Bootstrapping Iterative Demodulation and Decoding Without Pilot Symbols

Simone Pecorino, Silvio Mandelli, Luca Barletta, *Member, IEEE*, Maurizio Magarini, *Member, IEEE*, and Arnaldo Spalvieri

Abstract— The iterative demodulation and decoding algorithm introduced in 2005 by Colavolpe, Barbieri, and Caire to cope with channels affected by phase noise needs pilot symbols to bootstrap. However, pilot symbols reduce the spectral efficiency of the system and, consequently, system's throughput. The aim of the paper is to show that trellis-based demodulation can be used to bootstrap the iterative process without the need of pilot symbols. Also, the complexity issue of trellis-based demodulation is addressed in the paper. The result is that the performance of iterative demodulation and decoding after the iterations is virtually unaffected by complexity reduction, provided that the reduced-complexity demodulator guarantees cycle-slip-free operation. From the numerical results presented in the paper we show that cycle-slip-free operation can be achieved with substantial complexity reduction also for phase noise associated with linewidths of practical interest.

Index Terms— Coherent communication, Phase noise, Coded modulation, Demodulation, Information rate.

I. Introduction

The context of the paper is coherent demodulation of coded modulations with high spectral efficiency in the presence of phase noise. The goal of demodulation is to work out the sequence of probability distributions of the input bits given the entire received signal. Demodulation can become a challenging task when transmission takes place over a channel with memory. Prominent examples of channels with memory are the intersymbol interference channel, the phase noise channel, and the fading channel. Among channels with memory, the phase noise channel is considered here, a recent tutorial on the subject being [1].

Demodulation of signals affected by the phase noise is a classical and hot topic in the context of point-to-point radio backhauling systems [2], [3] and coherent optical systems [4]–[7]. In these systems, coded constellations with high spectral efficiency are adopted, and the performance of demodulation in the presence of phase noise becomes one of the major issues in the rush to higher and higher spectral efficiency.

In optical systems, phase noise is due to both laser oscillators used for up- and down-conversion [8], and cross-phase modulation that arises in wavelength-division-multiplexing systems [9]. Several schemes have been proposed to estimate

the received carrier phase for arbitrary phase-shift keying (PSK) and quadrature amplitude modulation (QAM) constellations in presence of phase noise. Among these methods, the *blind* feedforward scheme of [10] addresses the constraints imposed by high speed parallel processing. Pilot-aided carrier phase recovery schemes have recently gained attention as candidate phase recovery approaches for systems affected by strong phase noise. Papers [11], [12] are based on the insertion of a pilot tone in a notch of the transmitted signal spectrum, while in papers [13]–[15] pilot symbols are inserted in time domain. Papers [16]–[21] discuss coding and demodulation techniques based on pilot symbols aimed at combating the cycle-slip phenomenon. The drawback of pilot symbols is that they are introduced at the expense of spectral efficiency. To mitigate the loss of spectral efficiency one can resort to the approach of [22], [23]. Basically, strongly encoded symbols are periodically inserted in the sequence of modulated symbols. Thanks to their strong encoding, these *quasi-pilot* symbols allow for robust decision-directed phase detection, as it happens in conventional decision-directed phase detection schemes operating at high signal-to-noise ratio (SNR). An advanced approach to demodulation is based on the union of pilot symbols and iterative demodulation and decoding [24], [25], where pilot symbols allow for bootstrapping the receiver before the first decoding, then the following demodulation stages take advantage from the estimates of source symbol's log-likelihood ratios (LLRs) obtained by soft-in soft-out (SISO) decoding of the error correcting code. The aim of this paper is to show that trellis-based demodulation can be used to bootstrap the iterative process of [24] without the need of pilot symbols, leading to throughput increment.

To show the benefits of trellis-based demodulation we rely upon the analysis of the information rate transferred through the phase noise channel. It is strongly intuitive that inaccuracies that are introduced in tracking the time-varying phase process can impair the information rate between the source and channel's output. This topic has been deeply investigated in the past. Specifically, the information rate transferred over the Wiener phase noise single input single output channel is considered in [26]–[36], while the information rate transferred over the more general autoregressive moving-average phase noise channel is studied in [37], [38]. White phase noise with uncorrelated samples is studied in [39]. In [2], [40], [41] the impact of phase noise on the capacity of the multiple input multiple output channel is considered.

The source/channel information rate is a limit that can be achieved only by ideal demodulation, while, with suboptimal
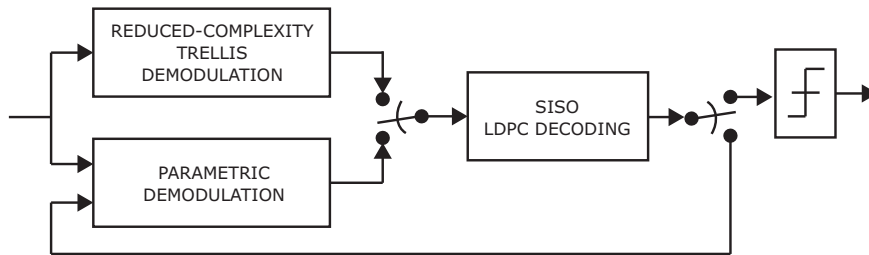
Fig. 1: Block diagram of the proposed receiver with hybrid iterative demodulation and decoding. In the receiver proposed in [24] the switch between demodulation and decoding is always connected to the block "parametric demodulation".

