

Approximate ML Decoding of Short Convolutional Codes over Phase Noise Channels

Lorenzo Gaudio, Balázs Matuz, Tudor Ninacs, Giulio Colavolpe, and Armando Vannucci

Abstract—We propose a decoding algorithm for tail-biting convolutional codes over phase noise channels. It can be seen as a reduced complexity approximation of maximum-likelihood decoding. We target short blocks and extend the wrap-around Viterbi algorithm to trellises describing the random evolution of the phase impairment, for which we adopt two different models: a blockwise non-coherent and a blockwise Wiener channel model. Numerical results show that the performance of the proposed algorithm is within a few tenths of dB or less from maximum likelihood decoding for the setup studied in this paper.

Index Terms—Tail-biting convolutional codes; wrap-around Viterbi algorithm; phase noise; short blocks; machine-to-machine communications

I. INTRODUCTION

In the contexts of the 5G mobile standard and internet of things, one of the defined application scenarios is that of machine to machine (M2M)-type communications, in which low-cost sensors transmit small amounts of data, e.g., measurements [1], [2]. The number of connected devices is expected to reach 50 billion by 2025 [1], so that economies of scale become a relevant issue. To limit the cost of sensors, low-cost/low-end oscillators might be employed, which gives rise to significant phase noise (PN), especially at low datarates, i.e., with symbol periods that are not short enough for the limited phase coherence time. The technical literature on digital receivers is rich of algorithms that perform exceptionally well in the presence of PN (see, e.g., [3], [4] and references therein). However, whenever blocks in the range of a few hundred bits or less are considered, the aforementioned schemes may show a loss in performance, since the employed channel codes and decoding algorithms were not explicitly tailored to short blocks. An exception is the work in [5] that focuses on M2M-type communications with block lengths of a few hundreds of bits. In absence of pilots, [5] derives an approach based on non-binary low-density parity-check (LDPC) codes with differential phase-shift keying (PSK). The authors show that an iterative joint detection and decoding algorithm can be

employed to achieve results within 1.2 dB from finite-length performance bounds.

In this work, we propose a novel decoding approach for tail-biting (TB) convolutional codes (CCs) over PN channels. Such codes were shown to be a good choice on additive white Gaussian noise (AWGN) channels when the block length is short [6], [7]. We characterize the PN process as a blockwise non-coherent model, where the phase changes randomly and independently after a block of transmitted symbols. In addition we also model a more severe phase distortion, considering an evolution of the phase inside each block, following the usual Wiener process. In order to deal with such a general channel model we derive a maximum likelihood (ML) approach that is able to cope with all the configurations of PN discussed herein, and show its performance gap with respect to a finite-length performance benchmark. Furthermore, we propose a reduced complexity approach by extending the wrap-around Viterbi algorithm (WAVA) [8], showing its minor loss w.r.t. ML decoding.

II. SYSTEM MODEL

We consider a trellis code transmitted over an AWGN channel perturbed by PN. An information word \mathbf{a} is mapped to a modulated codeword \mathbf{c} by an encoder. The coded symbols are denoted by $\{c_k\}_{k=0}^{K-1}$, with K being the codeword length.¹ The received samples $\{r_k\}_{k=0}^{K-1}$ can thus be expressed as

$$r_k = c_k e^{j\theta_k} + w_k \quad (1)$$

where $w_k \sim \mathcal{CN}(0, 2\sigma^2)$ and θ_k is the discrete-time PN process.

