List Decoding of Short Codes for Communication over Unknown Fading Channels

N.B. When citing this work, cite the original published paper.

(article starts on next page)

List Decoding of Short Codes for Communication over Unknown Fading Channels

Marvin Xhemrishi[*][†], Mustafa Cemil Coşkun[*][†], Gianluigi Liva[*], Johan Östman[‡] and Giuseppe Durisi[‡]

[*] Institute of Communications and Navigation, German Aerospace Center, Weßling, Germany.
[†] Institute for Communications Engineering, Technical University of Munich, Munich, Germany
[‡]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

Abstract—In this paper, the advantages of list decoding for short packet transmission over fading channels with an unknown state are illustrated. The principle is applied to polar codes (under successive cancellation list decoding) and to general short binary linear block codes (under ordered-statistics decoding). The proposed decoders assume neither a-priori knowledge of the channel coefficients, nor of their statistics. The scheme relies on short pilot fields that are used only to derive an initial channel estimate. The channel estimate is required to be accurate enough to enable a good list construction, i.e., the construction of a list that contains, with high probability, the transmitted codeword. The final decision on the message is obtained by applying a non-coherent decoding metric to the codewords composing the list. This allows one to use very few pilots, thus reducing the channel estimation overhead. Numerical results are provided for the Rayleigh block-fading channel and compared to finite-length performance bounds. The proposed technique provides (in the short block length regime) gains of 1 dB with respect to a traditional pilot-aided transmission scheme.

I. Introduction

Recently, there has been an increasing interest in designing wireless communication systems with short information blocks, up to a few tens of bits, due to emerging applications with strict latency constraints [1]. As capacity is far beyond what is achievable for such message lengths, the fundamental limits of communications for finite-length messages have received renewed attention [2]–[5]. Code designs [6]–[8] and sophisticated decoding algorithms [9], [10] closing the gap to those limits in the moderate- and short-length regimes have been proposed. It is possible to operate within a few tenth of a dB from the finite length bounds as illustrated in a recent survey [11] comparing various code classes and finite-length bounds. While most of the attention has been focused on communication over additive white Gaussian noise (AWGN) channels, some applications require communicating with short packets over a fading channel where no *a priori* channel state information (CSI) is available at the transmitter and receiver [1]. In fact, classic pilot-assisted transmission (PAT) methods [12] become highly sub-optimal when short blocks are used [13]. The rates achievable over fading channels, when the CSI is not *a priori* available have been investigated in [14], [15] for

a fixed blocklength and error probability. Bounds on the error probabilities are provided in [16] not only for non-coherent transmission but also for various PAT strategies.

Recently, a novel decoding strategy for PAT schemes has been proposed [17], which aims at approaching the bounds developed in [16] without significantly increasing the complexity with respect to the so-called coherent setup where perfect CSI is available at the receiver. The decoder, in particular, assumes neither *a priori* knowledge of the channel coefficients, nor of their statistics. The technique relies on the use of very few pilot symbols to obtain a channel estimate, which is used to initialize an efficient list decoder. The list decoder provides a set of candidate codewords, which includes, with a sufficiently high probability, the transmitted codeword. For the final decision, a non-coherent decision metric is adopted. The effectiveness of the method was illustrated for a quasi-cyclic code under ordered-statistics decoder (OSD) [9], [18]. Although the gains were remarkable, the list sizes required by OSD to provide such gains are relatively large. In this paper, we apply the same principle to polar codes with successive cancellation list (SCL) decoding. We show that larger gains can be achieved in this setting, with much smaller list sizes than for the OSD, over a single-input single-output (SISO) Rayleigh block-fading channel.

II. Preliminaries

We use capital letters, e.g., X, for random variables (RVs) and their lower case counterparts, e.g., x, for their realizations. We denote the random vectors via capital bold letters, e.g., $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$, and their vector realizations via the lower case counterparts, e.g., $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$. As an exception, $\boldsymbol{I}_a$ refers to the $a \times a$ identity matrix. The probability density function of the continuous RV X is denoted as p_X. We use $\|\cdot\|$ for the l^2-norm, $\langle \cdot, \cdot \rangle$ for the inner product of two vectors, $\ln(\cdot)$ for the natural logarithm, and $\mathbb{E}[\cdot]$ for the expectation. We write $\mathcal{CN}(\mu, \sigma^2)$ to denote a complex Gaussian distribution with mean μ and variance σ^2.

