Fast Blind Recognition of Channel Codes

Reza Moosavi and Erik G. Larsson

Reza Moosavi and Erik G. Larsson, Fast Blind Recognition of Channel Codes, 2013, IEEE Transactions on Communications, (62), 5, 1393-1405.
http://dx.doi.org/10.1109/TCOMM.2014.050614.130297
Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-102536

Fast Blind Recognition of Channel Codes

Reza Moosavi and Erik G. Larsson

***Abstract*— We present a fast algorithm that, for a given input sequence and a linear channel code, computes the *syndrome posterior probability* (SPP) of the code, i.e., the probability that all parity check relations of the code are satisfied. According to this algorithm, the SPP can be computed blindly, i.e., given the soft information on a received sequence we can compute the SPP for the code without first decoding the bits. We show that the proposed scheme is efficient by investigating its computational complexity.**

We then consider two scenarios where our proposed SPP algorithm can be used. The first scenario is when we are interested in finding out whether a certain code was used to encode a data stream. We formulate a statistical hypothesis test and we investigate its performance. We also compare the performance of our scheme with that of an existing scheme. The second scenario deals with how we can use the algorithm for reducing the computational complexity of a blind decoding process. We propose a heuristic sequential statistical hypotheses test to use the fact that in real applications, the data arrives sequentially, and we investigate its performance using system simulations.

***Index Terms*—Blind code detection; Blind decoding; Control signaling; Sequential probability ratio test.**

I. Introduction

TODAY'S wireless access systems need to offer high throughput. At the same time, as more and more users join the system, the channel resources (such as bandwidth) have become scarce. Many sophisticated algorithms have been devised to cope with the situation. A common approach is to use adaptive modulation and coding (AMC) [2], i.e., instead of using fixed transmission parameters, the transmitter changes the modulation format and coding rate on the fly in order to adapt to a changing channel quality. To support this, there is typically a need for a control channel on which the AMC parameters are signaled. However, it has recently been shown that AMC can be achieved without explicitly signaling the parameters, using so-called *blind decoding*. The idea is that the receiver tries to blindly identify the AMC parameters from the data collected from the channel. For instance in [3]–[5] novel schemes for blind classification of modulation formats have been proposed. In [6]–[9], blind identification of encoder parameters have been studied. In these studies, the receiver uses some properties of the channel code such as algebraic properties of the parity check matrix or a recursive structure of the encoder (that happens for example for convolutional codes) to blindly identify the encoder parameters. For an illustration of how blind decoding is implemented in practice, see [10, Section 16.4], where the procedure for the physical downlink control channel (PDCCH) decoding in LTE is described.

Adaptive modulation and coding using blind decoding comes at the price of a decoding delay and more importantly energy consumption in the decoder on the receiver side. Given that the receiver is a mobile device with limited battery capacity, the latter is of some concern and any reduction in the decoding complexity incurred by the blind decoding strategy would be valuable.

In this paper, we are concerned with finding which channel code out of a possible set of general linear channel codes was used to encode the data. This problem is therefore different from [3]–[9] in the sense that (i) we assume that the modulation format is known a priori, and (ii) the objective is to *recognize* or *verify* which one of the channel codes out of the possible codes (which is denoted by the "candidate set" from hereon) was used to encode the data stream. To the best of our knowledge, there is very little work in the literature addressing this problem in its generality. For instance in [11], [12], two blind schemes for recognition of space-time block codes and LDPC codes were proposed, respectively. The proposed algorithm therein, after intercepting a number of code blocks, computes the likelihood of each code candidate and picks the most likely one. However, these schemes can only be applied for recognition of specific codes. In [13], an algorithm for blind recognition of a linear code in a binary symmetric channel (BSC) was proposed. In order to determine if a certain code with a given parity check matrix was used to encode the data, the author therein proposed to first take hard decisions on the received data to get a rough estimate of the transmitted codeword, and then to find the inner products between the estimated codeword and the rows of the parity check matrix and use this quantity to determine whether the data was encoded with the given channel code or not. We use this latter scheme as a benchmark in our comparison.

A. Contribution

We present a fast algorithm that, for a given input sequence and a given linear channel code, computes the probability that all the parity check relations of the code are satisfied. We call this probability the *syndrome posterior probability* (SPP) of the code. Using this algorithm, the SPP can be computed blindly, i.e., given the soft information on a received sequence we can compute the SPP for the code without first decoding

The associate editor coordinating the review of this paper and approving it for publication was Prof. T.-K. Truong. Manuscript received April 22, 2013; revised November 15, 2013, January 23 and March 20, 2014.

R. Moosavi was with the Dept. of Electrical Engineering (ISY), Linköping University, Linköping, Sweden. He is now with Ericsson Research, Linköping, Sweden (e-mail: reza.moosavi@ericsson.com).

E. G. Larsson is with the Dept. of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (e-mail: egl@isy.liu.se).

This work was supported in part by the Swedish research council (VR) and the Excellence Center at Linköping-Lund in Information Technology (ELLIIT).

Parts of the material in this paper were presented at the IEEE GLOBECOM 2011 conference [1].

Digital Object Identifier 10.1109/TCOMM.2014.09.130297

the bits. We show that the proposed scheme for obtaining the SPP is efficient by investigating the computational complexity of the scheme. We also show that under some conditions, we can approximate the log-likelihood ratio (LLR) of SPP with a normal distribution and we then compute the mean and the variance of it. This approximation gives us quantitative insight on how the SPP behaves.

We then consider two scenarios where our proposed SPP algorithm can be used. The first scenario is when we are interested in finding out whether a certain code was used to encode a data stream. We formulate a statistical hypothesis test and we investigate the performance of the proposed test. The second scenario deals with how we can use the proposed algorithm for reducing the computational complexity of the blind decoding process. We propose a heuristic sequential statistical hypotheses test (SSHT) to use the fact that in real applications, the data arrives in a sequential manner, and we investigate its performance using system simulations. The paper extends our conference paper [1], among others by proposing the optimum statistical hypothesis test for the detection problem and also the proposed SSHT test.

II. Computing the Syndrome Posterior Probability

In this section, we provide a fast algorithm for computing the syndrome posterior probability (SPP) for a given code. More precisely, given the soft information for an arbitrary sequence of encoded bits $\boldsymbol{c}$, we compute the probability that all the parity check relations for the code are satisfied. A preliminary, slightly different version of the proposed algorithm was introduced initially in our conference paper [1] for detecting the presence of a channel code. A modified version of this algorithm has subsequently been used to detect an additional lonely bit piggybacked on a linearly encoded data stream [14]. Also, a similar algorithm has been used in [15], [16] in the context of blind frame synchronization. Since this scheme provides a basis for our further discussions, we will present it again in this section. We will also use the same technique as used in [16] to find analytical closed-form expressions for the probabilities of false alarm and missed detection in Section III. However, in addition to the work in [16], we also compute the cross-correlations that we need in using the central limit theorem (see Appendices A and B).

Consider a general communication link depicted in Figure 1. The information bits $\boldsymbol{b} = [b_1, \ldots, b_K]$ are first encoded to obtain a sequence of coded bits $\boldsymbol{c} = [c_1, \ldots, c_N]$, (generally $N > K$). The coded bits are then transmitted using a certain modulation scheme. Upon reception of the received vector $\boldsymbol{r}$, the receiver computes the soft information $\boldsymbol{\ell} = [\ell_1, \ldots, \ell_N]$ for the transmitted bits $\boldsymbol{c}$. The soft information ℓ_i for the ith encoded bit c_i is usually presented as the posterior log-likelihood ratio (LLR), that is,

$$\ell_i = \log\left(\frac{\Pr(c_i = 0|\boldsymbol{r})}{\Pr(c_i = 1|\boldsymbol{r})}\right).$$

The proposed scheme for computing the SPP for a given code uses the fact that any codeword $\boldsymbol{c}$ obtained from the channel code with parity check matrix $\boldsymbol{H}$, satisfies $\boldsymbol{Hc} = \boldsymbol{0}$.[1] That is, all the codewords of this code satisfy $N - K$ parity check relations[2] of the form

$$\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0, \quad \text{for} \;\; i = 1, \ldots N - K,$$

where h_{il} is the index of the lth nonzero element of the ith row of the parity check matrix $\boldsymbol{H}$. Since the encoded bits may be corrupted during the transmission, if we would take hard decisions on $\boldsymbol{\ell}$ to obtain estimates $\hat{\boldsymbol{c}}$ of the coded bits, $\boldsymbol{H}\hat{\boldsymbol{c}}$ may not be a zero vector even if the channel code with parity check matrix $\boldsymbol{H}$ was used to encode the data. However, given the soft information on $\boldsymbol{c}$, we can compute the SPP as follows. We are interested in finding

$$\begin{aligned}\Gamma &\triangleq \Pr\left(\text{all syndrome checks are satisfied}|\boldsymbol{r}\right)\\ &= \Pr\left(\bigcap_{i=1}^{N-K} \underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0\Big|\boldsymbol{r}\right) \approx \prod_{i=1}^{N-K}\Pr\left(\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0\Big|\boldsymbol{r}\right),\end{aligned} \tag{1}$$

where we assumed in the last step that the syndrome checks are independent in the sense that the two events $\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0$ and $\underset{l}{\oplus}\boldsymbol{c}_{h_{i'l}} = 0$ are independent for $i \neq i'$, given $\boldsymbol{r}$. This independence assumption should be justifiable for long observation sequences and for channel codes with sparse parity check matrices, but in any case it is not crucial for the upcoming discussion.

The LLR associated with the ith syndrome check of the code is given by

$$\gamma_i \triangleq \log\left(\frac{\Pr\left(\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0\Big|\boldsymbol{r}\right)}{1 - \Pr\left(\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0\Big|\boldsymbol{r}\right)}\right) = \underset{l}{\boxplus}\ell_{h_{il}} \tag{2}$$

where $\boxplus$ denotes the *box-plus* operation [17].[3] From (2), we have that

$$\Pr\left(\underset{l}{\oplus}\boldsymbol{c}_{h_{il}} = 0\Big|\boldsymbol{r}\right) = \log\left(\frac{e^{\gamma_i}}{1 + e^{\gamma_i}}\right). \tag{4}$$

Using (4) in (1) and taking the logarithm yields

$$\log(\Gamma) \approx \sum_{i=1}^{N-K}\log\left(\frac{e^{\gamma_i}}{1 + e^{\gamma_i}}\right) = -\sum_{i=1}^{N-K}\log(1 + e^{-\gamma_i}). \tag{5}$$

We call Γ the syndrome posterior probability (SPP) of the code.