demodulation, the actual information rate extracted by the receiver depends on the specific demodulation algorithm. An observation of major practical importance that we can make by studying the information rate is that, as it will be clear from the results to be hereafter presented, the information rate achieved by non-iterative and non-pilot-aided demodulation can be close to the information rate of ideal demodulation in many channels of practical interest, provided that demodulation fully exploits the information that can be obtained from the received signal about the carrier phase. To fully exploit the information carried by the received signal, one can adopt non-parametric Bayesian tracking methods as the trellis-based demodulation scheme of [15] which relies upon the classical forward/backward recursion on the trellis of [42]. The main difference between the conventional parametric approach and non-parametric Bayesian tracking is the following. In the conventional parametric approach, the demodulation algorithm produces an estimate of two parameters that aims to represent the unknown phase: its mean value and variance given the available observations. In contrast, in trellis-based demodulation the entire probability distribution of the phase is computed and used in an integral equation to produce the wanted statistics.

After having observed that, in many cases of practical interest, non-iterative and non-data-aided non-parametric demodulation can virtually extract all the information that can be extracted by ideal demodulation, our next step, and main contribution of the paper, is to propose non-pilot-aided trellis demodulation to bootstrap the iterative demodulation and decoding algorithm of Colavolpe Barbieri and Caire [24] (CBC in the following). We show that, once properly initialized, the CBC algorithm does not need the aid of pilot symbols, hence the combination of a first trellis-based demodulation followed by the iterative algorithm allows for saving the bit rate that would be elsewhere sacrificed to pilot symbols. However, while the complexity of the parametric iterative CBC algorithm is moderate, the complexity in terms of number of operations per symbol of trellis-based demodulation can be large, making it not so attractive for those systems where, due to the very high symbol rate, the number of operations per symbol can become a stringent constraint. This motivates the second contribution of the paper, that is the introduction of techniques that allow to reduce the complexity of the full trellis where, with full trellis, we mean a trellis parametrized in such a way that it achieves virtually optimal performance. The block diagram of the hybrid non-parametric/parametric

iterative demodulation and SISO decoding receiver that we propose is shown in Fig. 1.

The outline of the paper is as follows. In Section II the notation and the channel model are given. Section III deals with the achievable information rate of general channels with memory, and presents formulas for the information rate achieved by modulations based on the specific bit mapping that will be adopted in the Section devoted to simulation results. In Section IV ideal demodulation, that is, information-rate achieving demodulation, and the practical demodulation schemes of [15] and CBC, are illustrated. Section V is devoted to complexity reduction of the trellis-based algorithm, while in Section VI simulation results for the achievable information rate and for the bit error rate (BER) performance of the proposed iterative demodulation and decoding scheme and of competitor schemes are presented. Finally, conclusive remarks are given in Section VII.

II. NOTATION AND CHANNEL MODEL

Let u_i^k indicate the vector $(u_i, u_{i+1}, \cdots, u_k)$ with $u_i^k \in \mathcal{U}_i^k$. When $k < i$ we assume that u_i^k is void. Let U indicate a stationary and ergodic process, $U = (U_1, U_2, \cdots)$, whose generic realization is the sequence $(u_1, u_2, \cdots)$. When $\mathcal{U}_i^k$ is a continuous set, $p(u_i^k)$ is used to indicate the multivariate probability density function of U_i^k evaluated in u_i^k, while when $\mathcal{U}_i^k$ is a discrete set $p(u_i^k)$ indicates the multivariate mass probability $\Pr(U_i^k = u_i^k)$ and $|\mathcal{U}_i|$ denotes the number of elements in $\mathcal{U}_i$.

Assuming that the chromatic dispersion has been compensated before demodulation and that laser phase noise is constant during a symbol interval, the k-th output of the phase noise channel is

$$y_k = x_k e^{j\phi_k} + w_k, \quad k = 1, 2, \cdots, \quad (1)$$

where $j = \sqrt{-1}$ is the imaginary unit, Φ is the phase noise process, W is a complex-valued additive white Gaussian noise (AWGN) process with two-dimensional variance SNR^{-1}, and the input modulation process X is made by points drawn from a two-dimensional M-ary constellation $\mathcal{X}$ carved out from the grid of two-dimensional integers $\mathcal{Z}^2$, scaled in order to have a unitary average input power. For the first-order continuous-state Markov phase process Φ we take a discrete-time Wiener process folded in the range $[0, 2\pi)$:

$$\phi_k = [\phi_{k-1} + \gamma v_{k-1}]_{\text{mod } 2\pi}, \quad k = 1, 2, \cdots \quad (2)$$

where the frequency noise V is a sequence of independent and identically distributed Gaussian random variables with zero mean and unit variance, the step size γ is a real scalar, and ϕ_0 is uniformly distributed in $[0, 2\pi)$. The phase process (2) is obtained by sampling at symbol frequency the phase of a continuous-time complex exponential $\{e^{j\phi(t)}\}$ whose power spectral density is the Lorentzian function

$$\mathcal{L}(f) = \frac{4\gamma^2 T}{\gamma^4 + 16\pi^2 f^2 T^2},$$

where T is the symbol repetition interval and f is the frequency. The parameter γ^2 is often expressed as

$$\gamma^2 = 2\pi \Delta \nu T,$$

where $\Delta \nu$ is the full-width half-maximum bandwidth of the spectral line.