A. System Model

We consider a SISO Rayleigh block-fading channel, i.e., the random fading coefficient is constant for n_c channel uses and changes independently across ℓ coherence blocks, which are also called diversity branches. Therefore, the packet size is

$n = \ell n_c$. Such a setup is relevant for OFDM systems, e.g., LTE and 5G (see [19]). The input-output relationship of the channel for the ith coherence block is

$$\boldsymbol{y}_i = h_i \boldsymbol{x}_i + \boldsymbol{n}_i, \quad i = 1, \ldots, \ell$$

where $\boldsymbol{x}_i \in \mathcal{X}^{n_c}$ and $\boldsymbol{y}_i \in \mathbb{C}^{n_c}$ denote the transmitted and received vectors, h_i is the realization of the channel coefficient, which is distributed as $H_i \sim \mathcal{CN}(0, 1)$ and $\boldsymbol{n}_i$ is the corresponding AWGN term, which is distributed as $\boldsymbol{N}_i \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n_c})$. The mutually independent RVs H_i and $\boldsymbol{N}_i$ are assumed to be independent over i. We will focus on quadrature phase shift keying (QPSK) signalling where the energy per symbol is normalized to 1.

B. Decoding Metrics over the Rayleigh Fading Channel

If the channel coefficients are known to the receiver, then the (coherent) maximum likelihood (ML) decoding rule is

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{C}} p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h})$$

$$= \arg\min_{\boldsymbol{x} \in \mathcal{C}} \sum_{i=1}^{\ell} ||\boldsymbol{y}_i - h_i \boldsymbol{x}_i||^2$$

where $\mathcal{C}$ is the set of transmitted signal vectors induced by the chosen channel code and modulation. When $||\boldsymbol{x}_i||$ is constant across codewords and blocks, we have

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \sum_{i=1}^{\ell} \Re\{\langle \boldsymbol{y}_i, h_i \boldsymbol{x}_i \rangle\}. \quad (1)$$

This is the case, for instance, if the modulation is QPSK. The idealized setting described above is often approximated by including pilot symbols in the transmitted sequence, which are used to obtain an estimate $\hat{\boldsymbol{h}}$ of the channel coefficients. The estimate is treated as ideal by the decoder, yielding the mismatched decoding rule (1) where h_i is replaced by $\hat{h}_i$.

Assume next that the decoder does not have access to the channel coefficients and to their distribution, and that no pilots are embedded in the transmitted sequence. In this case, the problem of decoding can be tackled, for instance, by designing a generalized likelihood ratio test (GLRT) [20], yielding

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \sup_{\boldsymbol{h}} p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h})$$

$$= \arg\min_{\boldsymbol{x} \in \mathcal{C}} \sum_{i=1}^{\ell} \inf_{h_i} ||\boldsymbol{y}_i - h_i \boldsymbol{x}_i||^2 \quad (2)$$

$$= \arg\max_{\boldsymbol{x} \in \mathcal{C}} \sum_{i=1}^{\ell} \frac{|\langle \boldsymbol{y}_i, \boldsymbol{x}_i \rangle|^2}{||\boldsymbol{x}_i||^2}. \quad (3)$$

The last step follows because the estimate $\hat{h}_i$ achieving (2) is $\hat{h}_i = \langle \boldsymbol{y}_i, \boldsymbol{x}_i \rangle / ||\boldsymbol{x}_i||^2$. Note that, under the assumption that the signals in $\mathcal{C}$ have the same energy over each coherence block, (3) reduces to

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \sum_{i=1}^{\ell} |\langle \boldsymbol{y}_i, \boldsymbol{x}_i \rangle|^2. \quad (4)$$

