

FER Estimation in a Memoryless BSC with Variable Frame Length and Unreliable ACK/NAK Feedback

Alberto Rico-Alvariño, *Member, IEEE*, Roberto López-Valcarce, *Member, IEEE*,
Carlos Mosquera, *Senior Member, IEEE*, Robert W. Heath Jr., *Fellow, IEEE*

Abstract

We consider the problem of estimating the frame error rate (FER) of a given memoryless binary symmetric channel by observing the success or failure of transmitted packets. Whereas FER estimation is relatively straightforward if all observations correspond to packets with equal length, the problem becomes considerably more complex when this is not the case. We develop FER estimators when transmissions of different lengths are observed, together with the Cramer-Rao Lower Bound (CRLB). Although the main focus is on Maximum Likelihood (ML) estimation, we also obtain low complexity schemes performing close to optimal in some scenarios. In a second stage, we consider the case in which FER estimation is performed at a node different from the receiver, and incorporate the impairment of unreliable observations by considering noisy ACK/NAK feedback links. The impact of unreliable

A. Rico-Alvariño was with the Signal Theory and Communications Department, University of Vigo, Spain. He is now with Qualcomm Technologies Inc., San Diego, CA, USA (e-mail: albertor@qti.qualcomm.com).

R. López-Valcarce and Carlos Mosquera are with the Signal Theory and Communications Department, University of Vigo, Spain. (e-mail: {valcarce, mosquera}@gts.uvigo.es).

R. W. Heath Jr. is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, TX, USA. (e-mail: rheath@utexas.edu).

The work of A. Rico-Alvariño, R. López-Valcarce and C. Mosquera was partially funded by the Agencia Estatal de Investigación (Spain) and the European Regional Development Fund (ERDF) under projects COMPASS (TEC2013-47020-C2-1-R), MYRADA (TEC2016-75103-C2-2-R), WINTER (TEC2016-76409-C2-2-R) and COMONSENS (TEC2015-69648-REDC), and also by the Xunta de Galicia (Agrupación Estratégica Consolidada de Galicia accreditation 2016-2019, Red Temática REDTEIC 2017-18) and ERDF. A. Rico-Alvariño was supported in part by a Fundación Pedro Barrié de la Maza graduate scholarship. R. W. Heath Jr. was supported in part by the National Science Foundation under Grant No. NSF-CCF-1319556.

A preliminary version of some parts of this paper was presented at ICASSP 2014 [1].

feedback is analyzed by means of the corresponding CRLB. In this setting, the ML estimator is obtained by applying the Expectation-Maximization algorithm to jointly estimate the error probabilities of the data and feedback links. Simulation results illustrate the benefits of the proposed estimators.

Index Terms

FER Estimation, Maximum Likelihood, Unreliable Feedback, Link Adaptation, Expectation-Maximization.

I. INTRODUCTION

Frame error rate (FER) estimation is a crucial step in many problems related to the design and deployment of wireless communication systems. FER estimation is used in PHY layer abstraction for system-level simulation [2]–[4], as well as in rate adaptation algorithms where a modulation and coding scheme (MCS) is selected according to the state of the channel [5]–[8]. Some of these rate adaptation schemes rely only on the estimation of the channel state by exploiting binary feedback information about the success or failure in the decoding of previously transmitted packets [9], [10]. Also, network analyzers for different wireless protocols (IEEE 802.11 [11], 3GPP LTE [12], DVB-T [13]) provide FER estimates as a result of the observation of the current channels.

The core of any FER estimation method is the observation of the outcome of the transmission process of several frames over a channel. For the case of constant length codewords, the FER can be readily estimated as the sample mean of the error events. In many communication standards, however, frames have a length that varies with different parameters, and depends on the size of the protocol data units (PDU) delivered by the MAC layer. For instance, standards such as those in the IEEE 802.11 family [14] or 3GPP LTE [15] operate with codewords of variable length; as a particular example, in the LTE Physical Downlink and Uplink Shared Channels (PDSCH and PUSCH) the PHY layer adds a 24-bit Cyclic Redundancy Check (CRC) to data blocks whose sizes may vary between 40 and 6144 bits including CRC bits [16]. Estimating the FER for a given frame length based only on those observations corresponding to received frames with that same length is clearly suboptimal, due to an inefficient usage of available data. However, generalizing FER estimation in order to account for observations corresponding to packets with different lengths is not straightforward, as it will be shown. The goal of this paper is to study

and build novel FER estimators for this setting, based on statistical estimation theory.

Previous work on FER estimation dealt with the issue of variable frame length in different ways. In [5]–[7], [17], [18] constant frame length was assumed, which is not realistic for current communication systems, as discussed above. In other examples [19]–[22], perfect knowledge of the bit error rate (BER) was assumed, and the corresponding FER is obtained from that value. This approach, however, has two significant drawbacks. First, it is not clear how BER estimation errors would affect the quality of the resulting FER estimate. Second, in many communication scenarios (link adaptation or network analysis) the BER is not directly measurable, since it would require knowledge of the transmitted bits. Alternatively, if the transmitter makes use of an error detection code (e.g. CRC) as it is usually the case, then it becomes feasible to measure packet error events at the receiver, so that FER estimation becomes feasible based on these observations even if the BER is not available or observable.

Our approach hinges on a memoryless binary symmetric channel (BSC) abstraction for bit-level transmission. Previous works have considered estimation of the BSC parameter in several scenarios, e.g., in [23], [24], where estimation is based on the observation of the BSC output and the assumption of a nonrandom input with finite complexity; or in a distributed source coding or channel coding framework [25], [26], based on the availability of the received individual bits, including the syndrome, of a single codeword. These approaches can be regarded as "PHY-layer estimators." In contrast, the FER estimators considered in this paper are based on the observation of multiple binary (success/failure) packet error events alone (albeit with packets of different lengths), and therefore they can be directly applied at the MAC layer.

FER estimation is further complicated when an observer different from the receiver (a transmitter, or a third node acting as a network analyzer) tries to estimate the FER experienced by a receiver. In that case the FER-related information is obtained from the ACK/NAK sequence reported by the receiver. If this feedback channel is unreliable, then the estimation procedure has to be modified, since treating the ACK/NAK information as true may result in performance degradation. In this paper, we will consider unreliable feedback links and develop estimators for the error probability of the feedback channel as well as the FER of the forward data channel.

The contributions of this paper are summarized as follows:

- 1) *FER estimation from perfect measurements.* We derive estimators for the FER of a given frame length from ACK/NAK observations which may correspond to frames of different

lengths. In this first step, we assume that the ACK/NAK information is correct, i.e., either the feedback channel can be assumed error-free, or the estimation process is directly carried out by the receiver itself. We obtain the Maximum Likelihood estimator (MLE) in this setting by means of an iterative procedure. We also obtain reduced complexity approximations to the MLE in some particular scenarios (e.g. large number of observations, small number of errors) and compare their performance to the Cramer-Rao Lower Bound (CRLB).

- 2) *FER estimation with an unreliable feedback link.* In this second setting, the ACK/NAK observations are no longer assumed to be perfect. The feedback link is modeled as a BSC with error probability ϵ . The impact of unreliable feedback is analyzed by studying the corresponding CRLB. We obtain joint estimators for the general case in which the error probabilities in both links are unknown. Interestingly, these estimators only exist for the case of having codewords of different length: for the constant codeword length, the parameters become unidentifiable. The joint estimator is obtained by applying the Expectation-Maximization (E-M) algorithm.

The remaining of the paper is organized as follows. Section II presents the system model, addressing the difficulty of FER estimation with codewords of different lengths. The CRLB and the estimators for the case of reliable feedback are derived in Section III. The estimation problem with an unreliable feedback channel is presented and analyzed in Section IV, and conclusions are presented in Section V. Simulation results for the different estimators are included at the end of each of the corresponding sections.

II. SYSTEM MODEL AND PRELIMINARIES

Consider a transmitter-receiver pair communicating through a noisy channel. The transmitter builds blocks of bits $[b_1, \dots, b_L]$, $b_i \in \{0, 1\}$ of variable bit length L , that we will refer to as frames. The receiver observes $[\hat{b}_1, \dots, \hat{b}_L]$ at the output of the channel, with $\hat{b}_i \in \{0, 1\}$. We assume a memoryless BSC with BER p , i.e., $p \triangleq \mathbb{P}[\hat{b}_i = 1 | b_i = 0] = \mathbb{P}[\hat{b}_i = 0 | b_i = 1]$. We also denote $q \triangleq 1 - p$ for the sake of simplicity. The transmitter makes use of an error detection encoder, such that the receiver is able to identify any received block with at least one erroneous bit. We also assume that the undetected error probability of the error detection code is negligible¹. If we denote by θ_L the probability of receiving an erroneous block of length L ,

¹The undetected error probability of a CRC code is approximately 2^{-r} , where r is the number of check bits. For example, for the PDSCH and PUSCH of LTE, $r = 24$ and $2^{-r} \approx 6 \times 10^{-8}$.

i.e., the FER for frame length L , then we have that

$$\theta_L = 1 - \mathbb{P} \left[\bigcap_{i=1}^L \{b_i = \hat{b}_i\} \right] = 1 - q^L. \quad (1)$$

If the FER for a frame length L was perfectly known, then it would be possible to obtain the FER for a different frame length \tilde{L} as

$$\theta_{\tilde{L}} = 1 - (1 - \theta_L)^{\tilde{L}/L}. \quad (2)$$

Unfortunately, in practice, the exact FER is unlikely to be available for any length, and its value has to be estimated based on the observation over a time period of the success and failures of the transmission of frames of a certain length. Moreover, in a realistic environment, frames of different length may be transmitted during the observation window. In the following, we formalize the problem of FER estimation from observations of frames of various sizes.

Consider a communication system over a BSC with ℓ different frame sizes $L_1 < L_2 < \dots < L_\ell$. During an observation period, a receiver observes n_i transmissions of size L_i , out of which m_i are received with errors. The random variable m_i is binomially distributed with parameters n_i and θ_{L_i} , so that its probability mass function (PMF) is given by

$$f(m_i; \theta_{L_i}) = \binom{n_i}{m_i} \theta_{L_i}^{m_i} (1 - \theta_{L_i})^{n_i - m_i}. \quad (3)$$

The mean value of m_i is given by

$$\mathbb{E}[m_i] = n_i(1 - q^{L_i}). \quad (4)$$

The MLE of θ_{L_i} given the observation $\{n_i, m_i\}$ is the empirical FER, i.e.,

$$\hat{\theta}_{L_i} = \frac{m_i}{n_i}, \quad (5)$$

which is unbiased, and whose variance is given by

$$\sigma_i^2 \triangleq \mathbb{E} \left[\left(\theta_{L_i} - \hat{\theta}_{L_i} \right)^2 \right] = \frac{(1 - \theta_{L_i}) \theta_{L_i}}{n_i}. \quad (6)$$

In view of (2), when only observations from equal length frames are available (i.e., $\ell = 1$), the MLE of $\theta_{\tilde{L}}$ can be obtained from that of θ_{L_1} by virtue of the invariance property of the MLE [27, Th. 7.2] as $\hat{\theta}_{\tilde{L}} = 1 - \left(1 - \hat{\theta}_{L_1} \right)^{\tilde{L}/L_1}$.

