

Blind Estimation of Bit and Block Error Probabilities Using Soft Information

Andreas Winkelbauer and Gerald Matz

Institute of Telecommunications, Vienna University of Technology
 Gusshausstrasse 25/389, 1040 Vienna, Austria
 email: {andreas.winkelbauer, gerald.matz}@tuwien.ac.at

Abstract—We consider the problem of estimating the bit error probability (BEP) and the block error probability (BLEP) in digital communication systems in a blind manner, i.e., without knowledge of the transmitted data. We propose estimators based on soft information and show that they outperform conventional non-blind estimators, which know the transmitted data. We study the performance of these estimators and provide conditions that guarantee that our BEP estimator is the minimum variance unbiased estimator. For the Gaussian case we derive the Cramér-Rao lower bound for BEP estimation and show that an efficient estimator does not exist. Furthermore, we derive simple lower and upper bounds for the BLEP estimator. Our results are numerically corroborated using Monte Carlo simulations.

I. INTRODUCTION

The bit error probability (BEP) and the block error probability (BLEP) are two fundamental performance metrics for digital communications and data storage. They are also crucial for techniques like adaptive modulation and coding [1], [2], coding for multimedia signals [3], and handoff algorithms in cellular networks [4]. Conventional error probability estimators use hard decisions and assume knowledge of the transmitted data. In contrast, the estimators that we propose in this paper use soft outputs and operate blindly, i.e., do not require the transmitted data. Our results go significantly beyond previous work on soft-information based BEP estimation [5], [6]. Specifically, our contributions are as follows:

- We propose estimators for the BEP and BLEP which are based on soft information and require neither knowledge of the transmitted data nor training overhead.
- We give a condition for the unbiasedness of the BLEP estimator and we derive simple upper and lower bounds for the BLEP estimate.
- We prove that the mean square error (MSE) of our estimators is smaller by at least a factor of two, compared to the MSE of conventional estimators.
- We derive a series expansion for the estimator variance that can be used to evaluate the MSE.
- We provide conditions for our BEP estimator to be the minimum-variance unbiased (MVU) estimator. For BEP estimation in the Gaussian case, we derive the Cramér-Rao lower bound (CRLB) and show that an efficient estimator does not exist.
- We demonstrate the effectiveness of our estimators and confirm our results using Monte Carlo simulations.

The remainder of this paper is organized as follows. In Section II we introduce the necessary background. Sections III and IV deal with BEP estimation and BLEP estimation, respectively. In Section V, we study MVU estimation and the CRLB for BEP estimation. A numerical example is provided in Section VI and conclusions are given in Section VII.

Notation: Boldface and uppercase symbols denote vectors and random variables, respectively (except for the deterministic integers B, K, M, N). $\sum_{\sim x_n}$ denotes summation over all elements of the vector \mathbf{x} except x_n . We use $\text{sign}(X) \in \mathcal{B} \triangleq \{+1, -1\}$ to denote hard decisions; if $X = 0$, $\text{sign}(X)$ depends on the prior distribution of X . $\mathbb{I}\{\cdot\}$ is the indicator function, which equals 1 if the argument is true and 0 otherwise. Expectation and probability are denoted by $\mathbb{E}\{\cdot\}$ and $\mathbb{P}\{\cdot\}$, respectively. The natural and binary logarithms are denoted by $\log(\cdot)$ and $\log_2(\cdot)$, respectively. Finally, $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 .

II. BACKGROUND AND DEFINITIONS

We consider the generic digital communication system depicted in Fig. 1. The source emits length- N blocks of bits $\mathbf{u} \in \mathcal{B}^N$ with known distribution $p(\mathbf{u})$. The encoder ϕ maps the data to the length- M channel input sequence $\mathbf{x} = \phi(\mathbf{u}) \in \mathcal{X}^M$, which is then transmitted over the channel $(\mathcal{X}^M, p(\mathbf{y}|\mathbf{x}), \mathcal{Y}^K)$. By allowing $M \neq K$ we can cover for example deletion and insertion channels. We focus on continuous-output channels, i.e., $\mathcal{Y} = \mathbb{R}$, even though with obvious changes our results also hold for discrete-output channels. The decoder ψ is a soft-output maximum a posteriori (MAP) decoder that computes bit-wise a posteriori log-likelihood ratios (LLRs) $\ell = \psi(\mathbf{y})$ as $(n = 1, 2, \dots, N)$

$$\ell_n = \psi_n(\mathbf{y}) \triangleq \log \frac{p(U_n = +1|\mathbf{y})}{p(U_n = -1|\mathbf{y})} = \ell_{ch,n} + \ell_{a,n}, \quad (1)$$

where

$$\ell_{ch,n} \triangleq \log \frac{p(\mathbf{y}|U_n = +1)}{p(\mathbf{y}|U_n = -1)} \quad \text{and} \quad \ell_{a,n} \triangleq \log \frac{p(U_n = +1)}{p(U_n = -1)}$$

are due to the channel and the a priori source statistics, respectively. LLRs are a convenient representation of the posterior probabilities $p(u_n|\mathbf{y})$. The MAP hard-decision equals $\hat{u}_n = \text{sign}(\ell_n)$ with the associated reliability $r_n = |\ell_n|$.

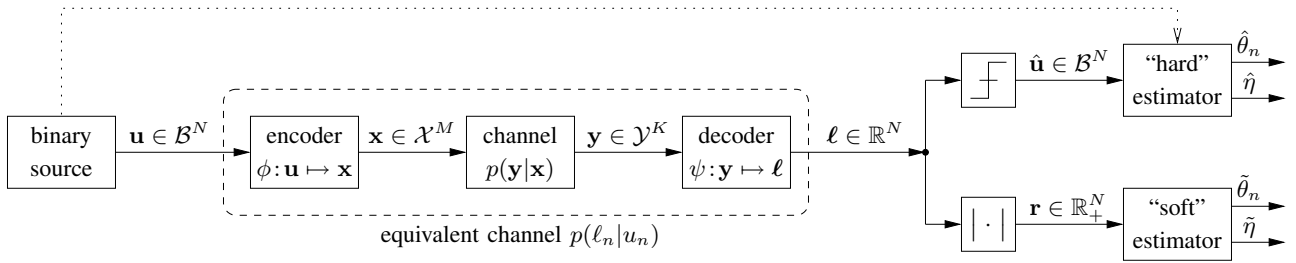


Figure 1: Digital communication system with “hard” and “soft” estimation of BEP (θ) and BLEP (η).