The *state transition probability* of the hidden phase is

$$p(\phi_k | \phi_{k-1}) = \sum_{i=-\infty}^{\infty} g(\phi_{k-1} + 2i\pi, \gamma^2; \phi_k), \qquad (3)$$

where $g(\mu, \sigma^2; x)$ indicates a Gaussian probability density function of x with mean μ and variance σ^2. The *channel probability* is

$$\begin{aligned} p(y_k | \phi_k) &= \sum_{x_k \in \mathcal{X}} p(x_k, y_k | \phi_k) \\ &= \sum_{x_k \in \mathcal{X}} p(x_k) p(y_k | \phi_k, x_k), \qquad (4) \end{aligned}$$

where the last step is due to independence between X_k and Φ_k, and from the channel model (1), since the additive noise is Gaussian and with circular symmetry, one has

$$p(y_k | \phi_k, x_k) = g_c(x_k e^{j\phi_k}, \text{SNR}^{-1}; y_k), \qquad (5)$$

where $g_c(\mu, \sigma^2; x)$ indicates a two-dimensional Gaussian probability density function with circular symmetry over the complex plane spanned by x with mean μ and two-dimensional variance σ^2.

III. ACHIEVABLE INFORMATION RATE

Consider a transmission scheme where the source is the vector X_1^N of N independent random variables, and the channel output is Y_1^N. The achievable information rate between the source and channel's output can be written by chain rule as

$$\begin{aligned} I(X; Y) &\triangleq \lim_{N \to \infty} \frac{1}{N} I(X_1^N; Y_1^N) \\ &= \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} I(X_k; Y_1^N | X_1^{k-1}). \qquad (6) \end{aligned}$$

The above formula is very general and can be used as a starting point when one studies the information rate transferred through a channel with memory. In (6), the information extracted about X_k takes advantage from the knowledge of past source symbols X_1^{k-1}. A scheme that allows to make use of past source symbols is the joint equalization and decoding scheme proposed in [43] for the inter-symbol interference channel, which, in principle, could be used also in the phase noise

channel. However, practical use of this scheme, that is based on interleaving codewords with interleaving depth equal to the memory of the channel, is limited by the high latency that it introduces, which turns out to be too high in many practical systems.

When demodulation and decoding are disjoint, demodulation cannot exploit the knowledge of past symbols, that is, the conditions appearing in the argument of the mutual information in the right side of (6) are dropped. In many cases, one uses preliminary decisions taken from some signal processing made on Y_1^N in place of the actual past source symbols, but we know that this approach can lead to unsatisfactory performance when the error probability on preliminary decisions is high. Whatever use we make of Y_1^N, renouncing to the observations X_1^{k-1} one renounces to some information rate, as it is apparent from the following:

$$\begin{aligned} I(X_k; Y_1^N | X_1^{k-1}) &= H(X_k | X_1^{k-1}) - H(X_k | Y_1^N, X_1^{k-1}) \\ &= H(X_k) - H(X_k | Y_1^N, X_1^{k-1}) \qquad (7) \\ &\geq H(X_k) - H(X_k | Y_1^N) \qquad (8) \\ &= I(X_k; Y_1^N) \end{aligned}$$

where the inequality is because conditioning does not increase entropy.

Bit Mapping

Consider the case of disjoint demodulation and decoding where $m = \log_2 M$ independent bits are mapped onto a constellation with M points, and let $B_{1;k}^m = (B_{1;k}, B_{2;k}, \cdots, B_{m;k})$ be the vector of bits mapped onto X_k. In the following, we will consider a mapping scheme that is based on the partition $\mathcal{Z}^2 / 4\mathcal{Z}^2$. In this mapping, the generic constellation point X is decomposed as

$$X = 4X' + X'',$$

where $\mathcal{X}''$ is 16-QAM with Gray mapping, $\mathcal{X}'$ is $(M/16)$-QAM with Gray mapping, and the minimum squared Euclidean distance in $\mathcal{X}'$ and $\mathcal{X}''$ is $6(M^2 - 1)^{-1}$, thus obtaining unitary power for $\mathcal{X}$. The four bits that select the point in $\mathcal{X}'$ are coded by a unique error correcting code and their LLRs are independently computed from the demodulated signal, while the bits of $\mathcal{X}''$ are left uncoded and are demodulated after the decision of the error correcting code [44], [45]. The information rate brought by this LLRs computation scheme is

$$\sum_{i=1}^{4} I(B_{i,k}; Y_1^N) + \sum_{i=5}^{m} I(B_{i;k}; Y_1^N | B_{1;k}^4) \qquad (9)$$

$$\leq I(B_{1;k}^m; Y_1^N) = I(X_k; Y_1^N), \qquad (10)$$

where the inequality can be demonstrated by observing that intra-symbol chain rule is not exploited in the first sum, and that it is only partially exploited in the second sum. Actually, although source bits are assumed here to be independent, they are not conditionally independent given Y_1^N. However, in many practical systems, (9) is close to (10). Numerically one finds that the penalty payed by the first sum is negligible to all practical purposes, at least on the AWGN channel. Moreover,

with a redundancy below 1bit/2D,[1] the error probability on the uncoded bits given zero errors on the coded bits is virtually zero at the SNR corresponding to the constrained capacity, hence

$$I(B_{i;k}; Y_1^N | B_{1;k}^4) \approx 1, \quad i > 4,$$

meaning that also the penalty payed by the second sum compared to the case where intra-symbol chain rule is fully exploited is negligible.

