A decoding algorithm for terminated convolutional codes over the blockwise noncoherent Channel

A Decoding Algorithm for Terminated Convolutional Codes over the Blockwise Noncoherent Channel

Matteo Ferro*, Riccardo Schiavone†, Gianluigi Liva†, and Maurizio Magarini*
*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy
†Institute of Communications and Navigations, German Aerospace Center (DLR), Weßling, Germany

***Abstract*—A decoding algorithm for convolutional codes (CCs) is proposed to eliminate the need for pilot symbols in short-packet communication systems operating over a blockwise noncoherent channel, i.e., a channel that introduces a random phase rotation of the transmitted codeword block. The algorithm is tailored to M-PSK modulations and applied to zero-tail terminated CCs. It works by implementing a three-step Viterbi-based noncoherent decoder, where the performance of a first blind decoding stage is enhanced through a code-aided phase estimation (CPE) that allows for correction of previous errors. In this framework, it provides a gain of approximately 0.5 dB compared to a pilot-aided decoder, with a manageable increase in complexity. Furthermore, numerical results show a performance within a few tenths of a dB from finite-length achievability bounds with both BPSK and QPSK modulations.**

***Index Terms*—Convolutional codes, blockwise noncoherent channel, short-packet communication, Viterbi algorithm, code-aided phase estimation.**

I. Introduction

Numerous emerging wireless applications rely on the sporadic transmission of short data units. These include, for example, smart metering networks, logistics and asset tracking, remote command links, and messaging services [1]. As a result, there has been growing interest in optimizing short-packet communications, focusing on developing efficient coding and decoding techniques that approach finite-length performance limits [2]. Under short block-length constraints, CCs have been shown to provide excellent performance in coherent communications [3]–[5]. However, for short packets, computing a channel estimate through pilot symbols introduces significant overhead [6], [7]. To address this issue, we propose a noncoherent decoder for terminated CCs that operates without any pilot symbols, tailored to the blockwise noncoherent channel [8]. It implements a modified version of the Viterbi algorithm (VA) [9], where the transition metric is derived by developing the likelihood function with an unknown channel phase [10]. To further improve the performance of the decoder, a code-aided phase recovery step is employed. Once a codeword is received and decoded, the receiver decides whether an error has been committed. In that case, the channel phase is estimated through the incorrectly-decoded codeword, and a second decoding attempt is performed via VA with a mismatched likelihood based on the phase estimate. Numerical results show that the proposed approach is effective, yielding a performance that approaches finite blocklength benchmarks for the blockwise noncoherent channel, with remarkable gains over standard pilot-assisted transmission (PAT) schemes.

The paper is organized as follows. Section II introduces the basic notation for CCs, the blockwise noncoherent channel model, and reference decoder architectures. In Section III, the proposed noncoherent decoding algorithm is presented. A finite blocklength achievability bound for the blockwise noncoherent channel is described in Section IV. Section V presents simulation results for the proposed decoder. Conclusions are provided in Section VI.

II. Preliminaries

We use uppercase letters for random variables, e.g. X, and lowercase letters for their realizations, e.g. x. Bold letters indicate vectors, e.g. $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$. We use $||\cdot||$ for the l^2-norm, $\langle\cdot,\cdot\rangle$ for the inner product, $\angle\cdot$ for the phase of a complex number. We write $\mathcal{CN}(\mu, 2\sigma^2)$ to denote a complex Gaussian random variable with mean μ and variance $2\sigma^2$. We use $\mathcal{U}[a, b)$ for a uniform distribution over the right-open interval $[a, b)$. Moreover, $[M]$ denotes the integer set $\{0, 1, \ldots, M-1\}$.