The BEP for the n th bit and the BLEP achieved with MAP decoding are given by

$$\theta_n \triangleq \mathbb{P}\{\hat{U}_n \neq U_n\}, \quad \eta \triangleq \mathbb{P}\{\hat{\mathbf{U}} \neq \mathbf{U}\}. \quad (2)$$

Note that $\theta_n \in [0, 1/2]$, $\eta \in [0, 1]$. Assuming that $B \geq 1$ independent and identically distributed (iid) blocks are transmitted, conventional approaches use the transmitted data and the MAP hard decisions to estimate the BEP and BLEP respectively as

$$\hat{\theta}_n \triangleq \frac{1}{B} \sum_{b=1}^B \mathbb{I}\{\hat{u}_n^{(b)} \neq u_n^{(b)}\}, \quad (3)$$

$$\hat{\eta} \triangleq \frac{1}{B} \sum_{b=1}^B \mathbb{I}\{\hat{\mathbf{u}}^{(b)} \neq \mathbf{u}^{(b)}\}. \quad (4)$$

Here, the superscript (b) refers to the b th block. The BEP in general may depend on the bit position n . However, if $\theta_n \equiv \theta$ or only the average BEP is of interest, the BEP estimate $\hat{\theta}_n$ can further be averaged with respect to n . Similarly, in case of unequal error protection the BEP estimates can be averaged over the bits that have the same level of protection.

We note that the LLRs can often be computed only approximately, either because of complexity issues or because the channel law $p(\mathbf{y}|\mathbf{x})$ is imperfectly known. These issues and their implications for error probability estimation will be discussed in future work.

We next state a key property of LLR distributions.

Lemma 1. *Let $U \in \mathcal{B}$ be a binary random variable with a posteriori LLR L and a priori LLR ℓ_a . Assuming that all conditional moments of L are finite, the conditional distribution of L satisfies the consistency condition*

$$p(\ell|U=+1) = \exp(\ell - \ell_a)p(\ell|U=-1). \quad (5)$$

Proof: See Appendix A. ■

The assumptions that the conditional LLR distribution exists and that the conditional moments are finite is not overly restrictive and even captures some pathological cases. A still wider range of scenarios could be tackled by defining LLRs in terms of the Radon-Nikodym derivative.

The following results are direct consequences of (1) and (5).

Corollary 1. *The soft-output MAP decoder ψ is idempotent, i.e., $\psi(\psi(\mathbf{y})) = \psi(\ell) = \ell$.*

Corollary 2. *The conditional distribution of the posterior LLR L can be expressed in terms of unconditional distribution as*

$$p(\ell|u) = \frac{1}{1 + e^{-u\ell}} \frac{p(\ell)}{p(u)}, \quad u \in \mathcal{B}. \quad (6)$$

Through (5) and (6), the three distributions $p(\ell)$, $p(\ell|U=+1)$, and $p(\ell|U=-1)$ are connected such that any one of them is sufficient to express the other two.

Corollary 3. *The mutual information $I(U; L)$ [7] can be written in terms of $p(\ell)$ as follows:*

$$I(U; L) = - \sum_{u \in \mathcal{B}} \mathbb{E} \left\{ \frac{1}{1 + e^{-uL}} \log_2(p(u)(1 + e^{-uL})) \right\}. \quad (7)$$

If U is uniformly distributed, we can rewrite (7) as

$$I(U; L) = 1 - \mathbb{E} \left\{ h_2 \left(\frac{1}{1 + e^{|L|}} \right) \right\} \geq 1 - h_2(\theta) = C_{\text{BSC}}(\theta), \quad (8)$$

where θ denotes the BEP and $h_2(p)$ is the binary entropy function. We have equality in (8) if $p(\ell) = (\delta(\ell - \log \frac{1-\theta}{\theta}) + \delta(\ell + \log \frac{1-\theta}{\theta}))/2$, i.e., if the equivalent channel is a binary symmetric channel (BSC) with cross-over probability θ (here, $\delta(\cdot)$ denotes the Dirac delta distribution).

Corollary 3 gives us a simple and elegant lower bound on the mutual information $I(U; L)$ in terms of the BEP if U is uniform. Indeed, in this case among all equivalent channels with fixed BEP θ under soft-output MAP decoding, the BSC has the smallest mutual information $I(U; L)$. In passing we note that another interesting application of Corollary 3 is the unbiased estimation of $I(U; L)$ based on the decoder’s soft output only, e.g., by simply replacing the expectation in (7), (8) by the sample mean.

III. BIT ERROR PROBABILITY ESTIMATION

To derive our soft BEP estimator, we first use (6) to rewrite the BEP θ_n as

$$\begin{aligned} \theta_n &= \mathbb{P}\{L_n < 0|U_n=+1\}p(U_n=+1) \\ &\quad + \mathbb{P}\{L_n \geq 0|U_n=-1\}p(U_n=-1) \\ &= \int_{-\infty}^0 \frac{p(\ell_n)}{1 + e^{-\ell_n}} d\ell_n + \int_0^{\infty} \frac{p(\ell_n)}{1 + e^{\ell_n}} d\ell_n \\ &= \int_{-\infty}^{\infty} \frac{p(\ell_n)}{1 + e^{|\ell_n|}} d\ell_n = \mathbb{E}\{\Xi_n\}, \end{aligned} \quad (9)$$

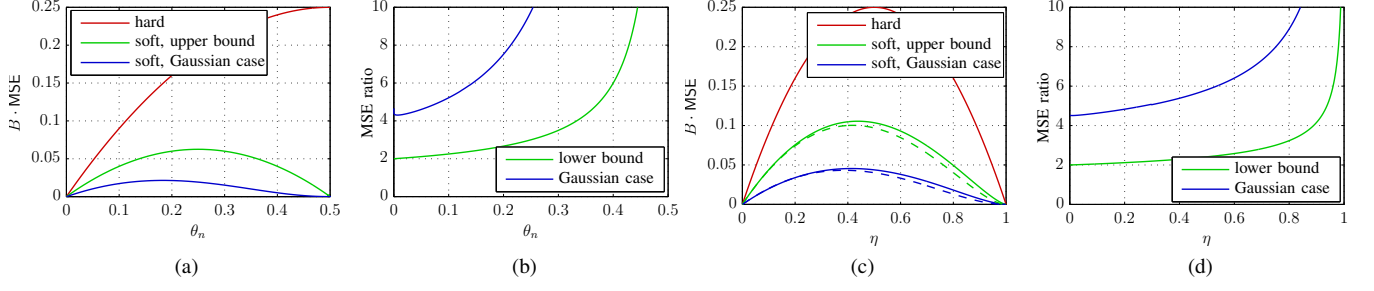


Figure 2: (a) Comparison of $\text{MSE}_{\hat{\theta}_n}$ in (18), upper bound (16), and $\text{MSE}_{\hat{\theta}_n}$ for the Gaussian case. (b) MSE ratio $\text{MSE}_{\hat{\theta}_n}/\text{MSE}_{\hat{\theta}_n}$ for the Gaussian case and lower bound (19). (c) Comparison of $\text{MSE}_{\hat{\eta}}$ (27), upper bounds (24), (25), and $\text{MSE}_{\hat{\eta}}$ for the Gaussian case for $N \rightarrow \infty$ (solid) and $N = 10$ (dashed). (d) MSE ratio $\text{MSE}_{\hat{\eta}}/\text{MSE}_{\hat{\eta}}$ for the Gaussian case and lower bound (29).