C. Successive Cancellation List Decoding of Polar Codes

Polar codes [21], [22] are the first class of provably capacity-achieving codes with low encoding and decoding complexity over any binary-input memoryless symmetric channel under successive cancellation (SC) decoding [22]. For a length-n polar code, a matrix $\boldsymbol{G}_n$ is constructed as

$$\boldsymbol{G}_n = \boldsymbol{B}_n \boldsymbol{K}_2^{\otimes m}$$

where $\boldsymbol{B}_n$ is the $n \times n$ bit-reversal permutation [22] and $\boldsymbol{K}_2^{\otimes m}$ denotes m-fold Kronecker product of $\boldsymbol{K}_2$ defined as

$$\boldsymbol{K}_2 \triangleq \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

with $m = \log_2 n$. A polar code of dimension k is designed by selecting the indexes of $n - k$ bit positions, to be included in a set $\mathcal{A}$. Then, an n-bit vector $\boldsymbol{u}$ is defined for encoding, where $u_i = 0$ for all $i \in \mathcal{A}$, yielding the so-called frozen bits, and the remaining k elements of $\boldsymbol{u}$ carry information bits. Encoding is performed as $\boldsymbol{c} = \boldsymbol{u} \boldsymbol{G}_n$. SC decoding [22] estimates u_i, $i = 1, 2, \ldots, n$, by using the channel observations and the previous decisions $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_{i-1}$. The code constraints imposed by the set $\mathcal{A}$ are taken into account by setting $\hat{u}_i$ to 0 if $i \in \mathcal{A}$. In SCL decoding [10], the two hypotheses 0/1 are kept active instead of making a hard decision for each bit u_i (if it is not a frozen bit). More specifically, several instances of an SC decoder run in parallel. Each relies on a different hypothesis, which we call path, on the previous information bits. At step i, each SC decoder instance computes the likelihoods for two new paths, corresponding to $\hat{u}_i = 0$ and $\hat{u}_i = 1$, arising from the same path. This doubles the number of paths for each $i \notin \mathcal{A}$. Whenever the number of paths exceeds a given maximum list size L, the least likely paths are pruned from the list. At the final stage, the decoder outputs the codeword maximizing the relevant decision metric.

D. Ordered-Statistics Decoding of Binary Linear Block Codes

An OSD represents an instance of a list decoding algorithm, which can be used for any binary linear block code [9], [18]. With a moderate complexity, it provides a very competitive performance at short blocklengths [11]. For a given (n, k) code with generator matrix $\boldsymbol{G}$, the algorithm starts by ordering the bit-wise channel log-likelihood ratios (LLRs) in decreasing order of reliability. This reordering is reflected in a permutation of the columns of the generator matrix $\boldsymbol{G}$. The permuted generator matrix is put into systematic form.[1] For a given parameter t, the decoder assumes that the maximum number of erroneous bits in the k most reliable bit positions is t. Based on this assumption, the decoder generates a list $\mathcal{L}$ of codewords with a size equal to $|\mathcal{L}| = \sum_{i=0}^{t} \binom{k}{i}$, by first taking hard decisions on the k most reliable positions and then by adding to them all the error patterns with Hamming weight up to t. Each vector is encoded via the systematic (permuted)

---

[1]This step might require an additional reordering of both the generator matrix and LLRs in case the first k columns of the permuted generator matrix are not linearly independent.

generator matrix to produce the candidate codewords in $\mathcal{L}$. Finally, the codeword in $\mathcal{L}$ that maximizes the decision metric of choice is selected. Typically, the likelihood metric is used if it is known and if it can be computed efficiently.

E. On the Complexity of Non-Coherent Decoding

Efficient maximum likelihood decoders (such as Viterbi decoding over the code trellis) as well as decoding algorithms that relies on the factorization of the channel likelihood (such as belief propagation decoding of low-density parity-check and turbo codes, SCL decoding of polar codes or OSD of a generic code) require to be initialized with bit-wise LLRs. However, the decoding metric in (4) does not admit a simple factorization, which prevents a straightforward use of such efficient decoders. A typical approach to address this issue relies on PAT. More specifically, n_p pilot symbols are embedded into each coherence block. For the ith coherence block, the vector of pilot symbols is denoted by $\boldsymbol{x}_i^{\mathsf{p}}$. The pilots are followed by $n_c - n_p$ coded symbols, denoted by $\boldsymbol{x}_i^{\mathsf{d}}$. The corresponding channel outputs are $\boldsymbol{y}_i^{\mathsf{p}}$ and $\boldsymbol{y}_i^{\mathsf{d}}$, respectively. The rate in bits per channel use (bpcu) is $R = k/(\ell n_c)$ where k is the number of information bits encoded by $\mathcal{C}$. The rate of the code $\mathcal{C}$ is $R_0 = k/(\ell(n_c - n_p))$. As a result, for a fixed rate R and a fixed blocklength ℓn_c, a large number of pilots comes at the cost of an increase in the code rate R_0, and thus a reduction of the error correction capability. This yields a trade-off between resources allocated to channel estimation and error correction [13].

