Frame-Error-Rate-wise Optimal Code-Aided Hypothesis Testing

Cédric Herzet, Xavier Wautelet, Valéry Ramon and Luc Vandendorpe

Communications Laboratory, Université catholique de Louvain,
Pl. du Levant 2, B1348 Louvain-la-Neuve, Belgium,
{herzet, wautelet, ramon, vandendorpe}@tele.ucl.ac.be

***Abstract*— This paper addresses the issue of code-aided hypothesis testing in communication systems, i.e., the estimation of discrete-valued nuisance parameters. A hypothesis testing procedure optimal in the sense of the minimization of the frame-error rate (FER) is derived. Its complexity is shown to be comparable to other recently-proposed code-aided hypothesis procedures. Moreover, when a conditional maximum a posteriori decision rule is used to make the decisions about the transmitted sequence, it is shown that the proposed hypothesis testing procedure combined with sequence detection reduces, in a good approximation, to a joint maximum-likelihood problem. Finally, as an illustrative example, we show the case of phase ambiguity resolution for a convolutionally-coded transmission.**

I. Introduction

Hypothesis testing is a common issue in digital burst communication systems. The most well-known examples of hypothesis testing include frame synchronization and phase ambiguity resolution. Conventional approaches to deal with such problems are based on the transmission of pilot symbols. The two most popular data-aided (DA) hypothesis testing methods are Massey's [1] and Lui's [2]. In these references, the authors derive the maximum-likelihood (ML) hypothesis testing procedure when the data symbols are assumed to be uncoded.

Since it implies the transmission of some pilot sequences, DA hypothesis testing leads to both a waste of power and bandwidth. In transmissions using powerful error-correcting codes [3], [4], the low signal-to-noise ratio (SNR) encountered by conventional DA hypothesis-testing devices may require the transmission of an unacceptable number of pilot symbols. As an extreme example of low operating SNR, [5] mentions a recent rate-1/31 code designed for deep-space communications and operating reliably at SNR's as low as -15dB! This problem has led several authors to propose hypothesis-testing procedures which take into account the structure of the code used to transmit the data symbols, see e.g. [6], [7], [8], [9], [10]. We can distinguish between two main approaches. In [6] and [7] for example, the authors propose hypothesis-testing methods based on the observation that the statistics of some decoder metrics varies as a function of the considered hypotheses. In other contributions, e.g. [8], [9] and [10], the authors place the coded-aided hypothesis-testing problem in the context of ML estimation. The exact ML solution being intractable, different suboptimal ML-based solutions are proposed in [8], [9] and [10].

In this paper, we place the hypothesis testing problem in the context of the sequence-wise optimal reception. Following this approach, we derive a hypothesis-testing procedure optimal in the sense of the minimization of the frame-error-rate (FER). Despite of the relative simplicity of the solution, the proposed sequence-wise optimal hypothesis-testing method has not been proposed to our knowledge in the literature yet.

The sequel of the paper is organized as follows. In section II, we set the model and the notations. In section III, the problem of hypothesis testing is placed in the context of optimal reception. The general equations of the FER-oriented hypothesis testing procedure are derived in section IV. In section V, we particularize our result to the case of a conditional maximum a posteriori decision rule on the transmitted sequence. Finally, in section VI we illustrate the performance of the proposed approach for phase ambiguity resolution in a convolutionally-coded transmission.

II. Model

We consider the transmission of an information sequence $\mathbf{u}$. The information sequence is coded and modulated, leading to a symbol sequence $\mathbf{a} \in \mathcal{A}$. We assume that the transmitted signal is corrupted by an additive white Gaussian noise (AWGN). Moreover, we assume that the channel can introduce an additional distortion, leading to the following received observation vector

$$\mathbf{r} = s(\mathbf{a}, \mathbf{b}) \; + \mathbf{w}, \tag{1}$$

where $\mathbf{w}$ is a complex zero-mean AWGN process with covariance $\sigma_w^2 \mathbf{I}$, $\mathbf{I}$ denoting the unity matrix, $\mathbf{b}$ is a discrete nuisance parameter vector, and $s()$ is the channel operator. Transmitted sequence $\mathbf{a}$ and nuisance parameter vector $\mathbf{b}$ are assumed to be independent and uniformly distributed over some sets $\mathcal{A}$ and $\mathcal{B}$ respectively.