A. Convolutional Codes

We consider $(2, 1, \nu)$ nonsystematic, nonrecursive CCs, where the value ν is the encoder memory, whereas 2 and 1 are the number of output and input bits to the encoder at the different time instants. We evaluate the behavior of CCs under zero-tail termination, denoting the corresponding block codes as zero-tail terminated convolutional codes (ZTCCs). We also explore the possibility of concatenating the CC with an outer cyclic redundancy check (CRC) code with polynomial generator defined as $g(\mathtt{x}) = g_0 + g_1\,\mathtt{x} + \ldots + g_m\,\mathtt{x}^m$, and degree m. We denote by K the number of information bits before the encoding procedure (before the outer encoder, when present), and by N the blocklength in bits. The number of sections in the trellis is denoted by $T = N/2$. When no outer code is used, the rate for $(2, 1, \nu)$ ZTCCs is

$$R_{\mathsf{ZT}} = \frac{K}{2(K+\nu)}.$$

R.S. and G.L. acknowledge the financial support by the Federal Ministry for Research, Technology and Space (BMFTR) in Germany in the programme of “Souverän. Digital. Vernetzt.” Joint project 6G-RIC, project identification number: 16KISK022.

When an outer CRC code is used, the rate is modified as

$$R_{\mathsf{ZT}} = \frac{K}{2(K+\nu+m)}.$$

We denote by $\mathcal{C}$ the binary linear block code resulting from the termination of the CC (including the concatenation with the outer code, when present). With a slight abuse of notation, $\mathcal{C}$ will be also used to denote the codebook *after* modulation.

B. Blockwise Noncoherent Channel

We consider the blockwise noncoherent additive white Gaussian noise (AWGN) channel model from [8], focusing on the setting where the transmitted word is rotated by a random phase $\Phi \sim \mathcal{U}[-\pi,\pi)$. The transmitted signal is M-PSK modulated, that is, the channel input alphabet is $\mathcal{X} = \{\exp(j2\pi\ell/M) : \ell \in [M]\}$, where M indicates the modulation order. We have

$$Y_i = \mathrm{e}^{j\Phi} X_i + Z_i \qquad i = 1, 2, \ldots, n$$

where $Z_i \sim \mathcal{CN}(0, 2\sigma^2)$. Moreover $n = N/\log_2 M$ is the blocklength after modulation. Note that the phase Φ is constant over a block of n channel uses. The channel law is

$$p_{\boldsymbol{Y}|\boldsymbol{X},\Phi}(\boldsymbol{y}|\boldsymbol{x},\phi) = \prod_{i=1}^{n} p_{Y|X,\Phi}(y_i|x_i,\phi). \tag{1}$$

The rate (including modulation) is $R = R_{\mathsf{ZT}} \log_2 M$, and it is measured in bits per channel use. We denote as E_s the energy of a symbol and $E_b = E_s/R$ the energy of an information bit. The signal-to-noise ratio (SNR) is defined as E_b/N_0, where N_0 is the single-sided noise power spectral density.

C. Decoding

We consider three types of decoders: a genie-aided (GA) decoder with perfect channel state information (CSI); a mismatched decoder that adopts (1) as the likelihood function, replacing the actual channel coefficient with a noisy estimate; and an optimal maximum likelihood (ML) decoder. The three decoders are discussed in the following.

1) Genie-Aided Decoder with Perfect CSI

This decoder is provided with perfect knowledge of the channel phase. According to the ML criterion, the decoder returns

$$\hat{\boldsymbol{x}}_{\mathsf{GA}} = \arg\max_{\boldsymbol{x}\in\mathcal{C}} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi}(\boldsymbol{y}|\boldsymbol{x},\phi). \tag{2}$$

For ZTCCs w/o outer code, (2) can be efficiently implemented by applying the VA to the code trellis. The GA decoder will be used as a performance reference for our solution.