where we have defined

$$\Xi_n \triangleq \frac{1}{1 + e^{|\ell_n|}} \in [0, 1/2].$$

By replacing the expectation of Ξ_n in (9) with the empirical mean over B iid blocks, we obtain the soft BEP estimator

$$\tilde{\theta}_n \triangleq \frac{1}{B} \sum_{b=1}^B \xi_n^{(b)}, \quad \text{with} \quad \xi_n^{(b)} = \frac{1}{1 + \exp(|\ell_n^{(b)}|)}. \quad (10)$$

This estimator does not involve the transmitted data. Since the blocks are iid, (9) and (10) imply the following result.

Proposition 1. *The soft BEP estimator is unbiased, i.e., $\mathbb{E}\{\tilde{\theta}_n\} = \theta_n$.*

Next, we study the MSE of $\tilde{\theta}_n$ (equal to the variance),

$$\text{MSE}_{\tilde{\theta}_n} = \mathbb{E}\{\tilde{\theta}_n^2\} - \theta_n^2 = \frac{1}{B} (\mathbb{E}\{\Xi_n^2\} - \theta_n^2). \quad (11)$$

The computation of (11) requires the distribution of Ξ_n , and thus, in turn the distribution of L_n . In Appendix B we prove the following result.

Proposition 2. *Assuming L_n has finite absolute moments, the mean power of Ξ_n equals*

$$\mathbb{E}\{\Xi_n^2\} = \frac{1}{4} + \sum_{m=1}^{\infty} \frac{\mathbb{E}\{|L_n|^m\}}{m!} \sum_{k=1}^m \frac{(-1)^k}{2^{k+2}} d_{k,m}, \quad (12)$$

where the coefficients $d_{k,m}$ are defined by the recursion

$$\begin{aligned} d_{k,m} &= (k+1)d_{k-1,m-1} + kd_{k,m-1} \quad \text{for } k \geq 2, m \geq 2, \\ d_{1,m} &= 2 \quad \text{for } m \geq 1, \quad d_{k,1} = 0 \quad \text{for } k \geq 2. \end{aligned} \quad (13)$$

The sign of the terms in the series (12) can be shown to change after every second term. Therefore, we can truncate the series after any pair of terms having the same sign and bound the error by the sum of the following two terms. We can further expand $\mathbb{E}\{|L_n|^m\}$ in (12) as

$$\begin{aligned} \mathbb{E}\{|L_n|^m\} &= \mathbb{E}\{|L_n|^m | U_n = +1\} p(U_n = +1) \\ &\quad + \mathbb{E}\{|L_n|^m | U_n = -1\} p(U_n = -1). \end{aligned} \quad (14)$$

For conditionally Gaussian LLRs, i.e., when the equivalent channel that comprises encoder, channel, and decoder (cf. Fig. 1) satisfies $p(\ell_n | u_n) \sim \mathcal{N}(\mu_n, \sigma_n^2)$, we have¹ [8]

$$\mathbb{E}\{|L_n|^m | U_n = u_n\} = 2^{\frac{m}{2}} \sigma_n^m \frac{\Gamma(\frac{m+1}{2})}{\sqrt{\pi}} {}_1F_1\left(-\frac{m}{2}; \frac{1}{2}; \frac{-\mu_n^2}{2\sigma_n^2}\right), \quad (15)$$

where

$$\begin{aligned} \Gamma(z) &= \int_0^{\infty} e^{-t} t^{z-1} dt, \quad \text{and} \\ {}_1F_1(\alpha; \gamma; z) &= \sum_{k=0}^{\infty} \frac{\alpha^{\bar{k}} z^k}{\gamma^{\bar{k}} k!}, \quad x^{\bar{k}} = \frac{\Gamma(x+k)}{\Gamma(x)}, \end{aligned}$$

are the gamma function and Kummer's confluent hypergeometric function. Hence, in this case we can express the MSE in terms of the conditional second-order LLR statistics.

The MSE can be bounded as

$$\text{MSE}_{\tilde{\theta}_n} \leq \frac{1}{B} \theta_n \left(\frac{1}{2} - \theta_n \right) \leq \frac{1}{16B}, \quad (16)$$

where the first inequality is satisfied with equality for $\theta_n = 0$ and $\theta_n = 1/2$. The (first) bound in (16) is tight; it is obtained using the inequality $\mathbb{E}\{\Xi_n^2\} \leq \mathbb{E}\{\Xi_n\}/2 = \theta_n/2$ or equivalently $\text{var}\{\Xi_n\} \leq \theta_n/2 - \theta_n^2$ (which follows from $0 \leq \Xi_n \leq 1/2$). A distribution for which this upper bound is met with equality for all θ_n is given by $p(\xi_n) = (1 - 2\theta_n)\delta(\xi_n) + 2\theta_n\delta(\xi_n - 1/2)$, which corresponds to a binary erasure channel (BEC) with erasure probability $2\theta_n$.

Let us next compare our proposed estimator to the conventional hard estimator (3). To this end, note that the mean and variance of $\hat{\theta}_n$ equal

$$\mathbb{E}\{\hat{\theta}_n\} = \mathbb{E}\{\mathbb{I}\{\hat{U}_n \neq U_n\}\} = \theta_n, \quad (17)$$

$$\begin{aligned} \text{MSE}_{\hat{\theta}_n} &= \mathbb{E}\{\hat{\theta}_n^2\} - \theta_n^2 = \frac{1}{B} (\mathbb{E}\{\mathbb{I}^2\{\hat{U}_n \neq U_n\}\} - \theta_n^2) \\ &= \frac{1}{B} \theta_n (1 - \theta_n) \leq \frac{1}{4B}. \end{aligned} \quad (18)$$

Hence, the hard estimator is also unbiased. From (11), (18) it follows that both estimators are consistent, i.e., they converge

¹Here, μ_n and σ_n^2 depend on the value of the bit u_n .

in probability to θ_n . Fig. 2a depicts the MSE in (18), the MSE upper bound (16), and the MSE for conditionally Gaussian distributed LLRs with $\mu_{u_n} = \pm\sigma_{u_n}^2/2$ (numerically evaluated using (12)-(15)), all versus $\theta_n = Q(\sigma_{u_n}/2)$ (here, $Q(\cdot)$ denotes the Q -function). Combining (16) and (18) allows us to relate the MSEs of the soft and hard BEP estimators.