If $\ell > 1$, however, the derivation of the MLE of $\theta_{\tilde{L}}$ is more involved. Due again to the invariance property, the MLE of $\theta_{\tilde{L}}$ could be readily obtained if that of q were available: $\hat{\theta}_{\tilde{L}, \text{ML}} =$

$1 - (\hat{q}_{\text{ML}})^{\tilde{L}}$. We will see that although the MLE cannot be expressed in closed form in general, it is possible to derive some of its properties, as well as an efficient numerical method for its computation. We will also analyze some asymptotic cases (small p , large number of observations for each length), for which closed-form approximations will be exposed.

III. FER ESTIMATION WITH OBSERVATIONS OF DIFFERENT LENGTH

Given a vector of observed errors $\mathbf{m} \triangleq [m_1, \dots, m_\ell]^T$, the goal is to estimate the FER corresponding to a frame length \tilde{L} . Since the channel is assumed memoryless, observations are independent, and therefore the PMF of \mathbf{m} parameterized by the FER $\theta_{\tilde{L}}$ can be obtained from (2) and (3) as

$$f(\mathbf{m}; \theta_{\tilde{L}}) = \prod_{i=1}^{\ell} \binom{n_i}{m_i} \left(1 - (1 - \theta_{\tilde{L}})^{\frac{L_i}{\tilde{L}}}\right)^{m_i} (1 - \theta_{\tilde{L}})^{\frac{L_i}{\tilde{L}}(n_i - m_i)}. \quad (7)$$

Alternatively, we can rewrite (7) in terms of q as

$$f(\mathbf{m}; q) = \prod_{i=1}^{\ell} \binom{n_i}{m_i} (1 - q^{L_i})^{m_i} q^{(n_i - m_i)L_i}. \quad (8)$$

Depending on the formulation of the problem we may resort to using (7) or (8). In the following, we derive different estimators for the problem under study. First, we derive the CRLB (which constitutes a bound on the variance of any unbiased estimator) to benchmark the performance of the different proposed estimators.

A. CRLB

We now derive the CRLB for the estimation of q from the observations \mathbf{m} ; the CRLB for the estimation of $\theta_{\tilde{L}}$ can be obtained by applying a suitable transformation to the final result [27]. The log-likelihood function (LLF) of q is obtained from (8) as

$$\mathcal{L}(q) = \sum_{i=1}^{\ell} [m_i \log(1 - q^{L_i}) + (n_i - m_i) L_i \log q], \quad (9)$$

where a constant term was omitted. Its derivative is readily obtained as

$$\mathcal{L}'(q) = \sum_{i=1}^{\ell} \left[\frac{(n_i - m_i) L_i}{q} - \frac{m_i L_i q^{L_i - 1}}{1 - q^{L_i}} \right]. \quad (10)$$

It can be checked that the regularity condition $\mathbb{E}[\mathcal{L}'(q)] = 0$ (necessary and sufficient for the application of the CRLB) is satisfied, just by substituting (4) after taking expectation in (10). It

is also seen that an efficient estimator for q (or, equivalently, p) does not exist for this problem, as it is not possible to find functions $g(\mathbf{m})$ and $I(q)$ such that $\mathcal{L}'(q) = I(q)(g(\mathbf{m}) - q)$ [27, Th. 3.1]. The second derivative of the LLF is

$$\mathcal{L}''(q) = - \sum_{i=1}^{\ell} \frac{L_i \left(n_i (1 - q^{L_i})^2 + m_i ((1 + L_i) q^{L_i} - 1) \right)}{q^2 (1 - q^{L_i})^2}. \quad (11)$$

Therefore, using (4), the Fisher information is found:

$$I(q) = -\mathbb{E}[\mathcal{L}''(q)] = \frac{1}{q^2} \sum_{i=1}^{\ell} \frac{n_i L_i^2 q^{L_i}}{1 - q^{L_i}}. \quad (12)$$

As the observations are independent, the Fisher information is the sum of the corresponding contributions for each frame length L_i , weighted by the number of observations n_i . The dominant individual term in (12) depends on the actual value of the parameter: for very small p , the term with largest frame length is dominant, but eventually the situation is reversed as p increases. Note that $I(q)$ increases without bounds as $p \rightarrow 0$, and goes to zero as $p \rightarrow 1$. This is due to the fact that we can only observe if a packet is in error or not, but not how many errors there are in an erroneous packet. For $p \ll \frac{1}{L_i}$, multiple bit errors within a packet become unlikely, so that the BER becomes “easier” to estimate from the observations. When the probability of multibit error events is not negligible, however, estimating the BER based on observation of packet errors becomes much more difficult.

The CRLB, which bounds the variance of any unbiased estimator \hat{q} , is $\mathbb{V}\text{ar}[\hat{q}] \geq I^{-1}(q)$. In view of (1), and following [27, Sec. 3.6], the Normalized CRLB (NCRLB) of any unbiased FER estimate for packets of length \tilde{L} follows:

$$\frac{\mathbb{V}\text{ar}[\hat{\theta}_{\tilde{L}}]}{\theta_{\tilde{L}}^2} \geq \frac{\left(\frac{\tilde{L} q^{\tilde{L}}}{1 - q^{\tilde{L}}} \right)^2}{\sum_{i=1}^{\ell} \frac{n_i L_i^2 q^{L_i}}{1 - q^{L_i}}}. \quad (13)$$

It is insightful to examine the asymptotic behavior of (13). On the one hand, as $p \rightarrow 0$, using $1 - (1 - p)^L \approx pL$ in (13) yields the following low-BER approximation of the NCRLB:

$$\frac{\mathbb{V}\text{ar}[\hat{\theta}_{\tilde{L}}]}{\theta_{\tilde{L}}^2} \geq \frac{1}{p \sum_{i=1}^{\ell} n_i L_i}, \quad (14)$$

which is independent of \tilde{L} . On the other hand, for $p \rightarrow 1$, the corresponding asymptote is

$$\frac{\text{Var}[\hat{\theta}_{\tilde{L}}]}{\theta_{\tilde{L}}^2} \geq \frac{\tilde{L}^2}{n_1 L_1^2} (1-p)^{2\tilde{L}-L_1}. \quad (15)$$

The behavior of (15) clearly depends on the target length \tilde{L} . If $\tilde{L} > \frac{L_1}{2}$, then the NCRLB goes to zero as $p \rightarrow 1$; it settles to a constant when $\tilde{L} = \frac{L_1}{2}$, and goes to infinity if $\tilde{L} < \frac{L_1}{2}$. Note that, in the extreme case of packets with length $\tilde{L} = 1$ bit, the FER equals the BER and, from the discussion following (12), we know that the Fisher information for the problem of BER estimation goes to zero as $p \rightarrow 1$. Thus, it is reasonable to observe such behavior for short lengths \tilde{L} .

As an example, Fig. 1 shows the NCRLB (13) for a setting with $\ell = 3$, $\{L_1, L_2, L_3\} = \{1000, 2000, 3000\}$, for different target lengths, and with a total number of observations $n_1 + n_2 + n_3 = 60$ distributed in three different ways. It is clear that estimating the FER for packets whose length is much shorter than those for which observations are available is considerably difficult when the BER is high. In such regime, it is preferable to have more observations of shorter frames, as indicated by (15). This example also illustrates the fact that the NCRLB need not be a monotonic function of the BER.

B. Maximum Likelihood estimation via bisection

The MLE of q is obtained by maximizing the LLF (9): $\hat{q}_{\text{ML}} = \arg \max_{q \in [0,1]} \mathcal{L}(q)$. Consider first the following two particular cases. If $\mathbf{m} = \mathbf{0}$, then the MLE is readily obtained from (9) as $\hat{q}_{\text{ML}} = 1$; moreover, if $\hat{q}_{\text{ML}} = 1$ then $\mathbf{m} = \mathbf{0}$, since otherwise $\mathcal{L}(1) = -\infty$ cannot be the maximum value. Analogously, $\hat{q}_{\text{ML}} = 0$ iff $\mathbf{m} = \mathbf{n} \triangleq [n_1, \dots, n_\ell]^T$.

Therefore, let us assume that $\mathbf{m} \neq \mathbf{0}$ and $\mathbf{m} \neq \mathbf{n}$, so that $\hat{q}_{\text{ML}} \in (0, 1)$. Since (9) is continuous in q , we can obtain its maximum by finding the roots of its derivative. By rearranging the terms in (10) and performing some elemental operations, we arrive at

$$\mathcal{L}'(q) = -\frac{1}{q} \sum_{i=1}^{\ell} \left[\frac{m_i L_i}{1 - q^{L_i}} - n_i L_i \right], \quad (16)$$

so the MLE is a solution of

$$\Psi(q) \triangleq \sum_{i=1}^{\ell} \frac{m_i L_i}{1 - q^{L_i}} - \sum_{j=1}^{\ell} n_j L_j = 0. \quad (17)$$

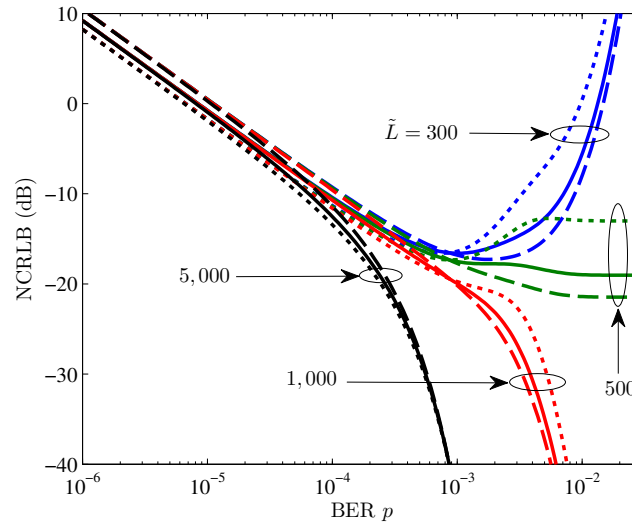


Fig. 1: Normalized CRLB (13) for FER estimation vs. BER. $\{\tilde{L}_1, \tilde{L}_2, \tilde{L}_3\} = \{1000, 2000, 3000\}$. Solid: $\{n_1, n_2, n_3\} = \{20, 20, 20\}$. Dashed: $\{n_1, n_2, n_3\} = \{35, 20, 5\}$. Dotted: $\{n_1, n_2, n_3\} = \{5, 20, 35\}$.