Consider the four coded bits, that is the first sum in (9). The achievable information rate can be evaluated by Monte-Carlo integration as

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{4} \log_2 \frac{p(b_{i;k}|y_1^N)}{p(b_{i;k})}. \quad (11)$$

The probability of the bits given the received signal is computed from the probability of constellation points as

$$p(b_{i;k} = 0 | y_1^N) = \sum_{x_k \in \mathcal{X}: b_{i;k}=0} p(x_k | y_1^N). \quad (12)$$

The second sum in (9) is treated in a similar manner, the only difference being that the four coded bits are now given as a condition.

IV. DEMODULATION

Ideal demodulation, which exploits the conditions X_1^{k-1} appearing in the conditional entropy (7), can be performed as

$$p(x_k|y_1^N, x_1^{k-1}) = \int_0^{2\pi} p(\phi_k, x_k | y_1^N, x_1^{k-1}) \,\mathrm{d}\phi_k$$

$$= \int_0^{2\pi} p(\phi_k | y_1^N, x_1^{k-1}) p(x_k | y_k, \phi_k) \,\mathrm{d}\phi_k$$

$$\overset{(a)}{\propto} \int_0^{2\pi} p(y_1^{k-1}, x_1^{k-1} | \phi_k) p(\phi_k, y_k^N) p(x_k | y_k, \phi_k) \,\mathrm{d}\phi_k \quad (13)$$

$$\propto \int_0^{2\pi} \frac{p(\phi_k | y_1^{k-1}, x_1^{k-1})}{p(\phi_k)} p(\phi_k | y_k^N) p(x_k | y_k, \phi_k) \,\mathrm{d}\phi_k$$

$$\propto \int_0^{2\pi} p(\phi_k | y_1^{k-1}, x_1^{k-1}) p(\phi_k | y_k^N) p(x_k | y_k, \phi_k) \,\mathrm{d}\phi_k, \quad (14)$$

where $\propto$ indicates that a factor independent of x_k and ϕ_k has been brought outside the integral and cancelled, the product $p(y_1^{k-1}, x_1^{k-1} | \phi_k) p(\phi_k, y_k^N)$ in step (a) is obtained from $p(\phi_k | y_1^N, x_1^{k-1})$ by passages similar to those of [42], the last step of (14) holds because $p(\phi_k)$ is uniform, and

$$p(x_k | y_k, \phi_k) = \frac{p(y_k | \phi_k, x_k) p(x_k)}{p(y_k | \phi_k)}.$$

The forward and backward probability distributions of the phase appearing inside the integral (14) are iteratively computed by Bayesian tracking:

$$p(\phi_{k+1} | y_1^k, x_1^k)$$
$$\propto \int_0^{2\pi} p(\phi_k | y_1^{k-1}, x_1^{k-1}) p(y_k, x_k | \phi_k) p(\phi_{k+1} | \phi_k) \,\mathrm{d}\phi_k,$$

[1]It is shown in [46] that, when the redundancy is 1bit/2D, the gap from the constrained capacity of the AWGN channel in the bandwidth constrained region is about 0.1 dB. Hence, the potential benefit of adding redundancy greater than 1bit/2D is negligible.

$$p(\phi_k | y_k^N) \propto p(y_k | \phi_k) \int_0^{2\pi} p(\phi_{k+1} | y_{k+1}^N) p(\phi_k | \phi_{k+1}) \,\mathrm{d}\phi_{k+1},$$

where the *blind* channel probability $p(y_k | \phi_k)$ is computed from (4) by taking for $p(x_k)$ a uniform distribution on $\mathcal{X}$, while the *data-aided* channel probability $p(y_k, x_k | \phi_k) = p(y_k | x_k, \phi_k) p(x_k)$ is computed by taking for the probability distribution of input data $p(x_k)$ an indicator function that is non-zero only for the input symbol visited by the realization.

When past input data are not available one is forced to resort to non-data-aided demodulation, that is similar to (14), but the conditions x_1^{k-1} are dropped:

$$p(x_k | y_1^N) \propto \int_0^{2\pi} p(\phi_k | y_1^{k-1}) p(\phi_k | y_k^N) p(x_k | y_k, \phi_k) \,\mathrm{d}\phi_k. \quad (15)$$

In iterative demodulation and decoding, demodulation is based on (15), but, after the first decoding, one can use the probability distribution $p(x_k)$ coming from the decoder as *extrinsic* information in (4) (not to compute the *intrinsic* term $p(x_k | y_k, \phi_k)$ inside the integral (15)) to get the transition metric of forward and backward recursions.