2) Mismatched Decoding with Pilot-Assisted Transmission

To deal with the uncertainty of the channel coefficient, the typical approach is to append a pilot sequence to the transmitted block. This knowledge is then used at the receiver end to obtain an estimate of the channel state. This technique is known by the name of PAT. We denote by L the length (in channel uses) of the pilot sequence. Its introduction decreases the rate to

$$R_{\mathsf{PAT}} = R\frac{n}{n+L}.$$

Denoting by $\boldsymbol{x}_p = (x_{p,1}, x_{p,2}, \ldots, x_{p,L})$ the pilot sequence and $\boldsymbol{y}_p = (y_{p,1}, y_{p,2}, \ldots, y_{p,L})$ its observation at the channel output, the ML phase estimate using the pilot symbols is

$$\hat{\phi} = \angle\hat{h}\ , \quad \hat{h} = \frac{\langle \boldsymbol{x}_p, \boldsymbol{y}_p\rangle}{||\boldsymbol{x}_p||^2}.$$

The mismatched decoder [11], [12] outputs

$$\hat{\boldsymbol{x}}_{\mathsf{PAT}} = \arg\max_{\boldsymbol{x}\in\mathcal{C}} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi}(\boldsymbol{y}|\boldsymbol{x},\hat{\phi}). \tag{3}$$

As for the GA decoder, for ZTCCs w/o outer code, (3) can be efficiently implemented by applying VA to the code trellis. Note that PAT underlies a non-trivial tradeoff [7], [13]. On one hand, a long preamble sequence may allow for an accurate estimate of the phase, limiting the performance degradation caused by the mismatched likelihood of (3). On the other hand, it may cause a significant decrease in the rate, which becomes utterly relevant when short packets are concerned. Indeed, the use of a preamble causes an energy loss of $10\log_{10}(1 + L/n)$ dB, which may already become unacceptable for moderate values of L and n.

3) Maximum-Likelihood Decoding

The ML decoder decides according to

$$\hat{\boldsymbol{x}}_{\mathsf{ML}} = \arg\max_{\boldsymbol{x}\in\mathcal{C}} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \tag{4}$$

where

$$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi}(\boldsymbol{y}|\boldsymbol{x},\phi)\,\mathrm{d}\phi.$$

Under M-PSK modulation, (4) is equivalent to [8], [14]

$$\hat{\boldsymbol{x}}_{\mathsf{ML}} = \arg\max_{\boldsymbol{x}\in\mathcal{C}} |\langle \boldsymbol{x}, \boldsymbol{y}\rangle|\,. \tag{5}$$

i.e., the ML decision maximizes the noncoherent correlation (NCC) between modulated codewords and channel observation. In contrast with (2) and (3), the rule defined by (4) – or, equivalently, by (5) – *cannot* be efficiently computed via VA. In particular, the impossibility of factoring the likelihood $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ into the product of marginals hinders the correct use of *any* decoder that requires operating with such marginals. This includes, among others, the VA [13].

III. Noncoherent Viterbi Algorithm

As discussed in Section II-C3, optimal ML decoding over the noncoherent channel defined in Section II-B requires evaluating (4), which cannot be performed by standard VA decoders. In the following, we address this issue by introducing a noncoherent decoding algorithm that consists of three steps. In the first step, inspired by [6], a suboptimal noncoherent VA is used to obtain a candidate solution (Section III-A). Secondly, this solution is validated either by an outer CRC or by a suitably defined threshold test (TT), applied to the cumulative metric of the Viterbi decoder (Section III-B). If the first decision is rejected, a third step is performed in an expectation-maximization fashion [15], [16]. This latter stage introduces in fact a code-aided channel phase estimation, followed by an additional parallel decoding attempt using M different phase offsets.

A. Step 1: Noncoherent Viterbi Decoder

We next introduce a noncoherent Viterbi decoding algorithm. This does not, in general, guarantee the solution of (4). Nevertheless, in practice, the proposed decoder often succeeds in recovering the transmitted sequence. Let us denote by $\lambda_t^{s\to i}$ the branch metric at time t for the state transition from state s (at trellis section $t-1$) to state i (at trellis section t), $t \in [1, T]$. Furthermore, we denote by Λ_t^i the cumulative metric in state i at trellis section t, and by $\mathcal{N}_i$ the set of states in section $t-1$ that are connected with state i at trellis section t. We introduce the partitioning $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T]$ and $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T]$, where $\boldsymbol{x}_t$ denotes the M-PSK modulated label of an edge in trellis section t, and $\boldsymbol{y}_t$ is its corresponding observation at the channel output.