Although the roots of Ψ cannot be obtained in closed form (unless $\ell = 1$), it is possible to characterize them by means of the following proposition.

Theorem 1: Assume that $\mathbf{m} \neq \mathbf{0}$ and $\mathbf{m} \neq \mathbf{n}$. Then the MLE \hat{q}_{ML} is the unique root of Ψ in $(0, 1)$, and can be obtained by application of the bisection algorithm in the interval $[0, 1]$.

Proof: As $\hat{q}_{\text{ML}} \in (0, 1)$, we examine the behavior of Ψ in this interval. Note that, since $m_i \leq n_i \forall i$ and $\mathbf{m} \neq \mathbf{n}$, one has

$$\Psi(0) = \sum_{i=1}^{\ell} L_i (m_i - n_i) < 0. \quad (18)$$

Also, $\Psi(1) = +\infty$; therefore, since Ψ is continuous, there exists at least one root in $(0, 1)$. The derivative of Ψ satisfies

$$\Psi'(q) = \sum_{i=1}^{\ell} \frac{q^{L_i-1} L_i^2 m_i}{(1 - q^{L_i})^2} > 0, \quad 0 < q < 1, \quad (19)$$

since $\mathbf{m} \neq \mathbf{0}$. Hence, $\Psi(q)$ is strictly monotonically increasing in $(0, 1)$, so that only one root exists in this interval, which can be found by applying the bisection method². ■

²The fact that $\Psi(1) = +\infty$ should not lead to numerical problems: since the sign of $\Psi(q)$ at $q = 0$ and $q = 1$ is known, there is no need to evaluate the function at these points.

Since $q \in [0, 1]$, the number of bisection iterations required to obtain an estimate \hat{p} satisfying $|\hat{p} - p| < \delta$ is $\log_2 \frac{1}{\delta}$. In the following sections, we present results for particular cases for which simpler FER estimators can be found.

C. Low Bit Error Rate regime

The intricate formulation of the MLE is due to the nonlinear dependency of the FER θ_L on the BER p (or, equivalently, on $q = 1 - p$). We now explore suboptimal alternatives in the low BER regime, which is of particular interest since most practical systems operate with low target FER values in the range 0.01–0.1 [28], corresponding to even lower BER values.

First, we develop the asymptotic expression of the MLE in the low-BER regime. Let

$$\Phi(q) \triangleq (1 - q)\Psi(q) \quad (20)$$

$$= \sum_{i=1}^{\ell} m_i L_i \frac{1 - q}{1 - q^{L_i}} - (1 - q) \sum_{i=1}^{\ell} n_i L_i, \quad (21)$$

so that the MLE is the root of $\Phi(q)$ in $q \in (0, 1)$. If we let z be a random variable denoting the total number of bits in error in all received frames, it can be readily shown that (21) satisfies

$$\Phi(q) = \mathbb{E}[z | \mathbf{m}] - \mathbb{E}[z], \quad (22)$$

and therefore the MLE is such that the *a priori* and *a posteriori* expected values of the total number of erroneous bits are equal.

Consider the first-order Taylor approximation of $\Phi(q)$ about $q = 1$, $\Phi(q) \approx \Phi(1) - \Phi'(1)(1 - q)$. Then the MLE approximately satisfies $\Phi(1) - \Phi'(1)\hat{p} = 0$. Using the fact that $\frac{1 - q^L}{1 - q} = \sum_{n=0}^{L-1} q^n$, one finds that

$$\Phi(1) = \sum_{i=1}^{\ell} m_i, \quad \Phi'(1) = \sum_{i=1}^{\ell} \left(n_i L_i - m_i \frac{L_i - 1}{2} \right), \quad (23)$$

and therefore, asymptotically as $p \rightarrow 0$, the MLE is given by

$$\hat{p} = \frac{\sum_{i=1}^{\ell} m_i}{\sum_{i=1}^{\ell} n_i L_i - \frac{1}{2} \sum_{i=1}^{\ell} m_i (L_i - 1)}. \quad (24)$$

An alternative estimator in the low-BER regime can be developed as follows. A first-order approximation of (1) around $p = 0$ yields $\theta_L = 1 - q^L \approx pL$, which amounts to neglecting the fact that bit errors are not mutually exclusive (i.e., multiple bits can be in error in the same packet):

$$\theta_L = \mathbb{P} \left[\bigcup_{i=1}^L (b_i \neq \hat{b}_i) \right] \approx \sum_{i=1}^L \mathbb{P} [b_i \neq \hat{b}_i] = pL. \quad (25)$$

Now, using $1 - q^{L_i} \approx pL_i$ in (17) yields

$$\hat{p} = \frac{\sum_{i=1}^{\ell} m_i}{\sum_{i=1}^{\ell} n_i L_i}. \quad (26)$$

This estimator divides the total number of frame errors by the total number of transmitted bits, in line with the low-BER assumption that neglects multi-bit errors within a frame. Observe that the estimators (24) and (26) differ only by the second term in the denominator of (24).

Note that (26) is linear in the observations. Its bias and variance are readily found:

$$p - \mathbb{E}[\hat{p}] = p - \frac{\sum_{i=1}^{\ell} n_i (1 - q^{L_i})}{\sum_{i=1}^{\ell} n_i L_i}, \quad (27)$$

$$\text{Var}[\hat{p}] = \frac{\sum_{i=1}^{\ell} n_i (1 - q^{L_i}) q^{L_i}}{\left(\sum_{i=1}^{\ell} n_i L_i\right)^2}. \quad (28)$$

When $\ell = 1$, (26) reduces to $\hat{p} = \frac{m_1}{n_1 L_1}$, and in fact, if we define $\hat{p}_i \triangleq \frac{m_i}{n_i L_i}$, i.e., the corresponding estimate based on the observations from length- L_i packets only, then (26) can be rewritten as

$$\hat{p} = \frac{\sum_{i=1}^{\ell} n_i L_i \hat{p}_i}{\sum_{i=1}^{\ell} n_i L_i}, \quad (29)$$

which is a convex combination of the ℓ individual estimators \hat{p}_i , with weights proportional to the number of observed bits $n_i L_i$. This suggests one potential means to extend the range of applicability of (29) to include larger values of p by substituting \hat{p}_i in (29) with the MLE based on the observations of length- L_i packets only, which is given by $\hat{p}_{\text{ML},i} \triangleq 1 - \left(1 - \frac{m_i}{n_i}\right)^{1/L_i}$:

$$\hat{p} = \frac{\sum_{i=1}^{\ell} n_i L_i \hat{p}_{\text{ML},i}}{\sum_{i=1}^{\ell} n_i L_i}. \quad (30)$$

D. Large number of observations

Inspired by (29) and (30), in this section we propose a computationally efficient approach by which the FER $\theta_{\bar{L}}$ is estimated as a linear combination of ℓ individual estimators, each obtained from the observations corresponding to packets of the same length. These individual estimators are chosen as the corresponding MLEs, which are given by $\hat{\theta}_{\bar{L}}(L_i) \triangleq 1 - \left(1 - \hat{\theta}_{L_i}\right)^{\bar{L}/L_i} = f_i(\hat{\theta}_{L_i})$, $i = 1, \dots, \ell$, with $\hat{\theta}_{L_i}$ as in (5), as can be seen by solving (17) when $\ell = 1$. Then, given a set of weights $\{\beta_i, i = 1, \dots, \ell\}$, the proposed estimator is given by

$$\hat{\theta}_{\bar{L}} = \sum_{i=1}^{\ell} \beta_i \hat{\theta}_{\bar{L}}(L_i). \quad (31)$$

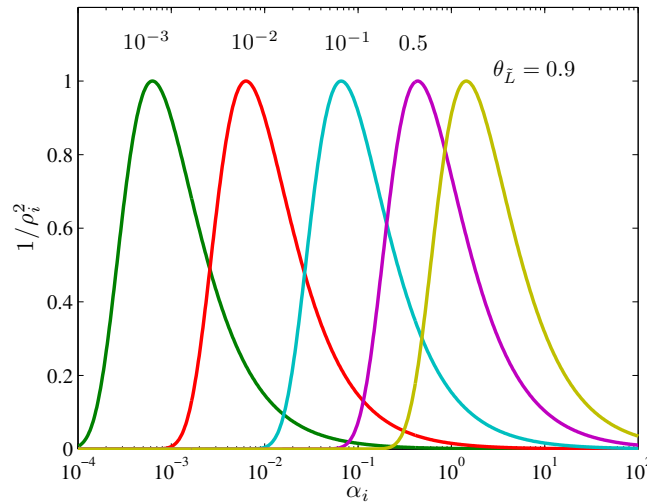


Fig. 2: Optimal weights as a function of α_i and for different FER values.

Ideally, one should select the weights β_i in order to optimize performance. This is difficult, however, because the bias and variance of $\hat{\theta}_{\tilde{L}}(L_i)$ depend on the unknown parameter. Next, we propose a method for weight selection based on the assumption that the number of observations $\{n_i, i = 1, \dots, \ell\}$ is sufficiently large. In that case, the MLEs $\hat{\theta}_{L_i}$ become (asymptotically) unbiased and normally distributed [27]: $\hat{\theta}_{L_i} \sim \mathcal{N}(\theta_{L_i}, \sigma_i^2)$, with σ_i^2 as in (6). Therefore, the proposed estimator (31) will be unbiased provided that $\sum_{i=1}^{\ell} \beta_i = 1$.