A. Computational Complexity of Computing SPP

In order to compute the SPP for a given code, we need to compute (2) and (5), respectively. We examine each computation separately. First note that the box-plus operation can be equivalently expressed as [18, Eq. (14)]

$$\begin{aligned}&\ell_1 \boxplus \ell_2 = \\ &\text{sign}(\ell_1)\text{sign}(\ell_2)\min(|\ell_1|, |\ell_2|) + f(|\ell_1 + \ell_2|) - f(|\ell_1 - \ell_2|),\end{aligned} \tag{6}$$

[1]The computations are carried out in binary field $\mathbb{F}_2$.

[2]since $\boldsymbol{H}$ is an $(N - K) \times N$ matrix.

[3]More precisely, the definition of $\boxplus$ is:

$$\ell_1 \boxplus \ell_2 \triangleq \log\left(\frac{1 + \tanh(\ell_1/2)\tanh(\ell_2/2)}{1 - \tanh(\ell_1/2)\tanh(\ell_2/2)}\right) \tag{3}$$

with $\ell \boxplus \infty = \ell$, $\ell \boxplus -\infty = -\ell$ and $\ell \boxplus 0 = 0$, see [17] for more details.

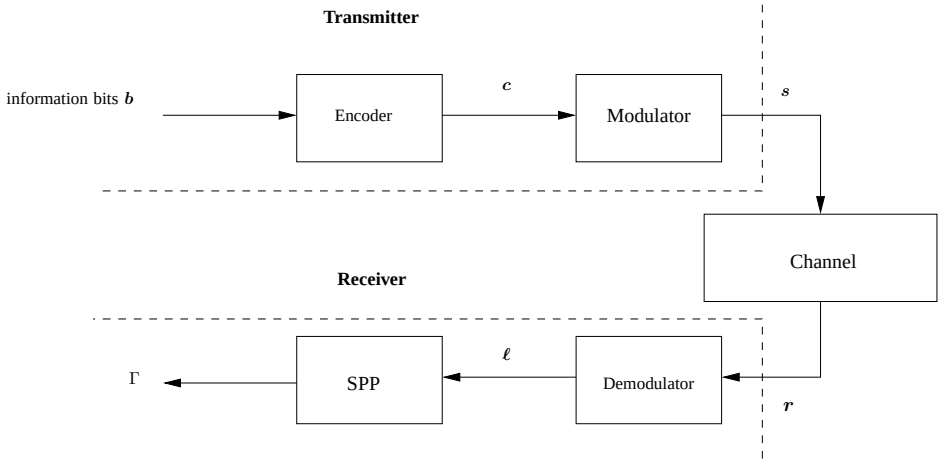


Fig. 1. Schematic for a general communication link. The proposed scheme for computing syndrome posterior probability works on the soft information $\boldsymbol{\ell}$ and computes the probability that all the parity check relations for a specific code is satisfied.

where $f(x) \triangleq \log(1+e^{-x})$. The function $f(x)$ is very well-behaved and can be computed efficiently.[4] Moreover, the box-plus operator has the associative property, that is

$$\ell_1 \boxplus \ell_2 \boxplus \ell_3 = (\ell_1 \boxplus \ell_2) \boxplus \ell_3,$$

which can be used to compute (2) recursively. This means that the computational complexity of computing γ_i is equal to the number of nonzero elements of the ith row of the parity check matrix, say J_i, and is generally small.[5]

Now looking at the second computation (5), we observe that it can also be computed efficiently using the same function $f(x)$ described above. That is, the second operation also has linear complexity with respect to the number of terms in the corresponding expression. Therefore, the total number of computations required to obtain the SPP is $\sum_{i=1}^{N(1-R)} \big(J_i+1\big)$, where $R = K/N$ is the code rate. The overall computational complexity of finding the SPP is thus $\mathcal{O}(N)$.

B. Parity Check Matrices and the SPP

We need to know the parity check matrix of the code in order to find the SPP for that code. The parity check matrix for block codes can be obtained from the corresponding generator matrix. For convolutional codes, the parity check matrix can be obtained from the syndrome former of the code [19]. For LDPC codes, the parity relations are obtained from the code graph.

Note that is some situations the transmitted bits $\boldsymbol{c}$ consist of several smaller codewords. This is especially true for linear block codes with low dimensions (such as Hamming codes). More specifically, consider an (n,k) linear block code with a given parity check matrix $\tilde{\boldsymbol{H}}$. For a received vector of length N consisting of N/n codewords, we can define the overall parity check matrix $\boldsymbol{H}$ consisting of N/n smaller and identical $(n-k)\times n$ matrices $\tilde{\boldsymbol{H}}$:

$$\boldsymbol{H} = \begin{bmatrix} \tilde{\boldsymbol{H}} & & & \\ & \tilde{\boldsymbol{H}} & & \\ & & \ddots & \\ & & & \tilde{\boldsymbol{H}} \end{bmatrix}. \tag{7}$$

This is very useful for the approximations in the next section.

C. Approximation of the SPP

Using (5), we have

$$\Gamma = \exp\left(\sum_{i=1}^{N-K} \log\left(\frac{e^{\gamma_i}}{1+e^{\gamma_i}}\right)\right) = \prod_{i=1}^{N-K} \frac{e^{\gamma_i}}{1+e^{\gamma_i}}, \tag{8}$$

and thus the LLR associated with the SPP Γ is given by,

$$\begin{aligned} \Lambda(\Gamma) &\triangleq \log\left(\frac{\Gamma}{1-\Gamma}\right) \\ &= \sum_{i=1}^{N-K} \gamma_i - \log\left(\prod_{i=1}^{N-K}(1+e^{\gamma_i}) - \exp\left(\sum_{i=1}^{N-K}\gamma_i\right)\right). \end{aligned} \tag{9}$$

If $e^{\gamma_i} \gg 0$ (which happens, for instance, when the operating SNR is high and the associated binary value is 0), the above expression can be approximated as

$$\Lambda(\Gamma) \approx \sum_{i=1}^{N-K} \gamma_i. \tag{10}$$

This approximation has been used in many works, mainly for reducing the computational complexity, since it does not significantly affect the performance [15], [16].

The box-plus operation can also be approximated using the well-known approximation [17]

$$\boxplus_{i=1}^{n} \ell_i \approx \left(\prod_{i=1}^{n} \text{sign}(\ell_i)\right) \min_{i=1,\ldots,n} |\ell_i|. \tag{11}$$

We will use the two approximations above later on to simplify some of our expressions, see Section III. It is worth noting that in [18], other better methods to approximate box-plus were proposed. These approximation methods can be used to compute the equations involving box-plus operations as in (2) more accurately.

[4]For $x>0$, $f(x)$ can be tabulated with arbitrary precision. For $x<0$, we can write $f(x) = -x + \log(1+e^{x})$ and then use the same table lookup.

[5]In practice, the number of nonzero elements in different rows need not be the same, that is, $J_i \neq J_{i'}$, for $i \neq i'$. However in many situations, all J_i are indeed equal (such as for regular LDPC codes, where J_i is given by the degree distribution of the code, or for convolutional codes).

III. Using SPP for Blindly Identifying a Channel Code

In this section, we will consider the problem of detecting whether a certain channel code was used to encode a sequence of received baseband data or not. That is, given a soft information $\boldsymbol{\ell}$ associated with coded bits $\boldsymbol{c}$, we would like to determine if a specific code, say the channel code $\mathcal{C}_t$ with parity check matrix $\boldsymbol{H}$, was used to encode the data or not. We study this problem by considering the following hypothesis test. We consider two a priori equally likely hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. Under hypothesis $\mathcal{H}_1$, we assume that the channel code $\mathcal{C}_t$ was used to encode the data stream, whereas under $\mathcal{H}_0$, we assume that the data stream is not encoded with any channel code but instead the transmitted bits are i.i.d. and may take 0 or 1 with equal probability. The specific choices of modulation on the communication link does not affect our discussion and thus we assume BPSK modulation over an AWGN channel throughout the rest of the paper[6], and we define the signal-to-noise ratio (SNR) to be the average transmitted power divided by the noise variance.

The supporting rationale behind the assumption of i.i.d. bits under hypothesis $\mathcal{H}_0$ lies in the fact that if $\boldsymbol{c}$ is obtained from some channel code $\mathcal{C}'$ not equal to $\mathcal{C}_t$ with parity check matrix $\boldsymbol{H}'$ ($\boldsymbol{H}' \neq \boldsymbol{H}$), and if we construct the vector $\boldsymbol{H}\boldsymbol{c}$, it will have almost no structure such that we can assume that it contains i.i.d. bits that take 0 or 1 with equal probability. The implicit assumption here is that the two channel codes are "distinct" in the sense that their corresponding parity check matrices do not have any identical rows. However, the proposed algorithm also works when we are interested in distinguishing between two channel codes with parity check matrices that share a few identical rows. The way to tackle these scenarios is to exclude the common rows since they impose the same parity check relations on the coded bits, and only consider the distinct parity check relations. It is worth noting that many codes in practice have indeed different (non-overlapping) parity check matrices.

Another way to resolve the situations with similar channel codes is to use different interleavers for each code which essentially results *almost always* in distinct parity check matrices. As an example, consider the case with two channel codes: (i) the channel code $\mathcal{C}$ with a given parity check matrix $\boldsymbol{H}$, and (ii) a randomly permuted version of $\mathcal{C}$ which is obtained by first encoding the information bits with $\mathcal{C}$ followed by a random interleaving of the coded bits. Let $\boldsymbol{H}'$ denote the parity check matrix of the interleaved code. Consider now the kth row of $\boldsymbol{H}$ and assume that the number of nonzero elements of this row is J_k, with the parity check relation $\bigoplus_l \boldsymbol{c}_{h_{kl}} = 0$. Assuming that all the $N!$ possible interleaver sequences are equally likely, then by the union bound the probability that $\boldsymbol{H}'$ has an identical row to the kth row of $\boldsymbol{H}$ is at most $I_{J_k} \frac{J_k!(N-J_k)!}{N!}$, where N denotes the number of coded bits $\boldsymbol{c}$, and I_{J_K} denotes the number or rows that have J_k nonzero elements. Since I_{J_k} is at most $N-K$, this probability is upper bounded by $(N-K)\frac{J_k!(N-J_k)!}{N!}$ and is vanishingly small[7], and hence we can assume that the two codes have no overlapping syndrome checks.

Having computed the SPP for the code Γ, the optimal test for detecting the presence of $\mathcal{C}_t$ is

$$\Lambda(\Gamma) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{12}$$

We can use the approximation (10) to simplify the test. Using this approximation, the resulting test becomes

$$\sum_{i=1}^{N-K} \gamma_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{13}$$

This suboptimal test coincides with the test that was proposed in our conference paper [1].