Under binary phase-shift keying (BPSK) modulation, $\boldsymbol{x}_t$ is composed by two BPSK symbols. Under quadrature phase-shift keying (QPSK) modulation, $\boldsymbol{x}_t$ is a single QPSK symbol.[1] The algorithm traverses the trellis, performing the following operations:

1) *Branch metric computation*

$$\lambda_t^{s\to i} = \langle \boldsymbol{x}_t^{s\to i}, \boldsymbol{y}_t \rangle$$

where $\boldsymbol{x}_t^{s\to i}$ is the (modulated) edge label connecting state s at section $t-1$ to state i at section t.

2) *Path selection*

$$\hat{s} = \arg\max_{s\in\mathcal{N}_i} \left| \Lambda_{t-1}^{s} + \lambda_t^{s\to i} \right|.$$

3) *Cumulative metric update*

$$\Lambda_t^i = \Lambda_{t-1}^{\hat{s}} + \lambda_t^{\hat{s}\to i}.$$

For ZTCCs, the algorithm is initialized with $\Lambda_0^0 = 0$, and it terminates by computing the cumulative metric Λ_T^0 (the trellis starts and ends in the zero state when we assume zero-tail termination). The suboptimality of the algorithm stems from the second step, where the survivor at the ith node in trellis section t is selected based on a local optimality criterion – namely, the partial NCC between $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t]$ and $[\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t]$ – which does not guarantee that $\hat{\boldsymbol{x}}_{\mathsf{ML}}$ will "survive" until the final trellis section.

B. Step 2: Error Detection

After the first decoding step, error detection is performed either (a) by checking whether the decoded word satisfies the outer CRC code (if an outer code is used), or (b) by applying a TT to the final cumulative metric. In the latter case, for ZTCCs, the test accepts the Viterbi decoder decision if $|\Lambda_T^0| > \Delta$, where Δ is a predefined threshold. Otherwise, an error is declared.

C. Step 3: Code-Aided Phase Estimation

The third step is invoked if the solution identified by the decoder in Step 1 fails the validation check of Step 2. Empirical evidence collected by Monte Carlo simulations shows that, under M-PSK modulation, solutions that fail the test in Step 2 are likely to be close, in terms of Euclidean distance, to either the transmitted codeword or a rotated version of it, where the phase rotation is $2\pi\ell$, $\ell \in [M]$. Based on this observation, a further decoding attempt can be performed by

1) Estimating the phase by means of the decoded codeword as $\hat{\phi} = \angle\hat{h}$, with $\hat{h} = \langle \hat{\boldsymbol{x}}, \boldsymbol{y} \rangle / ||\hat{\boldsymbol{x}}||^2$.
2) Performing M parallel mismatched decoding attempts via VA, using as a phase estimate the angle $\phi_\ell = \hat{\phi} + 2\pi\ell/M$, with $\ell \in [M]$. The output of the ℓth VA decoder is denoted as

$$\hat{\boldsymbol{x}}_\ell = \arg\max_{\boldsymbol{x}\in\mathcal{C}} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi}\left(\boldsymbol{y}|\boldsymbol{x}, \phi_\ell\right).$$

3) Taking the final decision as

$$\hat{\boldsymbol{x}} = \arg\max_{\ell\in[M]} |\langle \boldsymbol{y}, \hat{\boldsymbol{x}}_\ell \rangle|.$$

Remark 1: Unlike blind phase recovery estimation algorithms (e.g., the Viterbi-Viterbi algorithm [17]), followed by multiple M-ary Viterbi decoding instances to resolve channel phase ambiguity, our approach only resorts to the M-ary decoding instances when the initial decoding step fails. However, the first decoding stage already yields a correct solution in several cases. As a result, our method typically requires fewer Viterbi decoding instances on average.