III. Problem formulation

Our final goal is data detection, i.e., we want to minimize the probability of committing an error when making a decision, say $\hat{\mathbf{a}}$, on the transmitted sequence[1]. Under this criterion, the optimal decision consists in the determination of $\mathbf{a}$ irrespective of $\mathbf{b}$, i.e.,

$$\hat{\mathbf{a}}_{\mathrm{map}} = \arg\max_{\tilde{\mathbf{a}}} \, p(\mathbf{a} = \tilde{\mathbf{a}}|\mathbf{r}), \tag{2}$$

where $\tilde{\mathbf{a}}$ is a trial sequence. Unfortunately, the latter maximization problem does usually not have any simple analytical solution: in the general case, we are required to perform $|\mathcal{A}|$ evaluations of the objective function $p(\mathbf{a}|\mathbf{r})$ in order to find its maximum. Therefore, rather than implementing this quite complicated solution, conditional maximum a posteriori (CMAP) sequence estimation is often preferred, i.e.,

$$\hat{\mathbf{a}}_{\mathrm{cmap}} = \arg\max_{\tilde{\mathbf{a}} \in \mathcal{A}} \; p(\mathbf{a} = \tilde{\mathbf{a}}|\mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}), \tag{3}$$

where $\hat{\mathbf{b}}$ is an estimate of discrete synchronization parameter $\mathbf{b}$. In fact, the CMAP sequence estimator may be regarded as a particular case of a general family of sequence estimation algorithms operating as follows: they first compute an estimate $\hat{\mathbf{b}}$ of $\mathbf{b}$; then, they perform data detection assuming that $\mathbf{b}$ is equal to $\hat{\mathbf{b}}$, i.e.,

$$\hat{\mathbf{a}} = f(\mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}), \tag{4}$$

where $f(\cdot)$ denotes some sequence decision rule. In the sequel, we will use the short-hand notation $f(\mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}) = f(\mathbf{r}, \hat{\mathbf{b}})$.

From (4), the question of the choice of $\hat{\mathbf{b}}$ naturally arises. A common approach consists in considering the MAP (or ML[2]) solution, i.e.,

$$\hat{\mathbf{b}} = \arg\max_{\tilde{\mathbf{b}}} \, p(\mathbf{b} = \tilde{\mathbf{b}}|\mathbf{r}), \tag{5}$$

The ML solution is the one which minimizes the probability of committing an error on the decision made on $\mathbf{b}$. However, its computational complexity is intractable and ML-based approaches proposed in the literature [8], [9] and [10] are approximations of the ML solution (5).

In the remainder of this paper, we derive a decision rule for $\hat{\mathbf{b}}$ which is optimal in the sense of the minimization of the frame-error rate. We show that a tractable decision rule for $\hat{\mathbf{b}}$ may be designed in this way.

IV. FER-Oriented Discrete-Parameter Estimation

From (4) it is clear that the probability of making a wrong decision about the transmitted sequence depends on the decision $\hat{\mathbf{b}}$ made about $\mathbf{b}$. Let $e_{\mathbf{a}}$ denote the event of making a wrong decision on the transmitted sequence. In this paper we focus on the particular decision rule for $\hat{\mathbf{b}}$, say $g(\mathbf{r})$, such that the probability to make a wrong decision on $\mathbf{a}$ using (4) is minimized. In other words, we look for the function $g(\mathbf{r})$ such that

$$p\Big(e_{\mathbf{a}} \,\Big|\, \hat{\mathbf{a}} = f\big(\mathbf{r}, \hat{\mathbf{b}} = g(\mathbf{r})\big)\Big) \tag{6}$$

is minimum. For the sake of conciseness, probability (6) will simply be denoted as $p(e_{\mathbf{a}})$ in the sequel.

Let $e_{\mathbf{b}}$ (resp. $\bar{e}_{\mathbf{b}}$) denote the event of making a wrong (resp. correct) decision on discrete parameter $\mathbf{b}$. Taking into account that both $e_{\mathbf{a}}$ and $e_{\mathbf{b}}$ depend on observation vector $\mathbf{r}$, the objective function (6) may be rewritten as

$$p(e_{\mathbf{a}}) = \int \Big(p(e_{\mathbf{a}}, \bar{e}_{\mathbf{b}}, \mathbf{r}) + p(e_{\mathbf{a}}, e_{\mathbf{b}}, \mathbf{r}) \Big) \, d\mathbf{r}, \tag{7}$$

$$= \int \Big(p(e_{\mathbf{a}}, \bar{e}_{\mathbf{b}}|\mathbf{r}) + p(e_{\mathbf{a}}, e_{\mathbf{b}}|\mathbf{r}) \Big) \, p(\mathbf{r}) \, d\mathbf{r}. \tag{8}$$