A. Analysis of the Code Detection Performance

In order to analyze the performance of our detection algorithm, we need to know the probability distribution of $\Lambda(\Gamma)$ under each hypothesis. The distributions are not known in closed-form. However, using the suboptimal test (13) in combination with the approximations provided in Section II for the box-plus operation, we can use the central limit theorem (CLT) to approximate the probability distribution of $\Lambda(\Gamma)$ under the two hypotheses in some cases. More precisely, according to the CLT for a sequence of identically distributed weakly stationary random variables $\{Z_k\}$ with mean m_Z and variance σ_Z^2, if

$$V \triangleq \sigma_Z^2 + 2\sum_{i=2}^{\infty} \mathrm{cov}(Z_1, Z_i) < \infty, \tag{14}$$

then $\frac{1}{K}\sum_{k=1}^{K} Z_k$ approaches a Gaussian random variable with mean m_Z and variance V as K increases [20]. In the case of i.i.d. random variables $\{Z_k\}$ this reduces to the law of large number, i.e., $\frac{1}{K}\sum_{k=1}^{K} Z_k$ can be approximated with a Gaussian distribution with mean m_Z and variance σ_Z^2, as K increases.

Now consider the ith syndrome check constraint. Approximating γ_i using (11) and as our analysis in Appendix A shows, the probability distribution of γ_i depends only on the operating SNR and on the number of elements in the corresponding box-plus operator (2), i.e., the number of ones in the ith row of the parity check matrix J_i. That is, all the rows of the parity check matrix that have the same number of nonzero elements produce identically distributed syndrome check constraints. Therefore, if there are *sufficiently many* rows with J_i nonzero elements, then we can approximate their corresponding summation with a Gaussian distribution given that the condition for the CLT are satisfied. The idea is thus to split the γ_i into different sets where the γ_is in each set have the same number of terms in their corresponding box-plus operator (in other words, they corresponds to the rows with the same number of nonzero elements) and then apply the CLT to each set to approximate the summation with a Gaussian distribution. Note that the γ_is in each set

[6]This also simplifies the procedure that we will use later to derive approximate closed-form expressions for the probabilities of false alarm and missed detection.

[7]For instance for rate-1/2 code with $N = 100$ and $J_k = 7$, the probability is roughly 10^{-10}.

are identically distributed. However in order to use the CLT, two additional conditions are required: (i) the γ_is must be stationary (in the weak sense), and (ii) the inequality (14) must hold. As the analysis in Appendix B shows, $\text{cov}(\gamma_i, \gamma_{i'})$ depends on the SNR, J_i, $J_{i'}$ and on the number of common terms in the corresponding box-plus operations for γ_i and $\gamma_{i'}$, say $\lambda_{i,i'}$. As $\lambda_{i,i'}$ decreases, $\text{cov}(\gamma_i, \gamma_{i'})$ also decreases and hence, if $\lambda_{i,i'}$ is small, we can assume that γ_i and $\gamma_{i'}$ are uncorrelated.

As an example consider LDPC codes. For these codes, due to the sparsity of the parity check matrix, we can assume that the γ_is are independent[8] and hence based on the degree distribution, we can split them into different sets as mentioned above with each set containing the portion of γ_is that correspond to the rows with equal J_i. Let there be T such sets and let N_t, $t = 1, \ldots, T$ denote the number of elements in each set (obviously $N_1 + \ldots + N_T = N$). Then we can write

$$\Lambda(\Gamma) \approx \sum_{t=1}^{T} \sum_{j=1}^{N_t} \gamma_j^{(t)} \tag{15}$$

where $\gamma_j^{(t)}$, $j = 1, \ldots, N_t$ denotes γ_is in set t. If N_t is sufficiently large,[9] then we can approximate $\sum_{j=1}^{N_t} \gamma_j^{(t)}$ with a Gaussian distribution with mean zero and variance $N_t {\sigma_0^{(t)}}^2$ under $\mathcal{H}_0$ and with mean $N_t m_r^{(t)}$ and variance $N_t {\sigma_r^{(t)}}^2$ under $\mathcal{H}_1$ using the law of large numbers. Consequently, $\Lambda(\Gamma)$ can also be approximated as a Gaussian random variable with mean zero and variance

$$\sigma_0^2 = \sum_{t=1}^{T} N_t {\sigma_0^{(t)}}^2$$

under $\mathcal{H}_0$, and with mean and variance,

$$m_r = \sum_{t=1}^{T} N_t m_r^{(t)}, \qquad \sigma_r^2 = \sum_{t=1}^{T} N_t {\sigma_r^{(t)}}^2$$

under $\mathcal{H}_1$, respectively. Note that ${\sigma_0^{(t)}}^2$, $m_r^{(t)}$ and ${\sigma_r^{(t)}}^2$ can be obtained using the analysis in Appendix A. Also note that for regular LDPC codes, $T = 1$ which simplifies the expressions above.

As another example, let us consider convolutional codes. Since convolutional codes have memory, γ_i are not independent and hence the approximation is slightly more complicated. However, we still can approximate $\Lambda(\Gamma)$ with a Gaussian random variable in some situations. We explain this via an example. Consider the standard rate-1/2 convolutional code with constraint length $C = 4$, depicted in Figure 2. Using the syndrome former of the code, we get the following parity check matrix[10]

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & & & & \\ & & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & & \\ & & & \ddots & & \ddots & & \ddots & & \ddots & & \\ & & & & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \tag{16}$$

[8]In other words $\lambda_{i,i'}$ is small for any i and i', with $i \neq i'$.

[9]in the order of 100 as we have seen from our numerical experiments.

[10]The rest of the entries in the matrix are zero.

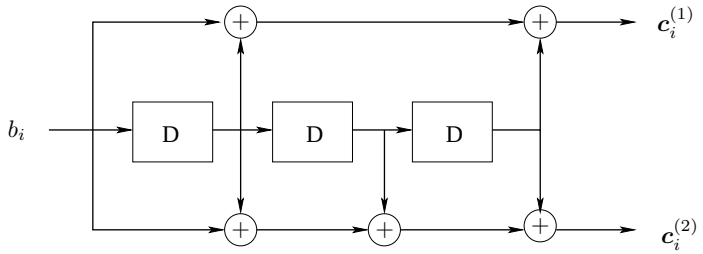


Fig. 2. The standard rate-1/2 convolutional code with constraint length 4. The generators for this code are $g_1 = 15$ and $g_2 = 17$ in octal.

and hence by arranging the coded bits and defining a window $\boldsymbol{S}_i$ as follows,

$$\begin{array}{ccccccccccc} & & & & & \triangleq \boldsymbol{S}_i & & & & & \\ \cdots & \boldsymbol{c}_{i-1}^{(1)} & \boldsymbol{c}_i^{(1)} & \boldsymbol{c}_{i+1}^{(1)} & \boldsymbol{c}_{i+2}^{(1)} & \boldsymbol{c}_{i+3}^{(1)} & \boldsymbol{c}_{i+4}^{(1)} & \cdots \\ \cdots & \boldsymbol{c}_{i-1}^{(2)} & \boldsymbol{c}_i^{(2)} & \boldsymbol{c}_{i+1}^{(2)} & \boldsymbol{c}_{i+2}^{(2)} & \boldsymbol{c}_{i+3}^{(2)} & \boldsymbol{c}_{i+4}^{(2)} & \cdots \end{array}$$

we get N_b syndrome check constraints

$$\bigoplus_{(i,j)\in\boldsymbol{S}_i} \boldsymbol{c}_i^{(j)} = 0, \qquad \text{for } i = 1, \ldots, N_b,$$

where N_b denotes the number of information bits. Thus, by defining

$$\gamma_i = \boxplus_{(i,j)\in\boldsymbol{S}_i} \ell_i^{(j)},$$

and using (5), the SPP for this code can be found directly. We use the following convention to specify the window $\boldsymbol{S}_i$, which will be used later in Section V,

$$\{\boldsymbol{c}_i^{(1)} \to (0,1,2,3), \quad \boldsymbol{c}_i^{(2)} \to (0,2,3)\} \tag{17}$$

meaning that the syndrome check constraint at time instant i consists of coded bits $\boldsymbol{c}_{i+j}^{(1)}$ and $\boldsymbol{c}_{i+j'}^{(2)}$ where $j \in \{0,1,2,3\}$ and $j' \in \{0,2,3\}$.

For this code, all the rows of the parity check matrix have $J = 7$ nonzero elements, and thus all γ_i are identically distributed. Moreover, due to the recursive nature of the code, $\text{cov}(\gamma_i, \gamma_{i'})$ depends only on the difference $|i - i'|$. In fact for this example, since $\boldsymbol{S}_i$ and $\boldsymbol{S}_{i'}$ share no common coded bits when $|i - i'| \geq 4$, and since the transmitted bits are statistically independent of each other, γ_i and $\gamma_{i'}$ are independent for $|i - i'| \geq 4$. That is, $\text{cov}(\gamma_1, \gamma_i) = 0$, for $i \geq 4$. This enables us to approximate $\Lambda(\Gamma)$ with a Gaussian random variable in this case too.[11]

As the examples above showed, we can in many situations approximate the distribution of $\Lambda(\Gamma)$ with a Gaussian distribution under the two hypotheses. According to Appendix A, under $\mathcal{H}_1$, $\mathbb{E}\{\gamma_i\}$ depends on the operating SNR and is inversely proportional to J_i (see (38)–(42)). This is intuitive, since both decreasing the SNR and having many terms in the ith parity check relation $\oplus_l \boldsymbol{c}_{h_{il}} = 0$ will increase the risk of error in the received sequence and hence $\Pr\left(\oplus_l \boldsymbol{c}_{h_{il}} = 0 \middle| \boldsymbol{r}\right)$ decreases. This is also seen from Figure 3, where we have plotted $\mathbb{E}\{\gamma_i\}$ under $\mathcal{H}_1$ as a function of the SNR for different values of J_i. Another important observation from equation

[11]Note that in this case, γ_is are stationary as discussed above.