Under a large number of observations, the variance σ_i^2 will be small, so we can use the first-order approximation

$$\begin{aligned} \hat{\theta}_{\tilde{L}}(L_i) &\approx f_i(\theta_{L_i}) + f'_i(\theta_{L_i})(\hat{\theta}_{L_i} - \theta_{L_i}) \\ &= \theta_{\tilde{L}} + \alpha_i(1 - \theta_{L_i})^{\alpha_i-1}(\hat{\theta}_{L_i} - \theta_{L_i}), \end{aligned} \quad (32)$$

where $\alpha_i \triangleq \tilde{L}/L_i$. Then the variance $\rho_i^2 \triangleq \mathbb{V}\text{ar}[\hat{\theta}_{\tilde{L}}(L_i)]$ asymptotically becomes

$$\begin{aligned} \rho_i^2 &= \alpha_i^2(1 - \theta_{L_i})^{2\alpha_i-2} \sigma_i^2 \\ &= \frac{\alpha_i^2}{n_i}(1 - \theta_{\tilde{L}})^2 \left(\frac{1}{(1 - \theta_{\tilde{L}})^{\frac{1}{\alpha_i}}} - 1 \right). \end{aligned} \quad (33)$$

Then, the weights minimizing the asymptotic variance of $\hat{\theta}_{\tilde{L}}$ in (31) under the constraint

$\sum_{i=1}^{\ell} \beta_i = 1$ are given by

$$\beta_i = \frac{\frac{1}{\rho_i^2}}{\sum_{j=1}^{\ell} \frac{1}{\rho_j^2}}, \quad i = 1, \dots, \ell. \quad (34)$$

Fig. 2 shows the optimal weights $1/\rho_i^2$ as a function of α_i and for different values of the FER $\theta_{\tilde{L}}$ (normalized by their peak value for each $\theta_{\tilde{L}}$), assuming n_i is the same for all i . The shape of these curves is due to the following: the transmissions of very short packets (small L_i , i.e., large α_i) are very likely to result in no errors, whereas those of very long packets (large L_i , i.e., small α_i) will likely result in a packet error *always*. Therefore, the information gathered by observing these events is low. Depending of the FER $\theta_{\tilde{L}}$, there is a “sweet spot” corresponding to the peak of the curves in Fig. 2, located at $\alpha_i \approx \log \frac{1}{\sqrt{1-\theta_{\tilde{L}}}}$ ($\approx \frac{\theta_{\tilde{L}}}{2}$ for small $\theta_{\tilde{L}}$). Since this is not necessarily near $\alpha_i = 1$, simply assigning more weight to observations with similar length $L_i \approx \tilde{L}$ can result in significant performance degradation. Note that the peak gradually shifts toward longer packets as the FER becomes smaller.

Since the optimum weights depend on the unknown parameter $\theta_{\tilde{L}}$, they cannot be directly computed. We propose a two-step procedure, in which an initial estimate $\hat{\theta}_{\tilde{L},0}$ as in (31) with weights $\beta_i = 1/\ell$ is first obtained. Then, this estimate $\hat{\theta}_{\tilde{L},0}$ is substituted in (33) in lieu of the true value $\theta_{\tilde{L}}$ to compute approximate values of ρ_i^2 , which are then used to obtain the weights β_i according to (34) for the final estimator³.

E. Simulation results

We evaluate the performance of the proposed schemes via Monte Carlo simulation⁴, in a setting with $\ell = 5$ in which observations of packets with lengths $\{L_1, \dots, L_5\} = \{100, 1000, 5000, 8000, 10000\}$ are available. Two different target lengths are considered: $\tilde{L} = 2000$ and $\tilde{L} = 50$, corresponding to long and short packets, respectively. We assume the same number of observations for each length, and consider two cases for each value of \tilde{L} : $n_i = 10$ and $n_i = 1000$. The total number of observed packets is therefore 50 and 5000, respectively.

³Although in principle this reweighting process could be further iterated, in practice the variation in performance after the first iteration seems to be negligible.

⁴For each BER value, a total of 20,000 Monte Carlo runs were executed.

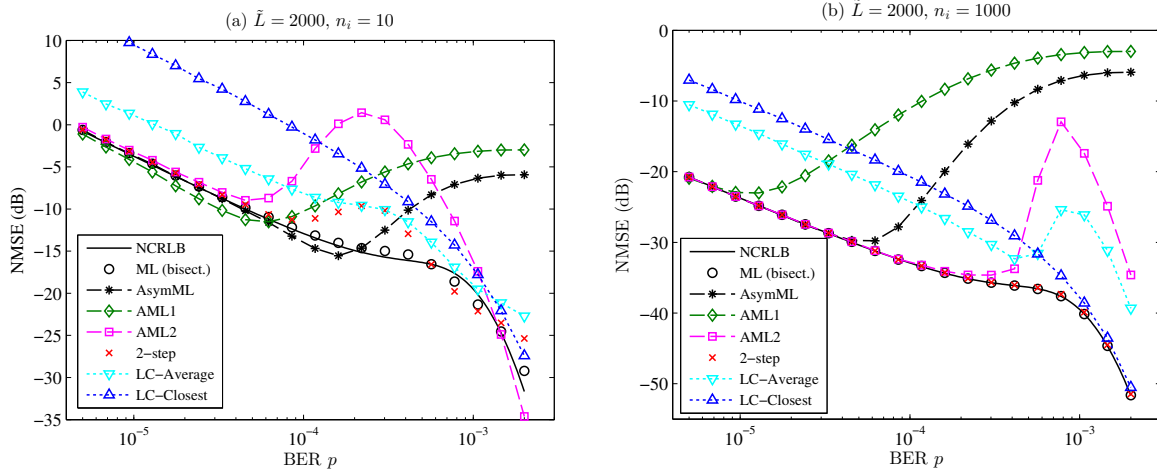


Fig. 3: NMSE for the different estimators vs. BER. $\tilde{L} = 2000$, $\ell = 5$. (a) $n_i = 10$, (b) $n_i = 1000$.

Figs. 3 and 4 show the CRLB (13) and the Normalized Mean Squared Error (NMSE) $\mathbb{E}[(\hat{\theta}_{\tilde{L}} - \theta_{\tilde{L}})^2]/\theta_{\tilde{L}}^2$ of the MLE (obtained via bisection) as a function of the BER p , together with that of several other estimators, namely:

- The asymptotic MLE for low BER (24) (*AsymML*);
- The MLE approximation (26) (*AML1*);
- The convex combination of MLEs (30) (*AML2*);
- The two-step estimator based on the asymptotically Gaussian property of the MLE, with weights as in (34) (*2-Step*);
- The convex combination (31) with weights $\beta_i = 1/\ell$ for all i (*LC-Average*);
- The convex combination (31) with weights $\beta_i = 1$ if $i = \arg \min_j |L_j - \tilde{L}|$, and 0 otherwise. (*LC-Closest*).

The true MLE presents a small gap to the CRLB for $n_i = 10$, but for $n_i = 1000$ it is efficient over the considered BER range. The performance of its low-BER approximations *AsymML* (24), *AML1* (26) and *AML2* (30) is close to optimal for small p as expected, but degrades significantly as p increases. With few observations, *AsymML* consistently outperforms both *AML1* and *AML2*; whereas with a larger number of observations, the BER range over which *AML2* performs close to the CRLB is significantly larger than those of *AsymML* and *AML1*. For $n_i = 10$, *AsymML* and *AML1* present some bias as evidenced by the fact that their NMSE

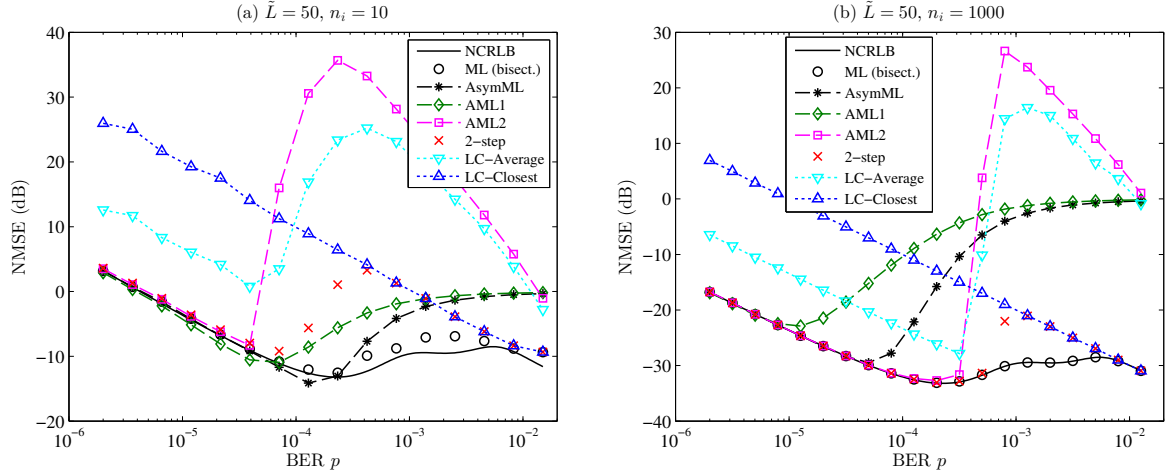


Fig. 4: NMSE for the different estimators vs. BER. $\tilde{L} = 50$, $\ell = 5$. (a) $n_i = 10$, (b) $n_i = 1000$.

is slightly below the NCRLB⁵.

Regarding the convex combination-based estimators, the simple *Average* and *Closest* approaches show a significant gap with respect to the CRLB, although *Closest* becomes close to efficient for large FER values (note that with $\tilde{L} = 2000$, a BER of $p = 10^{-3}$ translates into a FER value of $\theta_{\tilde{L}} = 0.865$, whereas for $\tilde{L} = 50$, a BER of $p = 10^{-2}$ yields $\theta_{\tilde{L}} = 0.395$). The reason is that in this setting, *Closest* picks the observations with $L_i = 1000$ and $L_i = 100$ for $\tilde{L} = 2000$ and $\tilde{L} = 50$ respectively, yielding $\alpha_i = 2$ and $\alpha_i = 0.5$, and in view of Fig. 2 these weights are close to the optimal ones when $\theta_{\tilde{L}}$ is large. The poor performance of *Average* is due to the large disparity of available packet lengths ($L_{\ell}/L_1 = 20$ dB), because in such situations the optimal linear combination weights will be far from being uniform as evidenced by the shape of the curves in Fig. 2.

The two-step estimator shows a good tradeoff between performance and complexity as long as the number of observations per length is sufficiently large: for example, it is seen to achieve efficiency for $\tilde{L} = 2000$, $n_i = 1000$. However, with a small number of observations per length, the Gaussian approximation of the individual MLEs underlying the derivation of this estimator is no longer valid, and its performance may be poor in such cases. This is clearly seen for $n_i = 10$ in Figs. 3 and 4, and is further illustrated by the following example. Suppose that $\ell = 100$, and

⁵Recall that the Cramer-Rao bound applies to the variance of *unbiased* estimators only.