From (8), we have that the decision rule $g(\mathbf{r})$ which minimizes $p(e_{\mathbf{a}})$ has to minimize $\big(p(e_{\mathbf{a}}, \bar{e}_{\mathbf{b}}|\mathbf{r}) + p(e_{\mathbf{a}}, e_{\mathbf{b}}|\mathbf{r})\big)$ for each realization of $\mathbf{r}$. Therefore, for a given value of $\mathbf{r}$, the value of $\hat{\mathbf{b}}$, say $\hat{\mathbf{b}}_{\mathrm{fer}}$, which minimizes $p(e_{\mathbf{a}})$ may be computed as

$$\hat{\mathbf{b}}_{\mathrm{fer}} = \arg\min_{\tilde{\mathbf{b}}} \Big\{ p(e_{\mathbf{a}}, \bar{e}_{\mathbf{b}}|\mathbf{r}) + p(e_{\mathbf{a}}, e_{\mathbf{b}}|\mathbf{r}) \,\Big|\, \hat{\mathbf{b}} = \tilde{\mathbf{b}} \Big\}, \tag{9}$$

where $\tilde{\mathbf{b}}$ denotes a trial value of $\hat{\mathbf{b}}$. In the appendix, we show that (9) may in fact be rewritten as

$$\hat{\mathbf{b}}_{\mathrm{fer}} = \arg\max_{\tilde{\mathbf{b}}} \; p\big(\mathbf{a} = f(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}})|\mathbf{r}\big). \tag{10}$$

We arrive therefore at the conclusion that the FER-oriented decision rule on $\mathbf{b}$ has to be such that the subsequent decision made on $\mathbf{a}$ via (4) maximizes the cost function which appears in the optimal receiver solution (2), namely $p(\mathbf{a}|\mathbf{r})$. The latter cost function, although difficult to maximize over $\mathbf{a}$, may be computed straightforwardly for a given value of $\mathbf{a}$ as

$$p(\mathbf{a}|\mathbf{r}) = \sum_{\mathbf{b}} p(\mathbf{a}|\mathbf{r}, \mathbf{b}) \, p(\mathbf{b}|\mathbf{r}), \tag{11}$$

$$\sim \sum_{\mathbf{b}} p(\mathbf{r}|\mathbf{a}, \mathbf{b}) \, p(\mathbf{a}) \, p(\mathbf{b}), \tag{12}$$

where $\sim$ denotes equality up to a normalization factor. Probabilities $p(\mathbf{a})$ and $p(\mathbf{b})$ are assumed to be uniformly distributed over sets $\mathcal{A}$ and $\mathcal{B}$ respectively, whereas $p(\mathbf{r}|\mathbf{a}, \mathbf{b})$ is easy to calculate due to the AWGN nature of the noise.

On the one hand, we see from (2) that, in order to find the maximum of $p(\mathbf{a}|\mathbf{r})$ over $\mathbf{a}$, the optimal receiver solution requires to test the $|\mathcal{A}|$ possible transmitted sequences, where $|\cdot|$ denotes cardinality. On the other hand, the suboptimal approach based on (4) and (10) only requires $|\mathcal{B}|$ evaluations of $p(\mathbf{a}|\mathbf{r})$. In fact, the sequence decision rule defined in (4) is such that there is only one sequence $\hat{\mathbf{a}}$ corresponding to each decision $\hat{\mathbf{b}}$. Consequently, imposing a decision rule such as (4) implicitly reduces the set of possible candidates for the final sequence decision to $|\mathcal{B}|$ sequences. Let us notice that $1 - p(\mathbf{a} = \hat{\mathbf{a}}|\mathbf{r})$ is the probability of committing an error if we decide that sequence $\hat{\mathbf{a}}$ has been transmitted. Therefore, the FER-wise optimal hypothesis-testing method (10) reduces to choosing $\hat{\mathbf{b}}$ such that the subsequent decision made on $\mathbf{a}$, i.e.,

[1]Note that there is a one-to-one mapping between symbol sequence $\mathbf{a}$ and information sequence $\mathbf{u}$. Therefore, the optimal receiver with respect to the detection of $\mathbf{a}$ is equivalent to the optimal receiver with respect to the detection of $\mathbf{u}$

[2]both approaches being equivalent since we assume that $p(\mathbf{b})$ is uniform.