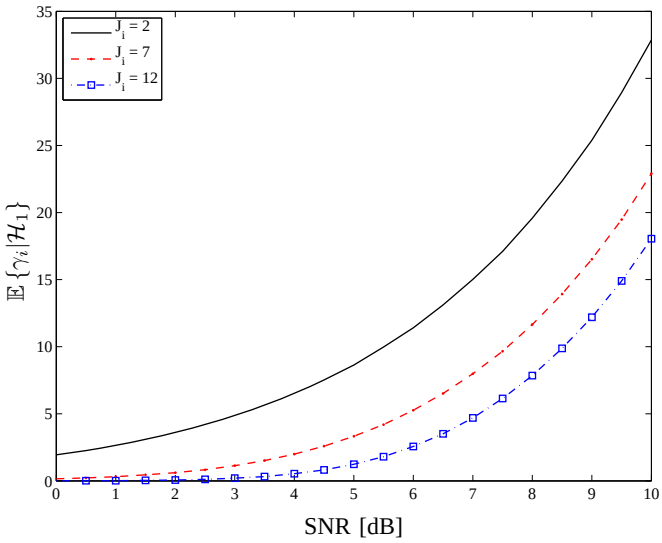


Fig. 3. $\mathbb{E}\{\gamma_i\}$ under $\mathcal{H}_1$ as a function of SNR for different values of J_i

(38) is that at high SNR regime, $\mathbb{E}\{\gamma_i\}$ scales linearly with SNR. This can also be seen from Figure 3.

Having computed the mean and variance under each hypothesis, we can approximate the false alarm and the detection probabilities as,

$$P_F \approx \Pr\left\{\sum_{i=1}^{K}\gamma_i > \eta|\mathcal{H}_0\right\} = \mathrm{Q}\left(\frac{\eta}{\sigma_0}\right), \tag{18}$$

$$P_D \approx \Pr\left\{\sum_{i=1}^{K}\gamma_i > \eta|\mathcal{H}_1\right\} = \mathrm{Q}\left(\frac{\eta - m_r}{\sigma_r}\right), \tag{19}$$

where σ_0^2 denotes the variance under $\mathcal{H}_0$ and m_r and σ_r^2 denote the mean and the variance under $\mathcal{H}_1$, and $\mathrm{Q}(x)$ is the Gaussian error integral (Q-) function, defined as

$$\mathrm{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\mathrm{d}t.$$

The receiver-operating-characteristics (ROC) is, therefore, given by

$$P_D = \mathrm{Q}\left(\frac{\sigma_0 \mathrm{Q}^{-1}(P_F) - m_r}{\sigma_r}\right). \tag{20}$$

Note that in the above expression, σ_0 and σ_r scale as $\sqrt{N}$ whereas m_r scales as N.

We stress that while the analysis provided in this section is restricted to some special cases, the proposed detection algorithm as such is applicable to any linear channel code without any modification including the tail-biting convolutional codes; the corresponding syndrome checks have to be written down in each specific case. The analysis provided in this section can provide, for instance, a rule-of-thumb for the required number of observations in order to achieve certain false alarm and detection probabilities. As for the performance with different rates, typically lower-rate codes are *easier* to detect than higher-rate codes, because there are more syndrome checks for lower-rate codes, and hence more reliable decisions can be made.

B. Application of the Code Detection Scheme to Convolutional Codes

In this section, we investigate the performance of our detection scheme for standard rate-1/2 convolutional codes. In particular, we consider three choices for the *true* channel code: (a) $\mathcal{C}_2$ with constraint length 4 (depicted in Figure 2), (b) $\mathcal{C}_5$ with constraint length $C_2 = 7$, and (c) $\mathcal{C}_7$ with constraint length 9. The generators for these codes are given in Table I. For each scenario, we let hypothesis $\mathcal{H}_0$ denote the hypothesis under which the transmitted bits are i.i.d. and hypothesis $\mathcal{H}_1$ denote the hypothesis under which the corresponding convolutional code is used to encode the data. We assume that both hypotheses are equally likely a priori and that the coded bits are transmitted over an AWGN channel using BPSK modulation. Let N_b denote the number of information bits. The syndrome check constraints for each code is given in Table I.

As a benchmark for comparison, we also implemented the test proposed in [13]. However, since the algorithm therein can only be used in binary symmetric channels (BSC), we consider an equivalent BSC channel with cross-over probability equal to the error probability of the AWGN channel in our model with BPSK modulation. Figures 4(a)–4(c) illustrate the ROC curves of the different schemes for the case where the SNR is 0 dB and for two different values of N_b for the three convolutional codes, respectively. More precisely, we have provided ROC curves for (i) the statistical test given by (12), (ii) the suboptimal test given by (13), (iii) the suboptimal test where in addition to (13), we also used the approximation given by (11) in computing γ_i, (iv) our analysis provided by (20), and (v) the test using the algorithm in [13].[12] As the results show, the proposed statistical test (i) has better performance compared to the other tests, specially compared to the test in [13]. This is reminiscent of hard-decision decoding in an AWGN channel which is known to be roughly 2 dB worse than soft-decision decoding [21, pp. 612]. Additionally, we see that the analysis provided by (20) is very close, specially when N_b is large, to the empirical performance of the corresponding suboptimal test (case (iii)). We also see, as expected that by increasing the number of information bits, a better detection performance is achieved. Also we see by comparing the ROC curves for the three scenarios that it is *easier* to recognize the convolutional code $\mathcal{C}_2$. This is expected since for this code the number of non-zero elements in the parity check matrix is smaller than for the other two convolutional codes with constraint lengths 7 and 9. The mean values according to our analysis in Section III-A are 0.154, 0.0411, and 0.017 for the three codes $\mathcal{C}_2$, $\mathcal{C}_5$ and $\mathcal{C}_7$, respectively.

IV. Using SPP for Reducing The Computational Complexity of Blind Decoding

In a system where blind decoding is used, the receiver is interested in blindly detecting the channel code used by the transmitter as early as possible with a given probability

[12] Similar simulation results, but without the curves representing the optimal test, the analysis and the comparison with [13], were provided in our conference paper [1]. However, the definition of SNR given there contained an error.

Channel Code	Const. Length	Generators	Syndrome Check Constraint
$\mathcal{C}_1$	3	(5,7)	$\{\boldsymbol{c}_i^{(1)} \to (0,1,2), \boldsymbol{c}_i^{(2)} \to (1,2)\}$
$\mathcal{C}_2$	4	(15,17)	$\{\boldsymbol{c}_i^{(1)} \to (0,1,2,3), \boldsymbol{c}_i^{(2)} \to (0,2,3)\}$
$\mathcal{C}_3$	5	(23,35)	$\{\boldsymbol{c}_i^{(1)} \to (0,2,3,4), \boldsymbol{c}_i^{(2)} \to (0,1,4)\}$
$\mathcal{C}_4$	6	(53,75)	$\{\boldsymbol{c}_i^{(1)} \to (0,2,3,4,5), \boldsymbol{c}_i^{(2)} \to (0,1,3,5)\}$
$\mathcal{C}_5$	7	(133,171)	$\{\boldsymbol{c}_i^{(1)} \to (0,3,4,5,6), \boldsymbol{c}_i^{(2)} \to (0,1,3,4,6)\}$
$\mathcal{C}_6$	8	(247,371)	$\{\boldsymbol{c}_i^{(1)} \to (0,3,4,5,6,7), \boldsymbol{c}_i^{(2)} \to (0,1,2,5,7)\}$
$\mathcal{C}_7$	9	(561,753)	$\{\boldsymbol{c}_i^{(1)} \to (0,1,3,5,6,7,8), \boldsymbol{c}_i^{(2)} \to (0,4,5,6,8)\}$

TABLE I
STANDARD RATE-1/2 CONVOLUTIONAL CODES WITH THEIR GENERATORS IN OCTAL REPRESENTATION AND ALSO WITH THEIR CORRESPONDING SYNDROME CHECK CONSTRAINTS.

of error. Assume that there are in total M candidate codes $\mathcal{C}_1, \ldots, \mathcal{C}_M$, with the corresponding parity check matrices $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_M$. If all soft channel symbols are presented to the receiver at once, then the optimal strategy would be to compute the syndrome probabilities $\gamma_1^{(m)}, \gamma_2^{(m)}, \ldots$ for all candidate codes $m = 1, \ldots, M$, and then compute the SPP for each candidate code and pick the candidate code that yields the maximum SPP. However, in practice the data arrives sequentially and the objective is to decide on the code candidate as soon as possible, subject to some constraints on the detection performance. The performance criteria may be expressed in terms of probabilities of detection and false alarm as follows:

(i). The probability of detection should be above a certain threshold, say $P_D^{\min}$.
(ii). The probability of false alarm should be smaller than a certain threshold, say $P_F^{\max}$.

This problem can be formulated as a sequential procedure for multiple hypothesis testing [22], [23]. When there are two alternative hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, that is, when there are two possible candidate codes ($M = 2$), then the optimal test is known and is found by the so-called sequential probability ratio test (SPRT) [24]. Here, the optimality is in the sense that among all hypothesis tests satisfying the above constraints on detection and false alarm probabilities, the SPRT requires the smallest number of observations N. However, when there are more than two alternative hypotheses ($M > 2$), then the optimal test is often not known, or the optimal test has a very complicated structure that limits its use in practice [22]. In these cases, one may opt for heuristic solutions. In what follows, we will propose a sequential test for our problem and evaluate its performance via simulations.

A. Proposed Sequential Statistical Hypothesis Test

Let $\mathcal{H}_m$, $m = 1, \ldots, M$ denote the hypothesis under which channel code $\mathcal{C}_m$ was used to encode the data, and let π_m^0 denote the prior probability of this hypothesis, i.e. the probability of picking channel code $\mathcal{C}_m$ by the transmitter. Having observed the received vector $\boldsymbol{r}$, the minimum probability-of-error detector chooses the hypothesis with the largest posterior probability $\Pr(H_m|\boldsymbol{r})$ [25]. That is, the decision is

$$k = \underset{m}{\operatorname{argmax}} \ \Pr(H_m|\boldsymbol{r}). \tag{21}$$

In our problem, we can find, for each candidate code $\mathcal{C}_m$, the syndrome posterior probability Γ_m using the corresponding parity check matrix $\boldsymbol{H}_m$;

$$\Gamma_m \triangleq P(\text{all syndrome checks of } \boldsymbol{H}_m \text{ are satisfied}|\boldsymbol{r}). \tag{22}$$

If the transmission was error-free, then Γ_m would be 1 for the true code and it would be zero for the rest of the code candidates. Due to errors introduced in the transmission, this will not be the case in practice. However, we know that the SPP for the true code is likely to be higher than that of the others, and that the difference will be larger as the length of the observed sequence increases. Therefore, we propose the following sequential probability ratio test to detect the channel code.