Proposition 3. *The BEP MSE ratio is bounded as*

$$\frac{\text{MSE}_{\hat{\theta}_n}}{\text{MSE}_{\tilde{\theta}_n}} \geq 2 \frac{1 - \theta_n}{1 - 2\theta_n} \geq 2, \quad (19)$$

with equality iff $\theta_n = 0$.

It is thus seen that the soft estimator outperforms the hard estimator in terms of MSE by a factor that is at least two and grows to infinity as $\theta_n \rightarrow 1/2$. Fig. 2b shows the lower bound (19) as well as the exact variance ratio in the Gaussian case versus θ_n . We see that $\text{MSE}_{\hat{\theta}_n}/\text{MSE}_{\tilde{\theta}_n} > 4$ in the Gaussian case. Therefore, in this case, the proposed estimator requires more than four times fewer blocks than the conventional hard estimator to achieve the same level of accuracy.

Finally, we note that the performance of the proposed estimator cannot be improved by using the source data. We thus conclude that (i) for the problem of BEP estimation precise channel knowledge (which is needed to compute the LLRs) is more important than knowing the transmitted data, and (ii) ‘‘soft is better than hard’’ is true also for BEP estimation.

IV. BLOCK ERROR PROBABILITY ESTIMATION

In this section we discuss estimation of the BLEP. We note that we could consider the error probability for any subset of data bits instead of the whole block. However, for the sake of simplicity we stick to the BLEP η defined in (2).

Let us first rewrite η using (9) as

$$\begin{aligned} \eta &= 1 - \text{P}\left\{\hat{U}_1=U_1 \cap \hat{U}_2=U_2 \cap \dots \cap \hat{U}_N=U_N\right\} \quad (20) \\ &= 1 - \prod_{n=1}^N \text{P}\left\{\hat{U}_n=U_n\right\} = 1 - \prod_{n=1}^N \left(1 - \text{P}\left\{\hat{U}_n \neq U_n\right\}\right) \quad (21) \end{aligned}$$

$$= 1 - \prod_{n=1}^N (1 - \text{E}\{\Xi_n\}) = 1 - \prod_{n=1}^N (1 - \theta_n),$$

where we have assumed independent bit errors in (21). With (21) we obtain the BLEP estimator

$$\tilde{\eta} \triangleq 1 - \frac{1}{B} \sum_{b=1}^B \prod_{n=1}^N (1 - \xi_n^{(b)}) = 1 - \frac{1}{B} \sum_{b=1}^B \prod_{n=1}^N \frac{1}{1 + \exp(-|\ell_n^{(b)}|)}, \quad (22)$$

which, like the BEP estimator, does not use the source data.

Proposition 4. *If the events $\{\hat{U}_n=U_n\}_{n=1}^N$ are independent, the soft BLEP estimator (22) is unbiased, i.e., $\text{E}\{\tilde{\eta}\} = \eta$.*

Proof: Given independent bit errors, the unbiasedness of (22) follows directly from (21). ■

In the following we assume that all bit error events are independent. The MSE (equivalently, the variance) of $\tilde{\eta}$ can

then be obtained as

$$\begin{aligned} \text{MSE}_{\tilde{\eta}} &= \text{E}\{\tilde{\eta}^2\} - \eta^2 \\ &= \frac{1}{B} \left(\prod_{n=1}^N \text{E}\{(1 - \Xi_n)^2\} - \prod_{n=1}^N (1 - \theta_n)^2 \right). \quad (23) \end{aligned}$$

Whether (23) can be computed in closed form depends again on the distribution of Ξ_n , and thus, on the distribution of L_n . As in the previous section, (12)-(15) can be used in the Gaussian case to express the MSE in terms of the second order statistics of the conditional LLR distributions. Using $\text{E}\{\Xi_n^2\} \leq \theta_n/2$ we can bound the MSE of $\tilde{\eta}$ as

$$\text{MSE}_{\tilde{\eta}} \leq \frac{1}{B} \left(\prod_{n=1}^N \left(1 - \frac{3}{2}\theta_n\right) - \prod_{n=1}^N (1 - \theta_n)^2 \right), \quad (24)$$

where equality holds for an equivalent BEC with erasure probability $2\theta_n$ (cf. (16)). For the special case $\theta_n \equiv \theta$, we have $\eta = 1 - (1 - \theta)^N$. Letting additionally $N \rightarrow \infty$, we obtain

$$\text{MSE}_{\tilde{\eta}} \leq \frac{1}{B} ((1 - \eta)^{3/2} - (1 - \eta)^2) \leq \frac{1}{B} \frac{27}{256}, \quad (25)$$

where the first inequality holds with equality if $\eta = 0$ or $\eta = 1$. The mean and variance of the hard estimator (4) are given by

$$\text{E}\{\hat{\eta}\} = \text{E}\{\text{I}\{\hat{\text{U}} \neq \text{U}\}\} = \eta, \quad (26)$$

$$\begin{aligned} \text{MSE}_{\hat{\eta}} &= \text{E}\{\hat{\eta}^2\} - \eta^2 = \frac{1}{B} \eta(1 - \eta) \\ &= \frac{1}{B} \left(\prod_{n=1}^N (1 - \theta_n) - \prod_{n=1}^N (1 - \theta_n)^2 \right) \leq \frac{1}{4B}. \quad (27) \end{aligned}$$

In contrast to the soft estimator, the hard estimator is always unbiased (cf. (26)). This is because it uses the transmitted data \mathbf{u} . It follows from (23) and (27) that both BLEP estimators are consistent. Fig. 2c depicts $\text{MSE}_{\tilde{\eta}}$ in (27), the upper bounds (24), (25) of $\text{MSE}_{\tilde{\eta}}$, and $\text{MSE}_{\hat{\eta}}$ for the Gaussian case, all versus η (assuming $\theta_n \equiv \theta$) for $N \rightarrow \infty$ (solid lines) and $N = 10$ (dashed lines), respectively. It is seen that the dependence of the MSE on N is very weak.

Combining (24) and (27) allows us to relate the MSEs of the soft and hard BLEP estimators.

Proposition 5. *The BLEP MSE ratio is bounded as*

$$\frac{\text{MSE}_{\hat{\eta}}}{\text{MSE}_{\tilde{\eta}}} \geq \frac{\prod_{n=1}^N (1 - \theta_n) - \prod_{n=1}^N (1 - \theta_n)^2}{\prod_{n=1}^N (1 - 3\theta_n/2) - \prod_{n=1}^N (1 - \theta_n)^2} \geq 2, \quad (28)$$

with equality iff $\theta_n = 0$, $n = 1, 2, \dots, N$.