The approach that will be used as reference in the following relies on a separation between the channel estimation and the channel decoding steps. In particular, upon observing the channel output, the pilot symbols are used in each coherence block to perform an ML estimation of the corresponding channel coefficient, i.e., we have

$$\hat{h}_i = \frac{\langle \boldsymbol{y}_i^{\mathsf{p}}, \boldsymbol{x}_i^{\mathsf{p}} \rangle}{\|\boldsymbol{x}_i^{\mathsf{p}}\|^2}, \quad i = 1, \ldots, \ell. \tag{5}$$

The channel estimates (5) are treated as perfect and the bit-wise LLRs based on the mismatched likelihoods $p_{\boldsymbol{Y}|\boldsymbol{X},H}(\boldsymbol{y}_i^{\mathsf{d}}|\boldsymbol{x}_i^{\mathsf{d}}, \hat{h}_i)$, with $i = 1, \ldots, \ell$, are fed to the (list) decoder in use. We refer to such approach as *pragmatic PAT* decoding. In the following, we demonstrate the effectiveness of a recently introduced technique [17] compared to the reference described above. It relies on a very limited number of pilot symbols (yielding a coarse initial channel estimate) to generate a list, from which the final decision is obtained according to a non-coherent metric.[2]

III. List Decoder with in-List GLRT

A simple modification of pragmatic PAT decoding can be devised for algorithms relying on list decoding. The modification works as follows. The initial channel estimates (5)

[2]An alternative approach to reduce the pilot overhead is to use iterative decoding and channel estimation algorithms [23]–[25]. A performance/complexity comparison between the technique outlined in this manuscript and iterative decoding and channel estimation approaches is beyond the scope of this paper.

are used by the list decoder to form the list $\mathcal{L}$ of codewords through the mismatched LLRs. Then, each codeword in the list is modified by re-inserting the pilot symbols, which yields a modified list $\mathcal{L}'$. We finally apply the GLRT rule (4) on the codewords in $\mathcal{L}'$ to obtain the decision as follows:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{L}'} \sum_{i=1}^{\ell} |\langle \boldsymbol{y}_i, \boldsymbol{x}_i \rangle|^2$$

$$= \arg\max_{\boldsymbol{x} \in \mathcal{L}'} \sum_{i=1}^{\ell} \Re\{\langle \boldsymbol{y}_i^{\mathsf{d}}, \hat{h}_i \boldsymbol{x}_i^{\mathsf{d}} \rangle\} + \frac{1}{2n_p} |\langle \boldsymbol{y}_i^{\mathsf{d}}, \boldsymbol{x}_i^{\mathsf{d}} \rangle|^2. \tag{6}$$

Note that the decoding metric has two contributions: a first term that resembles a coherent metric based on the estimate $\hat{\boldsymbol{h}}$, and a second term that is related to the non-coherent correlation. The second term is weighted by the inverse of the number of pilots; hence it becomes negligible when n_p is large (i.e., when the channel estimate is reliable).

