generality [22], but this is not necessarily true for $L > 1$.

We now turn to find the power error exponent for decoders from the family $\mathcal{F}$. For that matter, the pairwise minimal distances for each pair should be determined, depending on the selection of the matrix H.

The pairwise minimal distance $d_{ij}^{\Omega}(\boldsymbol{a})$ of a pairwise decoder Ω^{ij} is given by the solution of the following optimization problem:

$$\left|d_{ij}^{\Omega}(\boldsymbol{a})\right|^2 = \min_{\boldsymbol{y}} \|\boldsymbol{y} - X\tilde{\boldsymbol{a}}\|^2 \quad \text{s.t.} \quad \boldsymbol{y}^T H \boldsymbol{y} = 0$$

where H represents the pairwise decoding rule, and

$$\tilde{\boldsymbol{a}} = (\underbrace{a_0, 0, \ldots, 0}_{K}, \underbrace{a_1, 0, \ldots, 0}_{K}, \ldots\ldots, \underbrace{a_{L-1}, 0, \ldots, 0}_{K}).$$

Using the transformation (15) and the structure of the matrix H given in (16), the optimization problem can be stated as

$$\min_{\boldsymbol{u}} \|X(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \quad s.t. \quad \boldsymbol{u}^T Q \boldsymbol{u} = 0. \tag{17}$$

Taking the derivative of the lagrangian w.r.t. $\boldsymbol{u}$ and rearranging the terms, we get

$$(I + \mu(X^T X)^{-1} Q)\boldsymbol{u} = \tilde{\boldsymbol{a}} \tag{18}$$

where μ is the Lagrange multiplier. In order to find μ we have to substitute $\boldsymbol{u}$ into the constraint and we have

$$\tilde{\boldsymbol{a}}^T (I + \mu(X^T X)^{-1} Q)^{-T} Q (I + \mu(X^T X)^{-1} Q)^{-1} \tilde{\boldsymbol{a}} = 0$$

which results in a $2KL$ degree equation in μ that generally cannot be solved analytically, unless some unique situation occurs (such as per-band orthogonal codewords, for instance). Consequently, the minimax problem (14) cannot be stated explicitly. To allow further analysis, we replace the power error exponent in the minimax problem with a lower bound which we now derive.

B. Bounding the Power Error Exponent

We now turn to lower-bound the power error exponent for decoders in the family $\mathcal{F}$ by bounding the pairwise minimal distance. First, we transform the optimization problem (17) into a problem we can solve explicitly. Define a matrix $T = \Gamma X^{-1}$, for some diagonal matrix Γ. Now consider

$$\min_{\boldsymbol{u}} \|TX(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \quad \text{s.t.} \quad \boldsymbol{u}^T Q \boldsymbol{u} = 0$$

which is equivalent to

$$\min_{\boldsymbol{u}} \|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \quad \text{s.t.} \quad \boldsymbol{u}^T Q \boldsymbol{u} = 0. \tag{19}$$

For convenience of further calculations, let $Q = \Gamma^T \Delta \Gamma$ for some diagonal matrix Δ. The pairwise decoder Ω^{ij} is therefore assumed to be represented by a matrix $H = X^{-T}\Gamma^T \Delta \Gamma X^{-1}$, where the matrices Γ, Δ may both depend on i, j. Notice that selecting Γ, Δ for Ω^{ij} automatically defines a corresponding selection for Ω^{ji}. Now (19) becomes

$$\min_{\boldsymbol{u}} \|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \quad \text{s.t.} \quad \boldsymbol{u}^T \Gamma^T \Delta \Gamma \boldsymbol{u} = 0. \tag{20}$$

We will refer to the solution of the modified optimization problem above as the *pairwise modified distance*.

We now find the relation between the pairwise minimal distance and the pairwise modified distance, and for that matter we remind the reader of some facts from linear algebra, regarding the norm variations of vectors when multiplied by a matrix [13]. For any matrix $T \in \mathbb{C}^{N \times N}$, the matrix lower bound $l(T)$ and the (induced) *matrix norm* $\|T\|$ are defined as

$$l(T) \triangleq \inf_{\|\boldsymbol{v}\| \neq 0} \frac{\|T\boldsymbol{v}\|}{\|\boldsymbol{v}\|}, \qquad \|T\| \triangleq \sup_{\|\boldsymbol{v}\| \neq 0} \frac{\|T\boldsymbol{v}\|}{\|\boldsymbol{v}\|}$$

and the ratio $^{\|T\|}/_{l(T)}$ is referred to as the *condition number* of the matrix. Obviously, for any vector $\boldsymbol{v} \in \mathbb{C}^N$ we have

$$l^2(T)\|\boldsymbol{v}\|^2 \leq \|T\boldsymbol{v}\|^2 \leq \|T\|^2 \|\boldsymbol{v}\|^2$$

and, therefore, the matrix lower bound and the matrix norm provide lower and upper bounds on the norm variations of a vector multiplied by a matrix. The following lemma states a known relation between these quantities and the singular values of the matrix.

Lemma 1: For any matrix $T \in \mathbb{C}^{N \times N}$, $l(T)$ and $\|T\|$ correspond to the minimal and maximal singular values of T, respectively.

In our context, we have

$$\begin{aligned}\|X\boldsymbol{v}\|^2 &= \|T^{-1}TX\boldsymbol{v}\|^2 \geq l^2(T^{-1})\|TX\boldsymbol{v}\|^2 \\ &= l^2(T^{-1})\|\Gamma\boldsymbol{v}\|^2\end{aligned}$$

and, similarly, $\|X\boldsymbol{v}\|^2 \leq \|T^{-1}\|^2 \|\Gamma\boldsymbol{v}\|^2$. So, for every $\boldsymbol{u}$ we have

$$\begin{aligned}l^2(T^{-1})\|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 &\leq \|X(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \\ &\leq \|T^{-1}\|^2 \|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2\end{aligned}$$

minimizing over the separation surface

$$S = \{\boldsymbol{u} | \boldsymbol{u}^T \Gamma^T \Delta \Gamma \boldsymbol{u} = 0\}$$

we end up with

$$\begin{aligned}l^2(T^{-1}) \min_{\boldsymbol{u} \in S} \|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 &\leq \min_{\boldsymbol{u} \in S} \|X(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2 \\ &\leq \|T^{-1}\|^2 \min_{\boldsymbol{u} \in S} \|\Gamma(\boldsymbol{u} - \tilde{\boldsymbol{a}})\|^2.\end{aligned} \tag{21}$$

Equation (21) relates the pairwise minimal distance to the pairwise modified distance, in terms of lower and upper bounds provided using the matrix lower bound and matrix norm of T^{-1}. In light of that, we wish to find a matrix Γ so that the matrix $T^{-1} = X\Gamma^{-1}$ has a minimal condition number, thereby minimizing the upper-to-lower bound ratio.

In order to state our next results and to make further calculation more clear, we will denote the diagonal elements of Γ by $\gamma_{k,\ell}$, where $k = 0, \ldots, K-1$ and $\ell = 0, \ldots, L-1$. That is, for $\ell = 0$ and k running, we get the first K elements on the diagonal, for $\ell = 1$ we get the next K elements, and so forth. Respectively, we will mark the elements of the vector $\boldsymbol{u}$ by $u_{k,\ell}$, and of the diagonal matrix Δ by $\delta_{k,\ell}$.

Theorem 2: Consider the class of all matrices of the form $T^{-1} = X\Gamma^{-1}$ for some diagonal matrix Γ. The matrix $\hat{T}^{-1} = X\hat{\Gamma}^{-1}$ where the diagonal elements of $\hat{\Gamma}$ are given by

$$\gamma_{k,\ell} = \begin{cases} \sqrt{P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)}, & k=0 \\ \sqrt{P_\ell^{(j)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)}, & k=1 \\ 1, & k=2,\dots,K-1 \end{cases}$$

has a matrix lower bound $l(\hat{T}^{-1}) = 1$, and attains the minimal condition number in that class.

Proof: See the Appendix. □

After the optimal bounding matrix $\hat{T} = X\hat{\Gamma}^{-1}$ with $\hat{\Gamma}$ given above has been determined, we turn to explicitly find the pairwise modified distance by solving the optimization problem (20) substituting $\hat{\Gamma}$ for Γ

$$|\hat{d}_{ij}(\boldsymbol{a})|^2 \triangleq \min_{\boldsymbol{u}} \|\hat{\Gamma}(\boldsymbol{u}-\tilde{\boldsymbol{a}})\|^2 \quad \text{s.t.} \quad \boldsymbol{u}^T\hat{\Gamma}^T\Delta\hat{\Gamma}\boldsymbol{u} = 0 \quad (22)$$

where Δ is any diagonal matrix selected so that the resulting pairwise decoder satisfies the detectability condition (4). From (21) we have

$$\left|d_{ij}^{\Omega}(\boldsymbol{a})\right|^2 \geq l^2(\hat{T}^{-1})|\hat{d}_{ij}(\boldsymbol{a})|^2 = |\hat{d}_{ij}(\boldsymbol{a})|^2$$

and so $\hat{d}_{ij}(\boldsymbol{a})$ can be used to lower-bound the pairwise minimal distance (and consequently, to lower-bound the pairwise power error exponent). Generally, of course, the matrix Δ is dependent on the selection of $K \cdot L$ diagonal elements. Since our pairwise decoders consider only projections onto the codewords' subspaces in each band, we can narrow this selection down to $2L$ elements. As it turns out, a further reduction is possible, so that the matrix Δ is only dependent on the selection of a **single** parameter.

Theorem 3: For any selection of the matrix Δ, there exists another selection of the form

$$\delta_{k,l} = \begin{cases} 1, & k=0, \forall\ell \\ -\lambda, & k=1, \forall\ell \\ 0, & \text{o.w.} \end{cases}$$

for some $\lambda > 0$, for which both $\hat{d}_{ij}(\boldsymbol{a})$ and $\hat{d}_{ji}(\boldsymbol{a})$ are uniformly increased. Furthermore, for such a selection, the pairwise modified distance is given by

$$|\hat{d}_{ij}(\boldsymbol{a})|^2 = \frac{1}{1+\lambda}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2.$$

Proof: See the Appendix. □

We thus see that if we intend to use $\hat{d}_{ij}(\boldsymbol{a})$ instead of $d_{ij}^{\Omega}(\boldsymbol{a})$, it is sufficient to consider the family $\mathcal{F}^* \subset \mathcal{F}$ of decoders whose pairwise components Ω^{ij} are of the form $H = X^{-T}\hat{\Gamma}^T\Delta\hat{\Gamma}X^{-1}$, and are each dependent on a single parameter λ_{ij}. Notice that there is an inherent redundancy in the weights, and λ_{ij} is implicitly assumed to be equal to $\frac{1}{\lambda_{ji}}$, for the pairwise decoders Ω^{ij} and Ω^{ji} to describe the same decoding rule. Every decoder $\Omega \in \mathcal{F}^*$ is, therefore, determined by a set of $\frac{M(M-1)}{2}$ weights λ_{ij}.

Corollary 1: The power error exponent of a quadratic decoder $\Omega \in \mathcal{F}^*$ is lower-bounded by

$$E^{\Omega}(\boldsymbol{a}) \geq \frac{1}{2}\min_{i\neq j}\left\{\frac{1}{1+\lambda_{ij}}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2\right\} \triangleq \hat{E}^{\Omega}(\boldsymbol{a}). \quad (23)$$

Using Corollary 1, it is now straightforward to derive a lower bound for ξ^{Ω}, the guaranteed fraction attained by a decoder $\Omega \in \mathcal{F}^*$, by replacing the power error exponent with its lower bound

$$\xi^{\Omega} = \inf_{\|\boldsymbol{a}\|=1}\frac{E^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \geq \inf_{\|\boldsymbol{a}\|=1}\frac{\hat{E}^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \triangleq \hat{\xi}^{\Omega} \quad (24)$$

where the ML exponent $E^*(\boldsymbol{a})$ is given in (8). By definition, the QMM decoder for the family $\mathcal{F}^*$ (and $\mathcal{F}$) is the one maximizing ξ^{Ω}. In what follows, we shall find a decoder $\Omega \in \mathcal{F}^*$ maximizing the lower bound $\hat{\xi}^{\Omega}$ instead, as we lack an explicit expression for ξ^{Ω}. Although being suboptimal, this decoder will still be referred to as the QMM decoder, and its guaranteed fraction of the ML exponent will still be denoted ξ^*, in order to avoid to many notations. As we shall see in Section IV, there exists a simple procedure for finding the optimal weights λ_{ij}^* for that decoder, and it is described therein. But first, we describe the QMM decoding procedure, assuming the optimal weights are given.