In practice, the continuous-state channel model is intractable. To cope with the continuous-state model one can compute a non-parametric approximation to the wanted probability distribution by introducing an auxiliary channel where the state space is discretized into bins, leading to a trellis-based representation of the evolution of the phase in the discrete-state space, where trellis' states are the centroids of the bins, see for instance [15], [29] for details. However, trellis-based demodulation can be complex, and the complexity of demodulation increases by a factor equal to the number of iterations. Therefore we propose trellis-based demodulation at the first demodulation step followed by the classical CBC algorithm.

The CBC algorithm, where the probability distribution of the hidden phase given the observation is modelled as a Tikhonov distribution[2], is less complex than the trellis-based technique. In CBC, the incoming phase distribution is obtained by tracking the parameter of the Tikhonov distribution. To make the paper self-contained, we recall here the main steps of CBC. The two-step forward recursion for the tracked parameter is

$$\overline{a}_k^f = a_{k-1}^f + 2 \frac{y_{k-1} \alpha_{k-1}^*}{\mathrm{SNR}^{-1} + \beta_{k-1}^2 - |\alpha_{k-1}|^2}, \quad (16)$$

$$a_k^f = \frac{\overline{a}_k^f}{1 + \gamma^2 |\overline{a}_k^f|}, \quad (17)$$

where c^* denotes the complex conjugate of c and

$$\alpha_k = \sum_{x_k \in \mathcal{X}} x_k \, p(x_k),$$

$$\beta_k^2 = \sum_{x_k \in \mathcal{X}} |x_k|^2 \, p(x_k).$$

[2]Also known as Von Mises distribution.

The parametric approximation to the wanted probability distribution is computed as

$$p(x_k|y_1^N) \propto e^{-|x_k|^2 \text{SNR}} I_0 \left(\left| a_k^f + a_k^b + 2\,\text{SNR}\, y_k x_k^* \right| \right), \quad (18)$$

where a_k^b is obtained from a backward recursion similar to the forward one and $I_0(\cdot)$ is the zero-th order modified Bessel function of first kind. To prevent numerical problems that occur when the argument of the Bessel function becomes too large, one can use the exponential approximation

$$I_0(z) \approx \frac{e^z}{\sqrt{2\pi z}}, \quad z \gg 0, \quad (19)$$

leading to

$$p(x_k|y_1^N) \propto \frac{e^{-|x_k|^2 \text{SNR} + \left| a_k^f + a_k^b + 2\,\text{SNR}\, y_k x_k^* \right| - \lambda}}{\sqrt{2\pi \left| a_k^f + a_k^b + 2\,\text{SNR}\, y_k x_k^* \right|}}, \quad (20)$$

where the role of λ, which is up to the designer, is that of keeping under control the exponent of the exponential function. At the first iteration of the iterative demodulation and decoding algorithm, extrinsic information about the input symbols is not available, hence the parameters α_k and β_k are computed by assuming uniform distribution for $p(x_k)$. If the mean value of constellation's symbols is zero, then α_1^N is a vector of N zeros, hence from the numerator of (16) one realizes that the algorithm cannot bootstrap.

V. REDUCED COMPLEXITY TRELLIS-BASED FORWARD/BACKWARD DEMODULATION

In trellis-based demodulation, the continuous phase is discretized into $|\mathcal{S}|$ bins, where the discrete state s, takes its values in the set of bins' centroids:

$$\mathcal{S} = \{-(|\mathcal{S}|-1)\Delta, -(|\mathcal{S}|-3)\Delta, \cdots, (|\mathcal{S}|-1)\Delta\}, \quad (21)$$

where

$$\Delta = \frac{\pi}{|\mathcal{S}|} \quad (22)$$

is half the bin width. For the sake of simplicity, we limit here ourselves to uniform and time-invariant quantization. The state transition probability is

$$p(s_k|s_{k-1}) = \int_{\mathcal{R}(s_{k-1}^k)} \frac{p(\phi_k|\phi_{k-1})}{\Delta} \, d\phi_k \, d\phi_{k-1}, \quad (23)$$

where $\mathcal{R}(s_{k-1}^k)$ indicates the two-dimensional quantization region whose centroid is s_{k-1}^k. Since $p(s_k|s_{k-1})$ depends only on the difference $s_k - s_{k-1}$, it takes its values in a set of $|\mathcal{S}|$ numbers that can be pre-computed by equation (23). The probability $p(y_k|s_k, x_k)$ is similar to (5) with the centroid of the bin s_k in place of the continuous phase ϕ_k.

A. Reducing the State Transition Frequency

For small step size γ, an approximation to the wanted distribution can be computed by introducing an auxiliary discretized phase noise process obtained by merging n steps of the random walk. This leads to the following auxiliary state transition probability

$$p(s_k|s_{k-1}) = \int_{\mathcal{R}(s_{k-1}^k)} \frac{p(\phi_{kn}|\phi_{(k-1)n})}{\Delta} \, d\phi_{kn} \, d\phi_{(k-1)n}, \quad (24)$$

with

$$p(\phi_{kn}|\phi_{(k-1)n}) = \sum_{i=-\infty}^{\infty} g(\phi_{(k-1)n} + 2i\pi, n\gamma^2; \phi_{kn}). \quad (25)$$