An alternative derivation of the results is as follows. Observe that the channel estimate $\hat{h}_i$ provides the decoder with a statistical knowledge of the actual channel coefficient [26]. In particular, assuming the *a priori* distribution of the fading to be unknown to the decoder, the distribution of the channel coefficient for the ith coherence block given its estimate can be modeled as a complex Gaussian distribution with mean $\hat{h}_i$ given in (5) and variance $\sigma^2 \|\boldsymbol{x}_i^{\mathsf{p}}\|^{-2}$. Using this knowledge, we obtain

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \mathcal{C}'} \prod_{i=1}^{\ell} \mathbb{E}[p_{\boldsymbol{Y}^{\mathsf{d}}|\boldsymbol{X}^{\mathsf{d}}, H_i}(\boldsymbol{y}_i^{\mathsf{d}}|\boldsymbol{x}_i^{\mathsf{d}}, H_i)]$$

$$= \arg\max_{\boldsymbol{x} \in \mathcal{C}'} \sum_{i=1}^{\ell} \frac{|\langle \boldsymbol{y}_i^{\mathsf{d}}, \boldsymbol{x}_i^{\mathsf{d}} \rangle|^2 + 2\|\boldsymbol{x}_i^{\mathsf{p}}\|^2 \Re\{\langle \boldsymbol{y}_i^{\mathsf{d}}, \hat{h}_i \boldsymbol{x}_i^{\mathsf{d}} \rangle\}}{\|\boldsymbol{x}_i\|^2}$$

$$+ \frac{\|\boldsymbol{x}_i^{\mathsf{p}}\|^2 |\hat{h}_i|^2 \|\boldsymbol{x}_i^{\mathsf{d}}\|^2}{\|\boldsymbol{x}_i\|^2} + \sigma^2 \ln\left(\frac{\|\boldsymbol{x}_i^{\mathsf{p}}\|^2}{\|\boldsymbol{x}_i\|^2}\right) \tag{7}$$

where (7) follows because the conditional received vector $\boldsymbol{y}_i^{\mathsf{d}}$ per coherence block given the transmitted sequence $\boldsymbol{x}_i^{\mathsf{d}}$ is complex Gaussian with mean $\hat{h}_i \boldsymbol{x}_i^{\mathsf{d}}$ and covariance $\sigma^2 \left(\boldsymbol{I}_{n_c} + \|\boldsymbol{x}_i^{\mathsf{p}}\|^{-2} (\boldsymbol{x}_i^{\mathsf{d}})^{\mathsf{H}} \boldsymbol{x}_i^{\mathsf{d}}\right)$ and $\mathcal{C}'$ is the modified channel code obtained by re-inserting the pilot symbols to each codeword. By assuming QPSK and restricting the search space to $\mathcal{L}'$, we recover (6).

IV. Finite-blocklength Bounds

Next, we introduce the tools from finite-blocklength information theory that we shall use to benchmark our coding schemes. We provide an outer (converse) bound based on the metaconverse theorem in [5, Thm. 28] and for the inner (achievability) bound, we use a relaxed version of the random coding union (RCU) bound in [5, Thm. 16] that is commonly referred to as the RCUs bound [27, Thm. 1].

Let $q : \mathbb{C}^{n_c} \times \mathbb{C}^{n_c} \to \mathbb{R}^+$ be an arbitrary block-wise decoding metric and let $(\bar{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \sim p_{\boldsymbol{X}}(\bar{\boldsymbol{x}}_i) p_{\boldsymbol{X}}(\boldsymbol{x}_i) p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i)$, $i = 1, \ldots, \ell$, be independent

across coherence blocks. We define the generalized information density as

$$i_s(\boldsymbol{x}_i, \boldsymbol{y}_i) = \ln \frac{q(\boldsymbol{x}_i, \boldsymbol{y}_i)^s}{\mathbb{E}[q(\bar{\boldsymbol{X}}_i, \boldsymbol{y}_i)^s]} \qquad (8)$$

where $s > 0$ and the expectation is with respect to $\bar{\boldsymbol{X}}$. The RCUs achievability bound states that, for a given rate R, the average error probability is upper-bounded as

$$\epsilon \leq \inf_{s>0} \mathbb{E}\left[e^{-\left[\sum_{i=1}^{\ell} i_s(\boldsymbol{X}_i, \boldsymbol{Y}_i) - \ln(2^{Rn_c\ell}-1)\right]^+}\right]. \qquad (9)$$