IV. Finite Blocklength Achievability Bound

As a benchmark to assess the performance of the proposed decoding algorithms, we consider the random coding union (RCU) bound of [2], with the relaxation based on Markov's inequality (see, e.g., [18]). The bound has the form

$$\epsilon \le \mathbb{E}\left[\min\left\{1, 2^{-\iota(\boldsymbol{X},\boldsymbol{Y})+nR}\right\}\right] \tag{6}$$

where

$$\iota(\boldsymbol{x}, \boldsymbol{y}) = \log_2 \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{\mathbb{E}\left[p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{X}')\right]}$$

and

$$(\boldsymbol{X}', \boldsymbol{Y}, \boldsymbol{X}) \sim P_{\boldsymbol{X}}(\boldsymbol{x}') p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) P_{\boldsymbol{X}}(\boldsymbol{x}).$$

In our case, the codebook distribution $P_{\boldsymbol{X}}$ is uniform over $\mathcal{X}^n$. The bound of (6) represents an upper bound on the block error probability achievable by the best code with rate R, blocklength n (in channel uses), and codebook distribution $P_{\boldsymbol{X}}$. For its evaluation, we resort to the Monte Carlo method, replacing in (6) the statistical average with the sample average (see, e.g., [19], [20]).

[1] The general M-PSK case, with $M > 4$, can be treated by merging trellis sections, so that the resulting edge labels can be represented by an integer number of M-PSK symbols.

V. Numerical Results

In this section, we present numerical results from Monte Carlo simulations for the $[133, 171]$ ($\nu = 6$) and $[561, 753]$ ($\nu = 8$) ZTCCs, with both BPSK and QPSK modulations. The code dimension K is fixed at 64 bits. For PAT decoders, L is set to 14 symbols (this number is the result of an optimization performed by means of simulations).

The charts make use of the following shorthands: GA stands for the approach of Section II-C1, where the decoder is provided with the channel phase; PAT denotes the results with mismatched decoding based on the estimated channel phase (Section II-C2); NC-TT is used for the results of the proposed noncoherent decoding algorithm, where error detection is performed by TT; finally, NC-CRC denotes the results obtained with the proposed noncoherent decoder, where an outer CRC code is used for error detection. For NC-TT, the threshold Δ has been empirically optimized for each code and modulation, through a collection of values of $|\Lambda_T^0|$ via Monte Carlo simulations for both correct and incorrect recovery of the transmitted codeword. For NC-CRC decoders, employing the $\nu = 6$ and the $\nu = 8$ ZTCCs, the CRC codes used have generators

$$g_1(\mathtt{x}) = \mathtt{x}^4 + \mathtt{x}^3 + \mathtt{x} + 1 \quad \text{and} \quad g_2(\mathtt{x}) = \mathtt{x}^4 + 1,$$

respectively. The CRC codes have been optimized with the method in [21] to minimize the undetected error rate for the specific ZTCC adopted, and, for our code parameters, they correspond to those reported in [5, Table II].

Figure 1 shows the codeword error rate (CER) vs. E_b/N_0 for the different decoding schemes with ZTCCs and BPSK modulation. We note that the proposed scheme provides a coding gain of approximately 0.4 dB and 0.6 dB compared to PAT when the CRC code and the TT are used, respectively. Interestingly, the NC-TT decoder performs almost identically to the GA decoder. As expected, the use of a CRC code reduces the rate, leading to a smaller coding gain.

Figure 2 reports the results for ZTCCs with a QPSK modulation. Here, the proposed decoder with CRC-based error detection as well as the TT-based approach offer approximately a 0.5 dB coding gain over PATs. However, in this setting, the noncoherent decoder does not attain the GA decoder performance, maintaining a gap of around 0.5 dB at the error rates of interest, with a slight advantage in favor of the threshold-based solution. The result may be explained by the smaller number of channel uses under QPSK signaling, which intrinsically limits the accuracy of the phase recovery, and by the inherent reduced robustness of QPSK constellations to phase offsets when compared to BPSK constellations. It is unclear whether the gap may be recovered by resorting to more complex decoder architectures, i.e., list Viterbi decoders.