C. QMM Decoding Procedure

As described above, the decoder we propose depends on a set of optimal weights λ_{ij}^*, which maximize the lower bound of (24) over the family $\mathcal{F}^*$. These optimal weights can be found offline, and the procedure for that is given in the next subsection.

We now describe the decoding procedure assuming that λ_{ij}^* are given. Decoding is pairwise based, therefore, at each step a decision is made regarding two codewords, where the favored codeword survives and is then confronted with a new yet untested codeword. The last codeword to survive is the one declared as decoded. For some pair of codewords i and j, the pairwise decision rule is as follows. First, the observation vector is projected onto the subspace spanned by the two codewords in each band separately. Then, the projection is expressed as a linear combination of the two codewords in that band. Specifically, let $\boldsymbol{y}_\ell$ be the observation vector in band ℓ and let $\boldsymbol{r}_\ell$ be its projection, then

$$\boldsymbol{r}_\ell = \alpha_\ell\boldsymbol{x}_\ell^{(i)} + \beta_\ell\boldsymbol{x}_\ell^{(j)}.$$

The codeword i is favored over the codeword j if

$$\sum_\ell\left[P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)\right]\alpha_\ell^2 \geq \lambda_{ij}^*\sum_\ell\left[P_\ell^{(j)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)\right]\beta_\ell^2 \quad (25)$$

where $P_\ell^{(i)}$ is the transmitted power of codeword i in band ℓ, and $\rho_\ell^{(ij)}$ is the correlation coefficient for codewords i and j in band ℓ.

As for complexity, observe that apart from the preliminary step of finding the optimal weights λ_{ij}^* which can be preformed *a priori*, the QMM decoder has a computational complexity comparable to that of the GLRT. The only problem may be memory, since there are many weights that need to be stored for decoding. In some cases, memory consumption may be reduced at the cost of some online weight calculation, as mentioned in the next subsection.

D. Optimal Weights Selection

We return to the problem of determining the optimal weights λ^*_{ij} to be used in the QMM decoding procedure. Precisely, we are interested in solving the following minimax problem:

$$\xi^* \geq \sup_{\Omega \in \mathcal{F}^*} \hat{\xi}^\Omega = \sup_{\Omega \in \mathcal{F}^*} \inf_{\|\boldsymbol{a}\|=1} \frac{\hat{E}^\Omega(\boldsymbol{a})}{E^*(\boldsymbol{a})} \triangleq \hat{\xi}^* \tag{26}$$

where $\hat{\xi}^*$ denotes the maximal lower bound attained by a decoder solving the problem. We would be first interested in finding an expression dependent on the parameters λ_{ij} for the infimum, i.e., for $\hat{\xi}^\Omega$, and then maximize it w.r.t. those parameters. Unfortunately, since this involves finding the infimum of a ratio of two rather complicated functions, there is little hope of finding an analytical expression for $\hat{\xi}^\Omega$ in general. However, there is a way around this.

For a given decoder, the channels $\boldsymbol{a}$ for which the infimum in the right-hand side of (26) is attained will be referred to as *critical channels*. In these channels, the quadratic decoder has the worst value of the power error exponent bound relative to the ML exponent, and if $\boldsymbol{a}_0$ happens to be a critical channel, then the lower bound $\hat{\xi}^\Omega$ is just the ratio $\hat{E}^\Omega(\boldsymbol{a}_0)/E^*(\boldsymbol{a}_0)$. Therefore, for any specific decoder, the lower bound we seek can be found by determining any one of the critical channels. Generally, the critical channels are difficult to determine in closed form, and they may vary from one quadratic decoder to another, as they depend on the selection of the weights λ_{ij}. However, as we shall now show, there exists a finite set of channels independent of the weights, that always includes at least one critical channel. In order to show that, we need some definitions related to polyhedral sets and functions [2].

A set $P \subset \mathbb{R}^N$ is said to be a *polyhedra* if

$$P = \left\{\boldsymbol{x} \in \mathbb{R}^N | \boldsymbol{w}_j^T \boldsymbol{x} \leq c_j, j = 1, \dots, r\right\}$$

for some $\boldsymbol{w}_j \in \mathbb{R}^N$ and $c_j \in \mathbb{R}$. The *graph* of a function $f : P \mapsto \mathbb{R}$ is a set in $\mathbb{R}^{N+1}$ defined as

$$graph(f) = \{(\boldsymbol{x}, w) | \boldsymbol{x} \in P, w \in \mathbb{R}, f(\boldsymbol{x}) = w\}.$$

The function f is said to be a *polyhedral function* if $graph(f)$ is a polyhedra in $\mathbb{R}^{N+1}$. An *extreme point* of a polyhedral function $f : P \mapsto \mathbb{R}$ is defined as a point in P for which the corresponding point in $graph(f)$ has no line segment passing through it that is fully embedded in $graph(f)$.

Now, considering only vectors $\boldsymbol{a}$ with unity norm, introduce the following change of variables:

$$b_\ell = a_\ell^2, \qquad \ell = 0, \dots, L-2 \tag{27}$$

hereby reducing the dimension of the problem to $L-1$. With a slight abuse of notations, we shall refer to the ML power error exponent and to the lower bound for the power error exponent of the quadratic decoder as $E^*(\boldsymbol{b})$ and $\hat{E}^\Omega(\boldsymbol{b})$, respectively. The domain of interest for the error exponent expressions will be the $L-1$ dimensional simplex defined by

$$S = \left\{\boldsymbol{b} : b_\ell \geq 0, \sum_{\ell=0}^{L-2} b_\ell \leq 1\right\}. \tag{28}$$

It is easily seen that both the ML power error exponent $E^*(\boldsymbol{b})$ and the lower bound $\hat{E}^\Omega(\boldsymbol{b})$ are polyhedral functions over the simplex S, being the minima of affine functions (hyperplanes). With that in mind, we can now state our result.

Theorem 4: For any selection of weights λ_{ij}, the finite set of the extreme points of $E^*(\boldsymbol{b})$ over the simplex S includes at least one critical channel.

Proof: See the Appendix. □

Theorem 4 asserts that for a given codebook, the guaranteed fraction of the ML exponent achieved by any quadratic decoder from the family $\mathcal{F}^*$ can be found by searching over a finite *a priori* determined set of channels, independent of the decoder itself. Denote by $\boldsymbol{b}_0, \dots, \boldsymbol{b}_{N-1}$ the extreme points of the ML exponent. Then the bound $\hat{\xi}^\Omega$ is given by

$$\hat{\xi}^\Omega = \min_n \frac{\hat{E}^\Omega(\boldsymbol{b}_n)}{E^*(\boldsymbol{b}_n)} \tag{29}$$

and we now seek weights λ_{ij} that maximize it. Denote by $\boldsymbol{a}_0, \dots, \boldsymbol{a}_{N-1}$ the channels corresponding to the extreme points, and define the function

$$f_{ij}(\boldsymbol{a}) = \frac{1}{2} \sum_{\ell=0}^{L-1} P_\ell^{(i)} \left(1 - \left|\rho_\ell^{(ij)}\right|\right) a_\ell^2.$$

Now using (23) and (29) we have

$$\begin{aligned}\hat{\xi}^\Omega &= \min_n \frac{\min_{i\neq j} \frac{1}{1+\lambda_{ij}} f_{ij}(\boldsymbol{a}_n)}{E^*(\boldsymbol{a}_n)} \\ &= \min_{i\neq j} \left\{\frac{1}{1+\lambda_{ij}} \min_n \frac{f_{ij}(\boldsymbol{a}_n)}{E^*(\boldsymbol{a}_n)}\right\}\end{aligned}$$

but the *power fractions*, defined as

$$s^{(ij)} \triangleq \min_n \frac{f_{ij}(\boldsymbol{a}_n)}{E^*(\boldsymbol{a}_n)} \tag{30}$$

do not depend on the weighting, and can be calculated *a priori*. So finally the bound can be expressed as

$$\hat{\xi}^\Omega = \min_{i\neq j} \frac{s^{(ij)}}{1+\lambda_{ij}} \tag{31}$$

and we seek the weights λ_{ij} maximizing it, and thus attaining $\hat{\xi}^*$.

Proposition 1: The weights

$$\lambda^*_{ij} = \frac{s^{(ij)}}{s^{(ji)}}$$

maximize the lower bound for the guaranteed fraction of the ML exponent $\hat{\xi}^\Omega$.

Proof: See the Appendix. □

Corollary 2: The lower bound $\hat{\xi}^*$ for ξ^* attained using the weights in Proposition 1 is

$$\xi^* \geq \hat{\xi}^* = \min_{i\neq j} \frac{s^{(ij)} s^{(ji)}}{s^{(ij)} + s^{(ji)}}.$$

Summarizing the results of this subsection, the optimal weights λ_{ij}^* maximizing the lower bound for ξ^* are found through the following procedure.

1. Find the extreme points of the ML power error exponent given in (60) and (61) over the set S, which is either the simplex of (28) or the set defined in the next subsection in (35), derived by utilizing *a priori* known relations between the fading values. For an extensive reference on efficient extreme points enumeration algorithms, see [8].
2. Calculate the power fractions $s^{(ij)}$ using (30) and the set of extreme points found in step 1.
3. Find the optimal weights λ_{ij}^* using Proposition 1.

Notice that if the number of ML extreme points is relatively low, it may be preferable to calculate the weights online while decoding, to avoid storing them in memory.

E. Utilizing Relations Between Fading Values

In some cases, *a priori* information regarding the interdependency of the fading values may be available, and can be incorporated into our QMM decoding scheme. Specifically, assume it is known that the fading vector $\boldsymbol{a}$, rather than taking arbitrary values in $\mathbb{R}^L$, takes only values in a *constraint set* $\mathcal{C} \subseteq \mathbb{R}^L$. In this case, the minimax optimality criterion (12) will become

$$\inf_{\Omega\in\mathcal{F}} \sup_{\boldsymbol{a}\in\mathcal{C}} \frac{P_e(\Omega, \boldsymbol{a})}{(P_e^*(\boldsymbol{a}))^{\xi}}. \tag{32}$$

Under some conditions on the set $\mathcal{C}$, a suboptimal QMM decoder for the modified optimality criterion above can be derived using arguments similar to those presented in the previous subsections. First, in order for our SNR-asymptotic approach to apply, the set $\mathcal{C}$ must be scale invariant, so that any channel in it can be found with all possible gains, i.e.,

$$\boldsymbol{a} \in \mathcal{C} \quad \Leftrightarrow \quad r\boldsymbol{a} \in \mathcal{C}, \qquad \forall\, r \in \mathbb{R}. \tag{33}$$

Now, define the set

$$\mathcal{C}_1 \triangleq \{\boldsymbol{a} | \boldsymbol{a} \in \mathcal{C}, \|\boldsymbol{a}\| = 1\}. \tag{34}$$

We can now go through the same sequel of the previous subsections, with the infimum taken over the set $\mathcal{C}_1$ instead of over all $\|\boldsymbol{a}\| = 1$. The maximal lower bound for the guaranteed fraction of the ML power error exponent for a decoder Ω over the set $\mathcal{C}_1$ is then

$$\xi^*(\mathcal{C}_1) \geq \sup_{\Omega\in\mathcal{F}^*} \inf_{\boldsymbol{a}\in\mathcal{C}_1} \frac{\hat{E}^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \triangleq \hat{\xi}^*(\mathcal{C}_1)$$

and again we encounter the problem of determining the infimum. This time, the channels attaining it will be termed *constrained critical channels*. Using the same change of variables as in (27), define the set

$$S \triangleq \left\{\boldsymbol{b} | \boldsymbol{b} = \left(a_0^2, \ldots, a_{L-2}^2\right), \boldsymbol{a} \in \mathcal{C}_1\right\} \tag{35}$$

and a result similar to the one presented in Theorem 4 is valid, under a further condition on the set S.