We evaluate the bound in (9) for the following combinations of input distributions and decoding metrics [16, Sec. III.A-III.D]:

i) Input symbols uniformly distributed on a shell in $\mathbb{C}^{n_c}$, and ML decoding, i.e., $q(\boldsymbol{x}_i, \boldsymbol{y}_i) = p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i)$;

ii) a pilot-assisted scheme as in Section II-E with the $n_c - n_p$ data symbols uniformly distributed on a shell in $\mathbb{C}^{n_c - n_p}$ and ML decoding, i.e., $q(\boldsymbol{x}_i, \boldsymbol{y}_i) = p_{\boldsymbol{Y}^\mathsf{d}|\boldsymbol{X}^\mathsf{d}, \hat{H}}(\boldsymbol{y}_i^\mathsf{d}|\boldsymbol{x}_i^\mathsf{d}, \hat{h}_i)$;

iii) input distribution as in ii), and pragmatic PAT decoding, i.e., $q(\boldsymbol{x}_i, \boldsymbol{y}_i) = \exp(-\|\boldsymbol{y}_i^\mathsf{d} - \hat{h}_i \boldsymbol{x}_i^\mathsf{d}\|^2)$.

For the converse bound, we let $p_{\boldsymbol{X}}$ be as in i) and define

$$j_s(\boldsymbol{x}_i, \boldsymbol{y}_i) = \ln \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\boldsymbol{x}_i)}{q_{\boldsymbol{Y}}^s(\boldsymbol{y}_i)} \qquad (10)$$

where $q_{\boldsymbol{Y}}^s(\boldsymbol{y}_i) = \frac{1}{\mu(s)} \mathbb{E}[p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}_i|\bar{\boldsymbol{X}})^s]^{1/s}$ and $\mu(s)$ is chosen so that $q_{\boldsymbol{Y}}^s(\boldsymbol{y}_i)$ integrates to 1. Then, for a given rate R, the average error probability is lower-bounded as

$$\epsilon \geq \sup_{s>0} \sup_{\lambda>0} \mathbb{P}\left[\sum_{i=1}^{\ell} j_s(\boldsymbol{X}_i, \boldsymbol{Y}_i) \leq \lambda\right] - e^{\lambda - Rn_c\ell}. \qquad (11)$$

For more details on this bound, the reader is referred to [28].

V. NUMERICAL RESULTS

We present next the performance achieved by the modified list decoders proposed in Section III. We consider two coding schemes: a quasi-cyclic code and a polar code. The results are obtained by Monte Carlo simulations and are provided in terms of block error rate (BLER) vs. signal-to-noise ratio (SNR) with the SNR espressed as E_s/N_0, where E_s is the expected energy per symbol and N_0 the single-sided noise power spectral density. The results are compared with the bounds of Section IV. We consider a Rayleigh block-fading channel with 4 diversity branches. Each branch consists of 17 channel uses. This results in 68 channel uses per message. For the simulations, we transmit $k = 32$ information bits within each codeword, yielding a rate $R = 32/68 \approx 0.47$ bpcu. The symbols are taken from a QPSK constellation. The $(128, 32)$ quasi-cyclic code used in the simulations is obtained by tail-biting termination of a rate$-1/4$ non-systematic convolutional code with a memory 14 and generators $[47633, 57505, 66535, 71145]$ [29, Table 10.14]. A suitable number of codeword bits is punctured to accommodate the pilot symbols within the 68 channel uses. The minimum distance of the quasi-cyclic code is upper-bounded by the free distance of the underlying convolutional code, which is 37. In addition, we designed a $(128, 32)$ polar
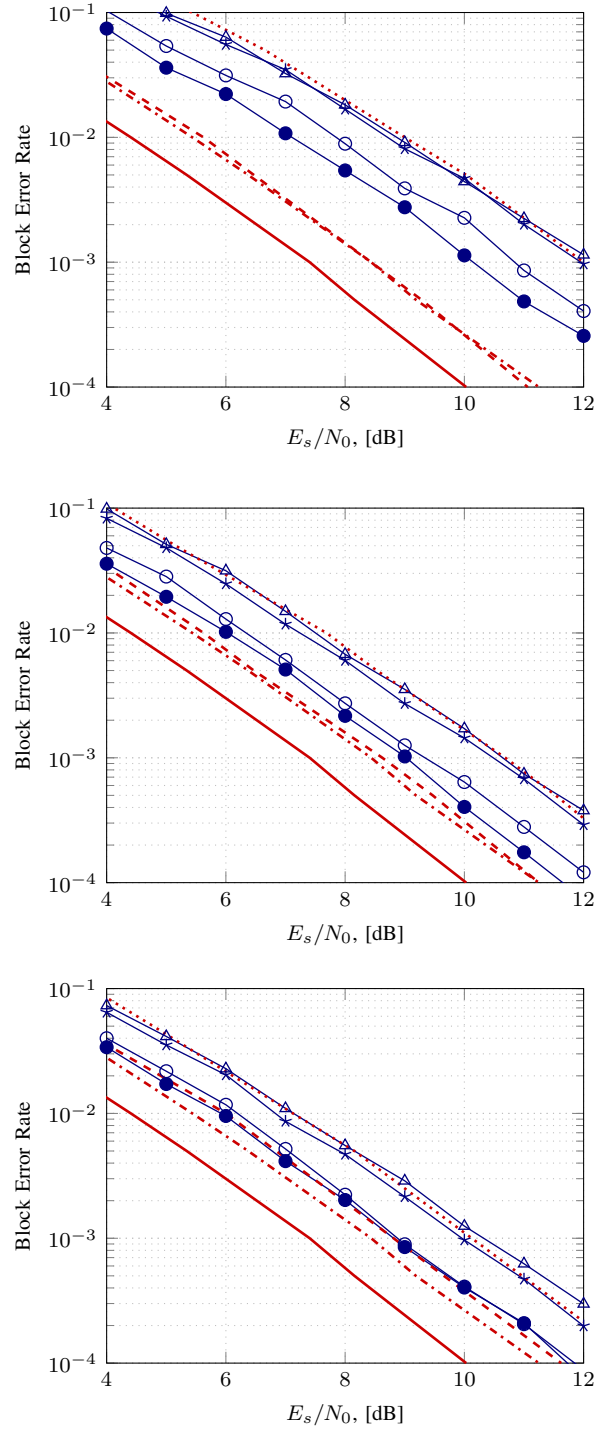