Proposed SSTH. *Let Γ_m^n denote the syndrome posterior probability for candidate code $\mathcal{C}_m$ obtained at stage n, i.e., after observing $\gamma_1^{(m)}, \gamma_2^{(m)}, \ldots, \gamma_n^{(m)}$, and define the vector $P^n = [p_1^n, \ldots, p_M^n]$, where*

$$p_m^n = \frac{\Gamma_m^n}{\sum_{i=1}^M \Gamma_i^n}.$$

Given a threshold ζ (to be explained below), we use the following rule as the stopping criterion

$$N_A = \text{ first } n \geq 1, \text{ such that } p_k^n \geq \zeta, \quad \text{for some } k = 1, \ldots, M. \tag{23}$$

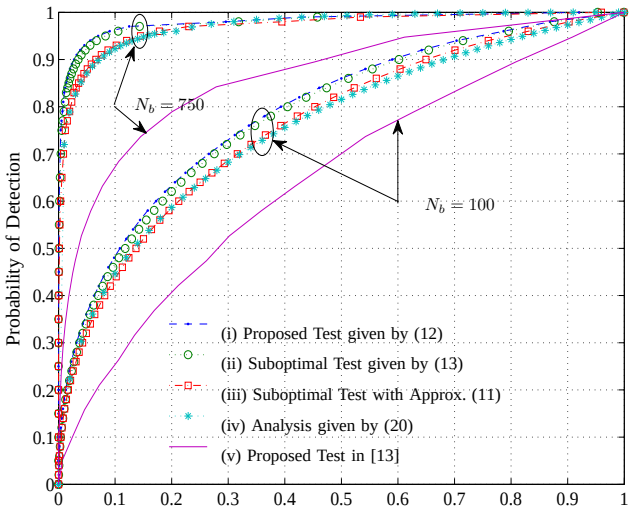
The decision rule at the stopping time is

$$k = \underset{m}{\operatorname{argmax}} \ p_m^{N_A}. \tag{24}$$

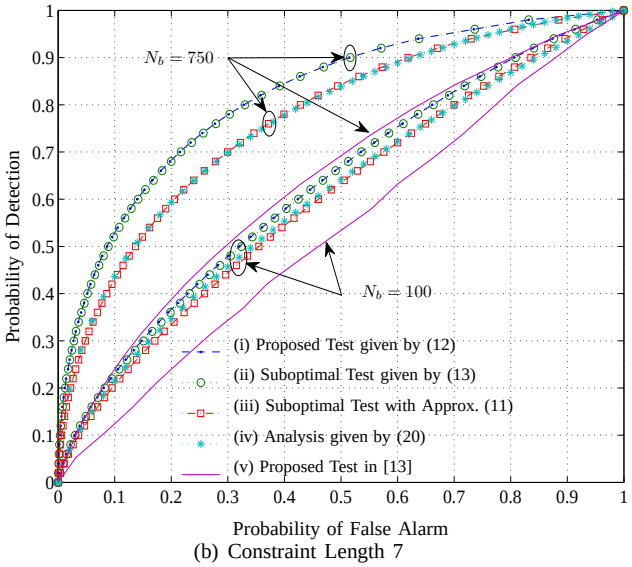
Note that according to the proposed SSHT, the required number of observations is finite. That is so because, we have

$$\begin{aligned} p_m^n = \frac{\Gamma_m^n}{\sum_{i=1}^M \Gamma_i^n} &= \frac{\exp\left(-\sum_{j=1}^n \log\left(1+e^{-\gamma_j^{(m)}}\right)\right)}{\sum_{i=1}^M \exp\left(-\sum_{j=1}^n \log\left(1+e^{-\gamma_j^{(i)}}\right)\right)} \\ &= \left(\sum_{i=1}^M \frac{\prod_{j=1}^n \left(1+e^{-\gamma_j^{(m)}}\right)}{\prod_{j=1}^n \left(1+e^{-\gamma_j^{(i)}}\right)}\right)^{-1}. \end{aligned} \tag{25}$$

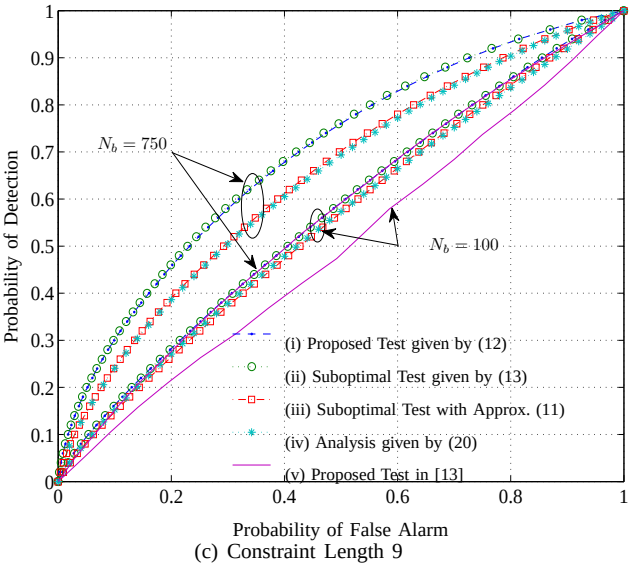
Assume that the mth code was used by the transmitter to encode the data. For the true code m, the syndrome probability constraints $\gamma_i^{(m)}$ are likely to be greater than zero (how much greater they are than zero generally depends on the operating SNR), whereas for a random code, the syndrome probability



(a) Constraint Length 4



(b) Constraint Length 7



(c) Constraint Length 9

Fig. 4. Receiver Operating Characteristic curves for the different schemes at an SNR of 0 dB.

constraints would take both positive and negative values with equal probability. This means that as n increases

$$\frac{\prod_{j=1}^{n}\left(1+e^{-\gamma_j^{(m)}}\right)}{\prod_{j=1}^{n}\left(1+e^{-\gamma_j^{(i)}}\right)} \to 0, \qquad \text{if } j \neq m,$$

and hence $p_m^n \to 1$ as n increases.

Remark 1. The overall error probability P_e of the proposed scheme depends on the value chosen for the threshold ζ. While it would be desirable to know the exact relation between P_e and ζ, due to the unknown probability distribution of the $\gamma_i^{(m)}$s, this relation is not known. The error probability is, however, in the order of $1-\zeta$. In Section V, we use a greedy search algorithm for choosing ζ such that the desired probability of error is achieved.

Remark 2. The implicit assumption is that the total number of available observations is infinite. In other words, we may continue sampling until one of p_k^n is greater than ζ. In many situations, however, there is a finite number of observations. Those situations can be considered as sequential hypothesis tests with finite horizon and the optimal solution may be found using the method of backward induction [26]. More precisely, if we reach the final stage N (where no more observations are available), then the optimal decision is known, resulting in a certain probability of error, say P_e^N. The decision in the previous stage $N-1$ is thus to stop sampling if the expected error probability in that stage is less than P_e^N and to continue and take the last sample, otherwise. The decision for the rest of the stages can be found similarly. As the simulation results in Section V show, the proposed SSHT scheme requires small number of observations for most operating SNR and thus the assumption of infinite available observations is not crucial in this study.

Remark 3. If the threshold ζ is greater than 1/2, then at the stopping time, only one of the codes can satisfy (23), since $\sum_m p_m^n = 1$. Also, the threshold can be chosen differently for different code candidates, i.e. if the cost of making a specific error is larger than the others, then the corresponding threshold may be chosen larger.

B. A Rule-of-Thumb for the Required Number of Observation

In this section, we provide an approximation of the error rate of the suboptimal test with a fixed number of observations. This approximation can be used to obtain a rule-of-thumb for the required number of observations according to the different schemes as explained below.

Let $\gamma_j^{(m)}$, $j = 1, \ldots, N$ denote the LLR associated with the jth parity check relation corresponding to the mth code candidate $\mathcal{C}_m$. Using the suboptimal test (13), the error probability under hypothesis $\mathcal{H}_m$ is,

$$P_{e|\mathcal{H}_m} = \Pr\left(\sum_{j=1}^{N}\gamma_j^{(i)} > \sum_{j=1}^{N}\gamma_j^{(m)},\ \text{for some } i \in I_{-m}|\mathcal{H}_m\right)$$
$$= 1 - \Pr\left(\sum_{j=1}^{N}\gamma_j^{(i)} \leq \sum_{j=1}^{N}\gamma_j^{(m)},\ \text{for all } i \in I_{-m}|\mathcal{H}_m\right)$$
$$\approx 1 - \prod_{\substack{i=1\\ i\neq m}}^{M}\Pr\left(\sum_{j=1}^{N}\gamma_j^{(i)} \leq \sum_{j=1}^{N}\gamma_j^{(m)}\right), \quad (26)$$

where $I_{-m} \triangleq \{1, 2, m-1, m+1, \ldots, M\}$. The approximation in the last step is due to the independence assumption that we have made among the events $\left(\sum_{j=1}^{N}\gamma_j^{(i)} > \sum_{j=1}^{N}\gamma_j^{(m)}\right)$ and $\left(\sum_{j=1}^{N}\gamma_j^{(i')} > \sum_{j=1}^{N}\gamma_j^{(m)}\right)$, for $i \neq i'$.[13] Using the analysis in Appendix A and B and applying the CLT, we can write

$$P_{e|\mathcal{H}_m} \approx 1 - \prod_{\substack{i=1\\ i\neq m}}^{M}\left[1 - \mathrm{Q}\left(\frac{m_r^{(m)}}{\sqrt{{\sigma_r^{(m)}}^2 + {\sigma_0^{(i)}}^2}}\right)\right] \quad (27)$$

where $m_r^{(m)}$ and ${\sigma_r^{(m)}}^2$ denote the mean and the variance of $\sum_{j=1}^{N}\gamma_j^{(m)}$ and ${\sigma_0^{(i)}}^2$ denotes the variance of $\sum_{j=1}^{N}\gamma_j^{(i)}$ under $\mathcal{H}_m$. Here also we assume that $\sum_{j=1}^{N}\gamma_j^{(m)}$ and $\sum_{j=1}^{N}\gamma_j^{(i)}$ are independent, which as explained before happens when the code candidates have distinct parity check matrices, see Section III. The overall error probability is therefore,

$$P_e \approx \frac{1}{M}\sum_{m=1}^{M} P_{e|\mathcal{H}_m}. \quad (28)$$

The ratio $m_r^{(m)}/\sqrt{{\sigma_r^{(m)}}^2 + {\sigma_0^{(i)}}^2}$ scales as $\sqrt{N}$ and hence we can use the above approximation to find a rough estimate of the required number of observations when applying the test (with a fixed number of observations). More particularly, for scenarios where we use a fixed given base code with different interleavers to obtain different channel codes, then

$$\frac{m_r^{(m)}}{\sqrt{{\sigma_r^{(m)}}^2 + {\sigma_0^{(i)}}^2}} = \sqrt{N}\cdot a, \quad \text{for all} \quad i, m = 1, \ldots, M$$

for some constant a that depends on J and on the SNR and hence we have,

$$N(M) \approx \frac{1}{a^2}\left[\mathrm{Q}^{-1}\left(1 - (1-P_e)^{\frac{1}{M-1}}\right)\right]^2. \quad (29)$$

Equation (29) in combination with our findings via simulations suggest a rule-of-thumb approximation for the number of observations required by the different schemes, for scenarios where we have a base code with different interleavers as the code candidates. More precisely, if $N_M(P_e)$ is the number of observations for a given candidate set size M and a given error probability P_e, then for an arbitrary candidate set size M' and arbitrary error probability P_e', we have $\frac{N_{M'}(P_e')}{N_M(P_e)} = d$, where

$$d \triangleq \left[\frac{\mathrm{Q}^{-1}\left(1 - (1-P_e')^{\frac{1}{M'-1}}\right)}{\mathrm{Q}^{-1}\left(1 - (1-P_e)^{\frac{1}{M-1}}\right)}\right]^2. \quad (30)$$

[13]Note that for the case with $M = 2$, this approximation is exact.