Proposition 2: Let $\mathcal{C}$ be a constraint set satisfying (33), for which the corresponding set S defined in (35) is a polyhedra. Then for any selection of weights λ_{ij}, the finite set of the extreme points of $E^*(\boldsymbol{b})$ over S includes at least one constrained critical channel.

Proof: Similar to the proof of Theorem 4, where S being a polyhedra guarantees that E^*, $\hat{E}^{\Omega}$ are polyhedral functions over S. □

It is now straightforward to see that the QMM optimal weights for the constrained case can be found in a manner similar to that of finding the optimal weights in the unconstrained case, where the only difference is that the set S defined in (35) replaces the simplex of (28) as the set over which ML extreme points are sought.

Example: Let $\boldsymbol{a}^0$ be some specific fading vector, and let the constraint set $\mathcal{C}$ be defined as

$$\mathcal{C} = \{\boldsymbol{a} | \boldsymbol{a} = r\boldsymbol{a}^0, r \in \mathbb{R}\}.$$

That is, the channel is known to be described by the fading vector $\boldsymbol{a}^0$ up to an unknown gain factor. In this case, assuming without loss of generality that $\|\boldsymbol{a}^0\| = 1$, we have $\mathcal{C}_1 = \{\boldsymbol{a}^0\}$, and $S = \{\boldsymbol{b}^0\}$, where $\boldsymbol{b}^0 = \left((a_0^0)^2, (a_1^0)^2, \ldots, (a_{L-2}^0)^2\right)$. The set S is a polyhedra (since any set of a finite number of points is a polyhedra), and therefore Proposition 2 applies. Obviously, the only extreme point of the ML exponent over S is the point $\boldsymbol{b}_0$, and using the results of the previous subsection, the optimal weights are given by

$$\lambda_{ij}^* = \frac{\sum P_\ell^{(i)} \left(1 - \left|\rho_\ell^{(ij)}\right|\right) (a_\ell^0)^2}{\sum P_\ell^{(j)} \left(1 - \left|\rho_\ell^{(ij)}\right|\right) (a_\ell^0)^2}.$$

Notice that this example is essentially equivalent to a flat-fading channel where the number of bands is $L = 1$, since the attenuation of each band here is known, up to a common scaling factor. An illustrative flat-fading channel example will be given in the next section.

For the complex OFDM setting, condition (33) for the constraint set $\mathcal{C}$ becomes

$$\boldsymbol{a} \in \mathcal{C} \quad \Leftrightarrow \quad \alpha\boldsymbol{a} \in \mathcal{C}, \qquad \forall\, \alpha \in \mathbb{C} \tag{36}$$

and

$$\begin{aligned} \mathcal{C}_1 &\triangleq \{\boldsymbol{a} | \boldsymbol{a} \in \mathcal{C}, \|\boldsymbol{a}\| = 1\} \\ S &\triangleq \{\boldsymbol{b} | \boldsymbol{b} = (|a_0|^2, \ldots, |a_{L-2}|^2), \boldsymbol{a} \in \mathcal{C}_1\}. \end{aligned} \tag{37}$$

Now assume, for instance, that the channel in use suffers from a multipath distortion with a maximal time spread of T_m, and that the interval between adjacent OFDM bands is δW. If $\delta W \ll \frac{1}{T_m}$, then adjacent frequency bands suffer fading that is not entirely independent, and an appropriate constraint set $\mathcal{C}$ can be found and incorporated into the QMM decoding scheme, as demonstrated in the following example.

Example: Consider a two-path propagation channel, modeled as a discrete time channel with an impulse response

$$h[n] = c_0\delta[n] + c_1\delta[n - m]$$

where m is the unknown delay between the paths, and c_0, c_1 are the unknown path gain coefficients, assumed for simplicity to be real. We further assume that the channel has a bounded delay spread, so that $0 < m < N$ for some N. Therefore, the channel can be converted into the complex OFDM setting with a block length of $L > N$ using a cyclic prefix and L-point IFFT/FFT. The fading coefficients a_ℓ are then given by

$$a_\ell = \sum_{n=0}^{L-1} h[n] \exp\left\{-\frac{j2\pi n\ell}{L}\right\} = c_0 + c_1 \exp\left\{-\frac{j2\pi m\ell}{L}\right\}.$$

We now establish a dependence between different fading coefficients, so hopefully a proper constraint set can be found.

$$\begin{aligned}\frac{|a_\ell|^2}{|a_k|^2} &= \frac{\left|c_0 + c_1 \exp\left\{-\frac{j2\pi m\ell}{L}\right\}\right|^2}{\left|c_0 + c_1 \exp\left\{-\frac{j2\pi mk}{L}\right\}\right|^2}\\ &= \frac{c_0^2 + 2c_0c_1\cos\left\{\frac{j2\pi m\ell}{L}\right\} + c_1^2}{c_0^2 + 2c_0c_1\cos\left\{\frac{j2\pi mk}{L}\right\} + c_1^2}\\ &= \frac{c^2 + 2\alpha_\ell^m c + 1}{c^2 + 2\alpha_k^m c + 1}\end{aligned}$$

where a change of variables $c = \frac{c_0}{c_1}$ has taken place and

$$\alpha_\ell^m \triangleq \cos\left\{\frac{2\pi m\ell}{L}\right\}. \quad (38)$$

To find the permitted range of values for the ratio let

$$\frac{c^2 + 2\alpha_\ell^m c + 1}{c^2 + 2\alpha_k^m c + 1} = d.$$

Rearranging terms results in the quadratic equation

$$(1-d)c^2 + 2\left(\alpha_\ell^m - \alpha_k^m d\right)c + (1-d) = 0.$$

A necessary and sufficient condition for the equation to have a solution is

$$\left(\alpha_\ell^m - \alpha_k^m d\right)^2 - (1-d)^2 \geq 0.$$

Solving for d provides us with the range of values we were looking for

$$\begin{aligned}\frac{1 - \alpha_\ell^m\alpha_k^m - |\alpha_\ell^m - \alpha_k^m|}{1 - (\alpha_k^m)^2} &\leq \frac{|a_\ell|^2}{|a_k|^2}\\ &\leq \frac{1 - \alpha_\ell^m\alpha_k^m + |\alpha_\ell^m - \alpha_k^m|}{1 - (\alpha_k^m)^2}\end{aligned}$$

where α_ℓ^m is given in (38). Define

$$\begin{aligned}\Upsilon_{k\ell}^L &\triangleq \min_{0<m<N} \frac{1 - \alpha_\ell^m\alpha_k^m - |\alpha_\ell^m - \alpha_k^m|}{1 - (\alpha_k^m)^2}\\ \Upsilon_{k\ell}^U &\triangleq \max_{0<m<N} \frac{1 - \alpha_\ell^m\alpha_k^m + |\alpha_\ell^m - \alpha_k^m|}{1 - (\alpha_k^m)^2}.\end{aligned}$$

Now, define the constraint set $\mathcal{C}$ to be the set of all fading vectors $\boldsymbol{a}$ satisfying

$$\Upsilon_{k\ell}^L \leq \frac{|a_\ell|^2}{|a_k|^2} \leq \Upsilon_{k\ell}^U.$$

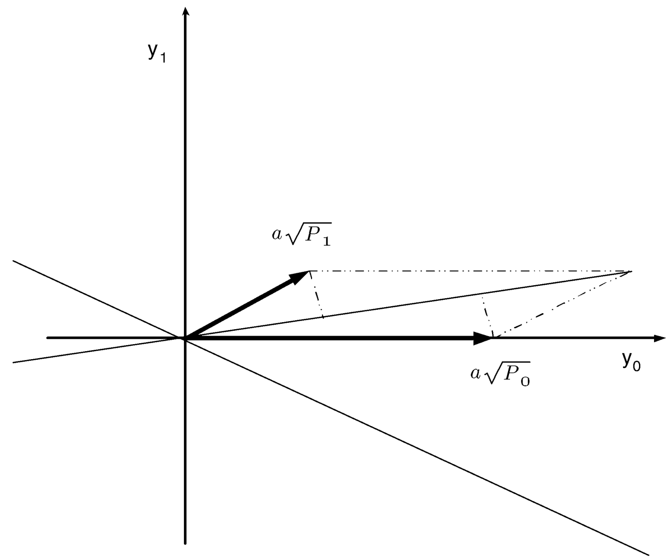

Fig. 1. QMM for simple fading.

This set obviously complies with the scale-invariance condition (36). The corresponding set S is easily verified to be the set of all vectors $\boldsymbol{b} \in \mathbb{R}^{L-1}$ satisfying the inequality conditions

$$\begin{aligned}b_\ell - \Upsilon_{k\ell}^U b_k \leq 0,\ \ \Upsilon_{k\ell}^L b_k - b_\ell \leq 0, &\qquad \ell, k = 0, \ldots, L-1\\ b_\ell \geq 0, &\qquad \ell = 0, \ldots, L-1\end{aligned}$$

where for convenience we define $b_{L-1} = 1 - \sum_{\ell=0}^{L-2} b_\ell$. Since S is obviously a polyhedra, the set $\mathcal{C}$ defines a proper constraint set satisfying the conditions of Proposition 2, and the QMM decoding scheme for the constrained case may be applied accordingly.

V. Some Simple Examples

In some cases, the weights of the QMM decoder can be explicitly determined without any numerical effort, as demonstrated in the following examples.

Example 1: Consider a simple flat-fading channel $(L = 1)$ with two codewords as depicted in the Fig. 1. In this case, any channel is a critical channel, since the simplex is zero-dimensional. Therefore, the optimal weight is just

$$\lambda^* = \frac{(1 - |\rho|)P_0}{(1 - |\rho|)P_1}$$

and we decide in favor of the first codeword if

$$(1 - |\rho|)P_0\alpha^2 \geq (1 - |\rho|)P_1 \frac{(1 - |\rho|)P_0}{(1 - |\rho|)P_1}\beta^2$$

which reduces to

$$\alpha^2 \geq \beta^2$$

where α, β are the coefficients for the representation of the observation as a linear combination of the two codewords. The resulting decision regions' boundaries are depicted by straight lines through the origin in Fig. 1. As can be seen, the minimal distance here is attained by both the codewords simultaneously, uniformly for all fading values. The resulting QMM decoder in this case is similar to the one derived for this example through a somewhat different reasoning in [10].

Notice that for orthogonal codewords ($\rho = 0$), the matrix $\hat{T}$ of Theorem 2 is just the identity matrix, and so the bound (23) is tight and the proposed decoder is an optimal QMM decoder for the family $\mathcal{F}$. Actually, in this simple setting, it can be shown that the proposed decoder is also a QMM decoder for the family of all quadratic decoders [22]. Observe further, that by setting the weight λ to unity, we end up with the GLRT, and since the optimal weight in this case is the codeword's power ratio, the GLRT turns out to coincide with the QMM (and thus to be optimal in the minimax sense) only when the codewords have equal power. This is due to the fact that the GLRT is insensitive to power, and takes into account only direction.

Example 2: Consider a 2×2 OFDM channel with two codewords

$$X^{(0)} = \begin{bmatrix} \sqrt{P_0} & 0 \\ \sqrt{P_2} & 0 \end{bmatrix}, \qquad X^{(1)} = \begin{bmatrix} 0 & \sqrt{P_1} \\ 0 & \sqrt{P_3} \end{bmatrix}$$

orthogonal on each band separately. In this case, the matrix $\hat{T}$ of Theorem 2 is again just the identity matrix, and therefore the lower bound in (23) is tight, and our proposed decoder is an optimal QMM decoder for the family $\mathcal{F}$. The simplex here is the segment $[0, 1]$ of the real line, and since there is only a single pair of codewords in the codebook, the ML extreme points are the two extreme points of this segment corresponding to the channels $\boldsymbol{a}_0 = (1, 0)$, $\boldsymbol{a}_1 = (0, 1)$, and the power fractions are given by

$$\begin{aligned} s^{(01)} &= \min\left\{\frac{f_{01}(\boldsymbol{a}_0)}{E^*(\boldsymbol{a}_0)}, \frac{f_{01}(\boldsymbol{a}_1)}{E^*(\boldsymbol{a}_1)}\right\} \\ &= 4\min\left\{\frac{P_0}{P_0+P_1}, \frac{P_2}{P_2+P_3}\right\} \\ s^{(10)} &= 4\min\left\{\frac{P_1}{P_0+P_1}, \frac{P_3}{P_2+P_3}\right\} \end{aligned}$$

so

$$\lambda^* = \frac{s^{(01)}}{s^{(10)}}, \qquad \xi^* = \frac{s^{(01)}s^{(10)}}{s^{(01)} + s^{(10)}}.$$

Notice again that by setting $\lambda = 1$ we end up with the GLRT, and therefore,

$$\xi^{\text{GLRT}} = \frac{1}{2}\min\{s^{(10)}, s^{(01)}\} \le \xi^*$$

with equality attained only when $\lambda^* = 1$, i.e., only when both codewords have equal power in both bands. It is therefore seen that even if the two codewords have an equal total power, the GLRT is not guaranteed to coincide with the QMM decoder.