Proof: See Appendix C. ■

The soft estimator thus outperforms the hard estimator in terms of the MSE by at least a factor of 2. In the special case $\theta_n \equiv \theta$ and $N \rightarrow \infty$, the MSE ratio of the soft and hard BLEP estimators satisfies the inequalities

$$\frac{\text{MSE}_{\hat{\eta}}}{\text{MSE}_{\tilde{\eta}}} \geq \frac{\eta(1 - \eta)}{(1 - \eta)^{3/2} - (1 - \eta)^2} \geq 2, \quad (29)$$

where the minimal ratio is achieved at $\eta = 0$ and the MSE ratio becomes unbounded as $\eta \rightarrow 1$.

Fig. 2d shows the lower bound (29) as well as the exact MSE ratio in the Gaussian case versus η for $\theta_n \equiv \theta$ and $N \rightarrow \infty$. We observe that $\text{MSE}_{\tilde{\eta}}/\text{MSE}_{\tilde{\eta}} > 4$ in the Gaussian case and, hence, the proposed estimator requires more than four times fewer blocks than the conventional hard estimator to achieve the same level of accuracy.

Computationally attractive bounds on $\tilde{\eta}$ are given by

$$\frac{1}{B} \sum_{b=1}^B \max_n \xi_n^{(b)} \leq \tilde{\eta} \leq \min \left\{ \tilde{\eta}_{\max}, \sum_{n=1}^N \tilde{\theta}_n \right\} \quad (30)$$

where

$$\tilde{\eta}_{\max} \triangleq 1 - \frac{1}{B} \sum_{b=1}^B \exp \left(-N \log \frac{1}{N} \sum_{n=1}^N \frac{\xi_n^{(b)}}{1 - \xi_n^{(b)}} \right)$$

The lower bound corresponds to the intuition that the BLEP is dominated by the least reliable bit in each block. The upper bounds are obtained by respectively applying the log-sum inequality to (22) and the union bound to (20). As $B \rightarrow \infty$, the union bound in (30) converges in probability to $\eta \leq \sum_n \theta_n$, i.e., to an upper bound on the true BLEP. The usefulness of (30) will be confirmed numerically in Section VI.

Finally, we note that whenever a cyclic redundancy check (CRC) is available, the hard BLEP estimate can be approximately computed without knowing the transmitted sequence. A CRC with p bits of redundancy guarantees an error detection probability of $1 - 2^{-p}$ [9] and can thus be viewed as partial knowledge of the transmit data. However, our proposed BLEP estimator requires no CRC and is more accurate (converges faster) than the hard estimator.

V. MINIMUM-VARIANCE UNBIASEDNESS AND THE CRAMÉR-RAO LOWER BOUND

In this section, we investigate whether an MVU estimator for the problem of BEP estimation exists. Furthermore, for the Gaussian case we derive the CRLB and show that in this case there exists no efficient estimator.

Let us first consider the case $B = 1$, i.e., the estimator $\tilde{\theta}$ uses a single reliability $r = |\ell|$. Let $p(r; \theta)$ denote the pdf of R , parametrized by the BEP θ . Clearly,

$$\xi(r) = \frac{1}{1 + e^r} \quad (31)$$

is a sufficient statistic for r since $\xi(r)$ is invertible. In what follows we have to exclude the special case where the pdf of $\Xi(R)$ is given by $p(\xi; \theta) = (1 - 2\theta)\delta(\xi) + 2\theta\delta(\xi - 1/2)$ (corresponding to a BEC). From the derivation in (9) we know that $\xi(r)$ is the unique function such that $\text{E}\{\xi(r)\} = \theta$ for all θ . Hence, $\xi(r)$ is a complete sufficient statistic and thus we can state the following theorem.

Theorem 1. *For $B = 1$, $\tilde{\theta} = \xi(r)$ is the MVU estimator of θ if $p(\xi; \theta) \neq (1 - 2\theta)\delta(\xi) + 2\theta\delta(\xi - 1/2)$.*

Proof: Since $\xi(r)$ in (31) is a complete sufficient statistic if $p(\xi; \theta) \neq (1 - 2\theta)\delta(\xi) + 2\theta\delta(\xi - 1/2)$, there is only one function $\rho(\cdot)$ such that $\text{E}\{\rho(\Xi)\} = \theta$ for all θ and by the Rao-Blackwell-Lehmann-Scheffé theorem [10] $\rho(\xi)$ is the MVU

estimator. From (9) we conclude that $\rho(\xi) = \xi$ and therefore $\tilde{\theta} = \xi(r)$ is the MVU estimator. ■

Next, we consider the case $B > 1$. Let $\mathbf{r} = (r_1 r_2 \dots r_B)^T$ denote the vector of iid reliabilities, whose pdf we assume to be from an exponential family with parameter θ , i.e., $p(r_b; \theta) = g(r_b)h(\theta) \exp(\zeta(\theta)\xi(r_b))$. It follows that

$$p(\mathbf{r}; \theta) = \left[\prod_{b=1}^B g(r_b) \right] h^B(\theta) \exp(B\zeta(\theta)\chi(\mathbf{r}))$$

with the complete sufficient statistic

$$\chi(\mathbf{r}) = \frac{1}{B} \sum_{b=1}^B \xi(r_b).$$

The MVU estimator of θ is characterized as follows.

Theorem 2. *Let $B > 1$ and assume that the elements of \mathbf{r} are iid, with distribution $p(r; \theta)$ from an exponential family with parameter θ . Then*

$$\tilde{\theta} = \chi(\mathbf{r}) = \frac{1}{B} \sum_{b=1}^B \xi(r_b) = \frac{1}{B} \sum_{b=1}^B \frac{1}{1 + e^{r_b}} \quad (32)$$

is the MVU estimator of θ if $p(\xi; \theta) \neq (1 - 2\theta)\delta(\xi) + 2\theta\delta(\xi - 1/2)$.

Proof: The proof uses the same arguments as the proof of Theorem 1. Here, if $p(\xi; \theta) \neq (1 - 2\theta)\delta(\xi) + 2\theta\delta(\xi - 1/2)$, the unique $\rho(\cdot)$ such that $\text{E}\{\rho(X)\} = \theta$ for all θ is given by $\rho(\chi) = \chi$. ■

Unfortunately, Theorem 2 does not cover the important case of conditionally Gaussian LLRs. In this case, the pdf $p(r; \theta)$ is a sum of exponential functions and thus cannot belong to an exponential family. The importance of conditionally Gaussian LLRs has two main reasons: (i) the binary-input AWGN channel leads to conditionally Gaussian LLRs; (ii) numerous receiver algorithms use Gaussian approximations to reduce computational complexity. We thus study the Gaussian case in more detail by analyzing the CRLB.