We assume the following model for the PN. Consider the codeword of K transmitted symbols decomposed into B blocks of length L , where $BL = K$. At the beginning of each block, the channel phase is assumed to be uniformly distributed in the interval $[0, 2\pi]$, then it evolves, within the block, according to the Wiener model, i.e.,

$$\theta_{k+1} = \theta_k + \Delta\theta_k \quad (2)$$

with increments $\Delta\theta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$. Once a block has ended, the phase changes randomly and independently at the beginning of the next block. Thus, the conditional distribution of the phase $p(\theta_k | \theta_{k-1})$ is either Gaussian, $\theta_k \sim \mathcal{N}(\theta_{k-1}, \sigma_\Delta^2)$, within a block, or is uniform, $\theta_k \sim \mathcal{U}[0, 2\pi]$, at the beginning of each block (including the initial PN condition $p(\theta_0) \sim \mathcal{U}[0, 2\pi]$). We refer to this scenario as blockwise Wiener

L. Gaudio, G. Colavolpe, and A. Vannucci are with Department of Engineering and Architecture, University of Parma, Italy.

B. Matuz is with the Institute of Communication and Navigation of the German Aerospace Center (DLR), 82234 Weßling, Germany.

T. Ninacs is with Apple, Munich, Germany.

The work of Lorenzo Gaudio and Giulio Colavolpe is supported by Fondazione Cariparma, under the TeachInParma Project.

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

¹We assume that the redundancy introduced by the code is inserted in the signal space, i.e., by expanding the alphabet size.

channel. A practical situation that justifies this PN model is, for instance, the transmission of a long message that is split in smaller blocks (to comply with system constraints) that are sent in different time- or frequency-slots (e.g. by frequency-hopping), i.e., through different subchannels, affected by independent realizations of PN.

In addition, we consider a simplified scenario where the channel phase remains constant within each block, while it changes randomly and independently from block to block. We call this scenario blockwise non-coherent channel.² Note that, in the limit case with $B = 1$, the received vector contains only a single block of length $L = K$ with an unknown initial phase that is constant over the entire codeword, so that the classical noncoherent channel arises. This is a simplified assumption, often made in the literature, for which well known algorithms can be applied to the associated model (see e.g., [9], [10]).

III. MAXIMUM LIKELIHOOD DECODING

Starting from the received samples in (1), a maximum a-posteriori probability (MAP) sequence decoder is described by

$$\begin{aligned}\hat{\mathbf{a}} &= \arg \max_{\mathbf{a}} P(\mathbf{a}|\mathbf{r}) \propto \arg \max_{\mathbf{a}} p(\mathbf{r}|\mathbf{a}) P(\mathbf{a}) \\ &= \arg \max_{\mathbf{a}} p(\mathbf{r}|\mathbf{c}) P(\mathbf{a}).\end{aligned}\quad (3)$$

For independent and uniformly distributed symbols $\{a_k\}$, as we assume here, this becomes the ML sequence decoder

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} p(\mathbf{r}|\mathbf{c}) \quad (4)$$

which can be efficiently implemented by using the Viterbi algorithm (VA). Let us denote the symbol index $k = 0, 1, \dots, K - 1$ through a pair of indices (n, ℓ) , such that

$$k \triangleq nL + \ell \quad (5)$$

where $n = 0, 1, \dots, B - 1$ is the block index, while $\ell = 0, 1, \dots, L - 1$ is the index of the sample inside each block. The n -th block of L received samples is given by the vector

$$\mathbf{r}^{(n)} = (r_{nL}, r_{nL+1}, \dots, r_{nL+L-1}). \quad (6)$$

The vector $\mathbf{c}^{(n)}$ of transmitted symbols is defined accordingly.

Owing to the PN model introduced in Section II, where the phase changes independently on a block basis, the probability of the full received vector of K samples, given the transmitted sequence of coded symbols, obeys the following factorization

$$p(\mathbf{r}|\mathbf{c}) = \prod_{n=0}^{B-1} p(\mathbf{r}^{(n)}|\mathbf{c}^{(n)}). \quad (7)$$

The trellis structure describing the overall sequence can be viewed as the concatenation of B blocks of L trellis sections each, which reflects the factorization in (7). Since a factorization similar to (7) does not hold on a symbol basis, for the general PN channel, the standard VA is not applicable within the blocks, and the path metrics must be computed for an entire sequence (see (8) and (9) below). When several blocks

are connected together, as the factors in (7), the standard VA can still be used for the overall minimization of the paths weight. Hence, the B blocks of the overall trellis are treated as if they were the sections of a super-trellis, each with its own branch metrics. For clearness, the steps of the ML decoding approach are:

- 1) For every block of length B , calculate the branch metrics of the paths from each possible initial state, at $k = nL$, to each possible final state, at $k = nL + L - 1$ (according to (8) and (9) below). Note that, as discussed in Sec. IV-B, this is computationally costly.
- 2) Use the standard VA to connect the B blocks of the overall trellis, which are treated as if they were the sections of a super-trellis, each with its own branch metrics.
- 3) Since the code is TB, at the end, select the most likely TB path, i.e., the one with minimum weight.

As to the expression of the branch metrics corresponding to (7), two separate cases arise, depending on the standard deviation σ_{Δ} of the phase increments $\Delta\theta_k$ in (2).

1) *Blockwise non-coherent* ($\sigma_{\Delta} = 0$): In this case, the PN is constant inside each block and a closed form for $p(\mathbf{r}^{(n)}|\mathbf{c}^{(n)})$ exists [10], i.e.,

$$\begin{aligned}p(\mathbf{r}^{(n)}|\mathbf{c}^{(n)}) &= \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\ell=0}^{L-1} \left[|r_{nL+\ell}|^2 + |c_{nL+\ell}|^2 \right] \right\} \\ &\cdot I_0 \left(\frac{1}{\sigma^2} \left| \sum_{\ell=0}^{L-1} r_{nL+\ell} c_{nL+\ell}^* \right| \right) \left(\frac{1}{2\pi\sigma^2} \right)^L.\end{aligned}\quad (8)$$

where $(\cdot)^*$ stands for the complex conjugate and I_0 is the Bessel function of the first kind of order zero.

2) *Blockwise Wiener* ($\sigma_{\Delta} > 0$): In this case, a closed form for $p(\mathbf{r}^{(n)}|\mathbf{c}^{(n)})$ does not exist. However, a good approximation is proposed in [11], i.e.,

$$p(\mathbf{r}^{(n)}|\mathbf{c}^{(n)}) \simeq \prod_{\ell=0}^{L-1} I_0(|z_{\ell}|) e^{-\frac{|c_{nL+\ell}|^2}{2\sigma^2}} \prod_{\ell=1}^{L-1} I_0^{-1} \left(\frac{|z_{\ell}|}{1 + \sigma_{\Delta}^2 |z_{\ell}|} \right) \quad (9)$$

where coefficients z_{ℓ} track the evolution of the PN inside each block and can be recursively computed as

$$z_{\ell} = \frac{z_{\ell+1}}{1 + \sigma_{\Delta}^2 |z_{\ell+1}|} + \frac{r_{nL+\ell} c_{nL+\ell}^*}{\sigma^2}, \quad \ell \in [L-1, \dots, 0] \quad (10)$$

with a fictitious $z_L = 0$ for the initialization.

In ML decoding, every super-trellis section has $S = 2^m$ states,³ where m is the code memory, so that there are S^L possible transitions, or paths, from initial to final states. The decoding complexity of this approach is strictly related to the number of transitions within a super-trellis of length L and grows exponentially with L and m , thus limiting the use of the optimal ML approach to cases of very short non-coherent block lengths and to CCs with small memory. The complexity analysis will be discussed in detail in Sec. IV-B.

²This is clearly a special case of the more general blockwise Wiener channel, that occurs whenever $\sigma_{\Delta} = 0$.