Using the above equation, if the required number of observations for specific choices of M and P_e is known, then we can use that as a basis to compute the required number of observations for arbitrary candidate set sizes and arbitrary error probabilities. The important observation is that the increase in the number of required observations when we increase the candidate set size from M to M' ($M' > M$), is independent of the choice of the base code, and in given by (30). Another important observation that we get from (30) is that the increase in the number of observations required when the candidate set size is increased is logarithmic. For instance, the increase in the required number of observations when we increase the candidate set size from say 2 to 4 is larger than that when the candidate set size is increased from 8 to 16. This might not be immediately clear from (30), however we can see this by first using the fact that for large x, $\mathrm{Q}(x) \approx \frac{1}{2}e^{-x^2/2}$. Therefore, since $\mathrm{Q}(\mathrm{Q}^{-1}(x)) = x$, we can write

$$\mathrm{Q}\big(\alpha\mathrm{Q}^{-1}(x)\big) \approx \frac{1}{2}(2x)^{\alpha^2}. \quad (31)$$

Now, using (31) in (30) in combination with the Taylor series approximation

$$1 - (1-P_e)^{\frac{1}{M-1}} \approx -\frac{\log(1-P_e)}{M-1},$$

we have

$$d \approx \frac{\log(M'-1) - \log\left(2\log(1-P_e')\right)}{\log(M-1) - \log\left(2\log(1-P_e)\right)}, \quad (32)$$

which confirms the above observation.

V. Simulation Results

In this section, we provide some simulation results for the performance of our proposed SSHT scheme. We assume as before BPSK transmission over an AWGN channel. We consider two scenarios for the code candidates: (i) where we use the standard rate-1/2 convolutional with constraint length 4 (depicted in Figure 2) in combination with different randomly chosen interleavers to obtain different code candidates, and (ii) where we consider a class of standard rate-1/2 convolutional codes with different constraint lengths as our candidate codes [21]. In this case, we consider 7 different rate 1/2 convolutional codes $\mathcal{C}_1, \ldots, \mathcal{C}_7$ with constraint lengths 3 to 9 respectively. The octal representations of the generator polynomials for different codes are given in Table I. Using syndrome formers of the codes, we get N_b syndrome check constraints for each of the codes (N_b is the number of information bits). The syndrome check constraints are also presented in Table I using the convention specified in Section III. All the codes are equally likely to be chosen by the transmitter. For the second scenario, the candidate set is constructed using the convolutional codes $\mathcal{C}_1, \ldots, \mathcal{C}_M$.

Table II and III show the average number of observations N_{SSHT} required by the proposed SSHT scheme for two values of the error probability at an SNR of 3 dB, for the first and the second scenario respectively. For comparison, the average number of observations N_{FIX} required to achieve the same error rate as with the optimal test with a fixed observation length and the corresponding reduction in the required number of observations are presented as well. As the results show, a

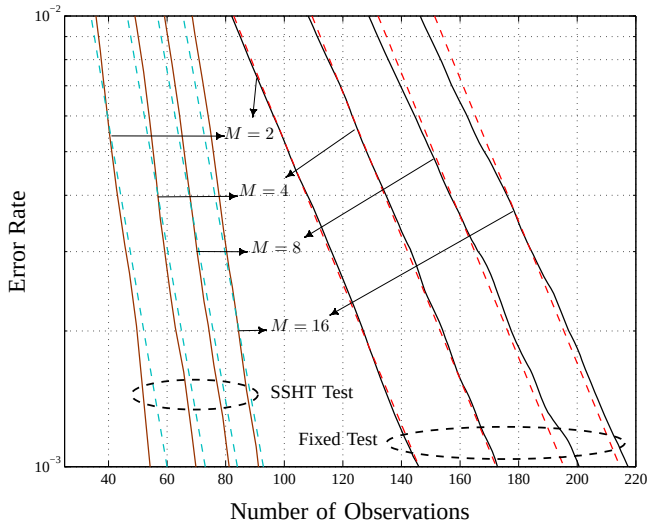


Fig. 5. Error rate as a function of the number of observations for different values of the candidate set size M, at an SNR of 3 dB. For comparison, the corresponding results predicted by (30) are also presented by dashed lines.

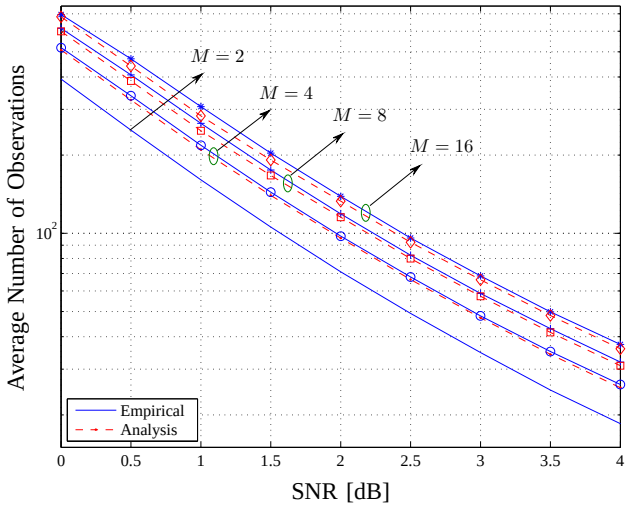


Fig. 6. Required number of observations as a function of SNR with our proposed SSHT scheme and for the first scenario with error probability of 1%. The corresponding analytical results obtained from (30) are also presented by dashed lines.

reduction of roughly 60% can be achieved with our proposed SSHT scheme. This reduction is in accordance with the expectations, as the number of required observations with SPRT is typically about one-half to one-third of that with the optimal test with a fixed number of observations [22]. For the first scenario, we can apply the proposed rule-of-thumb to obtain a rough estimates for N_{SSHT} and N_{FIX}. The corresponding normalized values are also presented in Table II. We can see that the results are very close to those predicted by equation (30), where we use the case with $M = 2$ and $P_e = 0.01$ as the basis for the computations.

For a more detailed comparison of the two schemes, we have plotted the empirical error rate curves as a function of the number of observations for the first scenario at SNR of 3 dB in Figure 5. Here also, the curves highlighted by dashed lines represent the corresponding predictions via (30), where for each M we take the case with $P_e = 0.005$ as the basis. We see that the proposed rule-of-thumb offers quite accurate predictions in all the cases. Also, we see that the proposed SSHT scheme requires significantly fewer observations on the average compared to the optimal test with a fixed number of observations to achieve a certain probability of error. Another observation is that as M increases, we need more observations on the average to achieve a given error probability for both tests, and that this increase is greater for smaller values of M, as predicted by (32).

To see the effect of the SNR on the performance, we have plotted the average required number of observations to achieve an error probability of 1% as a function of SNR for different values of M for the first and the second scenario in Figures 6 and 7, respectively. For the first scenario, we also plotted the corresponding predicted results using (30), where we have used the results for $M = 2$ as the basis. Again, we see a close match between the predicted results and the empirical results. Also we see that for higher SNR, the proposed SSHT requires very few observations in order to work.

We stress again that codes with parity check matrices that have a small number of nonzero elements in each row provide better operating conditions for our proposed SPP scheme. This is so because for such codes, the sum in (1) has fewer terms. Since in this sum, each additional term contributes an additional risk of making an error, having fewer terms means less overall probability of error. This explains the differences in the required number of observations for the two different scenarios in Figures 6 and 7. For instance, consider the case $M = 2$. In the first scenario, we have two randomly interleaved convolutional rate-1/2 codes with constraint length 4, whereas in the second scenario, we have two rate-1/2 convolutional codes with constraint lengths 3 and 4 respectively. The syndrome check constraints consist of 5 and 7 terms for the constraint lengths 3 and 4 respectively (see Table I). Therefore, we expect that the average required number of observations in the second scenario is smaller than in the first scenario, for a given error probability. This is also the case in the presented results (see Tables II and III).

VI. Conclusions and Future Work

In this paper, we presented a fast algorithm for blindly recognizing which channel code from a candidate set that was used to encode a data stream. The proposed algorithm uses the fact than any linear code satisfies a certain set of parity check relations, including convolutional codes (with and without tail-biting). Our algorithm obtains the probabilities that all parity check constraints are satisfied, called the syndrome posterior probability (SPP) of the code here, for all code candidates and then compares these probabilities. We also proposed a sequential hypothesis test that makes decisions before collecting all available data, hence saving computational complexity. Quantitatively, under typical operating conditions, the algorithm identifies the correct code (out of 16 candidates) in 99% of the cases by observing less than 50 samples, at an SNR of 4 dB.

The proposed scheme is potentially useful for complexity reduction of the PDCCH decoding in LTE. A detailed study

	$P_e = 0.01$				$P_e = 0.001$			
Set Size	$M = 2$	$M = 4$	$M = 8$	$M = 16$	$M = 2$	$M = 4$	$M = 8$	$M = 16$
N_{SSHT}	36	48	59	68	58	71	82	92
N_{FIX}	82	108	129	148	146	173	201	217
Reduction	56%	55%	54%	54%	60%	59%	59%	58%
Normalized N_{SSHT}	1	1.33	1.64	1.89	1.62	1.98	2.28	2.56
Normalized N_{FIX}	1	1.32	1.58	1.80	1.78	2.11	2.45	2.65
d	1	1.36	1.64	1.90	1.76	2.14	2.43	2.69

TABLE II

REQUIRED NUMBER OF OBSERVATIONS BY THE PROPOSED SSHT SCHEME N_{SSHT} AND BY THE OPTIMAL TEST WITH A FIXED OBSERVATION LENGTH N_{FIX} FOR DIFFERENT VALUES OF M AND AN SNR OF 3 DB, FOR THE FIRST SCENARIO. IN THIS TABLE, NORMALIZED VALUES OF N_{SSHT} AND N_{FIX} ARE ALSO GIVEN TO FACILITATE EASY COMPARISON WITH d IN (30). THE NORMALIZATION FACTOR IS THE CORRESPONDING VALUES FOR THE CASE WITH $M = 2$ AND $P_e = 0.01$.