$\hat{\mathbf{a}} = f(\mathbf{r}, \hat{\mathbf{b}})$, minimizes the probability of making a sequence error. In other words, decision rule (10) is implicitly equivalent to choosing, among $|\mathcal{B}|$ possible sequences, the one which minimizes the probability of wrong detection.

V. A PARTICULAR CASE: CONDITIONAL MAP SEQUENCE DECISION RULE

In this section, we consider a particular choice of sequence decision rule $f(\mathbf{r}, \hat{\mathbf{b}})$, namely a conditional maximum a posteriori (CMAP) decision rule, i.e,

$$\hat{\mathbf{a}}_{\mathrm{cmap}} = \arg\max_{\tilde{\mathbf{a}} \in \mathcal{A}} p(\mathbf{a} = \tilde{\mathbf{a}} | \mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}). \tag{13}$$

We will show that the FER-oriented optimal decision rule (10) combined with sequence decision rule (13) reduces in good approximation to a joint maximum-likelihood problem, i.e.,

$$(\hat{\mathbf{a}}_{\mathrm{cmap}}, \hat{\mathbf{b}}_{\mathrm{fer}}) \simeq \underset{(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) \in \mathcal{A}\times\mathcal{B}}{\operatorname{argmax}} \ p(\mathbf{r} | \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \tilde{\mathbf{b}}). \tag{14}$$

Indeed, using (10) with CMAP sequence decision rule, the search of the parameter value minimizing the sequence error probability, say $\hat{\mathbf{b}}_{\mathrm{fer}}$, becomes

$$\hat{\mathbf{b}}_{\mathrm{fer}} = \arg\max_{\tilde{\mathbf{b}} \in \mathcal{B}} \ p\big(\mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}) | \mathbf{r}\big), \tag{15}$$

where the notation $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}})$ recalls the fact that the decision made about the transmitted sequence both depends on $\mathbf{r}$ and $\hat{\mathbf{b}}$. Expanding then $p(\mathbf{a}|\mathbf{r})$ as

$$p(\mathbf{a}|\mathbf{r}) = \sum_{\mathbf{b}} p(\mathbf{a}|\mathbf{r}, \mathbf{b})\, p(\mathbf{b}|\mathbf{r}) \tag{16}$$

and taking into account that

$$p(\mathbf{a}|\mathbf{b}, \mathbf{r})\, p(\mathbf{b}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{a}, \mathbf{b}) p(\mathbf{a})\, p(\mathbf{b})}{p(\mathbf{r})}, \tag{17}$$

we get by dropping the terms independent of $\mathbf{b}$:

$$\hat{\mathbf{b}}_{\mathrm{fer}} = \arg\max_{\tilde{\mathbf{b}} \in \mathcal{B}} \Big\{ p(\mathbf{r} | \mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \tilde{\mathbf{b}}) + \sum_{\mathbf{b}^\star \neq \tilde{\mathbf{b}}} p(\mathbf{r} | \mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \mathbf{b}^\star) \Big\}.$$

In the same way, using (17) and dropping the terms independent of $\mathbf{a}$, the CMAP sequence decision rule (13) may be written as

$$\hat{\mathbf{a}}_{\mathrm{cmap}} = \arg\max_{\tilde{\mathbf{a}} \in \mathcal{A}} p(\mathbf{r} | \mathbf{a} = \tilde{\mathbf{a}}, \mathbf{b} = \hat{\mathbf{b}}). \tag{18}$$

Note that, due to the Gaussian nature of the noise, the CMAP decision rule (18) actually reduces to choosing the sequence minimizing the Euclidean distance with observation vector $\mathbf{r}$ under the assumption $\mathbf{b} = \hat{\mathbf{b}}$.

In the sequel, we make the assumption that $\forall\ \mathbf{b}^\star \neq \tilde{\mathbf{b}}$ we have

$$p(\mathbf{r} | \mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \tilde{\mathbf{b}}) \gg p(\mathbf{r} | \mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \mathbf{b}^\star) \tag{19}$$

i.e., we assume the Euclidean distance between observation vector $\mathbf{r}$ and CMAP sequence decision $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}})$ under the assumption $\mathbf{b} \neq \tilde{\mathbf{b}}$ is large. Assumption (19) holds in a number of cases and will be discussed in the next section in the case of phase ambiguity resolution. Taking then (19) into account, we have

$$\hat{\mathbf{b}}_{\mathrm{fer}} \simeq \arg\max_{\tilde{\mathbf{b}} \in \mathcal{B}} \ p(\mathbf{r} | \mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \tilde{\mathbf{b}}). \tag{20}$$

Combining (18) and (20), we get (14). As a conclusion, considering a CMAP sequence decision rule (13), we have under assumption (19) that the FER-wise optimal decision on $\mathbf{a}$ and $\mathbf{b}$ is the solution of the joint maximum-likelihood problem (14).