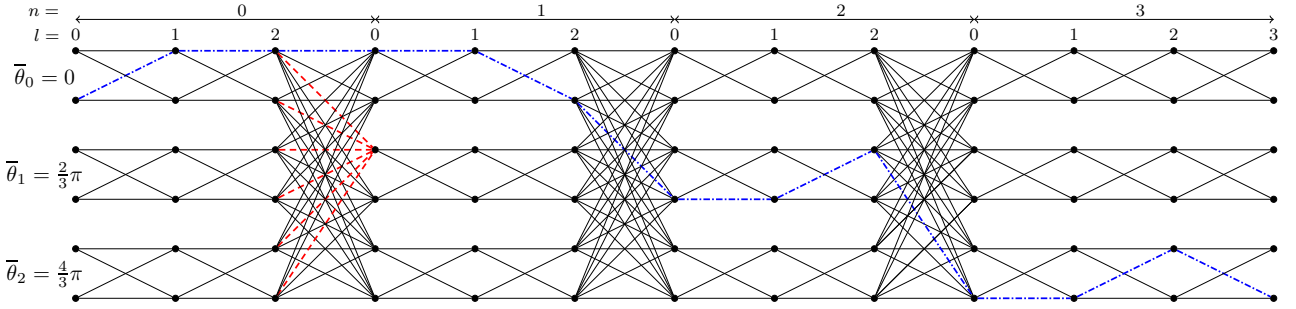


Fig. 1: Trellis for a 2-state code for the blockwise non-coherent channel, $Q = 3$ levels of quantized phase, block length $L = 3$, $B = 4$ blocks, and $P_\Delta = 0$. The trellis structure implemented in the decoder consists of a Q -fold replication of the basic binary-state trellis, where each replica corresponds to a different hypothesis on the phase.

IV. REDUCED COMPLEXITY DECODING

Motivated by the last observation, i.e., by the increased complexity of ML decoding on a super-trellis, and inspired by the detectors derived in [3], [4], we propose hereafter a receiver based on a novel reduced-complexity decoding approach. For the blockwise non-coherent channel, considering that the sequences of information symbols a_k , trellis states μ_k, μ_{k+1} , and code symbols c_k are in a one-to-one correspondence, according to generalized likelihood, a joint strategy for decoding and estimation of the channel phase can be expressed as a function of the joint distribution of symbols, encoder states, and PN samples $\theta = \{\theta_k\}$ as

$$\begin{aligned}
 (\hat{a}, \hat{\theta}) &= \arg \max_{(a, \theta)} P(a, c, \mu, \theta | r) \\
 &\propto \arg \max_{(a, \theta)} p(r | a, c, \mu, \theta) p(\theta) P(c, \mu | a) P(a) \\
 &= \arg \max_{(a, \theta)} \left[\prod_{k=0}^{K-1} p(r_k | c_k, \theta_k) \right] \left[p(\theta_0) \prod_{k=1}^{K-1} p(\theta_k | \theta_{k-1}) \right] \\
 &\quad \cdot \left[P(\mu_0) \prod_{k=1}^{K-1} I(a_k, c_k, \mu_k, \mu_{k+1}) \right] \quad (11)
 \end{aligned}$$

where $I(a_k, c_k, \mu_k, \mu_{k+1})$ is the code trellis indicator function, which is equal to 1 when a_k, c_k, μ_k , and μ_{k+1} obey the code constraints and 0 otherwise. The initial trellis state probability is $P(\mu_0) = 1/S$ and the term $P(a)$ has been neglected in (11), under the assumption that the information symbols are independent and uniformly distributed. The distribution of the k -th received sample given the k -th transmitted symbol and PN sample is

$$p(r_k | c_k, \theta_k) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} |r_k - c_k e^{j\theta_k}|^2 \right\}. \quad (12)$$

Let us uniformly quantize the channel phase with Q levels, $\theta_k \in \{0, \frac{2\pi}{Q}, \dots, \frac{2\pi}{Q}(Q-1)\}$, so that $\Delta\theta_k$ in (2) is also a multiple of $\frac{2\pi}{Q}$. We let the Wiener process be approximated as a

³Taking into account the possibility to have more bits as input of the encoder, i.e., $k_b > 1$, the more general definition of the number of states is $S = 2^{m \cdot k_b}$.

generalized random walk with fixed step, i.e., we approximate the conditional probability mass function (PMF)⁴

$$p(\theta_k | \theta_{k-1}) \simeq \begin{cases} P_\Delta & \text{if } \Delta\theta_k = \pm \frac{2\pi}{Q} \\ 1 - 2P_\Delta & \text{if } \Delta\theta_k = 0 \\ 0 & \text{otherwise} \end{cases} \quad \begin{matrix} k \neq nL \\ k = nL \end{matrix} \quad (13)$$

by assuming that either one of the two neighboring phase values is selected, or the current value. Although this assumption might seem inaccurate, in the presence of strong PN, results in Section V show that it yields a satisfactory approximation, even for large values of σ_Δ , provided that the transition probability between two adjacent quantized phase levels, P_Δ , is optimized as a design parameter.