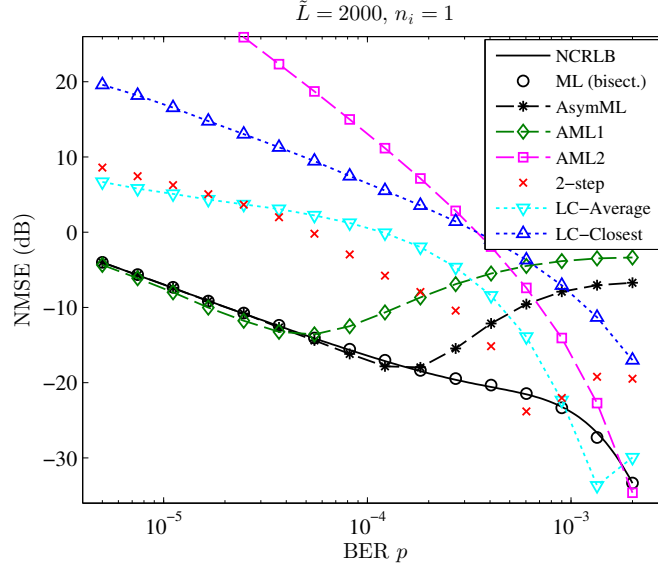


Fig. 5: NMSE for different estimators vs. BER. $\tilde{L} = 2000$, $\ell = 100$.

that we observe a total of 100 packets, with lengths $\{100, 200, \dots, 9900, 10000\}$ (thus, there is a *single* packet for each length, i.e., $n_i = 1$ for all i). Results for $\tilde{L} = 2000$ in this setting are shown in Fig. 5. The true MLE found via bisection remains efficient in this case. This example shows the importance of carefully choosing an appropriate estimator depending on the particular scenario.

IV. FER ESTIMATION WITH FEEDBACK ERRORS

So far we have assumed that it is possible to perfectly discriminate whether a packet was correctly decoded or not. In practice, however, this need not be the case when estimation is carried out either by the transmitter (rather than the receiver), or by a third node observing the network packet exchange. In that case, knowledge of successful/unsuccessful decoding is obtained by means of a feedback channel: the receiver sends a positive acknowledgement (ACK) each time a packet is correctly decoded, and a negative acknowledgement (NAK) otherwise. This information may be affected by channel errors. The feedback channel is modeled as a BSC with error probability ϵ :

$$\epsilon \triangleq \mathbb{P}[\text{rx ACK} | \text{tx NAK}] = \mathbb{P}[\text{rx NAK} | \text{tx ACK}]. \quad (35)$$

We assume that ϵ is unknown and does not depend on the packet length L_i , as ACK/NAK signaling is usually constant-length. Throughout this section, we denote by m_i the number of received NAKs for packets of length L_i (which reduces to the original definition of m_i in Sec. II when $\epsilon = 0$).

Let ξ_i be the probability of getting a NAK after transmission of a packet of length L_i . Then,

$$\xi_i = (1 - \epsilon)\theta_{L_i} + \epsilon(1 - \theta_{L_i}). \quad (36)$$

From (36), ξ_i is a convex combination of θ_{L_i} and $1 - \theta_{L_i}$; in the absence of errors, $\xi_i = \theta_{L_i}$. The observations m_i are binomially distributed with parameters n_i and ξ_i , and therefore

$$f(\mathbf{m}; \mathbf{v}) = \prod_{i=1}^{\ell} \binom{n_i}{m_i} \xi_i^{m_i} (1 - \xi_i)^{(n_i - m_i)}, \quad (37)$$

where $\mathbf{v} \triangleq [q \ \epsilon]^T$ is the vector of unknown parameters.

Ignoring the fact that the feedback channel is imperfect may significantly degrade performance. To see this, consider the simplest scenario in which the FER for a given length $\tilde{L} = L_1$ is to be estimated from observations of that same length. In the absence of feedback errors, the MLE of θ_{L_1} is the empirical FER (5), i.e., $\hat{\theta}_{L_1} = \frac{m_1}{n_1}$, which is unbiased and efficient. However, once feedback errors are present, this estimator becomes biased, with normalized bias $(\mathbb{E}[\hat{\theta}_{L_1}] - \theta_{L_1})/\theta_{L_1} = \left(\frac{1}{\theta_{L_1}} - 2\right)\epsilon$, which is worse for smaller FER values.

A. CRLB

The Fisher information matrix (FIM) $\mathbf{I}(\mathbf{v})$ has elements $\mathbf{I}_{ij}(\mathbf{v}) = -\mathbb{E} \left[\frac{\partial^2 \log f(\mathbf{m}; \mathbf{v})}{\partial v_i \partial v_j} \right]$, and boils down to

$$\mathbf{I}(\mathbf{v}) = \sum_{i=1}^{\ell} \frac{n_i}{\xi_i (1 - \xi_i)} \mathbf{u}_i \mathbf{u}_i^T, \quad \mathbf{u}_i \triangleq \begin{bmatrix} (1 - 2\epsilon) L_i q^{L_i - 1} \\ 1 - 2q^{L_i} \end{bmatrix}. \quad (38)$$

If $\ell = 1$ then \mathbf{I} is singular, i.e., the parameter \mathbf{v} is not identifiable [29]. This is because in that case, (37) depends on q and ϵ through ξ_1 alone, and different pairs (q, ϵ) can result in the same value of ξ_1 . Note also that for $\epsilon = 0$, the term $\mathbf{I}_{11}(\mathbf{v})$ reduces to (12). From (38), the NCRLB on the variance of any unbiased estimator of $\theta_{\tilde{L}}$ is seen to be

$$\frac{\text{Var}[\hat{\theta}_{\tilde{L}}]}{\theta_{\tilde{L}}^2} \geq \left(\frac{\tilde{L} q^{\tilde{L} - 1}}{1 - q^{\tilde{L}}} \right)^2 \frac{\mathbf{I}_{22}(\mathbf{v})}{\mathbf{I}_{11}(\mathbf{v}) \mathbf{I}_{22}(\mathbf{v}) - \mathbf{I}_{12}^2(\mathbf{v})}. \quad (39)$$

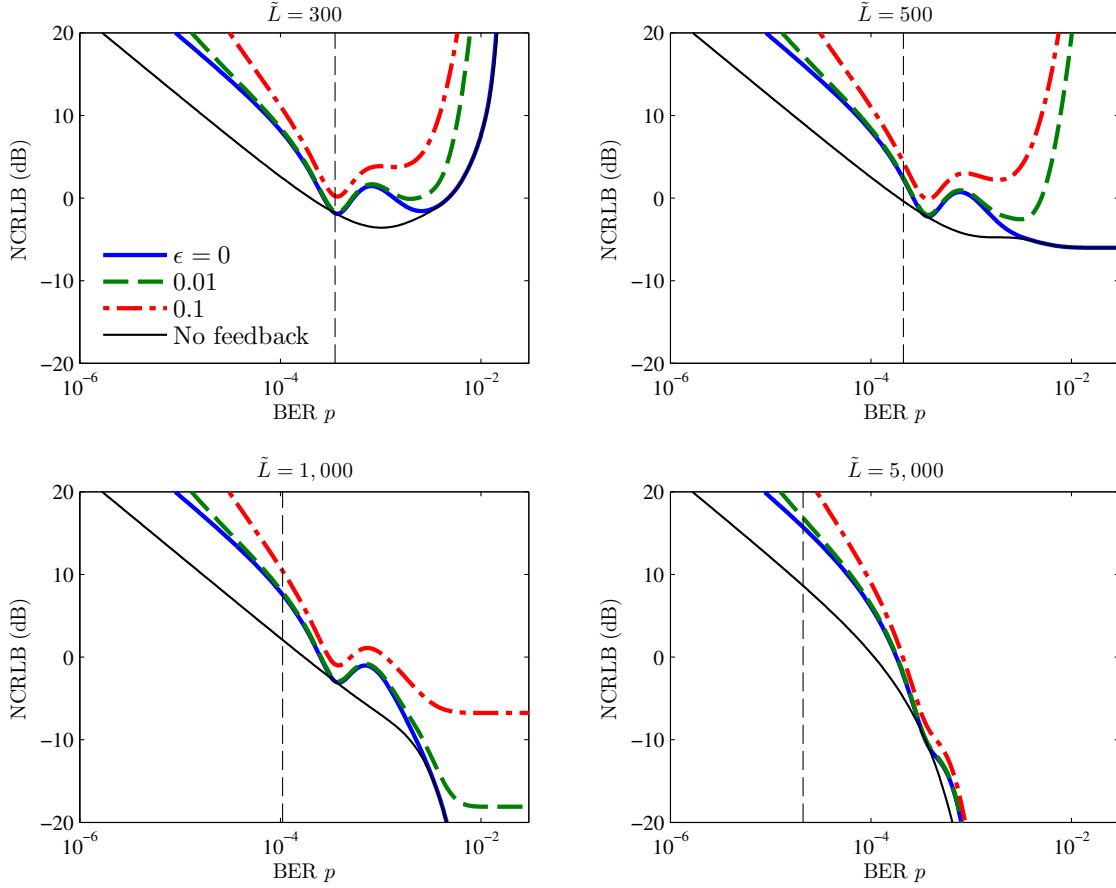


Fig. 6: Normalized CRLB for FER estimation with imperfect feedback channel vs. BER. $\{L_1, L_2, L_3\} = \{1000, 2000, 3000\}$, $n_1 = n_2 = n_3 = 1$. The curve labeled as 'No feedback' corresponds to the case in which ϵ is *known* and equal to zero. Vertical lines indicate the BER values corresponding to an FER $\theta_{\tilde{L}} = 0.1$ for each length \tilde{L} .

Note that if $\epsilon = \frac{1}{2}$ then $\mathbf{I}_{11}(\mathbf{v}) = \mathbf{I}_{12}(\mathbf{v}) = 0$, so that (39) goes to infinity; in that case, the input and output of the feedback channel become statistically independent, so that the received ACK/NAK do not carry any information about packet errors. Moreover, (39) does not change if ϵ is replaced by $1 - \epsilon$, as it should be since the error probability of a BSC can be changed from ϵ to $1 - \epsilon$ by deterministically flipping the channel output.

Fig. 6 shows the NCRLB (39) for the same setting as that of Fig. 1. The NCRLB (13), which applies when estimation is directly performed at the receiver, is also shown for comparison. A number of observations can be made:

- 1) The degradation incurred by reporting through a feedback channel with unknown but small error rate can be quantified by the ratio of the NCRLB (39) with $\epsilon = 0$ to the NCRLB (13). This ratio is larger than 1 and is given by $\frac{1}{1-r}$, with

$$r \triangleq \frac{\mathbf{I}_{12}^2(\mathbf{v})}{\mathbf{I}_{11}(\mathbf{v})\mathbf{I}_{22}(\mathbf{v})} \Big|_{\epsilon=0} \in [0, 1]. \quad (40)$$

Therefore, larger values of r imply larger degradation due to feedback errors. From (38), it is straightforward to show that, with $n \triangleq \sum_{i=1}^{\ell} n_i$,

$$\lim_{p \rightarrow 1} r = 0, \quad \lim_{p \rightarrow 0} r = \frac{\left(\sum_{i=1}^{\ell} \frac{n_i}{n} \frac{1}{L_i} \right)^{-1}}{\sum_{i=1}^{\ell} \frac{n_i}{n} L_i}, \quad (41)$$

meaning that, on the one hand, as the BER increases, the two bounds tend to agree; and on the other hand, for low BER values the degradation is quantified by the ratio of the (weighted) harmonic to arithmetic means of the lengths L_i , with weights $\frac{n_i}{n}$, $i = 1, \dots, \ell$. If the lengths L_i are close to each other, this ratio will be close to 1, resulting in larger degradation. For the setting of Fig. 6, one has $\lim_{p \rightarrow 0} r \approx 0.8182$, so that the degradation in the bound for low BER is ≈ 7.4 dB.