Example 3: Consider a repetition code, where in the first band we transmit symbols from an N-PAM constellation $\{\pm 1, \pm 3, \ldots\}$ over the K time points, and the other bands are just exact replicas of the first. We assume that the codewords in the first band are unrestricted as long as no two codewords are colinear. The ML exponent is dominated by the pair of codewords that differ in a single position by a value of two, and therefore,

$$E^*(\boldsymbol{a}) = \frac{1}{8}\sum_{\ell=0}^{L-1} a_\ell^2 \cdot 2^2 = \frac{1}{2}\sum_{\ell=0}^{L-1} a_\ell^2$$

and by the change of variables (27) we get

$$E^*(\boldsymbol{b}) = \frac{1}{2}$$

so the only extreme points of the ML exponent over the simplex S are at the vertices of the simplex, that is, for some $b_\ell = 1$ or $\sum b_\ell = 1$, which corresponds to a single nonzero fading coefficient. Therefore, we have

$$f_{ij}(\boldsymbol{a}_n) = \frac{1}{2}(1 - |\rho_n^{(i,j)}|)P_n^{(i)}, \qquad n = 0, \ldots, L-1$$

and so

$$s^{(ij)} = \min_n \frac{\frac{1}{2}\left(1 - \left|\rho_n^{(ij)}\right|\right) P_n^{(i)}}{\frac{1}{2}} = \left(1 - \left|\rho_0^{(ij)}\right|\right) P_0^{(i)}$$

where the last transition is due to the repetition. Finally

$$\lambda_{ij} = \frac{s^{(ij)}}{s^{(ij)}} = \frac{P_0^{(i)}}{P_0^{(j)}}$$

so the weights for the QMM decoder are just the energy ratio of the codewords in the first band (or in any other band).

Example 4: Consider an uncoded transmission in all the bands using N-PAM constellation, where again the symbols in each band are not allowed to be colinear, to make the codebook suited for universal decoding. The ML exponent in this case is dominated by codewords that differ only in a single band at a single time point, by a value of two. That is,

$$\begin{aligned} E^*(\boldsymbol{a}) &= \frac{1}{8}\min_\ell a_\ell^2 \cdot 2^2 = \frac{1}{2}\min_\ell a_\ell^2 \\ &\Rightarrow E^*(\boldsymbol{b}) = \frac{1}{2}\min_\ell b_\ell \end{aligned}$$

where for convenience we define $b_{L-1} = 1 - \sum b_\ell$ and take the minima for $\ell = 0, \ldots, L-1$. We now claim that the only extreme point of the ML exponent is the flat-fading channel $\boldsymbol{b}_{FF} = \frac{1}{L}(1, 1, \ldots, 1)$. It is obvious that $\boldsymbol{b}_{FF}$ is an extreme point since the ML exponent attains its maximal value $\frac{1}{2L}$ at this point. To prove that it is the only extreme point, assume that some $\boldsymbol{b} \ne \boldsymbol{b}_{FF}$ is an extreme point, and further assume it is not a simplex boundary point. Then for some band ℓ_1

$$E^*(\boldsymbol{b}) = \frac{1}{2}\min_\ell b_\ell = b_{\ell_1}$$

and there exists some $\epsilon > 0$ small enough so that $b_{\ell_2} - b_{\ell_1} > \epsilon$ for some band $\ell_2 \ne \ell_1$. Therefore, there exists a small enough neighborhood of $\boldsymbol{b}$ for which $E^*(\boldsymbol{b}) = b_{\ell_1}$, so this point cannot be an extreme point. To this end, notice that for the simplex boundary points, the ML exponent is zero. Therefore, the only extreme point is $\boldsymbol{b}_{FF}$, which corresponds to $\boldsymbol{a}_{FF} = \frac{1}{\sqrt{L}}(1,1,\ldots,1)$. Now the power fractions are just

$$s^{(ij)} = \frac{f_{ij}(\boldsymbol{a}_{FF})}{E^*(\boldsymbol{a}_{FF})} = \frac{\frac{1}{2}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)\left(\frac{1}{\sqrt{L}}\right)^2}{\frac{1}{2}\left(\frac{1}{\sqrt{L}}\right)^2} = \sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)$$

and the QMM weights are

$$\lambda_{ij}^* = \frac{\sum P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)}{\sum P_\ell^{(j)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)}.$$

VI. Comparison With the GLRT

As we have mentioned, the GLRT is a member of the family $\mathcal{F}$ from which the proposed QMM decoder of Section IV was selected. However, this fact does not guarantee that the latter outperforms the GLRT in the minimax sense, since it is suboptimal and does not necessarily attain the maximal value of ξ over the family $\mathcal{F}$. Nevertheless, we have the following result.

Theorem 5: For the OFDM setting, the QMM decoder of Section IV attains a guaranteed fraction ξ^* of the ML exponent that is equal or higher than the one attained by the GLRT.

Proof: The proof outline is as follows. First, we find an upper bound on the pairwise power error exponent of the GLRT. Then, we use it to upper-bound ξ^{GLRT}, the guaranteed fraction of the ML exponent attained by the GLRT. Finally, we show that this upper bound is smaller than or equal to the lower bound for ξ^* attained by the QMM decoder.

Using (9), and defining the per-band *normalized codewords* by

$$\hat{\boldsymbol{x}}_\ell^{(i)} \triangleq \frac{1}{\sqrt{P_\ell^{(i)}}}\boldsymbol{x}_\ell^{(i)}$$

the pairwise separation surface equation of the GLRT can be written as

$$\sum_\ell \left|\left\langle \hat{\boldsymbol{x}}_\ell^{(i)}, \boldsymbol{y}_\ell\right\rangle\right|^2 = \sum_\ell \left|\left\langle \hat{\boldsymbol{x}}_\ell^{(j)}, \boldsymbol{y}_\ell\right\rangle\right|^2. \tag{39}$$

A more restricting constraint is the per-band constraint

$$\left|\left\langle \hat{\boldsymbol{x}}_\ell^{(i)}, \boldsymbol{y}_\ell\right\rangle\right|^2 = \left|\left\langle \hat{\boldsymbol{x}}_\ell^{(j)}, \boldsymbol{y}_\ell\right\rangle\right|^2, \qquad \ell = 0,\ldots,L-1. \tag{40}$$

Therefore, the minimal distance of the codewords from the set given in (40) is an upper bound to the true minimal distance. Since the constraints given in (40) for each band are independent, we can find the minimal distance in each band separately. That is, for each band we seek a point $\boldsymbol{y}_\ell$ of minimal distance to each of the codewords, under the corresponding constraint. Instead of solving a constrained optimization problem, we take a geometric approach and observe that the constraint (40) is equivalent to the equal distance constraint

$$\left\|\boldsymbol{y}_\ell - \hat{\boldsymbol{x}}_\ell^{(i)}\right\|^2 = \left\|\boldsymbol{y}_\ell \pm \hat{\boldsymbol{x}}_\ell^{(j)}\right\|^2, \qquad \ell = 0,\ldots,L-1. \tag{41}$$

A point satisfying (41) is a point whose distance to the first normalized codeword is equal to its distance to the second normalized codeword or to its opposite. Therefore, in each band such a point has to be either on the line bisecting the angle between the two normalized codewords in that band (in the subspace spanned by the two normalized codewords), or on an orthogonal line (in the same subspace). Accordingly, the point $\boldsymbol{y}_\ell$ must satisfy either of the conditions

$$\boldsymbol{y}_\ell = \alpha\left(\hat{\boldsymbol{x}}_\ell^{(i)} \pm \hat{\boldsymbol{x}}_\ell^{(j)}\right) \tag{42}$$

for some α. For each of the possibilities, we seek a constant α so that the distance to the codeword is minimal. The distance from codeword i to points satisfying (42) with a plus sign is

$$\begin{aligned}
&\left\|\alpha\left(\hat{\boldsymbol{x}}_\ell^{(i)} + \hat{\boldsymbol{x}}_\ell^{(j)}\right) - a_\ell \boldsymbol{x}_\ell^{(i)}\right\|^2\\
&= \left\|\left(\alpha - a_\ell\sqrt{P_\ell^{(i)}}\right)\hat{\boldsymbol{x}}_\ell^{(i)} + \alpha\hat{\boldsymbol{x}}_\ell^{(j)}\right\|^2\\
&= \left(\alpha - a_\ell\sqrt{P_\ell^{(i)}}\right)^2 + \alpha^2 + 2\alpha\left(\alpha - a_\ell\sqrt{P_\ell^{(i)}}\right)\rho_\ell^{(ij)}\\
&= 2\alpha^2\left(1+\rho_\ell^{(ij)}\right) - 2\alpha a_\ell\sqrt{P_\ell^{(i)}}\left(1+\rho_\ell^{(ij)}\right) + a_\ell^2 P_\ell^{(i)}.
\end{aligned} \tag{43}$$

The minimum is attained for $\alpha = \frac{1}{2}\sqrt{P_\ell^{(i)}}a_\ell$ and the corresponding distance is

$$\left|d_\ell^{(i)+}\right|^2 = \frac{1}{2}P_\ell^{(i)}\left(1-\rho_\ell^{(ij)}\right)a_\ell^2.$$

Similarly, the minimal distance for a point satisfying (42) with a minus sign is

$$\left|d_\ell^{(i)-}\right|^2 = \frac{1}{2}P_\ell^{(i)}\left(1+\rho_\ell^{(ij)}\right)a_\ell^2$$

and therefore the minimal distance for codeword i in this band is

$$\left|d_\ell^{(i)}\right|^2 = \frac{1}{2}P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2.$$

Combining the per-band minimal distances, we end up with an upper bound for the true pairwise minimal distance from the separation surface

$$\left|d_{ij}^{\mathrm{GLRT}}(\boldsymbol{a})\right|^2 \le \frac{1}{2}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2.$$

So finally the power error exponent of the GLRT can be upper-bounded by

$$E^{\mathrm{GLRT}}(\boldsymbol{a}) \le \frac{1}{4}\min_{i\ne j}\left\{\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2\right\} \triangleq \tilde{E}(\boldsymbol{a}). \tag{44}$$

The preceding bound can be shown to be tight, by using the method of Section IV and providing a lower bound that coincides with it. However, we shall not dwell on this point, as it does not contribute to our arguments.