VI. PHASE AMBIGUITY RESOLUTION

As an illustrative example of the proposed hypothesis-testing method, we consider the following transmitted signal:

$$s(\mathbf{a}, \mathbf{b}) = \mathbf{a}\, e^{j\theta}, \tag{21}$$

where the only element of the nuisance-parameter vector $\mathbf{b}$, is the unknown carrier phase offset θ. The carrier phase offset may be written as

$$\theta = k_\theta \Psi + \epsilon_\theta \qquad \text{with } -\frac{\Psi}{2} \leq \epsilon_\theta \leq \frac{\Psi}{2}, \tag{22}$$

where k_θ is an integer and Ψ is the smallest angle of rotational symmetry of the constellation: for M-PSK constellation $\Psi = 2\pi/M$ whereas for M-QAM constellation we have $\Psi = 2\pi/4$. The continuous component ϵ_θ can be estimated using some standard phase estimation algorithm [11]. Hence, we may restrict our attention to the estimation of k_θ. We therefore consider $\mathbf{b} = k_\theta$. The corresponding hypothesis testing problem is known as phase ambiguity resolution.

In Fig. 1, we illustrate the performance in terms of frame-error rate (FER) of the proposed hypothesis testing procedure in the case of a convolutionally-coded transmission with BPSK modulation. We consider three rate-1/2 maximum-free-distance non-recursive convolutional codes with constraint lengths respectively equal to 3, 5 and 7. The length of the information word is 250 bits. We assume that a CMAP sequence decision rule is used and efficiently implemented via a Viterbi sequence decoder [12]. The continuous part of the phase offset ϵ_θ is estimated via a Viterbi&Viterbi synchronizer [11]. In Fig. 1, the dashed curves represent the performance of a perfectly-synchronized system, i.e., the FER achieved by a receiver which has a perfect knowledge of the carrier phase offset. The curves with circles represent the FER when phase-ambiguity decision rule (15) is used whereas the curves with squares correspond to a phase ambiguity resolution via (14).

Note that the approximation made in (19) makes sense in our illustrative example. Indeed, as mentioned above, the CMAP Viterbi sequence decoder delivers the codeword $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \tilde{\mathbf{b}})$ which has the smallest Euclidean distance with the received observation vector $\mathbf{r}$ under a given assumption for $\mathbf{b}$,

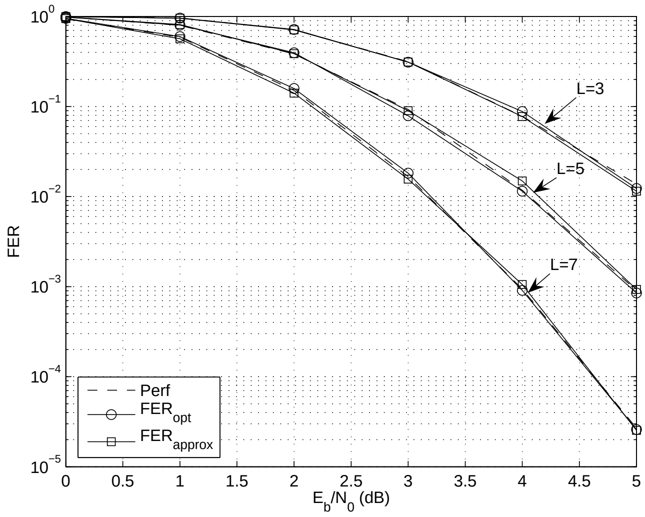

Fig. 1. FER for length-3, length-5 and length-7 rate-1/2 convolutional code.