- 2) For sufficiently small p , the NCRLB is independent of the frame length \tilde{L} ; this is because $\frac{\tilde{L}q^{\tilde{L}-1}}{1-q^{\tilde{L}}}$ goes to $\frac{q}{1-q}$ as $p \rightarrow 0$ for all \tilde{L} . In this regime, the effect of errors in the feedback channel is particularly detrimental: whereas the NCRLB behaves as $O(p^{-1})$ for small p when $\epsilon = 0$, as soon as errors are present in the feedback channel this asymptotic behavior becomes $O(p^{-2})$.
- 3) When $\epsilon > 0$, the NCRLB asymptote for $p \rightarrow 1$ is

$$\frac{\text{Var}[\hat{\theta}_{\tilde{L}}]}{\theta_{\tilde{L}}^2} \geq \frac{\epsilon(1-\epsilon)}{(1-2\epsilon)^2} \frac{n\tilde{L}^2}{(n-n_1)n_1L_1^2} (1-p)^{2(\tilde{L}-L_1)}, \quad (42)$$

which goes to zero, a constant, or infinity as $p \rightarrow 1$ for $\tilde{L} > L_1$, $\tilde{L} = L_1$, and $\tilde{L} < L_1$ respectively. This is clearly seen in Fig. 6 (in that setting, $L_1 = 1000$).

B. ML estimation via E-M algorithm

Maximizing (37) with respect to $\mathbf{v} = [q \ \epsilon]^T$ is not possible in closed form, so we propose to apply the E-M algorithm [30], [31] to this task. To this end, we introduce the variables $\{m_{ij}, j = 1, \dots, n_i, i = 1, \dots, \ell\}$, such that $m_{ij} = 1$ if a NAK is received for the j -th packet

TABLE I: E-M algorithm for FER estimation with feedback errors.

Initialize $0 < \hat{\epsilon}_1 \ll 1$ and $\hat{q}_1 = 1 - \frac{\sum_{i=1}^{\ell} m_i}{\sum_{j=1}^{\ell} n_j L_j - \frac{1}{2} \sum_{j=1}^{\ell} m_j (L_j - 1)}$.

For $t = 1, 2, \dots$

- 1) *E-step*: For $i = 1, \dots, \ell$, compute
$$\hat{\rho}_{i,t}^0 = \frac{\hat{\epsilon}_t(1 - \hat{q}_t^{L_i})}{\hat{\epsilon}_t(1 - \hat{q}_t^{L_i}) + (1 - \hat{\epsilon}_t)\hat{q}_t^{L_i}}, \quad \hat{\rho}_{i,t}^1 = \frac{(1 - \hat{\epsilon}_t)(1 - \hat{q}_t^{L_i})}{(1 - \hat{\epsilon}_t)(1 - \hat{q}_t^{L_i}) + \hat{\epsilon}_t\hat{q}_t^{L_i}}. \quad (44)$$
- 2) *M-step*: Obtain new estimates as
$$\hat{\epsilon}_{t+1} = \frac{\sum_{i=1}^{\ell} (m_i(1 - \hat{\rho}_{i,t}^1) + (n_i - m_i)\hat{\rho}_{i,t}^0)}{\sum_{j=1}^{\ell} n_j}, \quad (45)$$

$$\hat{q}_{t+1} = 1 - \frac{1 - \hat{q}_t}{\sum_{j=1}^{\ell} n_j L_j} \sum_{i=1}^{\ell} \frac{L_i (m_i \hat{\rho}_{i,t}^1 + (n_i - m_i) \hat{\rho}_{i,t}^0)}{1 - \hat{q}_t^{L_i}}. \quad (46)$$

After convergence, set $\hat{\theta}_{\bar{L}} = 1 - \hat{q}^{\bar{L}}$.

of length L_i , and $m_{ij} = 0$ otherwise. In that way, one has $m_i = \sum_{j=1}^{n_i} m_{ij}$. We also introduce the variables $\{z_{ij}, j = 1, \dots, n_i, i = 1, \dots, \ell\}$ such that $z_{ij} \in \{1, \dots, L_i\}$ denotes the number of bit errors actually taking place in the reception of the j -th packet of length L_i . Note that m_{ij} is a Bernoulli random variable with parameter ξ_i as in (36), whereas z_{ij} is binomially distributed with parameters L_i and p .

Clearly, estimation of the FER $\theta_{\bar{L}}$ is hindered by lack of knowledge about the z_{ij} 's. Therefore, we regard $\{m_{ij}\}$ as the *incomplete observations*, and $\{m_{ij}, z_{ij}\}$ as the *complete data*. Then the E-M algorithm is based on the iterative application of the following two steps:

- 1) *E-step*: Given the current estimate $\hat{\mathbf{v}}_t = [\hat{q}_t \ \hat{\epsilon}_t]^T$, compute the conditional expectation of the LLF

$$Q(\mathbf{v}, \hat{\mathbf{v}}_t) \triangleq \mathbb{E}[\log f(\bar{\mathbf{m}}, \bar{\mathbf{z}}; \mathbf{v}) | \bar{\mathbf{m}}; \hat{\mathbf{v}}_t], \quad (43)$$

where $\bar{\mathbf{m}}$ and $\bar{\mathbf{z}}$ are vectors of size $n = \sum_{i=1}^{\ell} n_i$ comprising $\{m_{ij}\}$ and $\{z_{ij}\}$, respectively, and \mathbf{v} denotes a trial value.

- 2) *M-step*: Obtain the next estimate as $\hat{\mathbf{v}}_{t+1} = \arg \max_{\mathbf{v}} Q(\mathbf{v}, \hat{\mathbf{v}}_t)$.

The final form of the algorithm is summarized in Table I. The detailed development is given in Appendix A. Although not obvious from the final expressions (45)-(46), it follows from the derivation in Appendix A that (i) the estimate $\hat{\epsilon}_{t+1}$ in (45) is given by the ratio of a *soft*

estimate of the number of errors in the feedback channel to the total number of transmissions $n = \sum_{j=1}^{\ell} n_j$; (ii) the BER estimate $1 - \hat{q}_{t+1}$ in (46) is obtained as the ratio of a soft estimate of the number of bits in error in all n packets to the total number of bits $\sum_{j=1}^{\ell} n_j L_j$ in these packets. As indicated in Table I, in order to initialize the iteration we suggest to set $\hat{\epsilon}_1$ to a small positive value, and to use the asymptotic MLE in the low BER regime with no feedback errors, given by (24). Incidentally, note that if we set $\hat{\epsilon}_1 = 0$, then it follows from Table I that $\hat{\epsilon}_t = 0$ for all t , whereas

$$\hat{q}_{t+1} = 1 - \frac{1 - \hat{q}_t}{\sum_{j=1}^{\ell} n_j L_j} \sum_{i=1}^{\ell} \frac{m_i L_i}{1 - \hat{q}_t^{L_i}}. \quad (47)$$

It is readily checked that any fixed point of the iteration (47) must satisfy (17), so that (47) provides an alternative means to numerically compute the MLE assuming no feedback errors to the bisection method of Sec. III-B.

C. Simulation results

Consider a setting similar to that in Sec. III-E ($\ell = 5$, $\{L_1, \dots, L_5\} = \{100, 1000, 5000, 8000, 10000\}$), assuming $n_i = 100$ for each observed length, and with a target length $\tilde{L} = 2000$. We evaluate the performance of the estimators from Sec. III-E, which are oblivious to potential feedback errors, and also the E-M estimator from Table I, initialized with $\hat{\epsilon}_1 = 10^{-5}$ and run for 10^3 iterations. Two values of the error probability in the feedback channel were considered, namely $\epsilon = 0.01$ and $\epsilon = 0.1$.

The normalized bias $(\mathbb{E}[\hat{\theta}_{\tilde{L}}] - \theta_{\tilde{L}})/\theta_{\tilde{L}}$ of the different methods is shown in Fig. 7. As the fraction of feedback errors increases, the estimators which do not take this effect into account become significantly biased, particularly for low BER values. The E-M estimator, in contrast, is practically unbiased in this setting, except for a slight bias in the low BER region, which is reduced if the number of observations increases (this is not the case for the other estimators, which suffer from an irreducible bias). Fig. 8 shows the NMSE of the estimators. As expected, the E-M estimator is the only method which remains close to being efficient in the whole BER range. Although the performance of the bisection-based and 2-step estimators is acceptable in this setting if the feedback channel is sufficiently reliable ($\epsilon = 0.01$), the degradation is significant for $\epsilon = 0.1$.

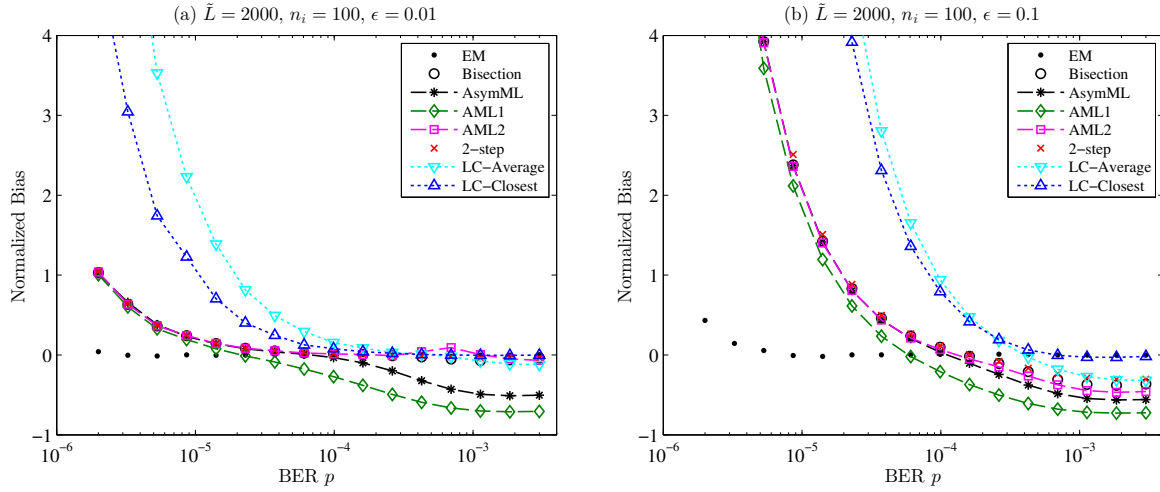


Fig. 7: Normalized bias for the different estimators vs. BER, with errors in the feedback channel.
 (a) $\epsilon = 0.01$, (b) $\epsilon = 0.1$.