	$P_e = 0.01$			$P_e = 0.001$		
Set Size	N_{SSHT}	N_{FIX}	**Reduction**	N_{SSHT}	N_{FIX}	**Reduction**
$M = 2$	23	56	59%	36	100	64%
$M = 3$	33	81	59%	49	140	65%
$M = 4$	48	114	58%	69	196	65%
$M = 5$	62	157	60%	87	274	68%
$M = 6$	80	205	61%	109	348	69%
$M = 7$	99	270	63%	140	450	69%

TABLE III

REQUIRED NUMBER OF OBSERVATIONS WITH OUR PROPOSED SSHT SCHEME AND ACCORDING TO THE OPTIMAL TEST WITH A FIXED OBSERVATION LENGTH FOR DIFFERENT VALUES OF M AND AN SNR OF 3 DB, FOR THE SECOND SCENARIO.

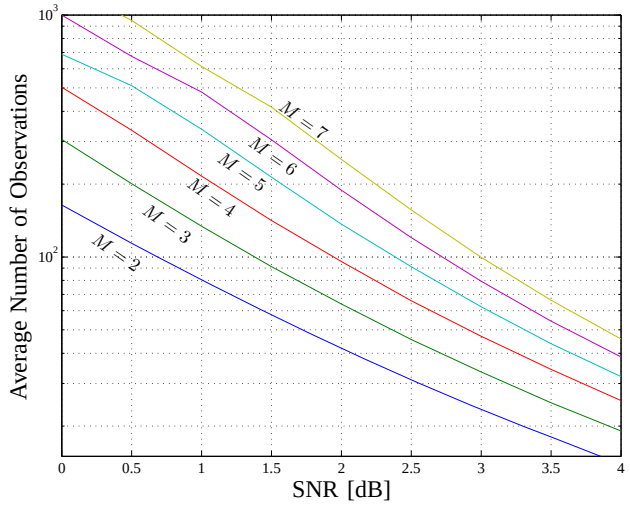


Fig. 7. Average number of observations required as a function of SNR with our proposed SSHT scheme for different values of the candidate set size M, at error probability of 1%, for the second scenario.

of this topic is a possible direction for the extension of this work. Another potential application of our algorithm, that may also be studied in future work, is to facilitate entirely blind multiple access based on the terminals blindly recognizing their *payload* data. In this case, the base station would not signal any explicit control information or AMC parameters at all. This may be facilitated either by assigning different terminals different codes, or different interleaving sequences. Since our algorithm tends to perform better for codes with low variable node degrees, in this foreseen application appropriate consideration has to be made when choosing the channel codes.

APPENDIX A
COMPUTING THE MEAN AND THE VARIANCE OF γ_k

Assume, without loss of generality, that the modulation scheme is BPSK, and consider the transmission of J bits $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_J$ over an AWGN channel with noise variance $N_0/2$ per real dimension. The received symbol at time instance i is

$$r_i = s_i + n_i,$$

where s_i denotes the BPSK symbol (binary "0" is mapped to +1, and binary "1" is mapped to -1) and n_i is the additive white Gaussian noise with mean zero and variance $N_0/2$. We consider two hypotheses:

- $\mathcal{H}_1$ under which we know that $\overset{J}{\underset{i=1}{\oplus}} \boldsymbol{c}_i = 0$, and
- $\mathcal{H}_0$ under which the transmitted bits are i.i.d. and take 0 or 1 with equal probability, which consequently means that $\overset{J}{\underset{i=1}{\oplus}} \boldsymbol{c}_i$ may take 0 or 1 with equal probability (no structure).

We are interested in computing the mean and the variance of

$$\gamma = \left(\prod_{i=1}^{J} \text{sign}(\ell_i) \right) \min_{i=1}^{J} |\ell_i| ,$$

under the two hypotheses, where ℓ_i denotes the posterior conditional LLR of $\boldsymbol{c}_i$. Let

$$X \triangleq \prod_{i=1}^{J} \text{sign}(\ell_i), \qquad Y \triangleq \min_{i=1}^{J} |\ell_i| .$$

Since we assume BPSK transmission over an AWGN channel, we have [21]

$$\ell_i = \Lambda(\boldsymbol{c}_i | r_i) = \frac{4 r_i}{N_0}. \tag{33}$$

According to our system model, r_i has a mixture Gaussian distribution. This allows us to use the following lemma to simplify the computations for finding $\mathbb{E}\{Y\}$.

Lemma. *Consider two random variables:*

1) *W with a mixture Gaussian probability distribution of the form $p\mathcal{N}(m, \sigma^2) + (1-p)\mathcal{N}(-m, \sigma^2)$, for some given $0 \leq p \leq 1$, $m \geq 0$ and σ^2.*

2) A zero-mean Gaussian random variable Z with variance σ^2.

Then, $|W|$ and $|Z+m|$ have the same probability distribution.

Proof. Let $f_Z(z)$ denote the probability distribution function (pdf) of Z, i.e.,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right), \tag{34}$$

and let $F_Z(z)$ denote its cumulative distribution function (cdf). We start by finding the cdf for $|W|$. We have,

$$\begin{aligned} F_{|W|}(w) &= \Pr\{|W| < w\} \\ &= \Pr\{-w < W < w\} = F_W(w) - F_W(-w) \end{aligned} \tag{35}$$

for $w \geq 0$, and for $w < 0$, $F_{|W|}(w) = 0$. Thus, for $w \geq 0$ the pdf of $|W|$ is given by

$$\begin{aligned} f_{|W|}(w) &= \frac{\mathrm{d}}{\mathrm{d}w} F_{|W|}(w) = f_W(w) + f_W(-w) \\ &= p f_Z(w+m) + (1-p) f_Z(w-m) \\ &+ p f_z(-w+m) + (1-p) f_Z(-w-m) \\ &= f_Z(w+m) + f_Z(w-m), \end{aligned} \tag{36}$$

since $f_Z(z) = f_Z(-z)$. Therefore,

$$f_{|W|}(w) = \begin{cases} f_Z(w+m) + f_Z(w-m), & w \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

It is straight forward to check that $|Z+m|$ has the same pdf too, which completes the proof. □

A direct conclusion of this lemma is that $|n_i+1|$ and $|n_i-1|$ have the same pdf, so

$$\mathbb{E}\{Y\} = \frac{4}{N_0}\mathbb{E}\left\{\min_{i=1}^{J} |r_i|\right\} = \frac{4}{N_0}\mathbb{E}\left\{\min_{i=1}^{J} |n_i+1|\right\}, \tag{38}$$

under both hypotheses. That is, we may work with n_i rather than r_i. The important observation is that r_i, $i = 1, \ldots, J$ are i.i.d. under $\mathcal{H}_0$ while under $\mathcal{H}_1$, they are not independent. However as we will see later, the same technique can be used to simplify the computations for finding the statistical properties under $\mathcal{H}_1$. Before we continue further, it is worth noting that since $\mathbb{E}\{X^2\} = 1$, $\mathbb{E}\{\gamma^2\}$ is also the same under both hypotheses and is given by

$$\begin{aligned} \mathbb{E}\{\gamma^2\} = \mathbb{E}\{Y^2\} &= \frac{16}{N_0^2}\mathbb{E}\left\{\left(\min_{i=1}^{J} |r_i|\right)^2\right\} \\ &= \frac{16}{N_0^2}\mathbb{E}\left\{\min_{i=1}^{J} |r_i|^2\right\}. \end{aligned} \tag{39}$$

Now, we are ready to compute the means and the variances of γ under the two hypotheses. Under hypothesis $\mathcal{H}_0$, since $\bigoplus_i \boldsymbol{c}_i$ is equally likely to be 0 or 1 (no structure), X and Y are independent. More specifically, X will be a binary random variable that takes one of the values $\{-1, +1\}$ with equal probability, and thus

$$\mathbb{E}\{\gamma|\mathcal{H}_0\} = \mathbb{E}\{XY|\mathcal{H}_0\} = \mathbb{E}\{X|\mathcal{H}_0\}\mathbb{E}\{Y|\mathcal{H}_0\} = 0, \tag{40}$$

and therefore

$$\sigma_0^2 \triangleq \mathbb{E}\{\gamma^2|\mathcal{H}_0\} = \mathbb{E}\{Y^2\} = \frac{16}{N_0^2}\mathbb{E}\left\{\min_{i=1}^{J} |r_i|^2\right\}. \tag{41}$$

Under hypothesis $\mathcal{H}_1$, X and Y are not independent. Indeed, if there are no errors, then $X = 1$. To compute the mean of γ under hypothesis $\mathcal{H}_1$, we can write

$$\begin{aligned} m_t \triangleq \mathbb{E}\{\gamma|\mathcal{H}_1\} &= \mathbb{E}\{Y|\mathcal{H}_1, X=1\}\Pr\{X=1|\mathcal{H}_1\} \\ &- \mathbb{E}\{Y|\mathcal{H}_1, X=-1\}\Pr\{X=-1|\mathcal{H}_1\}. \end{aligned} \tag{42}$$

The event $\{X = 1\}$ implies that either there have been no errors or there have been an even number of errors in the received sequence. Similarly, the event $\{X = -1\}$ implies that there have been an odd number of errors in the received sequence. Therefore,

$$\Pr\{X=1\} = \sum_{i=0}^{\lfloor \frac{J}{2} \rfloor} \binom{J}{2i} P_e^{2i}(1-P_e)^{J-2i}, \tag{43}$$

where P_e is the bit error probability of the channel. Since, we assume BPSK modulation, an error in the received sequence r_i occurs, when (i) $n_i < -1$, and $\boldsymbol{c}_i = 0$, or (ii) $n_i > 1$, and $\boldsymbol{c}_i = 1$. Using this, and the results from the Lemma, we can write

$$\mathbb{E}\{Y|\mathcal{H}_1, X=1\} = \frac{4}{N_0}\mathbb{E}\left\{\min_{i=1}^{J} |n_i+1| \,\middle|\, \mathcal{B}_1\right\} \tag{44}$$

$$\mathbb{E}\{Y|\mathcal{H}_1, X=-1\} = \frac{4}{N_0}\mathbb{E}\left\{\min_{i=1}^{J} |n_i+1| \,\middle|\, \mathcal{B}_2\right\} \tag{45}$$

where the event $\mathcal{B}_1$ ($\mathcal{B}_2$) is defined as the event that among J noise samples, none or an even (an odd, respectively) number of the samples are smaller than -1. We can finally write

$$\sigma_t^2 \triangleq \mathbb{E}\{\gamma^2|\mathcal{H}_1\} - m_t^2 = \frac{16}{N_0^2}\mathbb{E}\left\{\min_{i=1}^{J} |r_i|^2\right\} - m_t^2. \tag{46}$$

Note that the above quantities all depend only J and on the noise variance $N_0/2$ and can be found numerically. Once they are computed, they can be saved in a look-up table for future use. Also note that by increasing J, $\Pr\{X=1\}$ decreases and hence m_t decreases too.