say $\mathbf{b} = \tilde{\mathbf{b}}$. Considering the probability of $\mathbf{r}$ given $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \tilde{\mathbf{b}})$ under the complementary assumption, i.e., $\mathbf{b} = \tilde{\mathbf{b}}\, e^{j\pi}$, we have

$$p(\mathbf{r}|\mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}}), \mathbf{b} = \tilde{\mathbf{b}}\, e^{j\pi}) = p(\mathbf{r}|\mathbf{a} = \hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \hat{\mathbf{b}} = \tilde{\mathbf{b}})\, e^{j\pi}, \mathbf{b} = \tilde{\mathbf{b}}), \tag{23}$$

where (23) directly derives from the particular structure of the received signal (21). Now, $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})\, e^{j\pi}$ is the sequence which has the largest Hamming distance with respect to sequence $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})$. Since $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})$ is close to $\mathbf{r}$ in terms of Euclidean distance, $\hat{\mathbf{a}}_{\mathrm{cmap}}(\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})\, e^{j\pi}$ is proportionally quite far. This implies (19).

We see from Fig. 1 that both the FER-wise optimal approach (15) and its approximated version (14) enable to recover the performance of the perfectly synchronized system without the use of any pilot symbols. These approaches enable therefore to improve the power and/or the spectral efficiency of the system at the expense of an increase of the receiver complexity. The amount of complexity increase depends on the number of hypotheses to be tested. For instance, in our illustrative example there are two possible ambiguities, namely $k_\theta = 0$ and $k_\theta = 1$. We are therefore required to perform two decoding operations since, as shown by (13) and (15), each hypothesis test requires one decoding operation. This fact is a direct consequence of the discrete nature of the nuisance parameter and appears in most of the code-aided hypothesis methods, see e.g. [9], [10]. The gain in power and spectral efficiency is a function of the considered scenario and will therefore not be discussed here. We just mention that the power/spectral efficiency gain basically depends on parameters such as the frame length, the code rate, the code length, the nature of the nuisance parameter, the operating point, etc.

Conclusion

A novel approach based on a FER-minimization criterion is proposed to design code-aided hypothesis testing methods. The optimal solution in that sense is derived and is shown to differ from the classical ML-based approach. The proposed hypothesis testing method has the common feature to require one decoding operation per hypothesis to test but its complexity remains tractable. We particularize our result to the case of a sequence conditional maximum a posteriori (CMAP) decision rule. It is shown that the combination of CMAP sequence detection and FER-optimal hypothesis testing reduces in good approximation to a joint ML problem. The validity of the proposed approach is illustrated via a simple illustrative example of phase ambiguity resolution.

Appendix

Let us first notice that the function to minimize in (9) may be rewritten as

$$p(e_{\mathbf{a}}, \bar{e}_{\mathbf{b}}|\mathbf{r}) + p(e_{\mathbf{a}}, e_{\mathbf{b}}|\mathbf{r}) = p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r})\, p(\bar{e}_{\mathbf{b}}|\mathbf{r}) + p(e_{\mathbf{a}}|e_{\mathbf{b}}, \mathbf{r})\, p(e_{\mathbf{b}}|\mathbf{r}), \tag{24}$$

where equality follows from the Bayes rule. To prove (10), we calculate an alternative expression of the probabilities appearing in (24), namely $p(\bar{e}_{\mathbf{b}}|\mathbf{r})$, $p(e_{\mathbf{b}}|\mathbf{r})$, $p(e_{\mathbf{a}}|e_{\mathbf{b}}, \mathbf{r})$ and $p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r})$.

A. *Computation of* $p(\bar{e}_{\mathbf{b}}|\mathbf{r})$ *and* $p(e_{\mathbf{b}}|\mathbf{r})$

The computation of $p(\bar{e}_{\mathbf{b}}|\mathbf{r})$ and $p(e_{\mathbf{b}}|\mathbf{r})$ is straightforward as

$$p(\bar{e}_{\mathbf{b}}|\mathbf{r}) = p(\mathbf{b} = \hat{\mathbf{b}}|\mathbf{r}), \tag{25}$$

$$p(e_{\mathbf{b}}|\mathbf{r}) = 1 - p(\mathbf{b} = \hat{\mathbf{b}}|\mathbf{r}). \tag{26}$$

Indeed, the probability of committing an error when the nuisance parameter estimate is set to $\hat{\mathbf{b}}$ is simply equal to 1 minus the probability that the actual nuisance parameter value is equal to $\hat{\mathbf{b}}$.