Fig. 1: BLER vs. SNR for the proposed scheme adopting quasi-cyclic code under OSD (—○—) and polar code under SCL decoding (—●—) with $n_p = 1$ (top), $n_p = 2$ (middle) and $n_p = 3$ (bottom). Finite length performance bounds given by the converse bound of (11) (——), the achievability of (9) for a non-coherent setup with ML decoding (—·—·), for PAT under ML decoding (— — —) and for PAT under pragmatic PAT decoding (·······). The performance of the pragmatic PAT scheme of Section II-E adopting both quasi-cyclic and polar codes, (—✶—) and (—△—), respectively, are provided as a reference.

code using the Gaussian approximation of density evolution with a design SNR of $E_s/N_0 = 3$ dB. As for the quasi-cyclic code, puncturing is applied according to the number of channel uses available after pilot insertion. For the polar code, quasi-uniform puncturing [30] is adopted, while the quasi-cyclic code is punctured randomly. For both codes, a random interleaver is applied to the codeword bits after encoding. The receiver uses an OSD to decode the quasi-cyclic code. The OSD order is set to 3, which provides a reasonable trade-off between performance and decoding complexity. With this choice, the OSD builds a list $\mathcal{L}$ of 5489 candidate codewords. For the polar code, the list size of the SCL decoder is set to 1024.

In Fig. 1, we compare the performance of the proposed modified list decoders to the performance of the pragmatic decoders described in Section II-E for different numbers of pilot symbols ($n_p \in \{1, 2, 3\}$) per coherence block. For the tested cases, the gains achieved by the proposed technique is no less than 1 dB compared to pragmatic PAT decoding at a BLER $\approx 10^{-3}$. Remarkably, the polar code under SCL decoding outperforms the quasi-cyclic code under OSD despite the much smaller list size, attaining a performance close to the RCUs achievability bound for PAT under ML decoding.