Using (44), we can upper-bound ξ^{GLRT}

$$\xi^{\mathrm{GLRT}} = \inf_{\|\boldsymbol{a}\|=1}\frac{E^{\mathrm{GLRT}}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \le \inf_{\|\boldsymbol{a}\|=1}\frac{\tilde{E}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \triangleq \tilde{\xi}. \tag{45}$$

Now, we remind the reader of the lower bound for the power error exponent found for the family $\mathcal{F}^*$ of quadratic decoders from which the suboptimal QMM decoder was selected, given in (23)

$$E^{\Omega}(\boldsymbol{a}) \ge \frac{1}{2}\min_{i\ne j}\left\{\frac{1}{1+\lambda_{ij}}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2\right\} = \hat{E}^{\Omega}(\boldsymbol{a}).$$

Fortunately, this bound has a structure very similar to the bound in (44), and if one sets $\lambda_{ij}=1$ for all of the pairs i, j, the bounds coincide for all fading values. Precisely, denote by $\tilde{\Omega}\in\mathcal{F}^*$ the quadratic decoder obtained by using weights all equal to one, then we have

$$\hat{E}^{\tilde{\Omega}}(\boldsymbol{a}) \equiv \tilde{E}(\boldsymbol{a}).$$

However, the suboptimal QMM decoder uses the weights λ_{ij}^* given in Proposition 1, which are guaranteed to maximize the the lower bound for ξ^*. The guaranteed fraction attained by the suboptimal QMM can be lower-bounded as follows:

$$\begin{aligned}\xi^* &\ge \sup_{\Omega\in\mathcal{F}^*}\inf_{\|\boldsymbol{a}\|=1}\frac{\hat{E}^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \ge \inf_{\|\boldsymbol{a}\|=1}\frac{\hat{E}^{\tilde{\Omega}}(\boldsymbol{a})}{E^*(\boldsymbol{a})}\\ &= \inf_{\|\boldsymbol{a}\|=1}\frac{\tilde{E}(\boldsymbol{a})}{E^*(\boldsymbol{a})} = \tilde{\xi}.\end{aligned} \tag{46}$$

Finally, combining (45) and (46) gives the desired result

$$\xi^{\mathrm{GLRT}} \le \tilde{\xi} \le \xi^*. \qquad \square$$

VII. Simulation Results

In this section, we evaluate the performance of the proposed QMM decoder for various selections of codebooks, in terms of the fraction of the ML exponent attained over randomly selected channels. The QMM performance will be compared to that of the GLRT and to the training-ML decoder.

A. The Coding Scheme

In order to allow universal decoding, the codebook used has to be structured so that all codewords are distinguishable under different fading values, meaning that there are no two different fading values, which makes two different codewords coincide. We have further required that no pair of codewords is colinear in any of the OFDM bands, which is justified by noticing that colinearity in a band makes it useless in terms of deciding between the two codewords. This constraint on the codebook makes it difficult to use ordered constellations and algebraic coding, since that may result in many colinearities, and therefore many of the codewords would have to be eliminated in order to make the code robust for fading. To overcome this problem, we have employed a coding scheme for which almost no colinearity occurs. This scheme is based on the *complex field coding* (CFC) scheme suggested for Rayleigh-fading OFDM channels [24]. CFC is basically linear precoding with redundancy, and we have used it in a slightly different manner which is now described.

An encoder for a $K\times L$ OFDM channel is defined by a set of K matrices $\{T_k\}_{k=0}^{K-1}$, each of dimensions $L\times N$, where $N\le L$. The input to the encoder at time point k is an N-dimensional column vector $\boldsymbol{s}_k$, whose components are normally taken from some constellation. This vector is multiplied by the matrix $T_{k(\mathrm{mod}\ K)}$, and the resulting L-dimensional vector is transmitted over the L bands at time point k. A codeword is therefore an $L\times K$ matrix of K consecutive transmissions, and the total number of codewords is determined by the number of different combinations of K consecutive input vectors. Effectively, one cannot allow all possible input combinations since that might result in colinearity, so some of the possible codewords cannot be used and have to be eliminated. For instance, using $N=2$, $K=3$, and a 4-PAM constellation, there are 4^6 possible codewords, but at least half of those are eliminated, because they have antipodal counterparts.

Since the encoder is entirely defined by the matrices T_ℓ, selecting them requires some consideration. Usually, we require each matrix to have orthonormal columns, so that the matrix multiplication retains power. Moreover, we choose the matrices so that the resulting number of ML extreme points is low, so that the weight's calculation is faster. In the following sections, different CFC codes were used for performance evaluation. It is stressed thought, that any other coding scheme may be used instead, provided that the codebook does not violate the colinearity constraint. Moreover, if colinearity does happen to occur for some pair of codewords in some band, the observation in that band may be disregarded when making the pairwise decision, and so QMM decoding is practically applicable for any selection of a codebook.

B. QMM Versus GLRT

A CFC code over a 3×4 real OFDM channel was used to evaluate the performance of the proposed QMM decoder and of the GLRT. We used a CFC code with $N=2$ and a 4-PAM constellations at the input which gives 256 possible codewords, but due to the colinearity constraint only 120 were used. Different channel realizations were used by randomly generating the fading values, and the symbol error rate was estimated by simulation for each realization for the QMM, GLRT, and ML decoder (the latter tuned to the channel). The corresponding estimations of the power error exponents were derived, and the fractions of the ML exponent attained by each decoder were found. In estimating the power error exponents, the receiver SNR for

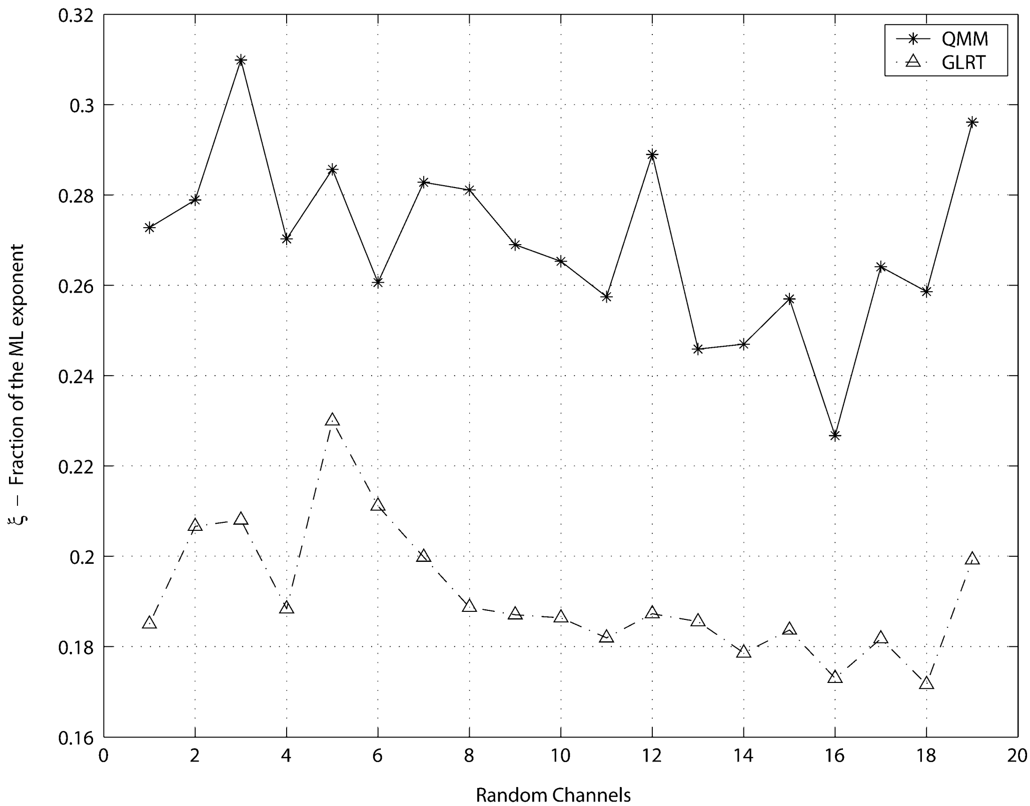

Fig. 2. QMM versus GLRT, 120 codewords CFC, 3×4 channel.

the realized channels was increased until the error probability's exponential behavior became practically constant. Fig. 2 depicts the results of the described simulations. The horizontal axis represents random channel realizations, and the vertical axis represents the attained fraction for the QMM and the GLRT. As can be seen, the QMM's worst case fraction over this random set of channels is approximately 0.23, while the GLRT's is around 0.17. Surprisingly, the QMM seems to outperform the GLRT uniformly, and not only in the worst case performance, which was not guaranteed, and is not true in general. Many simulations employing randomly selected CFC codes have exhibited the same behavior where the QMM uniformly outperforms the GLRT over the randomly selected channels.

C. QMM Versus Training Sequence Approach

A 3×3 CFC code was used in conjunction with training symbols in order to evaluate the performance of the proposed QMM decoder, the GLRT, and the training-ML decoder. Both 3×4 and 3×5 real OFDM channels were used, where the CFC codewords occupy only the last three time points, and the first time points contain constant training symbols. When employing the training-ML decoder, an ML estimate of the fading is found using these symbols, and the estimate is then used as the true fading for a standard ML decoding. The QMM and the GLRT consider the training symbols as an integral part of the codeword, and the corresponding decoding schemes apply. Notice that adding training symbols makes it inherently impossible for colinearity to exist in any of the bands, no matter the original codebook, and therefore makes the modified codebook more robust for fading and better suited for universal decoding.

The simulation results using random channel realizations for both settings are depicted in Figs. 3 and 4. As can be observed, the suboptimal QMM decoder outperforms both the GLRT and the training-ML decoder in the worst case performance. In the 3×4 channel, the superiority of the QMM seemed to be uniform (which was not guaranteed).

VIII. Summary and Future Research

In this paper, we have considered the problem of universal decoding for an unknown frequency-selective fading channel, using an OFDM signaling scheme, and a block-fading model. We have presented the minimax criterion which, for a given codebook, seeks a decoder guaranteeing the highest fraction of the ML power error exponent uniformly over all fading values, out of a given family of decoders. Specifically, we were interested in families of quadratic decoders, defined as decoders for which the pairwise decision rule can be represented using a quadratic form, and the optimal decoder selected from a family of quadratic decoders was termed the QMM decoder for thatfamily. A specific family $\mathcal{F}$ of well-structured quadratic decoders was selected and most of our efforts were dedicated to exploring the minimax decoding problem for that family.

The problem of explicitly determining the QMM decoder for the family $\mathcal{F}$ turned out to be difficult, as the power error exponent has no analytical solution in general. Nevertheless, a lower bound on the power error exponent has been derived, and it was further shown that considering a family $\mathcal{F}^* \subset \mathcal{F}$ of decoders

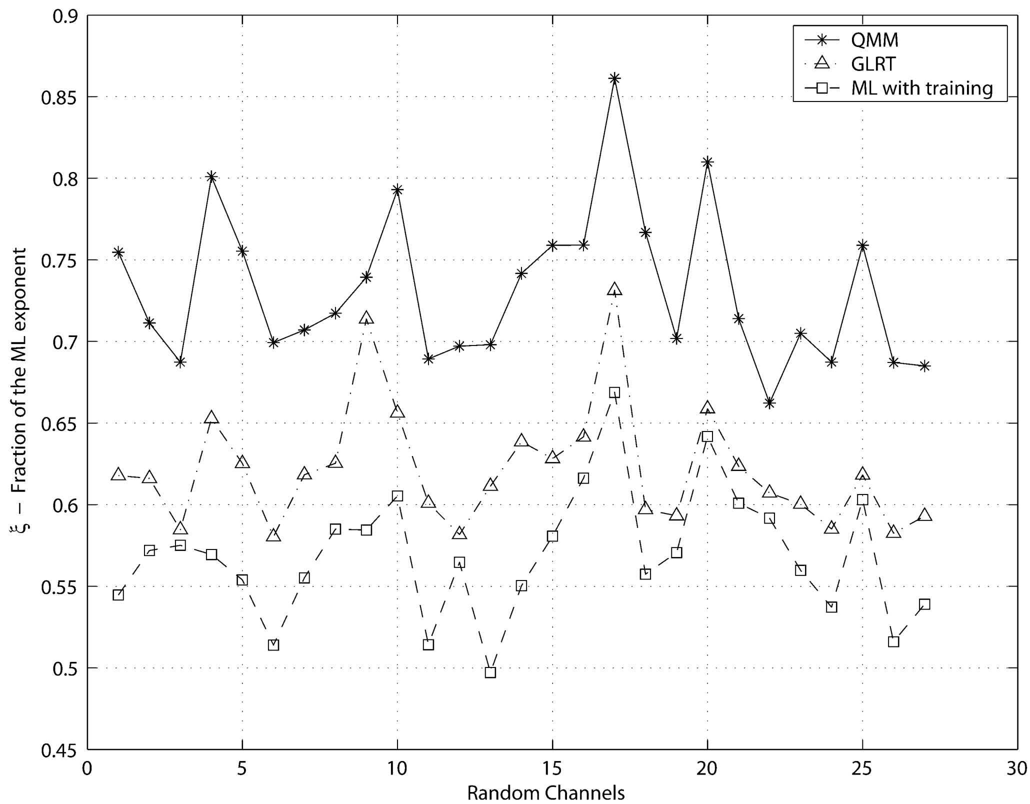

Fig. 3. QMM versus GLRT and ML-training, 28 codewords CFC, 3×4 channel, one training symbol per block.