B. *Computation of* $p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r})$

Probability $p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r})$ may be expanded as

$$p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}) = \sum_{\tilde{\mathbf{b}} \in \mathcal{B}} p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})\, p(\mathbf{b} = \tilde{\mathbf{b}}|\mathbf{r}, \bar{e}_{\mathbf{b}}),$$

or, using the Bayes rule on $p(\mathbf{b} = \tilde{\mathbf{b}}|\mathbf{r}, \bar{e}_{\mathbf{b}})$:

$$p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}) = \sum_{\tilde{\mathbf{b}} \in \mathcal{B}} p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}})\, \frac{p(\bar{e}_{\mathbf{b}}|\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}}) p(\mathbf{b} = \tilde{\mathbf{b}}|\mathbf{r})}{\sum_{\tilde{\mathbf{b}} \in \mathcal{B}} p(\bar{e}_{\mathbf{b}}|\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}}) p(\mathbf{b} = \tilde{\mathbf{b}}|\mathbf{r})}. \tag{27}$$

Let $\mathbb{I}\{\cdot\}$ denote the indicator function, which is equal to 1 if the statement between the braces is true and 0 otherwise. Since the decision $\hat{\mathbf{b}}$ made on $\mathbf{b}$ is univocally defined by the observation vector $\mathbf{r}$, we have

$$p(\bar{e}_{\mathbf{b}}|\mathbf{r}, \mathbf{b} = \tilde{\mathbf{b}}) = \mathbb{I}\{\hat{\mathbf{b}} = \tilde{\mathbf{b}}\}, \tag{28}$$

and therefore (27) simply reduces to

$$p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}) = p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}).$$

Then, for a given decision rule $f(\mathbf{r}, \hat{\mathbf{b}})$, we finally have

$$p(e_{\mathbf{a}}|\bar{e}_{\mathbf{b}}, \mathbf{r}) = 1 - p(\mathbf{a} = f(\mathbf{r}, \hat{\mathbf{b}})|\mathbf{r}, \mathbf{b} = \hat{\mathbf{b}}). \tag{29}$$

C. Computation of $p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r})$

Using the same kind of expansion as (27), we have

$$p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r}) = \sum_{\tilde{\mathbf{b}}\in\mathcal{B}} p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) \frac{p(e_{\mathbf{b}}|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}})p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})}{\sum_{\tilde{\mathbf{b}}\in\mathcal{B}} p(e_{\mathbf{b}}|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}})p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})},$$

and taking into account that

$$p(e_{\mathbf{b}}|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) = \mathbb{I}\{\hat{\mathbf{b}}\neq\tilde{\mathbf{b}}\}, \tag{30}$$

we get

$$p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r}) = \sum_{\tilde{\mathbf{b}}\neq\hat{\mathbf{b}}} p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) \frac{p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})}{\sum_{\tilde{\mathbf{b}}\neq\hat{\mathbf{b}}} p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})}. \tag{31}$$

Let us now expand $p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}})$ as

$$p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) = \sum_{\tilde{\mathbf{a}}\in\mathcal{A}} p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}},\mathbf{a}=\tilde{\mathbf{a}}) \times p(\mathbf{a}=\tilde{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}). \tag{32}$$

Let us notice that the decision $\hat{\mathbf{a}}$ made on $\mathbf{a}$ is also a univocal function of the observation vector $\mathbf{r}$ and therefore

$$p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}},\mathbf{a}=\tilde{\mathbf{a}}) = \mathbb{I}\{\hat{\mathbf{a}}\neq\tilde{\mathbf{a}}\}. \tag{33}$$

Using (33), (32) reduces to

$$\begin{aligned} p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) &= \sum_{\tilde{\mathbf{a}}\neq\hat{\mathbf{a}}} p(\mathbf{a}=\tilde{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}), \\ &= \sum_{\tilde{\mathbf{a}}\neq\hat{\mathbf{a}}} p(\mathbf{a}=\tilde{\mathbf{a}}|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}), \qquad (34) \\ &= 1 - p(\mathbf{a}=\hat{\mathbf{a}}|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}), \qquad (35) \end{aligned}$$

where (34) follows from the fact that the transmitted sequence is independent of the event of making a wrong decision on the value of $\mathbf{b}$. Plugging (35) into (31) and taking into account that $\hat{\mathbf{a}} = f(\mathbf{r},\hat{\mathbf{b}})$, we finally get

$$p(e_{\mathbf{a}}|e_{\mathbf{b}},\mathbf{r}) = 1 - \sum_{\tilde{\mathbf{b}}\neq\hat{\mathbf{b}}} p(\mathbf{a}=f(\mathbf{r},\hat{\mathbf{b}})|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}}) \frac{p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})}{\sum_{\tilde{\mathbf{b}}\neq\hat{\mathbf{b}}} p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r})}. \tag{36}$$