Let us assume $p(\ell|u; \theta) \sim \mathcal{N}(u\sigma^2/2, \sigma^2)$. We note that $\text{E}\{L|U = u\} = u\sigma^2/2$ is due to the consistency condition (5), which further implies $p(-\ell|u; \theta) = p(\ell|-u; \theta)$. The BEP is given by $\theta = Q(\sigma/2)$. Consequently, the pdf of $r = |\ell|$ is given by

$$p(r; \theta) = (p_L(r; \theta) + p_L(-r; \theta))u(r)$$

where $u(r)$ is the unit step function and the unconditional LLR pdf $p(\ell; \theta)$ equals

$$p_L(\ell; \theta) = p(\ell|U = +1; \theta)p(U = +1) + p(\ell|U = -1; \theta)p(U = -1).$$

In what follows we assume that $B \geq 1$ and \mathbf{r} is drawn iid from $p(r; \theta)$. To derive the Fisher information, we first compute the score (see Appendix D for details),

$$v \triangleq \frac{\partial}{\partial \theta} \log p(\mathbf{r}; \theta) = \sum_{b=1}^B v_b, \quad (33)$$

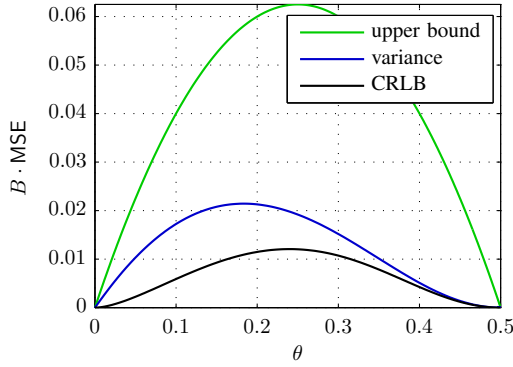


Figure 3: Comparison between CRLB (37), upper bound (16), and true variance of the soft BEP estimator in the case of conditionally Gaussian LLRs.

where

$$\begin{aligned} v_b &= \frac{\partial}{\partial \theta} \log p(r_b; \theta) \\ &= \sqrt{\frac{\pi}{2}} \exp\left(\frac{Q^{-2}(\theta)}{2}\right) Q^{-1}(\theta) \left(2 - \frac{r_b^2 - 4Q^{-2}(\theta)}{2Q^{-4}(\theta)}\right) u(r_b), \end{aligned} \quad (34)$$

and $Q^{-m}(\theta)$ is shorthand for $(Q^{-1}(\theta))^m$, with $Q^{-1}(\cdot)$ denoting the inverse of the Q -function. Now, the Fisher information $J(\theta) \triangleq -E\left\{\frac{\partial}{\partial \theta} v\right\}$ can be shown to be given by

$$J(\theta) = -\sum_{b=1}^B \int_{\mathbb{R}} \left(\frac{\partial^2}{\partial \theta^2} \log p(r_b; \theta)\right) p(r_b; \theta) dr_b = B J_1(\theta), \quad (35)$$

where (see Appendix D for details)

$$\begin{aligned} J_1(\theta) &= -\int_{\mathbb{R}} \left(\frac{\partial^2}{\partial \theta^2} \log p(r; \theta)\right) p(r; \theta) dr \\ &= 4\pi \exp(Q^{-2}(\theta)) \frac{1 + 2Q^{-2}(\theta)}{Q^{-2}(\theta)}. \end{aligned} \quad (36)$$

The following theorem is based on (33)-(36).

Theorem 3. *The CRLB for BEP estimation with iid conditionally Gaussian distributed LLRs is given by*

$$\text{MSE}_{\hat{\theta}} = \text{var}\{\tilde{\theta}\} \geq \frac{1}{B} \frac{Q^{-2}(\theta)}{4\pi \exp(Q^{-2}(\theta)) (1 + 2Q^{-2}(\theta))}. \quad (37)$$

Furthermore, for this problem, there exists no efficient estimator, i.e., the CRLB cannot be attained uniformly.

Proof: See Appendix D. ■

Fig. 3 shows the CRLB (37), the upper bound (16), and the true MSE (numerically evaluated using (12)-(15)) versus θ . We observe that the MSE is closer to the CRLB than to the upper bound.

VI. NUMERICAL EXAMPLE

In this section, we illustrate the usefulness of the proposed error probability estimators using a numerical example.

Specifically, we consider the Monte Carlo simulation of coded transmission over a noisy intersymbol interference (ISI) channel.

General Setup. The source transmits blocks of $N = 2^{10}$ independent and uniformly distributed bits which are encoded by a rate-1/2 recursive, systematic convolutional code with generator polynomial $[1 \ 13/15]_8$ (in octal notation). The encoded bits are then transmitted using BPSK, i.e., $\mathbf{x} \in \mathcal{B}^N$, over a noisy ISI channel with impulse response $\mathbf{h} = [1 \ -1]^T/\sqrt{2}$ and additive white Gaussian noise. The transmit signal is normalized to unit power, the noise is zero-mean with variance N_0 and, hence, the SNR is $\gamma = 1/N_0$. The receive signal $\mathbf{y} \in \mathbb{R}^N$ is processed by a BCJR [11] equalizer and subsequently by a BCJR channel decoder which outputs the *a posteriori* LLRs $\ell \in \mathbb{R}^N$. Finally, we employ hard and soft estimators for the *average* BEP and the BLEP. We have performed simulations over $B = 1000$ blocks.

BEP Estimation. Fig. 4a shows the estimated BEP versus the channel SNR in dB. We observe that the hard estimate $\hat{\theta}$ and the soft estimate $\tilde{\theta}$ coincide, confirming the unbiasedness of $\tilde{\theta}$. Moreover, we can see that the soft estimate is accurate also at higher SNR values, while the hard estimator did not observe any error events for $\gamma > 7$ dB. We note that we have verified the result of the soft estimator at high SNR using the hard estimator with larger values of B . However, we do not show this result because we want to emphasize the difference between the two estimators for fixed B . Similarly, we could use the soft estimator to obtain results that are as accurate as those of the hard estimator while using fewer blocks, i.e., smaller B .