Figure 3 shows the minimum E_b/N_0 required to achieve a target CER of 10^{-3} as a function of the rate (in bits per channel use). The figure includes the achievable information rates predicted by the bound (6). The results show that PAT always requires a higher SNR to achieve the desired CER, and that it has a lower information rate compared to the proposed algorithm, when CRC or TT-based error detection is used. This is particularly noticeable for QPSK modulation. It is important to observe that the noncoherent decoder with TT-based error detection performs, with memory-8 ZTCCs, within 0.3 dB from the RCU bound under BPSK modulation, and within 0.6 dB from the RCU bound under QPSK modulation.

Remark 2: The trade-off between accuracy and complexity is of the uttermost relevance in assessing the quality of the decoder performance. The TT method, although providing a lower CER than the CRC-aided one, requires a careful

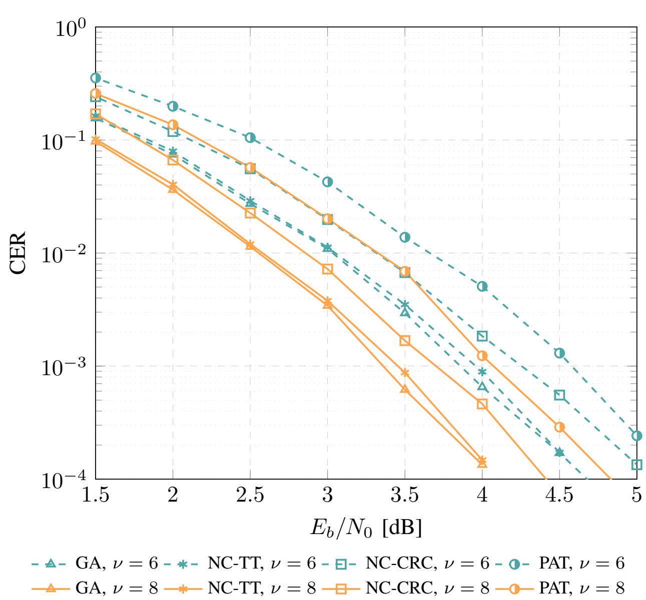

Fig. 1. CER vs. E_b/N_0 for ZTCCs with BPSK modulation.

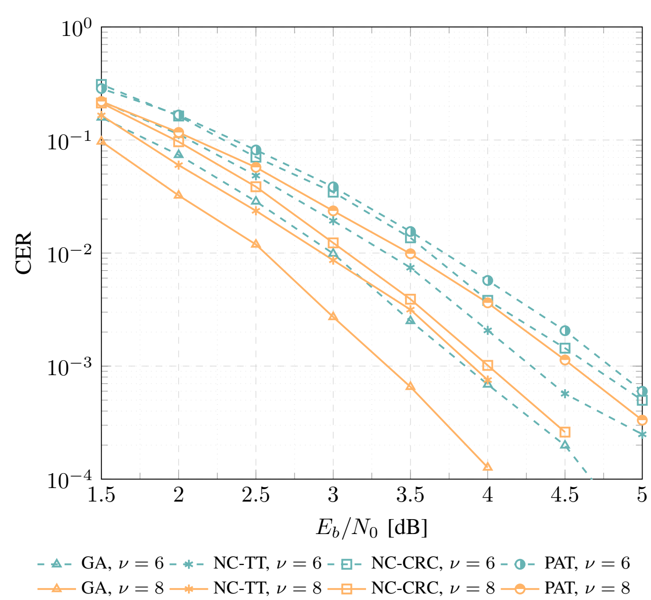

Fig. 2. CER vs. E_b/N_0 for ZTCCs with QPSK modulation.