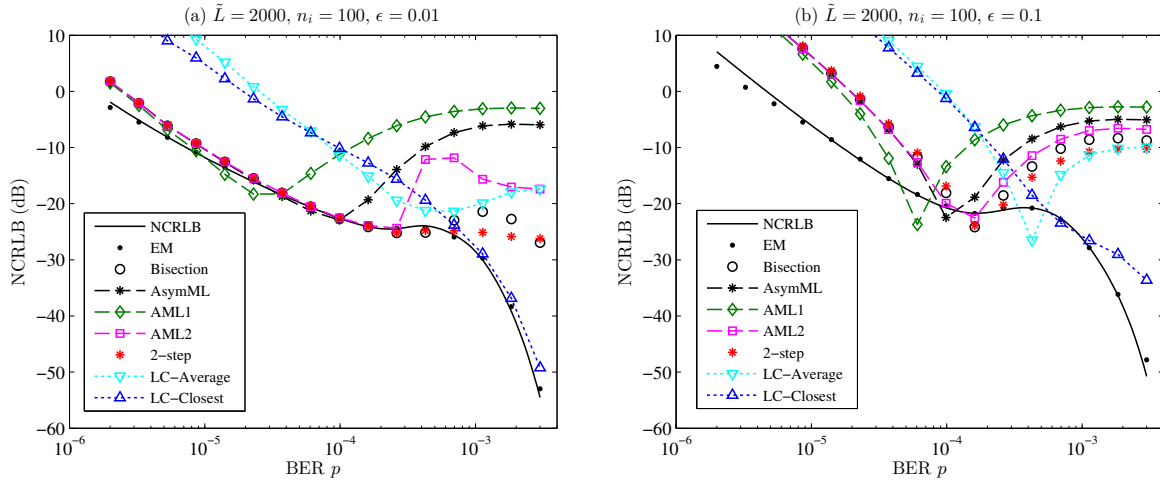


Fig. 8: Normalized MSE for the different estimators vs. BER, with errors in the feedback channel.

V. CONCLUSIONS

We have analyzed the problem of FER estimation from the observation of packet error events corresponding to the transmission of packets with disparate lengths; the set of observed packet lengths may not even include the target length. The CRLB for this problem was obtained and analyzed, and several estimators were proposed. Although the MLE admits no closed-form expression, it can be unambiguously obtained by numerical means. Alternatives to the MLE

based on low-BER and large sample size approximations were also presented, with varying performance depending on the setting. As a general rule, the MLE comes out as a good performance/complexity tradeoff.

The impact of having unreliable observations due to an imperfect feedback link has also been studied. The corresponding CRLB reveals the importance of having observations from packets with lengths as diverse as possible, and in fact when only measurements corresponding to a single length are available, the FER is not identifiable. Estimators that implicitly assume perfect measurements may degrade significantly if the error probability in the feedback link is sufficiently large. The E-M algorithm can be applied in this setting to obtain the MLE. The computational cost of this approach is relatively high, and thus developing simpler FER estimators robust to unreliable feedback remains an open issue. Extending the framework to systems employing forward error correction (FEC) encoders and to time-varying channels is also an interesting avenue for future work.

APPENDIX A

E-M ALGORITHM DERIVATION

Since observations are independent, the joint PMF $f(\bar{\mathbf{m}}, \bar{\mathbf{z}}; \mathbf{v})$ is given by

$$f(\bar{\mathbf{m}}, \bar{\mathbf{z}}; \mathbf{v}) = \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} f(m_{ij}, z_{ij}; \mathbf{v}). \quad (48)$$

Let us introduce the variable e_{ij} taking the value 1 if the j -th packet of length L_i is received with errors, and 0 otherwise. Since z_{ij} takes integer values, we can write $e_{ij} = 1 - \delta[z_{ij}]$, with $\delta[k]$ the unit impulse. Note that e_{ij} is a Bernoulli random variable with $\mathbb{P}[e_{ij} = 1; \mathbf{v}] = 1 - q^{L_i}$. Then we can write the joint PMF of m_{ij} and z_{ij} as

$$\begin{aligned} f(m_{ij}, z_{ij}; \mathbf{v}) &= f(m_{ij}, z_{ij} | e_{ij} = 0; \mathbf{v}) q^{L_i} \\ &\quad + f(m_{ij}, z_{ij} | e_{ij} = 1; \mathbf{v}) (1 - q^{L_i}). \end{aligned} \quad (49)$$

For a binomial random variable Z with parameters L and p , let us denote

$$B(k; L, p) \triangleq \mathbb{P}[Z = k] = \binom{L}{k} p^k (1 - p)^{L-k}, \quad 0 \leq k \leq L. \quad (50)$$

Then, the conditional PMFs in (49) are given by

$$f(m_{ij}, z_{ij} | e_{ij} = 0; \mathbf{v}) = (1 - \epsilon)^{1-m_{ij}} \epsilon^{m_{ij}} \delta[z_{ij}] \quad (51)$$

$$= (1 - \epsilon)^{1-m_{ij}} \epsilon^{m_{ij}} (1 - e_{ij}), \quad (52)$$

$$f(m_{ij}, z_{ij} | e_{ij} = 1; \mathbf{v}) = (1 - \epsilon)^{m_{ij}} \epsilon^{1-m_{ij}} \times \frac{B(z_{ij}; L_i, p)}{1 - B(0; L_i, p)} (1 - \delta[z_{ij}]) \quad (53)$$

$$= (1 - \epsilon)^{m_{ij}} \epsilon^{1-m_{ij}} \frac{B(z_{ij}; L_i, p)}{1 - q^{L_i}} e_{ij}. \quad (54)$$

Then (49) can be written as

$$f(m_{ij}, z_{ij}; \mathbf{v}) = (1 - \epsilon)^{1-m_{ij}} \epsilon^{m_{ij}} q^{L_i} (1 - e_{ij}) + (1 - \epsilon)^{m_{ij}} \epsilon^{1-m_{ij}} B(z_{ij}; L_i, p) e_{ij} \quad (55)$$

$$= [(1 - \epsilon)^{1-m_{ij}} \epsilon^{m_{ij}} q^{L_i}]^{1-e_{ij}} \times [(1 - \epsilon)^{m_{ij}} \epsilon^{1-m_{ij}} B(z_{ij}; L_i, p)]^{e_{ij}}. \quad (56)$$

Therefore, up to terms not depending on \mathbf{v} , the logarithm of (56) is given by

$$\begin{aligned} \log f(m_{ij}, z_{ij}; \mathbf{v}) &= [(1 - e_{ij})m_{ij} + e_{ij}(1 - m_{ij})] \log \epsilon \\ &+ [(1 - e_{ij})(1 - m_{ij}) + e_{ij}m_{ij}] \log(1 - \epsilon) \\ &+ [(1 - e_{ij})L_i + e_{ij}(L_i - z_{ij})] \log q \\ &+ e_{ij}z_{ij} \log(1 - q). \end{aligned} \quad (57)$$

Note that the random variable $x_{ij} \triangleq (1 - e_{ij})m_{ij} + e_{ij}(1 - m_{ij})$ takes the value 1 if $e_{ij} \neq m_{ij}$ (i.e., if an error event took place in the feedback channel), and 0 otherwise. Note also that

$e_{ij}z_{ij} = (1 - \delta[z_{ij}])z_{ij} = z_{ij}$. Hence,

$$\begin{aligned} \log f(\bar{\mathbf{m}}, \bar{\mathbf{z}}; \mathbf{v}) &= \log \epsilon \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} x_{ij} \\ &\quad + \log(1 - \epsilon) \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} (1 - x_{ij}) \\ &\quad + \log q \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} (L_i - z_{ij}) \\ &\quad + \log(1 - q) \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} z_{ij}. \end{aligned} \quad (58)$$

In the E-step, the conditional expectation of (58) has to be computed. Thus, let us introduce

$$\hat{e}_{ij,t} \triangleq \mathbb{E}[e_{ij} | m_{ij}; \hat{\mathbf{v}}_t], \quad \hat{z}_{ij,t} \triangleq \mathbb{E}[z_{ij} | m_{ij}; \hat{\mathbf{v}}_t]. \quad (59)$$

It follows that $\hat{x}_{ij,t} \triangleq \mathbb{E}[x_{ij} | m_{ij}; \hat{\mathbf{v}}_t] = (1 - \hat{e}_{ij,t})m_{ij} + \hat{e}_{ij,t}(1 - m_{ij})$, and therefore, taking the conditional expectation of (58) yields

$$\begin{aligned} Q(\mathbf{v}, \hat{\mathbf{v}}_t) &= \hat{X}_t \log \epsilon + (n - \hat{X}_t) \log(1 - \epsilon) \\ &\quad + \left[\left(\sum_{i=1}^{\ell} n_i L_i \right) - \hat{Z}_t \right] \log q + \hat{Z}_t \log(1 - q), \end{aligned} \quad (60)$$

where

$$\hat{X}_t \triangleq \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \hat{x}_{ij,t}, \quad \hat{Z}_t \triangleq \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \hat{z}_{ij,t}. \quad (61)$$

Note that \hat{X}_t and \hat{Z}_t are soft estimates (based on the observations $\{m_{ij}\}$ and the current estimate $\hat{\mathbf{v}}_t$) of the total number of errors in the feedback channel and the total number of bits in error, respectively. Now, maximizing (60) with respect to ϵ and q , the estimates are updated as

$$\hat{\epsilon}_{t+1} = \frac{\hat{X}_t}{n}, \quad \hat{q}_{t+1} = 1 - \frac{\hat{Z}_t}{\sum_{i=1}^{\ell} n_i L_i}. \quad (62)$$