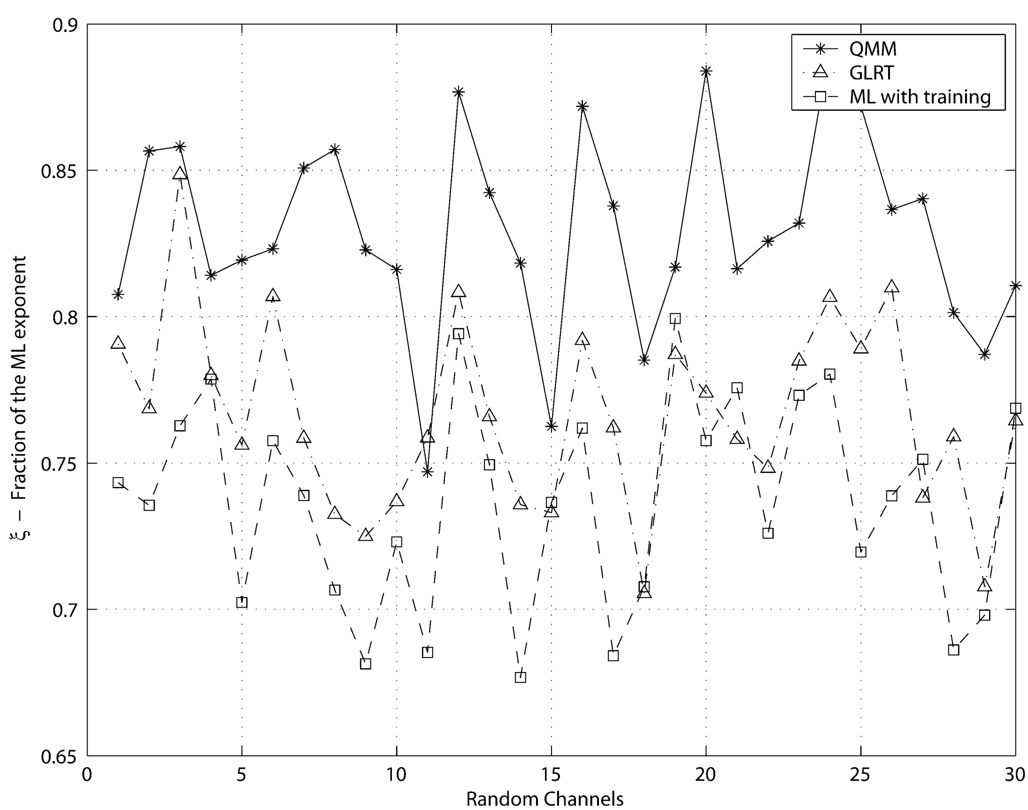

Fig. 4. QMM versus GLRT and ML-training, 28 codewords CFC, 3×5 channel, two training symbols per block.

whose pairwise components each depends on a single weight parameter, is possible without any loss of generality. Using the lower bound, a suboptimal QMM decoder was explicitly derived, as the decoder maximizing the corresponding lower bound for ξ (the guaranteed fraction of the ML exponent) over the family. A procedure for finding the optimal weights was described, as well as the decoding scheme that follows, and the corresponding (maximal) lower bound was given explicitly. A minor modification of the weights determining procedure was also presented, covering the interesting case where the fading coefficients are known not to be entirely independent but rather to satisfy some general linear inequality constraints.

The complexity of the suggested QMM decoder is comparable to that of the GLRT, with the exceptions of the offline effort for computing the weights, and of memory issues resulting from storing them. Online computation of the weights was suggested as means of reducing the memory consumption when the number of extreme points of the ML exponent is relatively low.

Despite the suboptimality of the proposed QMM decoder, it was shown to attain a higher fraction ξ than that attained by the GLRT, thus outperforming it in the minimax sense. Simulations performed over randomly selected channels and different codebooks verified the superiority of the QMM decoder over the GLRT and the commonly used training sequence approach.

Although the QMM decoder has a complexity comparable to that of the GLRT, it is still unsatisfactory for practical applications, as it requires going over all the codewords. Therefore, a possible direction for future research would be exploring efficient QMM decoding methods, possibly suboptimal methods. One idea may be using convolutional codes and running a QMM-based Viterbi algorithm, where the metrics are modified appropriately.

Universal codes may be explored by means of finding conditions on the codebook or seeking specific codes so that the QMM decoder has good performance relative to the ML decoder. Notice thought, that the QMM attaining a high value of ξ for some codebook does not necessarily guarantee the absolute performance of the code to be good at all.

The QMM decoder for the constrained fading case takes into account only a subset of the set of possible channels, and therefore is guaranteed to have a higher value of ξ than that attained by the QMM decoder for the unconstrained fading case, over those channels. The behavior of ξ as a function of the constraint set can be investigated. It would be interesting to see whether there exists a "large" constraint set for which the value of ξ is significantly improved, or even approaches one, which will make the corresponding QMM decoder optimal for "almost" all channels.

In most of our simulations, the QMM decoder seemed to uniformly outperform the GLRT. This property is not true in general, and simple counterexamples can be constructed in which the GLRT is better than the QMM for some channels and worse for others (although the QMM still wins in the worst case). Nevertheless, it may be interesting to seek general conditions on the codebook for this property to hold.

The QMM decoder was derived from a minimax optimality criterion. Still, it may be insightful to investigate its performance under commonly used statistical fading models, such as the Rayleigh or the Rician models, and compare its performance to decoders derived under these statistical assumptions, in terms of diversity gain for instance.

Finally, notice that by considering general matrices instead of diagonal ones, our channel model becomes a model for an unknown multiple-input multiple-output (MIMO) channel. An interesting direction for future research would be extending the results of this work and find a (possibly suboptimal) QMM decoder for the MIMO setting.

APPENDIX

Proof of Theorem 1

Let $\hat{H}$ be any symmetric $KL \times KL$ matrix describing a pairwise decoder $\hat{\Omega}^{ij}$ with a separation surface $\hat{S} = \{\boldsymbol{y} | \boldsymbol{y}^T \hat{H} \boldsymbol{y} = 0\}$ and let $\mathcal{W}$ be the subspace spanned by codewords i and j in all of the bands, i.e., the minimal subspace containing each of the two codewords undergoing all possible fading scenarios. Now, let $\boldsymbol{y}_{\mathcal{W}}$ be the projection of a point $\boldsymbol{y}$ onto the subspace $\mathcal{W}$, and let $\boldsymbol{y}_{\mathcal{W}^\perp}$ be its projection onto the orthogonal complementary subspace $\mathcal{W}^\perp$. Then for any value of the fading vector $\boldsymbol{a}$ with a corresponding matrix representation $\bar{A}$, the distance of codeword i from the separation surface $\hat{S}$ can be bounded as follows:

$$\begin{aligned}\min_{\boldsymbol{y}\in\hat{S}} \|\boldsymbol{y} - \bar{A}\boldsymbol{x}^{(i)}\|^2 &= \min_{\boldsymbol{y}\in\hat{S}} \left(\|\boldsymbol{y}_{\mathcal{W}^\perp}\|^2 + \|\boldsymbol{y}_{\mathcal{W}} - \bar{A}\boldsymbol{x}^{(i)}\|^2\right) \\ &\le \min_{\boldsymbol{y}\in\hat{S}\cap\mathcal{W}} \left(\|\boldsymbol{y}_{\mathcal{W}^\perp}\|^2 + \|\boldsymbol{y}_{\mathcal{W}} - \bar{A}\boldsymbol{x}^{(i)}\|^2\right) \\ &= \min_{\boldsymbol{y}\in\hat{S}\cap\mathcal{W}} \|\boldsymbol{y} - \bar{A}\boldsymbol{x}^{(i)}\|^2. \end{aligned} \tag{47}$$

A similar derivation is true for codeword j, and we thus see that the minimal distance of the decoder is upper-bounded by the minimal distance from the separation surface within the subspace $\mathcal{W}$. To this end, we now show that there exists another decoder of the required form, whose minimal distance equals this bound for all fading values.

Define the $KL \times KL$ diagonal matrix

$$V = \mathrm{diag}\{V_0, V_1, \ldots, V_{L-1}\}$$
$$V_\ell = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & 0 \end{bmatrix}_{K\times K}.$$

It is easy to verify that XVX^{-1} is a projection matrix and

$$\boldsymbol{y}_{\mathcal{W}} = XVX^{-1}\boldsymbol{y}, \qquad \mathcal{W} = \{\boldsymbol{w} | \boldsymbol{w} = \boldsymbol{y}_{\mathcal{W}}, \boldsymbol{y} \in \mathbb{R}^{KL}\}.$$

Now, express the matrix $\hat{H}$ as

$$\hat{H} = X^{-T}\hat{Q}X^{-1}$$

and define a new matrix

$$H = X^{-T}QX^{-1}, \qquad Q = V\hat{Q}V \tag{48}$$

which is exactly in the required form. The matrix H describes a decision rule of a pairwise decoder Ω^{ij} with a separation surface $S = \{\boldsymbol{y} | \boldsymbol{y}^T H \boldsymbol{y} = 0\}$ satisfying

$$\begin{aligned} S &= \{\boldsymbol{y} | \boldsymbol{y}^T X^{-T} V\hat{Q}VX^{-1}\boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} | \boldsymbol{y}_{\mathcal{W}}^T X^{-T}\hat{Q}X^{-1}\boldsymbol{y}_{\mathcal{W}} = 0\} = \{\boldsymbol{y} | \boldsymbol{y}_{\mathcal{W}} \in \hat{S}\} \\ &= (\hat{S}\cap\mathcal{W}) \oplus \mathcal{W}^\perp \end{aligned}$$

and thus for any value of the fading vector $\boldsymbol{a}$ and a corresponding matrix representation $\bar{A}$, the distance of codeword i from the separation surface S is

$$\min_{\boldsymbol{y}\in S}\|\boldsymbol{y}-\bar{A}\boldsymbol{x}^{(i)}\|^2 = \min_{\boldsymbol{u}\in\hat{S}\cap W}\min_{\boldsymbol{v}\in W^\perp}(\|\boldsymbol{v}\|^2+\|\boldsymbol{u}-\bar{A}\boldsymbol{x}^{(i)}\|^2) = \min_{\boldsymbol{u}\in\hat{S}\cap W}\|\boldsymbol{u}-\bar{A}\boldsymbol{x}^{(i)}\|^2 \quad (49)$$

and a similar derivation is true for codeword j. It is now straightforward from (47) and (49) that indeed the new decoder Ω^{ij} represented by the matrix H in (48), achieves the upper bound for the decoder $\hat{\Omega}^{ij}$ represented by the matrix $\hat{H}$, for all fading values. □

Proof of Theorem 2

In order to determine the condition number, we need to find the minimal and maximal singular values of T^{-1}. Note that scaling of T^{-1} has no effect on the condition number, so we will assume with no loss of generality that the matrix lower bound of T^{-1} is equal to one, and try to minimize the matrix norm under this constraint. Since the matrix X is block diagonal and Γ is diagonal, the matrices T, T^{-1} are also block diagonal, with the same structure as X. Denote by T_ℓ the $K\times K$ matrix corresponding to X_ℓ, extracted from the diagonal of T. Similarly, denote by Γ_ℓ the corresponding matrix extracted from the diagonal of Γ. We then have

$$T^{-1}=\operatorname{diag}\left\{T_0^{-1},T_1^{-1},\ldots,T_{L-1}^{-1}\right\}$$

and the singular values of T^{-1} are just the union of the singular values T_ℓ^{-1}, for $\ell=0,\ldots,L-1$. T_ℓ^{-1} is given by the block-wise relation

$$T_\ell^{-1}=\left(\Gamma_\ell X_\ell^{-1}\right)^{-1}=X_\ell\Gamma_\ell^{-1}$$

and its singular values are the square roots of the eigenvalues of $T_\ell^{-T}T_\ell^{-1}$. Specifically, see the equation at the bottom of the page. The eigenvalues of this matrix are merely those of the 2×2 matrix on the top left, and the rest of the $K-2$ elements on the diagonal. Note that the matrices $T_\ell^{-T}T_\ell^{-1}$ are independent, and therefore the eigenvalues optimization can be performed separately for each matrix.