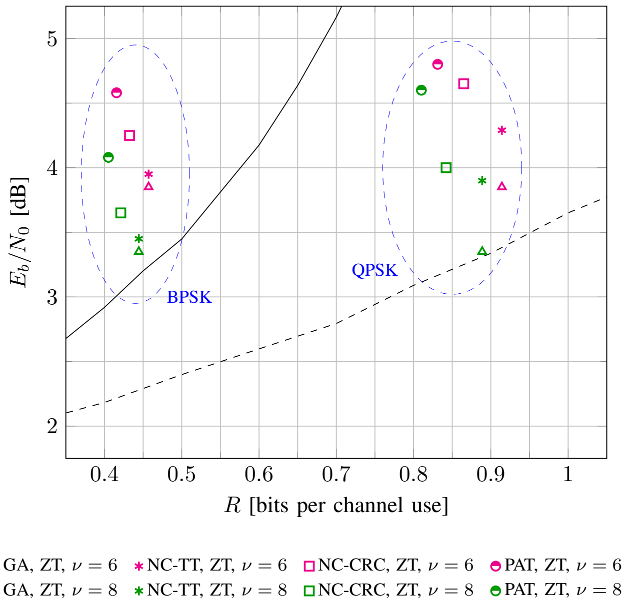

Fig. 3. Minimum E_b/N_0 required to achieve a CER of 10^{-3} for the different transmission schemes for both ZTCCs with BPSK and QPSK modulations. The RCU achievability bound (6) is depicted by the solid black line (for BPSK) and by the dashed back line (for QPSK).

selection of the threshold Δ. Its error detection is less accurate than what a sufficiently high-order CRC guarantees, so that the performance improvement has to be paid through a slightly higher complexity, where the latter is intended as the number of Viterbi instances per received word. Therefore, both options being advantageous with respect to the PAT case, one may not always be preferable to the other, and the choice has to be made carefully considering each use case's requirements.

VI. Conclusions

We proposed a novel decoding algorithm for convolutional codes over the blockwise noncoherent channel, eliminating the need for a pilot sequence to estimate the channel phase. The proposed decoder is a modified version of the Viterbi algorithm (VA), where the transition metric is derived by developing the likelihood function with an unknown channel phase. To further improve the performance of the decoder, a code-aided phase recovery step is employed. When an error detection algorithm declares an error after the first decoding stage, a second decoding attempt is performed via VA with a mismatched likelihood based on the phase estimate. Numerical results show that the proposed approach is effective, yielding a performance that approaches finite blocklength benchmarks for the block-wise noncoherent channel, with remarkable gains over standard pilot-aided schemes. The performance of tail-biting terminated convolutional codes under this decoding algorithm may be object of future studies.

References

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

[2] 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, 2010.

[3] L. Gaudio, T. Ninacs, T. Jerkovits, and G. Liva, "On the performance of short tail-biting convolutional codes for ultra-reliable communications," in *Proc. ITG Int. Conf. Syst., Commun. and Coding (SCC)*. Hamburg, Germany: VDE, Feb. 2017.

[4] 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," *Physical Communication*, vol. 34, pp. 66–79, 2019.

[5] H. Yang, E. Liang, M. Pan, and R. D. Wesel, "CRC-aided list decoding of convolutional codes in the short blocklength regime," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, pp. 3744–3766, 2022.

[6] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1874–1888, 2005.

[7] J. Östman, G. Durisi, E. 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.

[8] M. Peleg and S. Shamai, "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 603–609, May 1998.

[9] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[10] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300–308, Mar. 2002.

[11] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[12] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[13] M. Xhemrishi, M. C. Coskun, G. Liva, J. Östmanl, and G. Durisi, "List decoding of short codes for communication over unknown fading channels," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2019, pp. 810–814.

[14] B. Matuz, G. Liva, E. Paolini, M. Chiani, and G. Bauch, "Low-rate non-binary ldpc codes for coherent and blockwise non-coherent awgn channels," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4096–4107, Oct. 2013.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, Nov. 1977.

[16] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[17] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of psk-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, 1983.

[18] Y. Lomnitz and M. Feder, "A simpler derivation of the coding theorem," *arXiv preprint arXiv:1205.1389*, 2012.

[19] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-Based Computation of Information Rates for Channels With Memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[20] T. Ninacs, B. Matuz, G. Liva, and G. Colavolpe, "Short Non-Binary Low-Density Parity-Check Codes for Phase Noise Channels," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4575–4584, Jul. 2019.

[21] C.-Y. Lou, B. Daneshrad, and R. D. Wesel, "Convolutional-code-specific CRC code design," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3459–3470, Oct. 2015.