BLEP Estimation. Fig. 4b depicts the estimated BLEP as well as the upper and lower bounds (30) versus the SNR. First we can see that the hard and soft estimates do not overlap, i.e., the soft estimate is biased. This is because the memory of the channel and the convolutional code introduce dependencies between the bit errors, violating the assumption in (21). However, we can see that the soft estimate respects the bounds and we have verified that also the (unbiased) hard estimate is within the bounds (for large enough B). In fact, the true BLEP matches the lower bound in (30) at high SNR, thus confirming our intuition that in this case the least reliable bit in each block determines the BLEP. Furthermore, we note that the lower and upper bounds are within a few tenth of a dB for BLEP values of interest. We can moreover observe that the soft estimate almost coincides with the upper bound and we note that both upper bounds in (30) are equal for medium to high SNR (the union bound does not work well at low SNR). For BLEP estimation the advantage of the soft estimator is even more pronounced than in BEP estimation, except that we have to accept a (small) bias if the bit errors are not independent.

VII. CONCLUSIONS

We have studied blind estimation of the BEP and BLEP. The proposed “soft” estimators do not use the source data and outperform their “hard” counterparts by at least a factor of 2

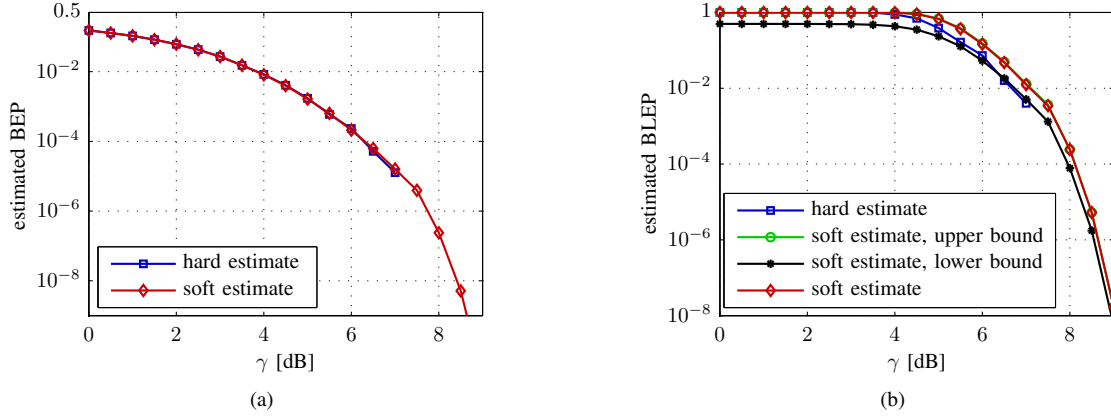


Figure 4: (a) Comparison of hard and soft BEP estimation. (b) Comparison of hard and soft BLEP estimation, including lower and upper bounds.

in terms of the MSE (in the Gaussian case the performance is improved by more than a factor of 4). Our estimators are fundamentally based on the consistency property of LLRs, which connects the two conditional pdfs $p(\ell|u)$ and the unconditional pdf $p(\ell)$ such that any one of the three is sufficient to express the other two. Besides error probability estimation this allows us for example to blindly estimate the mutual information $I(U; L)$. We have given conditions under which the proposed BEP estimator is the MVU estimator and we have studied BEP estimation with conditionally Gaussian distributed LLRs in detail. Our numerical results confirm that the proposed estimators entail significantly increased accuracy and simulation speed-up. An extension of our work to more general binary hypothesis testing problems seems to be possible and shall be studied in future work.

ACKNOWLEDGEMENTS

The authors are grateful to GÜNTHER KOLIANDER for his help with the proof of Proposition 5. This work was supported by the FWF Grant S10606 “Information Networks” and the WWTF Grant ICT08-44.

APPENDIX A: PROOF OF LEMMA 1

Let $\Lambda = e^L$ be the likelihood ratio. We write the conditional moments of Λ as

$$\begin{aligned}
 \mathbb{E}\{\Lambda^k(\mathbf{Y})|U=+1\} &= \int_{\mathbb{R}} \lambda^k p(\lambda|U=+1) d\lambda \\
 &= \int_{\mathbb{R}^K} \lambda^k(\mathbf{y}) p(\mathbf{y}|U=+1) d\mathbf{y} \\
 &= \int_{\mathbb{R}^K} \frac{p^k(U=+1|\mathbf{y})}{p^k(U=-1|\mathbf{y})} p(\mathbf{y}|U=+1) d\mathbf{y} \\
 &= \lambda_a^{-1} \int_{\mathbb{R}^K} \lambda^{k+1}(\mathbf{y}) p(\mathbf{y}|U=-1) d\mathbf{y} \\
 &= \lambda_a^{-1} \int_{\mathbb{R}} \lambda^{k+1} p(\lambda|U=-1) d\lambda \\
 &= \lambda_a^{-1} \mathbb{E}\{\Lambda^{k+1}(\mathbf{Y})|U=-1\},
 \end{aligned}$$

where $\lambda_a = e^{\ell_a} = p(U=+1)/p(U=-1)$. Since this holds for all $k \in \mathbb{N}$, it follows that $p(\lambda|U=+1) = p(\lambda|U=-1) \lambda/\lambda_a$. Now, Lemma 1 follows by noting that $p_\Lambda(\lambda) = p_L(\log(\lambda))/\lambda$.

APPENDIX B: PROOF OF PROPOSITION 2

We first write $\mathbb{E}\{\Xi_n^2\}$ as

$$\begin{aligned}
 \mathbb{E}\{\Xi_n^2\} &= \mathbb{E}\left\{\left(\frac{1}{1+\exp(|L_n|)}\right)^2\right\} \\
 &= \int_{-\infty}^0 \left(\frac{1}{1+\exp(-\ell_n)}\right)^2 p(\ell_n) d\ell_n \\
 &\quad + \int_0^{\infty} \left(\frac{1}{1+\exp(\ell_n)}\right)^2 p(\ell_n) d\ell_n. \quad (38)
 \end{aligned}$$

By induction it can be shown that the m th ($m \in \mathbb{N}$) derivative of $(1+\exp(\ell_n))^{-2}$ is

$$\frac{d^m}{d\ell_n^m} \left(\frac{1}{1+\exp(\ell_n)}\right)^2 = \sum_{k=1}^m (-1)^k d_{k,m} \frac{e^{k\ell_n}}{(1+e^{\ell_n})^{k+2}}, \quad (39)$$

where the coefficients $d_{k,m}$ are defined in (13). With (39) and (13) we can write the Taylor series expansion of $(1+\exp(\ell_n))^{-2}$ around $\ell_n = 0$ as

$$\left(\frac{1}{1+\exp(\ell_n)}\right)^2 = \frac{1}{4} + \sum_{m=1}^{\infty} \frac{\ell_n^m}{m!} \sum_{k=1}^m \frac{(-1)^k}{2^{k+2}} d_{k,m}. \quad (40)$$

Combining (38) and (40) yields (12).