Since we have assumed that the matrix lower bound of T^{-1} is equal to one, it is only natural to set all lower $K-2$ values on the diagonal to one, since we cannot set any of them to be lower than one, and there is no reason to set them to any value higher if one is to minimize the matrix norm. All that is left now is to minimize the maximal eigenvalue of the 2×2 matrix in the upper left corner, under the constraint that the minimal eigenvalue is equal to one.

The two eigenvalues of the 2×2 matrix are the roots of the second-degree polynomial

$$\begin{aligned}&\left(s-\gamma_{0,\ell}^{-2}P_\ell^{(i)}\right)\left(s-\gamma_{1,\ell}^{-2}P_\ell^{(j)}\right)-\gamma_{0,\ell}^{-2}\gamma_{1,\ell}^{-2}P_\ell^{(i)}P_\ell^{(j)}\left(\rho_\ell^{(ij)}\right)^2\\&\quad=s^2-\left(\gamma_{0,\ell}^{-2}P_\ell^{(i)}+\gamma_{1,\ell}^{-2}P_\ell^{(j)}\right)s\\&\qquad+\left(\gamma_{0,\ell}^{-2}\gamma_{1,\ell}^{-2}P_\ell^{(i)}P_\ell^{(j)}\right)\left(1-\left(\rho_\ell^{(ij)}\right)^2\right).\end{aligned}$$

Defining $b=\gamma_{0,\ell}^{-2}P_\ell^{(i)}$, $c=\gamma_{1,\ell}^{-2}P_\ell^{(j)}$, and $\rho=\rho_\ell^{(ij)}$, the eigenvalues λ_0, λ_1 are the roots of

$$s^2-(b+c)s+bc(1-\rho^2). \quad (50)$$

The minimal eigenvalue is then forced to be equal to one

$$\lambda_0=\frac{b+c-\sqrt{(b+c)^2-4bc(1-\rho^2)}}{2}=1; \quad (51)$$

under this constraint, the maximal eigenvalue is

$$\lambda_1=\frac{b+c+\sqrt{(b+c)^2-4bc(1-\rho^2)}}{2}=b+c-1.$$

Rearranging the terms in (51) we end up with a more compact constraint

$$bc(1-\rho^2)-b-c+1=0. \quad (52)$$

We now find b, c to minimize the λ_1, under the constraint (52), and under additional inequality constrains that guarantee that b, c are nonnegative and that the resulting λ_1 is indeed the larger eigenvalue

$$\min_{b,c}(b+c-1)\quad\text{s.t.}\quad\begin{cases}bc(1-\rho^2)-b-c+1=0\\ b\geq 0\\ c\geq 0\\ b+c-1\geq 1.\end{cases} \quad (53)$$

Noticing that none of the inequality conditions are active, it is easily verified the the solution to (53) is given by

$$b=c=\frac{1}{1-|\rho|}$$

$$T_\ell^{-T}T_\ell^{-1}=\Gamma_\ell^{-1}X_\ell^TX_\ell\Gamma_\ell^{-1}=\Gamma_\ell^{-1}\begin{bmatrix}P_\ell^{(i)} & \sqrt{P_\ell^{(i)}P_\ell^{(j)}}\rho_\ell^{(ij)} & 0 & \cdots & 0\\ \sqrt{P_\ell^{(i)}P_\ell^{(j)}}\rho_\ell^{(ij)} & P_\ell^{(j)} & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & & 1\end{bmatrix}\Gamma_\ell^{-1}$$

$$=\begin{bmatrix}\gamma_{0,\ell}^{-2}P_\ell^{(i)} & \gamma_{0,\ell}^{-1}\gamma_{1,\ell}^{-1}\sqrt{P_\ell^{(i)}P_\ell^{(j)}}\rho_\ell^{(ij)} & 0 & \cdots & 0\\ \gamma_{0,\ell}^{-1}\gamma_{1,\ell}^{-1}\sqrt{P_\ell^{(i)}P_\ell^{(j)}}\rho_\ell^{(ij)} & \gamma_{1,\ell}^{-2}P_\ell^{(j)} & 0 & \cdots & 0\\ 0 & 0 & \gamma_{2\ell}^{-2} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & & \gamma_{K-1,\ell}^{-2}\end{bmatrix}.$$

and by substituting

$$\gamma_{0,\ell}^2 = \frac{P_\ell^{(i)}}{b} = P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)$$
$$\gamma_{1,\ell}^2 = \frac{P_\ell^{(i)}}{c} = P_\ell^{(j)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)$$

we end up with the desired result. □

Proof of Theorem 3

Let Δ be some diagonal matrix with diagonal elements $\delta_{k,l}$, used to represent the pairwise decoder Ω^{ij}. Using Theorem 1, we can assume without any loss of generality, that $\delta_{k,l} = 0$ for any $k > 1$. We also require the resulting pairwise decoder $H = X^{-T}\Gamma^T\Delta\Gamma X^{-1}$ to satisfy the detectability (4) condition. Specifically, a necessary and sufficient condition for detectability of codeword i is

$$\tilde{\boldsymbol{a}}^T\Gamma^T\Delta\Gamma\tilde{\boldsymbol{a}} > 0, \qquad \forall\, \|\boldsymbol{a}\| \neq 0.$$

Writing the condition explicitly gives

$$\sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\delta_{0,\ell}a_\ell^2 > 0, \qquad \forall\, \|\boldsymbol{a}\| \neq 0. \tag{54}$$

A similar condition for detectability of codeword j is

$$\sum_{\ell=0}^{L-1}\gamma_{1,\ell}^2\delta_{1,\ell}a_\ell^2 < 0, \qquad \forall\, \|\boldsymbol{a}\| \neq 0. \tag{55}$$

Finally, the necessary and sufficient conditions for the detectability property of Ω^{ij} are given by

$$\delta_{0,\ell} > 0, \delta_{1,\ell} < 0, \qquad \ell = 0,\dots,L-1. \tag{56}$$

Returning to the optimization problem (22) and taking the derivative of the Lagrangian, we get

$$(I+\mu\Delta)\boldsymbol{u} = \tilde{\boldsymbol{a}}$$

where μ is the Lagrange multiplier. We immediately get $u_{k,l} = 0$ for $k > 1$, and we further have the following $2L$ equations:

$$(1+\mu\delta_{k,\ell})u_{k,\ell} = \begin{cases} a_\ell, & k=0 \\ 0, & k=1. \end{cases}$$

Due to the constraint, the coefficients in the equations with a zero right-hand side cannot all vanish together. Therefore, for some ℓ^* we have

$$\mu = -\frac{1}{\delta_{1,\ell^*}}$$

and by substituting that into the equations for $k = 0$ we get

$$u_{0,\ell} = \frac{a_\ell}{1-\frac{\delta_{0,\ell}}{\delta_{1,\ell^*}}} = \frac{\delta_{1,\ell^*}}{\delta_{1,\ell^*}-\delta_{0,\ell}}a_\ell.$$

We now lack the values of $u_{1,\ell}$. Notice that if $\delta_{1,\ell} \neq \delta_{1,\ell^*}$, then the corresponding $u_{1,\ell}$ must be zero. So until now we have

$$u_{k,\ell} = \begin{cases} \frac{\delta_{1,\ell^*}}{\delta_{1,\ell^*}-\delta_{0,\ell}}a_\ell, & k=0 \\ 0, & k=1,\ \delta_{k,\ell}\neq\delta_{1,\ell^*} \\ 0, & k>1 \\ ? & \text{o.w.} \end{cases}$$

denote by J the set of indices ℓ for which $\delta_{1,\ell} = \delta_{1,\ell^*}$. Using the constraint we get

$$\begin{aligned}\boldsymbol{u}^T\Gamma^T\Delta\Gamma\boldsymbol{u} &= \sum_{k=0}^{K-1}\sum_{\ell=0}^{L-1}\gamma_{k,\ell}^2\delta_{k,\ell}u_{k,\ell}^2 \\ &= \sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\delta_{0,\ell}u_{0,\ell}^2 + \sum_{\ell\in J}\gamma_{1,\ell}^2\delta_{1,\ell^*}u_{1,\ell}^2 = 0 \\ &\Downarrow \\ \delta_{1,\ell^*}\sum_{\ell\in J}\gamma_{1,\ell}^2u_{1,\ell}^2 &= -\sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\delta_{0,\ell}\left(\frac{\delta_{1,\ell^*}}{\delta_{1,\ell^*}-\delta_{0,\ell}}a_\ell\right)^2 \\ &= -\sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\frac{\delta_{0,\ell}\delta_{1,\ell^*}^2}{(\delta_{1,\ell^*}-\delta_{0,\ell})^2}a_\ell^2\end{aligned}$$

so finally

$$\sum_{\ell\in J}\gamma_{1,\ell}^2u_{1,\ell}^2 = -\delta_{1,\ell^*}\sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\frac{\delta_{0,\ell}}{(\delta_{1,\ell^*}-\delta_{0,\ell})^2}a_\ell^2. \tag{57}$$

Returning to the distance expression, and using (57) we get

$$\begin{aligned}\|\Gamma(\boldsymbol{u}-\tilde{\boldsymbol{a}})\|^2 &= \sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\left(\frac{\delta_{1,\ell^*}}{\delta_{1,\ell^*}-\delta_{0,\ell}}-1\right)^2a_\ell^2 \\ &\quad+\sum_{\ell\in J}\gamma_{1,\ell}^2u_{1,\ell}^2 \\ &= \sum_{\ell=0}^{L-1}\frac{\gamma_{0,\ell}^2\delta_{0,\ell}^2}{(\delta_{1,\ell^*}-\delta_{k,\ell})^2}a_\ell^2 - \delta_{1,\ell^*} \\ &\quad\times\sum_{\ell=0}^{L-1}\gamma_{0,\ell}^2\frac{\delta_{0,\ell}}{(\delta_{1,\ell^*}-\delta_{0,\ell})^2}a_\ell^2 \\ &= \sum_{\ell=0}^{L-1}\frac{\delta_{0,\ell}\gamma_{0,\ell}^2}{\delta_{0,\ell}-\delta_{1,\ell^*}}a_\ell^2.\end{aligned}$$

The pairwise modified distance is therefore given by

$$|\hat{d}_{ij}(\boldsymbol{a})|^2 = \sum_{\ell=0}^{L-1}\frac{\delta_{0,\ell}\gamma_{0,\ell}^2}{\delta_{0,\ell}-\min\{\delta_{1,\ell}\}}a_\ell^2 \tag{58}$$

and a similar derivation results in

$$|\hat{d}_{ji}(\boldsymbol{a})|^2 = \sum_{\ell=0}^{L-1}\frac{\delta_{1,\ell}\gamma_{1,\ell}^2}{\delta_{1,\ell}-\max\{\delta_{0,\ell}\}}a_\ell^2. \tag{59}$$

Now, setting all $\delta_{0,\ell}$ equal to $\max\{\delta_{0,\ell}\}$ uniformly increases $\hat{d}_{ij}(\boldsymbol{a})$, while not affecting $\hat{d}_{ji}(\boldsymbol{a})$. Similarly, setting all $\delta_{1,\ell}$ equal to $\min\{\delta_{1,\ell}\}$ uniformly increases $\hat{d}_{ji}(\boldsymbol{a})$, while not affecting $\hat{d}_{ij}(\boldsymbol{a})$. We therefore conclude that we can restrict ourselves to decoders for which $\delta_{0,\ell}$ are all equal to some (positive) value λ_0 for all $\ell = 0,\dots,L-1$, and $\delta_{1,\ell}$ are all equal to some (negative) value $-\lambda_1$, without any loss in performance. However, since our decoder is insensitive to scaling of the matrix Δ, we can assume that $\lambda_0 = 1$ and $\lambda_1 = \lambda$ for some $\lambda > 0$, and the first claim in the theorem is proved.