VI. CONCLUSIONS

In this paper, the performance of short block codes over the Rayleigh block fading channel have been analyzed under list decoding. The analysis deals with the case where no channel state information is available at the transmitter/receiver, and where no information on the fading distribution is available at the decoder. Focusing on pilot-aided transmission, it is shown how (modified) list decoders can provide an efficient solution to the reduction of the pilot overhead. The approach has been applied to polar codes under successive cancellation list decoding and quasi-cyclic codes with an ordered statistics decoder, and it was shown that the former operate within ~ 0.25 dB from the best known random coding achievability bounds at a block error rate $\approx 10^{-4}$, with a remarkably smaller list size than the ordered statistics decoder.

REFERENCES

[1] G. Durisi, T. Koch, and P. Popovski, "Towards massive, ultra-reliable, and low-latency wireless communications with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[2] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," Jet Propulsion Laboratory, Pasadena, CA, USA, TMO progress report 42-133, May 1998.

[3] A. Valembois and M. Fossorier, "Sphere-Packing Bounds Revisited for Moderate Block Lengths," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2998 – 3014, Dec. 2004.

[4] I. Sason and S. Shamai, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial*. Now Publisher Inc., Jul. 2006, vol. 3, no. 1–2.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[6] C. Poulliat, M. Fossorier, and D. Declercq, "Design of regular $(2, d_c)$-LDPC codes over GF(q) using their binary images," *IEEE Trans. Commun.*, vol. 56, no. 10, pp. 1626–1635, 2008.

[7] G. Liva, E. Paolini, B. Matuz, S. Scalise, and M. Chiani, "Short turbo codes over high order fields," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2201–2211, Jun. 2013.

[8] L. Dolecek, D. Divsalar, Y. Sun, and B. Amiri, "Non-binary protograph-based LDPC codes: Enumerators, analysis, and designs," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3913–3941, Jul. 2014.

[9] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *Trans. on Inf. Theory*, vol. 41, no. 5, pp. 1379–1396, Sep. 1995.

[10] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[11] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient Error-Correcting Codes in the Short Blocklength Regime," *Elsevier Phys. Commun.*, vol. 34, pp. 66–79, Jun. 2019.

[12] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[13] G. Liva, G. Durisi, M. Chiani, S. S. Ullah, and S. C. Liew, "Short codes with mismatched channel state information: A case study," in *IEEE Int. Workshop on Signal Process. Adv. in Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[14] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[15] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.

[16] J. Östman, G. Durisi, E. G. Ström, M. C. Coşkun, and G. Liva, "Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1521–1536, Feb. 2019.

[17] M. C. Coşkun, G. Liva, J. Östman, and G. Durisi, "Low-complexity joint channel estimation and list decoding of short codes," in *Proc. ITG Int. Conf. Syst., Commun. and Coding*, Feb 2019.

[18] B. Dorsch, "A decoding algorithm for binary block codes and j-ary output channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 391–394, May 1974.

[19] J. Östman, G. Durisi, E. G. Ström, J. Li, H. Sahlin, and G. Liva, "Low-latency ultra-reliable 5G communications: Finite block-length bounds and coding schemes," in *Int. ITG Conf. Sys. Commun. Coding (SCC)*, Hamburg, Germany, Feb. 2017.

[20] D. Warrier and U. Madhow, "Spectrally efficient noncoherent communication," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 651–668, Mar. 2002.

[21] N. Stolte, "Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung," Ph.D. dissertation, TU Darmstadt, 2002.

[22] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[23] H. Wymeersch, *Iterative Receiver Design*. Cambridge: Cambridge University Press, 2007.

[24] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe, "Code-aided turbo synchronization," *Proc. of the IEEE*, vol. 95, no. 6, pp. 1255–1271, 2007.

[25] M. Khalighi and J. J. Boutros, "Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3165–3173, Nov. 2006.

[26] G. Taricco and E. Biglieri, "Spacetime decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, June 2005.

[27] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random–coding bounds," in *Proc. Inf. Theory Applicat. Workshop (ITA)*, San Diego, CA, U.S.A., Feb. 2011.

[28] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for short-packet wireless communications," 2019. [Online]. Available: https://arxiv.org/abs/1612.01276

[29] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, 2nd ed. Piscataway, NJ, USA: Wiley-IEEE Press, 2015.

[30] V. Bioglio, F. Gabry, and I. Land, "Low-complexity puncturing and shortening of polar codes," in *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Mar. 2017, pp. 1–6.