It remains to compute the conditional expectations in (59). We can write

$$\hat{e}_{ij,t} = (1 - m_{ij})\hat{\rho}_{i,t}^0 + m_{ij}\hat{\rho}_{i,t}^1, \quad (63)$$

where

$$\hat{\rho}_{i,t}^0 \triangleq \mathbb{E}[e_{ij} | m_{ij} = 0; \hat{\mathbf{v}}_t] = \frac{\hat{\epsilon}_t(1 - \hat{q}_t^{L_i})}{\hat{\epsilon}_t(1 - \hat{q}_t^{L_i}) + (1 - \hat{\epsilon}_t)\hat{q}_t^{L_i}}, \quad (64)$$

$$\hat{\rho}_{i,t}^1 \triangleq \mathbb{E}[e_{ij} | m_{ij} = 1; \hat{\mathbf{v}}_t] = \frac{(1 - \hat{\epsilon}_t)(1 - \hat{q}_t^{L_i})}{(1 - \hat{\epsilon}_t)(1 - \hat{q}_t^{L_i}) + \hat{\epsilon}_t\hat{q}_t^{L_i}}, \quad (65)$$

which do not depend on j . On the other hand, observe that z_{ij} can be written as $z_{ij} = e_{ij}b_{ij}$, where b_{ij} is a random variable statistically independent of e_{ij} and m_{ij} , taking values in the set $\{1, 2, \dots, L_i\}$ with probabilities

$$\mathbb{P}[b_{ij} = k; p] = \frac{B(k; L_i, p)}{1 - B(0; L_i, p)}, \quad 1 \leq k \leq L_i. \quad (66)$$

Therefore,

$$\begin{aligned} \hat{z}_{ij,t} &= \mathbb{E}[e_{ij}b_{ij} \mid m_{ij}; \hat{\mathbf{v}}_t] \\ &= \hat{e}_{ij,t} \mathbb{E}[b_{ij}; \hat{\mathbf{v}}_t] = \hat{e}_{ij,t} \frac{L_i(1 - \hat{q}_t)}{1 - \hat{q}_t^{L_i}}. \end{aligned} \quad (67)$$

Using (64)-(65) in (63) and (67) and then substituting in (62), the iteration is completed.

REFERENCES

- [1] A. Rico-Alvarino, R. W. Heath Jr., and C. Mosquera, "FER prediction with variable codeword length," in *Proc. IEEE ICASSP*, May 2014, pp. 1931–1935.
- [2] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link error prediction methods for multicarrier systems," in *Proc. IEEE VTC Fall*, vol. 6, Los Angeles, CA, Sep. 2004, pp. 4175–4179.
- [3] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. IEEE PIMRC*, vol. 4, Berlin, Germany, Sep. 2005, pp. 2306–2311.
- [4] A. Cipriano, R. Visoz, and T. Salzer, "Calibration issues of PHY layer abstractions for wireless broadband systems," in *Proc. IEEE VTC Fall*, Calgary, Canada, 2008, pp. 1–5.
- [5] S. Wahls and H. V. Poor, "Link adaptation for BICM-OFDM through adaptive kernel regression," in *Proc. IEEE ICASSP*, Vancouver, Canada, 2013, pp. 5136–5140.
- [6] Z. Puljiz, M. Park, and R. W. Heath Jr., "A machine learning approach to link adaptation for SC-FDE system," in *Proc. IEEE GLOBECOM*, Houston, TX, Dec. 2011, pp. 1–5.
- [7] M. Lampe, T. Giebel, H. Rohling, and W. Zirwas, "PER prediction for PHY mode selection in OFDM communication systems," in *Proc. IEEE GLOBECOM*, vol. 1, San Francisco, CA, 2003, pp. 25–29 Vol.1.
- [8] T. Jensen, S. Kant, J. Wehinger, and B. Fleury, "Mutual information metrics for fast link adaptation in IEEE 802.11n," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 4910 – 4915.
- [9] C. Koksall and P. Schniter, "Robust rate-adaptive wireless communication using ACK/NAK-feedback," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1752–1765, Apr. 2012.
- [10] A. Kamerman and L. Monteban, "WaveLAN-II: a high-performance wireless LAN for the unlicensed band," *Bell Labs Tech. J.*, vol. 2, no. 3, pp. 118–133, 1997. [Online]. Available: <http://dx.doi.org/10.1002/bltj.2069>
- [11] <http://www.flukenetworks.com/content/anytime-anywhere-wlan-monitoring-and-troubleshooting>.
- [12] http://www.rohde-schwarz.com/en/product/tsmw-productstartpage_63493-10072.html.
- [13] http://www.rohde-schwarz.com/en/product/etc-productstartpage_63493-12060.html.

- [14] "IEEE approved draft standard for IT - Telecommunications and Information Exchange between Systems - LAN/MAN - specific requirements - part 11: Wireless LAN medium access control and physical layer specifications - Amd 4: Enhancements for very high throughput for operation in bands below 6 GHz," *IEEE P802.11ac/D7.0*, Sep 2013.
- [15] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.
- [16] "Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding," *3GPP TS 36.212, v 11.3.0*, Jun 2013.
- [17] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746 – 1755, Sep. 2004.
- [18] R. Daniels, C. Caramanis, and R. W. Heath Jr., "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 114–126, Jan. 2010.
- [19] C. Wang, "Dynamic ARF for throughput improvement in 802.11 WLAN via a machine-learning approach," *J. Network and Computer Applications*, vol. 36, no. 2, pp. 667 – 676, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1084804513000027>
- [20] G. Martorell, F. Riera-Palou, and G. Femenias, "Cross-layer fast link adaptation for MIMO-OFDM based WLANs," *Wireless Personal Commun.*, vol. 56, no. 3, pp. 599–609, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s11277-010-9992-9>
- [21] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Trans. Mobile Comput.*, vol. 1, no. 4, pp. 278–292, 2002.
- [22] P. Tan, Y. Wu, and S. Sun, "Link adaptation based on adaptive modulation and coding for multiple-antenna OFDM system," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1599–1606, 2008.
- [23] G. Simons, "Estimating distortion in a binary symmetric channel consistently," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1466–1470, 1991.
- [24] K. Podgórski, G. Simons, and Y.-W. Ma, "On estimation for a binary-symmetric channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1260–1272, 1998.
- [25] V. Toto-Zarasoia, A. Roumy, and C. Guillemot, "Maximum likelihood BSC parameter estimation for the Slepian-Wolf problem," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 232–234, Feb. 2011.
- [26] G. Lechner and C. Pacher, "Estimating channel parameters from the syndrome of a linear code," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2148–2151, Nov. 2013.
- [27] S. M. Kay, *Fundamentals of Statistical Signal Processing, vol. 1: Estimation Theory*. Prentice-Hall, 1993.
- [28] N. Tripathi and J. H. Reed, *Cellular Communications: A Comprehensive and Practical Guide*. Wiley-IEEE Press, 2014.
- [29] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.
- [30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Soc.*, vol. Series B 39, no. 1, pp. 1–38, 1977.
- [31] T. K. Moon, "The Expectation-Maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov 1996.



Alberto Rico-Alvariño (S'12-M'15) received his Ph.D. degree in electrical engineering from the University of Vigo, Vigo, Spain, in 2014. He joined Qualcomm Inc. (San Diego, CA) in 2014 to work in standardization aspects related to LTE-Advanced and LTE-Advanced Pro. He is actively involved in 3GPP RAN standardization of IoT technologies (NB-IoT, eMTC) and broadcast services (MBMS), among other topics. He is a recipient of the of the best professional paper award at AIAA ICSSC 2012 and the best paper award of ASMS/SPSC 2014.



Roberto López-Valcarce (S'95-M'01) received the Telecommunication Engineering degree from the University of Vigo, Vigo, Spain, in 1995, and the M.S. and Ph.D. degrees in electrical engineering from the University of Iowa, Iowa City, IA, in 1998 and 2000, respectively. During 1995 he was a Project Engineer with Intelsis. He was a *Ramón y Cajal* Postdoctoral Fellow of the Spanish Ministry of Science and Technology from 2001 to 2006. During that period, he was with the Signal Theory and Communications Department, University of Vigo, where he currently is an Associate Professor. From 2010 to 2013, he was Program Manager of the Galician Regional Research Program on Information and Communication Technologies, Department of Research, Development and Innovation. His main research interests include adaptive signal processing, digital communications, and sensor networks, having co-authored over 50 papers in leading international journals. He holds several patents in collaboration with industry.

Dr. López-Valcarce is the recipient of a 2005 Best Paper Award of the IEEE Signal Processing Society. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011, and as a member of the IEEE Signal Processing for Communications and Networking Technical Committee from 2011 to 2013.



Carlos Mosquera (S'93–M'98–SM'13), born in Vigo in 1969, is currently Professor at the Signal Theory and Communications Department, University of Vigo. He received the M.Sc. degree in Electrical Engineering from Stanford University, Stanford, CA, in 1994, and the Ph.D. degree in Telecommunications Engineering from University of Vigo, Vigo, Spain, in 1998. He held visiting positions with the European Space Agency (1999), the University of New Mexico (2007), and the University of York (2013). He holds different patents in collaboration with industry, and has co-authored two books and about 100 cited conference and journal papers. He has co-organized several international conferences and special sessions on communications, and is a reviewer for different European research agencies. He is member of the Satellite Network of Experts, funded by the European Space Agency, for which he has been involved in different research projects. In the field of Satellite Communications he was co-recipient of the best professional paper award at the 2012 AIAA International Communications Satellite Systems Conference and the best paper award at the 2014 7th Advanced Satellite Multimedia Systems Conference and 13th Signal Processing for Space Communications Workshop. He was one of the promoters and Director of the Communications Area at the Galician Research and Development Center for Advanced Telecommunications (Gradiant) for five years. He is currently the Director of the Atlantic Research Center on Information and Communication Technologies (AtlantTIC).



Robert W. Heath Jr. (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff then a Senior Consultant at Iospan Wireless Inc, San Jose, CA where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he is a Cullen Trust for Higher Education Endowed Professor, and is a Member of the Wireless Networking and Communications Group. He is also President and CEO of MIMO Wireless Inc. He authored "Introduction to Wireless Digital Communication" (Prentice Hall, 2017), co-authored "Millimeter Wave Wireless Communications" (Prentice Hall, 2014), and authored "Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP" (National Technology and Science Press, 2012).

Dr. Heath has been a co-author of fifteen award winning conference and journal papers including recently the 2010 and 2013 EURASIP Journal on Wireless Communications and Networking best paper awards, the 2012 Signal Processing Magazine best paper award, a 2013 Signal Processing Society best paper award, 2014 EURASIP Journal on Advances in Signal Processing best paper award, the 2014 Journal of Communications and Networks best paper award, the 2016 IEEE Communications Society Fred W. Ellersick Prize, and the 2016 IEEE Communications and Information Theory Societies Joint Paper Award. He was a distinguished lecturer in the IEEE Signal Processing Society and is an ISI Highly Cited Researcher. He is also an elected member of the Board of Governors for the IEEE Signal Processing Society, a licensed Amateur Radio Operator, a Private Pilot, and a registered Professional Engineer in Texas.