The auxiliary channel probability is obtained by approximating phase noise to a constant between time kn and time $kn+n-1$:

$$p(y_{kn}^{kn+n-1}|s_k) = \sum_{x_{kn}^{kn+n-1} \in \mathcal{X}^n} p(y_{kn}^{kn+n-1}, x_{kn}^{kn+n-1}|s_k)$$

$$= \sum_{x_{kn}^{kn+n-1} \in \mathcal{X}^n} \prod_{i=0}^{n-1} p(y_{kn+i}, x_{kn+i}|s_k)$$

$$= \prod_{i=0}^{n-1} \sum_{x_{kn+i} \in \mathcal{X}} p(y_{kn+i}, x_{kn+i}|s_k)$$

$$= \prod_{i=0}^{n-1} \sum_{x_{kn+i} \in \mathcal{X}} p(y_{kn+i}|x_{kn+i}, s_k) p(x_{kn+i}), \quad (26)$$

where

$$p(y_{kn+i}|x_{kn+i}, s_k) = g_c(x_{kn+i} e^{js_k}, \text{SNR}^{-1}; y_{kn+i}).$$

A transition on the trellis corresponds to a block of n source symbols, and, for $i = 0, 1, \cdots n-1$, the probability distribution of the i-th symbol in the k-th block is computed as

$$p(x_{kn+i}|y_1^{nN}) \propto \sum_{s_k \in \mathcal{S}} p(s_k|y_1^{kn-1}) p(s_k|y_{kn}^{Nn}) p(x_{kn+i}|y_{kn+i}, s_k), \quad (27)$$

where the forward and backward probabilities are recursively computed by the state transition probability (24) and by the channel probability (26). By reducing the state transition frequency one reduces the number of serial operations per second. This is crucial for feasibility of the serial algorithm in high-speed processing implemented in optical receivers. Anyway, if further reduction of serial operations per second is needed, it can be achieved by parallelizing the processing of codewords at the expense of latency.

B. Reducing the Computation per State Transition

By looking at the phase evolution in time domain, at time instant k only a small range of the phase domain has non-negligible probability, and the range with non-negligible probability can be predicted from the distribution at time $k-1$ and tracked by a feedback loop. Based on this observation, we propose a state reduction technique where the set of trellis states $\mathcal{S}'$ is a subset of the set $\mathcal{S}$ of (21) centered around phase zero, that is

$$\mathcal{S}' = \{-(|\mathcal{S}'|-2)\Delta, -(|\mathcal{S}'|-4)\Delta, \cdots, (|\mathcal{S}'|-2)\Delta\}, \quad (28)$$

where Δ is the same as (22) and $|\mathcal{S}'| \leq |\mathcal{S}|$ is even and up to the user.

Assuming that s_0 is known and starting from $\mu_0 = s_0$, $p(s'_0 = 0) = 1$, for $k = 1, 2, \cdots, N$ the forward recursion is

$$p(s'_k | u_1^{k-1}) = \sum_{s'_{k-1} \in \mathcal{S}'} p(s'_k | s'_{k-1}) p(s'_{k-1} | u_1^{k-1}), \qquad (29)$$

$$\mu_k = \mu_{k-1} + \arg \max_{s'_k \in \mathcal{S}'} p(s'_k | u_1^{k-1}), \qquad (30)$$

$$u_k = y_k e^{-j\mu_k}, \qquad (31)$$

$$p(s'_k | u_1^k) \propto p(u_k | s'_k) p(s'_k | u_1^{k-1}), \qquad (32)$$

and the distribution $p(s'_k | u_1^k)$ for s'_k outside the binned range is set to zero. Basically, between the classical predict (29) and update (32) steps of Bayesian tracking, we introduce equations (30) and (31) whose scope is that of tracking the hidden phase to maintain the phase error $\phi_k - \mu_k$ inside the binned range. Note that, since $p(s'_k | s'_{k-1})$ depends only on the difference $[s'_k - s'_{k-1}]_{\mathrm{mod}\ 2\pi}$, the distribution $p(s'_k | u_1^{k-1})$ in the left hand side of (29) is proportional to the convolution between the state transition probability, which is constant and can be pre-computed, and the distribution $p(s_{k-1} | u_1^{k-1})$. Therefore, for number of states greater than 32, (29) can be efficiently computed as: fast Fourier transform (FFT) of the state metrics $p(s'_{k-1} | u_1^{k-1})$, product in frequency domain by the FFT of the constant vector of state transition probabilities, inverse FFT.

Assuming that s_{N+1} is known, the backward recursion is performed in a similar way. To put together the contributions coming from forward and backward recursions, the correspondence between the absolute phase and the states of the trellis must be recovered. Specifically, the wanted probabilities of the actual state s_k are obtained from those of the state s'_k by the variable transformation

$$s_k = [s'_k + \mu_k]_{\mathrm{mod}\ 2\pi}, \qquad (33)$$

of course using the forward μ_ks for the forward recursion and the backward μ_ks for the backward one. The state reduction technique proposed here fails when the set of states with nonzero probability obtained from the forward recursion has zero overlap with the set of states with nonzero probability obtained from the backward recursion. When this happens, for instance due to a cycle slip occurred in one of the two recursions, (27) does not provide a meaningful distribution.