APPENDIX B
COMPUTING THE CORRELATION BETWEEN γ_k AND $\gamma_{k'}$

Consider again the same system model as presented in Appendix A and consider the transmission of $J+\alpha$ ($\alpha \geq 1$) bits $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_{J+\alpha}$. Let $z \triangleq \bigoplus_{i=1}^{J} \boldsymbol{c}_i$ and $\tilde{z} = \bigoplus_{i=\alpha+1}^{J+\alpha} \boldsymbol{c}_i$. We consider again two hypotheses:

- $\mathcal{H}_1$ under which both z and $\tilde{z}$ are zero, and
- $\mathcal{H}_0$ under which z and $\tilde{z}$ may take 0 or 1 with equal probability (no structure).

We are interested in computing the correlation between γ and $\tilde{\gamma}$ where as before,

$$\gamma = \left(\prod_{i=1}^{J} \mathrm{sign}(\ell_i)\right) \min_{i=1}^{J} |\ell_i|,$$

and

$$\tilde{\gamma} = \left(\prod_{i=\alpha+1}^{J+\alpha} \mathrm{sign}(\ell_i)\right) \min_{i=\alpha+1}^{J+\alpha} |\ell_i|.$$

Under hypothesis $\mathcal{H}_0$, since the transmitted bits are independent of each other, γ and $\tilde{\gamma}$ are independent and hence

$$\mathbb{E}\{\gamma\tilde{\gamma}|\mathcal{H}_0\} = \mathbb{E}\{\gamma|\mathcal{H}_0\}\mathbb{E}\{\tilde{\gamma}|\mathcal{H}_0\} = 0. \tag{47}$$

Under hypothesis $\mathcal{H}_1$, if $\alpha \geq J$, then γ and $\tilde{\gamma}$ are independent and hence for $\alpha \geq J$,

$$\mathbb{E}\{\gamma\tilde{\gamma}|\mathcal{H}_1\} = \mathbb{E}\{\gamma|\mathcal{H}_1\}\mathbb{E}\{\tilde{\gamma}|\mathcal{H}_1\} = m_t^2. \tag{48}$$

For $\alpha < J$, by defining

$$X \triangleq \prod_{i=1}^{J} \text{sign}(\ell_i), \quad \tilde{X} \triangleq \prod_{i=\alpha+1}^{J+\alpha} \text{sign}(\ell_i),$$

$$Y \triangleq \min_{i=1}^{J} |\ell_i|, \quad \tilde{Y} \triangleq \min_{i=\alpha+1}^{J+\alpha} |\ell_i|$$

we have

$$\begin{aligned}\mathbb{E}\{\gamma\tilde{\gamma}|\mathcal{H}_1\} &= \mathbb{E}\left\{X\tilde{X}Y\tilde{Y}|\mathcal{H}_1\right\} \\ &= \mathbb{E}\left\{Y\tilde{Y}|\mathcal{H}_1, X\tilde{X}=1\right\}\Pr\left\{X\tilde{X}=1|\mathcal{H}_1\right\} \\ &\quad - \mathbb{E}\left\{Y\tilde{Y}|\mathcal{H}_1, X\tilde{X}=-1\right\}\Pr\left\{X\tilde{X}=-1|\mathcal{H}_1\right\}\end{aligned}$$

Under hypothesis $\mathcal{H}_1$, since $\left(\text{sign}(x)\right)^2 = 1$, the event $\mathcal{B} \triangleq \{X\tilde{X}=1\}$ implies that among $J+\alpha$ received symbols, either there has been no error or there have been an even number of errors in the first and the last α (total 2α samples) symbols. Therefore,

$$\Pr\{\mathcal{B}\} = \sum_{i=0}^{\alpha} \binom{2\alpha}{2i} P_e^{2i}(1-P_e)^{2\alpha-2i}, \tag{49}$$

where as before P_e is the bit error probability of the channel. Using the result of the Lemma in Appendix A, we can write

$$\mathbb{E}\left\{Y\tilde{Y}|\mathcal{H}_1, \mathcal{B}\right\} = \tag{50}$$
$$\frac{16}{N_0^2}\mathbb{E}\left\{\left(\min_{i=1}^{J}|n_i+1|\right)\left(\min_{i=\alpha+1}^{J+\alpha}|n_i+1|\right)\middle|\mathcal{B}\right\},$$
$$\mathbb{E}\left\{Y\tilde{Y}|\mathcal{H}_1, \mathcal{B}^c\right\} = \tag{51}$$
$$\frac{16}{N_0^2}\mathbb{E}\left\{\left(\min_{i=1}^{J}|n_i+1|\right)\left(\min_{i=\alpha+1}^{J+\alpha}|n_i+1|\right)\middle|\mathcal{B}^c\right\},$$

where $\mathcal{B}^c$ denotes the complement of the event $\mathcal{B}$.

REFERENCES

[1] R. Moosavi and E. G. Larsson, "A fast scheme for blind identification of channel codes," in *Proc. IEEE Globecom*, Dec. 2011.

[2] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

[3] S. Xi and H. C. Wu, "Robust automatic modulation classification using cumulant features in the presence of fading channels," in *Proc. IEEE WCNC*, pp. 2094–2099, Apr. 2006.

[4] H. C. Wu, M. Saquib and Z. Yun, "Novel automatic modulation classification using cumulant features for communications via multipath channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3098–3105, Aug. 2008.

[5] F. Hameed, O. Dobre and D. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.

[6] M. Marazin, R. Gautier and G. Burel, "Blind recovery of k/n rate convolutional encoders in a noisy environment,", *EURASIP J. Wireless Commun. Netw.* pp. 1186–1687, 2011.

[7] ——, "Algebraic method for blind recovery of punctured convolutional encoders from an erroneous bitstream," *IET Signal Process.*, vol. 6, no. 2, pp. 122–131, Apr. 2012.

[8] J. Barbier, G. Sicot and S. Houcke, "Algebraic approach for the reconstruction of linear and convolutional error correcting codes," *International J. Applied Math. Comput. Sciences*, pp. 113–118, Nov. 2006.

[9] J. Dingel and J. Hagenauer, "Parameter estimation of a convolutional encoder from noisy observations," in *Proc. IEEE ISIT*, pp. 1776–1780, June 2007.

[10] E. Dahlman, S. Parkvall, J. Sköld and P. Beming, *3G Evolution HSPA and LTE for Mobile Broadband*, 2nd edition, Academic Press 2008.

[11] V. Choqueuse, M. Marazin and L. Collin, "Blind recognition of linear space-time block codes: a likelihood-based approach," *IEEE Trans. Signal Process.,*, vol. 58, No. 3, pp. 1290–1299, Mar. 2010.

[12] T. Xia and H. C. Wu, "Novel blind identification of LDPC codes using average LLR of syndrome a posteriori probability," in *Proc. 12th International Conf. ITS Telecommun.*, pp. 12–16, Nov. 2012.

[13] C. Chabot, "Recognition of a code in a noisy environment," in *Proc. IEEE ISIT*, pp. 2211–2215, June 2007.

[14] E. G. Larsson and R. Moosavi, "Piggybacking an additional lonely bit on linearly coded payload data," *IEEE Wireless. Commun. Lett.*, vol. 1, pp. 292–295, Aug. 2012.

[15] R. Imad, G. Sicot and S. Houcke, "Blind frame synchronization for error correcting codes having a sparse parity check matrix," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1574–1577, June 2009.

[16] R. Imad and S. Houcke, "Theoretical analysis of a MAP based blind frame synchronizer," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 5472–5476, Nov. 2009.

[17] J. Hagenauer, E. Offer and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inf. Theory*, vol. 42, pp. 429–445, Mar. 1996.

[18] T. Clevorn and P. Vary, "Low-complexity belief propagation by approximations with lookup-tables," *5th Internantional ITG Conf. Source Channel Coding (SCC)*, Erlangen, Germany, Jan. 2004.

[19] R. Johannesson and K. Sh. Zigangirov, *Fundamentals of Convolutional Coding*, IEEE Press, 1999.

[20] R. Durrett, *Probability: Theory and Examples*, Cambridge University Press, 2010.

[21] J. G. Proakis and M. Salehi, *Digital Communication*, 5th Edition, McGraw-Hill, 2008.

[22] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1994–2007, Nov. 1994.

[23] H. Chernoff, *Sequential Analysis and Optimal Design*, Philadelphia: SIAM, 1972.

[24] A. Wald, *Sequential Analysis*, NY: Wiley 1947.

[25] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, 2001.

[26] T. S. Ferguson, *Optimal Stopping and Applications*, [Online]. Available: http://www.math.ucla.edu/~tom/Stopping/Contents.html, Visited Dec. 2012.

Reza Moosavi received his B.Sc. degree in Electrical Engineering from Isfahan University of Technology, Isfahan, Iran in 2005, his M.Sc. degree from Chalmers University of Technology, Göteborg, Sweden, in 2008 and his Ph.D. degree from Linköping University, Linköping, Sweden in 2014.

Since December 2013, he is an experienced researcher with Ericsson Research at Linköping, Sweden. His research interests include resource allocation, signaling protocols in cellular systems, and future radio access technologies.

Erik G. Larsson received his Ph.D. degree from Uppsala University, Sweden, in 2002. Since 2007, he is Professor and Head of the Division for Communication Systems in the Department of Electrical Engineering (ISY) at Linköping University (LiU) in Linköping, Sweden. He has previously been Associate Professor (Docent) at the Royal Institute of Technology (KTH) in Stockholm, Sweden, and Assistant Professor at the University of Florida and the George Washington University, USA.

His main professional interests are within the areas of wireless communications and signal processing. He has published some 80 journal papers on these topics, he is co-author of the textbook *Space-Time Block Coding for Wireless Communications* (Cambridge Univ. Press, 2003) and he holds 10 patents on wireless technology.

He is Associate Editor for the *IEEE Transactions on Communications* and he has previously been Associate Editor for several other IEEE journals. He is a member of the IEEE Signal Processing Society SAM and SPCOM technical committees. He is active in conference organization, most recently as the Technical Chair of the Asilomar Conference on Signals, Systems and Computers 2012 and Technical Program co-chair of the International Symposium on Turbo Codes and Iterative Information Processing 2012.