We then build our reduced-complexity trellis structure, composed by Q interconnected trellises, one for each value of the quantized phase. A simple example of this structure for the blockwise non-coherent channel is depicted in Fig. 1. Within each of the Q replicas (trellises), the receiver assumes a different hypothesis on the phase and decoding occurs while compensating for that phase value. Since each trellis has its own fixed phase value θ_k , the branch metric in (12) depends on the Euclidean distance $d_E(r_k, c_k e^{j\theta_k}) = |r_k - c_k e^{j\theta_k}|$.

The idea of discretizing the unknown phase is *per se* not novel. It is however the new channel model considered here that makes the discrete phase estimate and the decoding of the TB sequence of states interplay in a novel way, such that the resulting m-WAVA Algorithm, discussed hereafter, is a generalization of previously known algorithms.

A. Proposed m-WAVA Algorithm

We are interested in TB CCs, that obey a termination condition requiring that the initial and final encoder states coincide. For TB CCs with short codeword lengths, the rate loss due to termination does play a role. The WAVA algorithm is a modification of the traditional VA that takes into account such a termination condition. It performs several iterations of the VA until hopefully the most likely path starts and ends in the same state [8]. We thus propose to modify the WAVA

⁴Note that, after the discretization, all probability density functions (PDFs) become PMFs.

B. Complexity

As discussed at the end of Section III, the complexity of the ML decoding strategy is exponential in L and m , thus becomes rapidly prohibitive whenever either the block length or the code memory increase. In fact, by taking into account the possibility to have more bits as input of the encoder, i.e., $k_b > 1$, the number of transitions within a block for ML decoding are

$$T_{\text{ML}} = 2^{k_b \cdot m \cdot L}. \quad (14)$$

The modified WAVA performs the decoding of TB CCs over the blockwise non-coherent channel, with complexity proportional to the number of transitions inside each block of length L . Suppose to have Q trellises with S states, a CC encoder with k_b bits as input, i.e., 2^{k_b} branches leave each state, by allowing transitions between adjacent trellises, i.e., letting $P_\Delta > 0$, the number of transitions within a block is

$$T_{\text{RC}} = 2^{k_b} S Q \left(\underbrace{1}_{\text{same trellis}} + \underbrace{2}_{\text{adjacent trellises}} \right) (L - 1) + \underbrace{2^{k_b} S Q^2}_{\text{phase changes randomly}}. \quad (15)$$

The complexity of the proposed approach grows linearly with L and not exponentially as in optimal ML decoding. Regarding Q , the dependence of (15) is quadratic, but we illustrate in Section V that often small values of Q are sufficient to perform close to theoretic benchmarks. In fact, the tradeoff for the choice of Q primarily depends on σ_Δ : a too large Q implies too many adjacent phase transitions to track the possibly rapid variations of PN, while a too small Q obviously implies a rough quantization of θ_k , that does not accurately represent the true PN values.

V. NUMERICAL RESULTS

Transmitted symbols are sent over the blockwise non-coherent channel (1) and performance is measured in terms of frame error rate (FER) versus E_b/N_0 (with and without Wiener PN). E_b is the received signal energy per information bit and N_0 the one-sided noise power spectral density. Fig. 2 shows simulation results when using the proposed modified WAVA (with two iterations) as well as under ML decoding, referring to the analysis of Sec. III (for a scenario where the complexity is still affordable, i.e., Fig. 2(a) and Fig. 2(b)). Simulations are compared against the random coding union bound with parameter s (RCUs) [12], computed for the blockwise non-coherent channel ($\sigma_\Delta = 0^\circ$).