Another saving can be made on the discrete-state version of (4) which, for QAM with large number of points, can be unnecessarily complex. Actually, at intermediate-to-high SNR only N_p constellation points give non-negligible contribution to the sum. To further simply signal processing, we suggest to associate with each constellation point a look-up table (LUT) that contains the coordinates of nearest constellation's points. After this, only the hard decision on the de-rotated signal in (31) and the points in the LUT associated with the hard-decision are considered in (4), thus avoiding the search of the nearest constellation points to the received signal.

VI. SIMULATION RESULTS

In the simulation results presented in this Section 16-QAM and 64-QAM are considered.
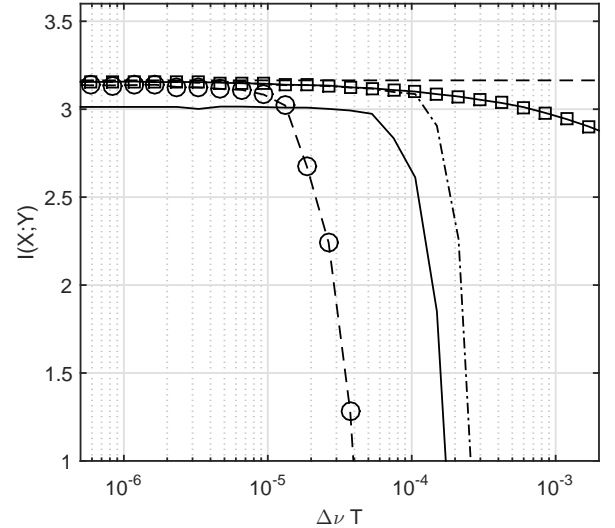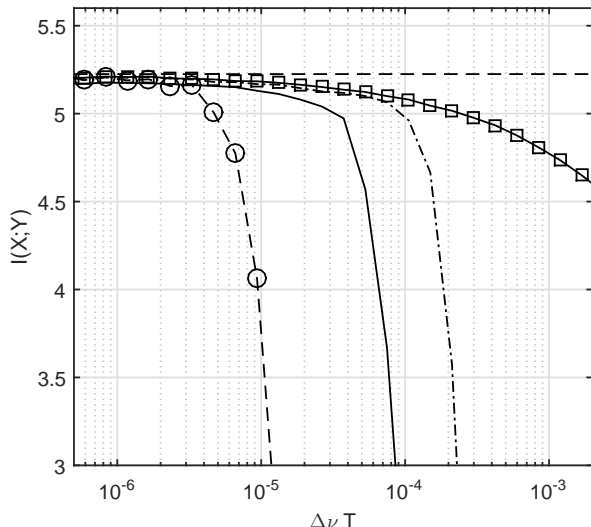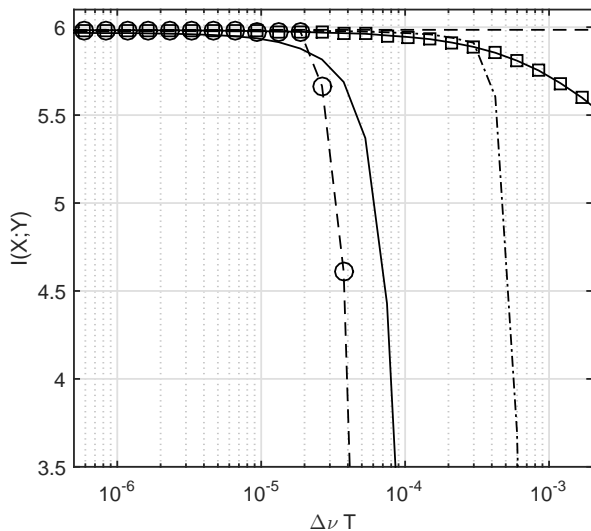

(a) SNR= 10 dB
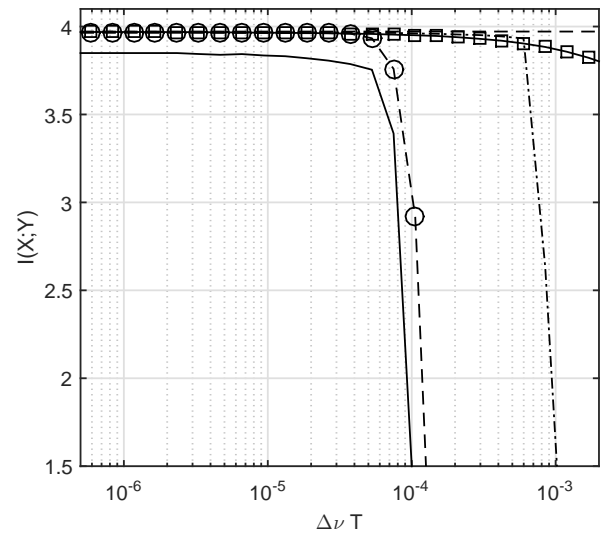
(b) SNR= 16 dB

Fig. 2: Achievable information rate versus $\Delta\nu T$ for the phase noise channel with 16-QAM and two values of SNR. Dashed line: pure AWGN ($\Delta\nu T = 0$). Solid line with squares: full trellis, forward data-aided recursion. Dash-dotted line: full trellis with forward/backward non-data-aided recursions. Solid line: reduced complexity trellis with forward/backward non-data-aided recursions. Dashed line with circles: Circular QAM with Viterbi-Viterbi carrier phase estimation [7].