D. Final expression

Replacing each probability in (9) by the corresponding expression (25), (26), (29) or (36) , we get by rearranging the terms:

$$g(\mathbf{r}) = \arg\min_{\tilde{\mathbf{b}}} \Big\{ 1 - \Big(p(\mathbf{a}=f(\mathbf{r},\hat{\mathbf{b}}=\tilde{\mathbf{b}})|\mathbf{r},\mathbf{b}=\tilde{\mathbf{b}})\, p(\mathbf{b}=\tilde{\mathbf{b}}|\mathbf{r}) + \sum_{\mathbf{b}^\star\neq\tilde{\mathbf{b}}} p(\mathbf{a}=f(\mathbf{r},\hat{\mathbf{b}}=\mathbf{b}^\star)|\mathbf{r},\mathbf{b}=\mathbf{b}^\star)\, p(\mathbf{b}=\mathbf{b}^\star|\mathbf{r}) \Big) \Big\}.$$

Taking into account the fact that $p(\mathbf{a}|\mathbf{r}) = \sum_{\mathbf{b}} p(\mathbf{a}|\mathbf{r},\mathbf{b})\, p(\mathbf{b}|\mathbf{r})$, we finally have

$$\begin{aligned} g(\mathbf{r}) &= \arg\min_{\tilde{\mathbf{b}}} \Big\{ 1 - p(\mathbf{a}=f(\mathbf{r},\hat{\mathbf{b}}=\tilde{\mathbf{b}})|\mathbf{r}) \Big\}, \\ &= \arg\max_{\tilde{\mathbf{b}}} \Big\{ p(\mathbf{a}=f(\mathbf{r},\hat{\mathbf{b}}=\tilde{\mathbf{b}})|\mathbf{r}) \Big\}. \end{aligned}$$

REFERENCES

[1] J. L. Massey. ”Optimum frame synchronization”. *IEEE Trans. on Commun.*, 20:pp. 115–119, Apr. 1972.

[2] J. L. Lui, and H. H. Tan. ”Frame Synchronization for Gaussian Channel”. *IEEE Trans. on Commun.*, 35(8):pp. 818–829, August 1987.

[3] C. Berrou and A. Glavieux. ”Near Optimum Error Correcting Coding and Decoding: Turbo Codes”. In *IEEE Trans. on Commun.*, volume 44, pages 1261–1271, October 1996.

[4] R. G. Gallager. ”Low Density parity-check codes”. *IRE Trans. Inform. Theory*, IT 8:pp. 21–29, Jan. 1962.

[5] J.R. Barry, A. Kavcic, S.W. McLaughlin, A. Nayak and W. Zeng. ”Iterative Timing Recovery”. *IEEE Trans. Signal Processing*, pages 89–102, January 2004.

[6] U. Mengali, A. Sandri and A. Spalvieri. ”Phase Ambiguity Resolution in Treillis-Coded Modulations”. *IEEE Trans. on Commun.*, 38(12):pp. 2087–2089, Dec. 1990.

[7] B. Mielczarek and A. Svensson. ”Timing error recovery in turbo coded systems on AWGN channels”. *IEEE Trans. Comm.*, 50(10):pp. 1584–1592, Oct. 2002.

[8] U. Mengali, R. Pellizzoni and A. Spalvieri. ”Soft-Decision-Based Node Synchronization for Viterbi Decoders”. *IEEE Trans. on Commun.*, 43(9):pp. 2532–2539, Sept. 1995.

[9] T. M. Cassaro and N. Georghiades. ”Frame Synchronization for Coded Systems Over AWGN Channels”. vol. 52:pp. 484–489, March 2004.

[10] H. Wymeersch and M. Moeneclaey. ”ML frame synchronization for turbo and LDPC codes”. In *Int. Symp. on Turbo Codes & Rel. Topics, ISTC'*, Brest, France, Sept. 2003.

[11] A. J. Viterbi and A. M. Viterbi. ”Nonlinear estimation of PSK-Modulated Carrier Phase with Application to Burst Digital Transmission”. *IEEE Trans. on Inform. Theory*, 29:pp. 543–551, Jul. 1983.

[12] A. J. Viterbi. ”Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”. *IEEE Trans. on Inform. Theory*, 13(2):pp. 260–269, 1967.