The code used in simulations is a rate-1/2 TB CC with two different memory lengths, i.e., $m = 8$ and $m = 11$. The code polynomials were taken from [7] and chosen to optimize the CC code free distance.⁵ They are specified in octal notation as [515, 677] (for $m = 8$) and [5537, 6131] (for $m = 11$). The information word length is 64 bits, while the codeword length is 128 bits. We consider different values of σ_Δ , describing the standard deviation of the Wiener PN process, up to $\sigma_\Delta = 6^\circ$, that is regarded as a very strong PN. The number of phase

⁵Note that the PN evolution may influence the optimality of the code, and a search for codes which are optimal in this context might be necessary. Nevertheless, simulation results show that the code performance associated to the chosen code polynomials is satisfactory.

quantization levels is set to $Q = 8$, which is shown to be a reasonable tradeoff between performance and complexity. A binary phase-shift keying (BPSK) modulation is used.

Comparing the performance of a ML receiver to that of m-WAVA with reduced complexity we see that there is little loss entailed in the phase quantization process, even with a small value of Q . For short blocks with $L = 8$ (Fig. 2(a)) performance is almost identical. For longer blocks $L = 16$ (and $m = 8$) a loss of 0.4 dB is visible (Fig. 2(b)). The loss increases with both L and the PN variance, since a larger σ_Δ emphasizes the discrepancies between the true blockwise Wiener PN and the approximation entailed in m-WAVA.

The close match between ML and m-WAVA performance is a key point to demonstrate the practical applicability of the proposed algorithm, since the complexity of the receiver is driven by the number of quantization levels, as per (15). For larger code memory length (i.e., $m = 11$), as expected, the performance of the m-WAVA algorithm shifts closer to the benchmark, which suggests that the performance loss w.r.t. the ML receiver with full complexity is further reduced.

As previously stated, according to (13), an optimization of the parameter P_Δ could improve the performance when $\sigma_\Delta > 0$. If the noncoherent block is short and if the standard deviation of the phase noise is low, the improvement from the optimization of P_Δ is negligible. In these cases, we decided not to consider optimization, i.e., we adopted $P_\Delta = 0$. On the other hand, when L is long and σ_Δ is high, the optimization is necessary to improve performance. This is the case of Fig. 2(c), in which the optimization gain is clearly visible.

VI. CONCLUSIONS

This work presents a novel decoding algorithm for tail-biting convolutional codes for blockwise non-coherent AWGN channels. The decoding algorithm is an approximation of ML decoding, which, due its complexity, may not be a viable solution. We show that for the proposed reduced complexity algorithm, for the current setup, the loss is limited to at most 0.4 dB. Due to its suitability for short codeword lengths, the proposed decoding algorithm might be of high interest for M2M type communications.

REFERENCES

- [1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.
- [2] 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.
- [3] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.
- [4] A. Barbieri and G. Colavolpe, "Soft-Output Decoding of Rotationally Invariant Codes Over Channels With Phase Noise," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2125–2133, Nov 2007.
- [5] 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, July 2019.
- [6] 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.

- [7] L. Gaudio, T. Ninnas, 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*, Feb. 2017, pp. 1–6.
- [8] R. Y. Shao, S. Lin, and M. P. C. Fossorier, "Two decoding algorithms for tailbiting codes," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1658–1665, Oct. 2003.
- [9] M. Peleg, S. Shamai, and S. Galan, "Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel," *IEE Proc. Commun.*, vol. 147, no. 2, pp. 87–95, Apr. 2000.
- [10] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1376–1385, Sep. 1999.
- [11] G. Colavolpe, "On LDPC codes over channels with memory," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1757–1766, July 2006.
- [12] A. Martinez and A. G. i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *2011 Information Theory and Applications Workshop*, Feb 2011, pp. 1–6.