Finally, using the above selection of the matrix Δ and substituting the optimal values of $\hat{\Gamma}$ given in Theorem 2 into (58) and (59), we get

$$|\hat{d}_{ij}(\boldsymbol{a})|^2 = \frac{1}{1+\lambda}\sum_{\ell=0}^{L-1} P_\ell^{(i)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2$$
$$|\hat{d}_{ji}(\boldsymbol{a})|^2 = \frac{\lambda}{1+\lambda}\sum_{\ell=0}^{L-1} P_\ell^{(j)}\left(1-\left|\rho_\ell^{(ij)}\right|\right)a_\ell^2.$$

Notice that the expression for d_{ji} can be attained from d_{ij} simply by using a weight $\frac{1}{\lambda}$ instead of λ, and swapping indices $i \leftrightarrow j$. This manifests the fact that selecting Δ for the decoder Ω^{ij} implicitly defines Δ for the decoder Ω^{ji}. □

Proof of Theorem 4

The ML power error exponent as a function of the new coordinates $\boldsymbol{b}$ is given by

$$E^*(\boldsymbol{b}) = \min_{i\neq j} E_{ij}^*(\boldsymbol{b}) = \min_{i\neq j}\left\{e_\ell^{(ij)} + \sum_{\ell=0}^{L-2} c_\ell^{(ij)} b_\ell\right\} \tag{60}$$

where

$$c_\ell^{(ij)} = \frac{1}{8}\left(P_\ell^{(i)} + P_\ell^{(j)} - 2\sqrt{P_\ell^{(i)}P_\ell^{(j)}}\rho_\ell^{(ij)} - P_{L-1}^{(i)} - P_{L-1}^{(j)} + 2\sqrt{P_{L-1}^{(i)}P_{L-1}^{(j)}}\rho_{L-1}^{(ij)}\right)$$
$$e_\ell^{(ij)} = \frac{1}{8}\left(P_{L-1}^{(i)} + P_{L-1}^{(j)} - 2\sqrt{P_{L-1}^{(i)}P_{L-1}^{(j)}}\rho_{L-1}^{(ij)}\right). \tag{61}$$

We see that $E^*(\boldsymbol{b})$ is a minima of a set of affine functions $E_{ij}^*(\boldsymbol{b})$ (hyperplanes), and therefore it is concave. Furthermore, its domain (the simplex S) is a polyhedra, and therefore $E^*(\boldsymbol{b})$ is a polyhedral function over S, since its graph can be represented by a set of linear inequality constraints. The same is true for $\hat{E}^\Omega(\boldsymbol{b})$, as it is of a similar structure.

The critical channels are the points attaining the infimum in

$$\hat{\xi}^\Omega = \inf_{\boldsymbol{b}\in S}\frac{\hat{E}^\Omega(\boldsymbol{b})}{E^*(\boldsymbol{b})}.$$

For any $\xi > 0$, define the function

$$f(\boldsymbol{b},\xi) \triangleq \hat{E}^\Omega(\boldsymbol{b}) - \xi E^*(\boldsymbol{b})$$

over the set S, with ξ as a parameter. We have

$$f(\boldsymbol{b},\hat{\xi}^\Omega) = \hat{E}^\Omega(\boldsymbol{b}) - \hat{\xi}^\Omega E^*(\boldsymbol{b}) = \hat{E}^\Omega(\boldsymbol{b}) - \left(\inf_{\boldsymbol{b}\in S}\frac{\hat{E}^\Omega(\boldsymbol{b})}{E^*(\boldsymbol{b})}\right)E^*(\boldsymbol{b}) \geq 0$$

with equality only for the critical channels. Therefore, the global minima points of $f(\boldsymbol{b},\hat{\xi}^\Omega)$ over S are the critical channels. Since $f(\boldsymbol{b},\xi)$ is a difference of two polyhedral functions, it is easily verified to be a polyhedral function itself. In addition, since S is closed and bounded, the global minima of $f(\boldsymbol{b},\xi)$ over S is attained on at least one of its extreme points, which are a subset of the union of the extreme points of $E^*(\boldsymbol{b})$ and $\hat{E}^\Omega(\boldsymbol{b})$. Therefore, we conclude that the extreme points of $E^*(\boldsymbol{b})$ and $\hat{E}^\Omega(\boldsymbol{b})$ over S contain at least one critical channel. To this end, we now show that the extreme points of $\hat{E}^\Omega(\boldsymbol{b})$ over S can never be critical channels.

Assume that some point $\boldsymbol{b}_0 \in S$ is a minima of $f(\boldsymbol{b},\xi)$, and assume it is an extreme point of $\hat{E}^\Omega(\boldsymbol{b})$ but not of $E^*(\boldsymbol{b})$. Since $\boldsymbol{b}_0$ is not an extreme point of $E^*(\boldsymbol{b})$, there exists a line segment going through the point $(\boldsymbol{b}_0, E^*(\boldsymbol{b}_0))$, corresponding to $\boldsymbol{b}_0$ in $graph(E^*)$, that is fully embedded in $graph(E^*)$. This line segment in turn corresponds to some direction $\boldsymbol{n}_0$ emanating from $\boldsymbol{b}_0$, along which $E^*(\boldsymbol{b})$ satisfies

$$E^*(\boldsymbol{b}_0 + \varepsilon\boldsymbol{n}_0) = E^*(\boldsymbol{b}_0) + \varepsilon C^*$$
$$E^*(\boldsymbol{b}_0 - \varepsilon\boldsymbol{n}_0) = E^*(\boldsymbol{b}_0) - \varepsilon C^*$$

some constant C^*, and for any $\varepsilon > 0$ small enough.

Now, since $\boldsymbol{b}_0$ is an extreme point of $\hat{E}^\Omega(\boldsymbol{b})$, the line segment above is not fully embedded in $graph(\hat{E}^\Omega)$, and therefore there exist two constants $C_1 \neq C_2$ so that

$$\hat{E}^\Omega(\boldsymbol{b}_0 + \varepsilon\boldsymbol{n}_0) = \hat{E}^\Omega(\boldsymbol{b}_0) + \varepsilon C_1$$
$$\hat{E}^\Omega(\boldsymbol{b}_0 - \varepsilon\boldsymbol{n}_0) = \hat{E}^\Omega(\boldsymbol{b}_0) - \varepsilon C_2$$

for any $\varepsilon > 0$ small enough. Since $\hat{E}^\Omega$ is also concave, we also have $C_1 < C_2$.

Finally, since we have assumed that $\boldsymbol{b}_0$ is a minima of $f(\boldsymbol{b},\xi)$, then $f(\boldsymbol{b},\xi)$ should be nondecreasing in any direction emanating from $\boldsymbol{b}_0$. Specifically, for the direction $\boldsymbol{n}_0$, we have

$$f(\boldsymbol{b}+\varepsilon\boldsymbol{n}_0,\xi) = \hat{E}^\Omega(\boldsymbol{b}_0) + \varepsilon C_1 - \xi(E^*(\boldsymbol{b}_0) + \varepsilon C^*) \geq \hat{E}^\Omega(\boldsymbol{b}_0) - \xi E^*(\boldsymbol{b}_0)$$

for any $\varepsilon > 0$ small enough, and so it is compulsory that $C_1 \geq C^*$. On the other hand

$$f(\boldsymbol{b}-\varepsilon\boldsymbol{n}_0,\xi) = \hat{E}^\Omega(\boldsymbol{b}_0) - \varepsilon C_2 - \xi(E^*(\boldsymbol{b}_0) - \varepsilon C^*) \geq \hat{E}^\Omega(\boldsymbol{b}_0) - \xi E^*(\boldsymbol{b}_0)$$

for any $\varepsilon > 0$ small enough, and so we have $C^* \geq C_2 > C_1$, which contradicts $C_1 \geq C^*$. Therefore, $\boldsymbol{b}_0$ is not a minima of $f(\boldsymbol{b},\xi)$ for any $\xi > 0$, let alone for $\hat{\xi}^\Omega$. Consequently, $\boldsymbol{b}_0$ cannot be a critical channel. □

Proof of Proposition 1

For any selection of the weights $\lambda_{ij} = \frac{1}{\lambda_{ji}}$, the lower bound can be divided into minimizations of pairs

$$\hat{\xi}^\Omega = \min_{i\neq j}\frac{s^{(ij)}}{1+\lambda_{ij}} = \min_{i>j}\min\left\{\frac{s^{(ij)}}{1+\lambda_{ij}}, \frac{s^{(ji)}}{1+\frac{1}{\lambda_{ij}}}\right\} = \min_{i>j}\min\left\{\frac{s^{(ij)}}{1+\lambda_{ij}}, \frac{s^{(ji)}\lambda_{ij}}{1+\lambda_{ij}}\right\}.$$

The left expression in the inner minima is decreasing with λ_{ij} and the right one is increasing with it, so the inside minima attains its maximal value when the two expression are equal, that is, for the selection

$$\lambda_{ij}^{*} = \frac{s^{(ij)}}{s^{(ji)}}.$$

Therefore, we can write

$$\hat{\xi}^{\Omega} = \min_{i>j} \min \left\{ \frac{s^{(ij)}}{1+\lambda_{ij}}, \frac{s^{(ji)}\lambda_{ij}}{1+\lambda_{ij}} \right\} \leq \min_{i>j} \min \left\{ \frac{s^{(ij)}}{1+\lambda_{ij}^{*}}, \frac{s^{(ji)}\lambda_{ij}^{*}}{1+\lambda_{ij}^{*}} \right\}.$$

To this end, we see that for any selection of weights λ_{ij}, the weights λ_{ij}^{*} result in a bound at least as good as the bound attained by λ_{ij}, and therefore, they maximize the bound. □

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their helpful comments.

REFERENCES

[1] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. Vehicular Technology Conf.*, vol. 2, Chicago, IL, Sep. 1995, pp. 815–819.

[2] D. P. Bertsekas, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[3] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[4] J. Choi, "A joint channel estimation and detection for frequency-domain equalization using an approximate EM algorithm," *Signal Process.*, vol. 84, no. 5, pp. 865–880, May 2004.

[5] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[6] I. Csiszár and J. Körner, *Information Theory: Coding for Discrete Memoryless Systems*. New York: Academic, 1981.

[7] E. Erez and M. Feder, "Improving the generalized likelihood ratio test for unknown linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 919–936, Apr. 2003.

[8] K. Fakuda, "Frequently asked questions in polyhedral computation," Dep. Math., Swiss Federal Inst. Technol., Zurich, Switzerland, Tech. Rep., 2000.

[9] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1726–1745, Sep. 1998.

[10] M. Feder and N. Merhav, "Universal composite hypothesis testing: A competitive minimax approach," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1504–1517, Jun. 2002.

[11] R. W. Heath and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Process.*, vol. 37, no. 3, pp. 848–856, Mar. 1999.

[12] *Supplement to IEEE Standard for Information Technology, Telecommunications and Information Exchange Between Systems–Local and Metropolitan Area Networks—Specific Requirements*, IEEE Std 802.11a-1999. Adopted by the ISO/IEC and Redesignated as ISO/IEC 8802-11:1999/Amd 1:2000(E).

[13] P. Lancaster and M. Tismenetsky, *The Theory of Matrices With Applications*, 2nd ed. Orlando, FL: Academic, 1985.

[14] A. Lapidoth and P. Narayan, "Reliable communications under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[15] A. Lapidoth and S. Shamai (Shitz), "A lower bound on the bit-error-rate resulting from mismatched viterbi decoding," *Europ. Trans. Telecommun.*, vol. 9, pp. 473–482, Nov./Dec. 1998.

[16] A. Lapidoth and İ. E. Teletar, private communication, 1997.

[17] A. Lapidoth and J. Ziv, "On the universality of the LZ-based decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1746–1755, Sep. 1998.

[18] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[19] N. Merhav, "Universal decoding for memoryless Gaussian channels with deterministic interference," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1261–1269, Jul. 1993.

[20] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[21] J. J. Proakis, *Digital Communications*, 3rd ed, ser. Electrical and Computer Engineering. New York: McGraw-Hill, 1995.

[22] O. Shayevitz, "Universal decoding for frequency selective channels," M.Sc. thesis, Tel-Aviv Univ., Tel-Aviv, Israel, 2004.

[23] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Vehicular Technology Conf.*, Phoenix, AZ, May 1997, pp. 1639–1643.

[24] Z. Wang and G. B. Giannakis, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 707–720, Mar. 2003.

[25] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun. Technol.*, vol. COM-19, no. 10, pp. 628–634, Oct. 1971.

[26] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992.

[27] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 453–460, Jul. 1985.