APPENDIX C: PROOF OF PROPOSITION 5

First, we note that (28) is equivalent to

$$\prod_{n=1}^N (1-\theta_n) + \prod_{n=1}^N (1-\theta_n)^2 \geq 2 \prod_{n=1}^N \left(1 - \frac{3}{2}\theta_n\right). \quad (41)$$

We have equality in (41) if all $\theta_n = 0$. Let us next assume that at least one $\theta_n > 0$; without loss of generality let $\theta_N > 0$. We now show that in this case

$$\prod_{n=1}^N (1-\theta_n) + \prod_{n=1}^N (1-\theta_n)^2 > 2 \prod_{n=1}^N \left(1 - \frac{3}{2}\theta_n\right) \quad (42)$$

holds. The inequality in (42) obviously holds for $N = 1$ and the right-hand side can be written as

$$2 \left(1 - \frac{3}{2}x_N\right) \prod_{n=1}^{N-1} \left(1 - \frac{3}{2}x_n\right) < \left(1 - \frac{3}{2}x_N\right) \left(\prod_{n=1}^{N-1} (1 - x_n) - \prod_{n=1}^{N-1} (1 - x_n)^2 \right).$$

We thus have to show that

$$\left(1 - \frac{3}{2}x_N\right) \left[1 + \prod_{n=1}^{N-1} (1 - x_n)\right] \prod_{n=1}^{N-1} (1 - x_n) < \left[1 + \prod_{n=1}^N (1 - x_n)\right] \prod_{n=1}^N (1 - x_n),$$

which is equivalent to

$$\left(1 - \frac{3}{2}x_N\right) \left[1 + \prod_{n=1}^{N-1} (1 - x_n)\right] < (1 - x_N) \left[1 + \prod_{n=1}^N (1 - x_n)\right]. \quad (43)$$

Combining the terms in (43) and dividing by x_N yields the true statement

$$\left(\frac{1}{2} - x_N\right) \prod_{n=1}^{N-1} (1 - x_n) < \frac{1}{2}.$$

APPENDIX D: PROOF OF THEOREM 3

We have

$$p(r; \theta) = \frac{u(r)}{\sqrt{2\pi\sigma^2(\theta)}} \left(\exp\left(-\frac{1}{2\sigma^2(\theta)}(r - \sigma^2(\theta)/2)^2\right) + \exp\left(-\frac{1}{2\sigma^2(\theta)}(r + \sigma^2(\theta)/2)^2\right) \right),$$

where

$$\theta = Q\left(\frac{\sigma}{2}\right), \quad \sigma^2(\theta) = 4Q^{-2}(\theta). \quad (44)$$

Next, we have

$$\frac{\partial}{\partial \theta} \log p(r; \theta) = \frac{\partial}{\partial \sigma^2} \log p(r; \sigma^2) \frac{d}{d\theta} \sigma^2(\theta), \quad (45)$$

where

$$\frac{\partial}{\partial \sigma^2} \log p(r; \sigma^2) = -u(r) \left(\frac{1}{8} - \frac{r^2 - \sigma^2}{2\sigma^4} \right), \quad (46)$$

$$\frac{d}{d\theta} \sigma^2(\theta) = -16\sqrt{\frac{\pi}{2}} \exp(Q^{-2}(\theta)/2) Q^{-1}(\theta). \quad (47)$$

Now, combining (44)-(47) yields (34). In order to compute the Fisher information, we need to calculate

$$\frac{\partial^2}{\partial \theta^2} \log p(r; \theta) = \frac{\partial^2}{\partial \sigma^4} \log p(r; \sigma^2) \left(\frac{d}{d\theta} \sigma^2(\theta) \right)^2 + \frac{\partial}{\partial \sigma^2} \log p(r; \sigma^2) \frac{d^2}{d\theta^2} \sigma^2(\theta), \quad (48)$$

where

$$\frac{\partial^2}{\partial \sigma^4} \log p(r; \sigma^2) = u(r) \frac{\sigma^2/2 - r^2}{\sigma^6}, \quad (49)$$

$$\frac{d^2}{d\theta^2} \sigma^2(\theta) = 16\pi \exp(Q^{-2}(\theta))(1 + Q^{-2}(\theta)). \quad (50)$$

With (48)-(50) and rewriting everything in terms of θ , we obtain

$$\frac{\partial^2}{\partial \theta^2} \log p(r; \theta) = \frac{\pi \exp(Q^{-2}(\theta))}{2 Q^{-4}(\theta)} (r^2 Q^{-2}(\theta) - 3r^2 + 4Q^{-2}(\theta) - 8Q^{-4}(\theta) - 4Q^{-6}(\theta)) u(r). \quad (51)$$

Finally, integrating (51) with respect to $p(r; \theta)dr$ yields (36) and thus also the CRLB in (37).

The negative result regarding the existence of an efficient estimator in the Gaussian case is due to the fact that an efficient estimator $\tilde{\theta}_{\text{eff}}(\mathbf{r})$ exists iff one can write the score as [10]

$$v = \frac{\partial}{\partial \theta} \log p(\mathbf{r}; \theta) \stackrel{!}{=} J(\theta) [\tilde{\theta}_{\text{eff}}(\mathbf{r}) - \theta], \quad (52)$$

which is not possible in our case (a counterexample to (52) is readily found).

REFERENCES

- [1] S. Catreux, D. Gesbert, V. Erceg, and R. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Comm. Mag.*, vol. 40, no. 6, June 2002.
- [2] A. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Comm.*, vol. 46, no. 5, pp. 595–602, May 1998.
- [3] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1764–1777, Oct. 1999.
- [4] A. Sgora and D. Vergados, "Handoff prioritization and decision schemes in wireless cellular networks: a survey," *IEEE Comm. Surveys*, vol. 11, no. 4, pp. 57–77, 2009.
- [5] P. Hoeher, I. Land, and U. Sorger, "Log-likelihood values and monte carlo simulation - some fundamental results," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics, Brest, France*, Sept. 2000, pp. 43–46.
- [6] I. Land and P. Hoeher, "New results on monte carlo bit error simulation based on the a posteriori log-likelihood ratio," in *Proc. Int. Symp. on Turbo Codes & Rel. Topics, Brest, France*, Sept. 2003, pp. 531–534.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [8] A. Winkelbauer, "Moments and absolute moments of the normal distribution," Sept. 2012. [Online]. Available: <http://arxiv.org/abs/1209.4340>
- [9] K. Witzke and C. Leung, "A comparison of some error detecting CRC code standards," *IEEE Trans. Comm.*, vol. 33, no. 9, pp. 996–998, Sept. 1985.
- [10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs (NJ): Prentice Hall, 1993.
- [11] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, March 1974.