Figures 2 and 3 report the achievable information rate versus normalized full-width half-maximum bandwidth $\Delta\nu T$ for various types of demodulation and two values of SNR per modulation. In addition to the information rate of the pure AWGN channel ($\Delta\nu T = 0$), for each pair SNR/modulation five curves are drawn. Two out of five curves refer to a full version of the trellis with $\Delta = \pi/512$ and $n = 1$. The state transition metric of the full trellis is based on 16 points with 16-QAM and on 49 points with 64-QAM. One of the two curves is obtained with the data-aided forward recursion and no non-data-aided backward recursion. Although the non-data-aided

(a) SNR = 17 dB

(b) SNR = 23 dB

Fig. 3: Achievable information rate versus $\Delta\nu T$ for the phase noise channel with 64-QAM and two values of SNR. Dashed line: pure AWGN ($\Delta\nu T = 0$). Solid line with squares: full trellis, forward data-aided recursion. Dash-dotted line: full trellis with forward/backward non-data-aided recursions. Solid line: reduced complexity trellis with forward/backward non-data-aided recursions. Dashed line with circles: Circular QAM with Viterbi-Viterbi carrier phase estimation [7].

TABLE I: Parameters of the trellis used to get the performance represented by the solid line curves in Figs. 2-5.

	16-QAM	64-QAM
Half bin width, Δ	$\pi/16$	$\pi/64$
Number of bins, $|\mathcal{S}'|$	8	8
Number of nearest constellation points, N_p	16	9
Number of merged steps, n	73	73

methods give virtually indistinguishable results with low-to-intermediate phase noise, while, with strong phase noise, the performance of non-data-aided demodulation suddenly drops after a threshold that depends on SNR, while the performance of data-aided demodulation degrades smoothly. At the threshold, non-data-aided demodulation starts to be affected by cycle slips, and, above the threshold, the only way to prevent cycle slips is of adding pilot symbols. The information rate extracted by the non-data-aided forward/backward trellis with reduced complexity is obtained with the dramatic reduction of complexity of the full trellis reported in Table I. The phase range spanned by the reduced trellis is $(-\pi/2, \pi/2)$ and $(-\pi/8, \pi/8)$ for 16-QAM and 64-QAM, respectively. We observe that the complexity reduction operated on the trellis affects the cycle-slip threshold, while the performance in the cycle-slip-free region is virtually unaffected by complexity reduction. The gap between the threshold of the reduced-complexity trellis and the threshold of the full trellis can be bridged by acting on the design parameters that have been introduced in the section devoted to complexity reduction. As a competitor of the reduced-complexity trellis, in Figs. 2 and 3 the information rate extracted by the Viterbi-Viterbi carrier phase estimator with circular QAM modulation proposed in [7] has been reported. For each value of $\Delta\nu T$ the duration of the sliding-window filter of [7] has been optimized. Note that, despite the strong complexity reduction reported in Tab. I, the threshold of trellis-based tracking outperforms the threshold of [7] at low SNR, that is at the SNR of interest for coded systems.

To show the impact of the results presented in Fig. 2 on the iterative demodulation and decoding algorithm, we design a coded system with a binary low-density parity-check (LDPC) code with $(n, k) = (4088, 3066)$ derived from the LDPC code with rate 7/8 in [47] by using the dummy bit insertion technique described in [48]. We hasten to point out that deeply investigating coding methods for channels affected by phase noise is out of the scope of the present paper, which is focused on demodulation. One pilot symbol is inserted at the beginning and at the end of every block. 16-QAM exploits Gray mapping of all the encoded bits, leading to a spectral efficiency of 3 bits per QAM symbol. Motivated by the discussion in Section III, 64-QAM is based on the $\mathcal{Z}^2/4\mathcal{Z}^2$ mapping, where four bits per symbol are encoded and two bits are left uncoded, leading to a spectral efficiency of 5 bits per QAM symbol. (The choice of a redundancy of one bit per dimension pair is discussed in footnote 2.) The normalized full-width half-maximum bandwidth of the spectral line $\Delta\nu T$ is $1.3 \cdot 10^{-5}$ for 16-QAM and $3.7 \cdot 10^{-5}$ for 64-QAM. The first demodulation and decoding step is performed by a reduced-
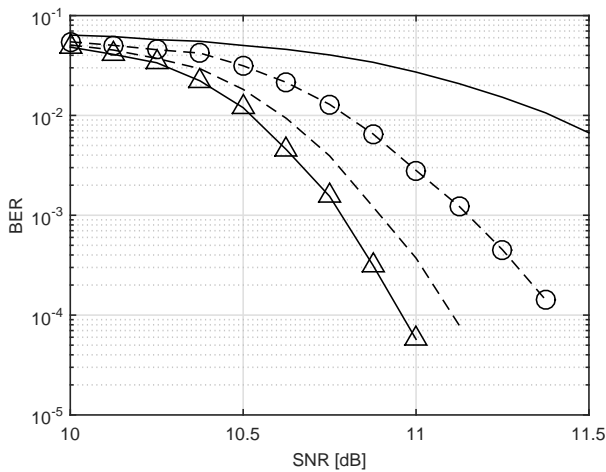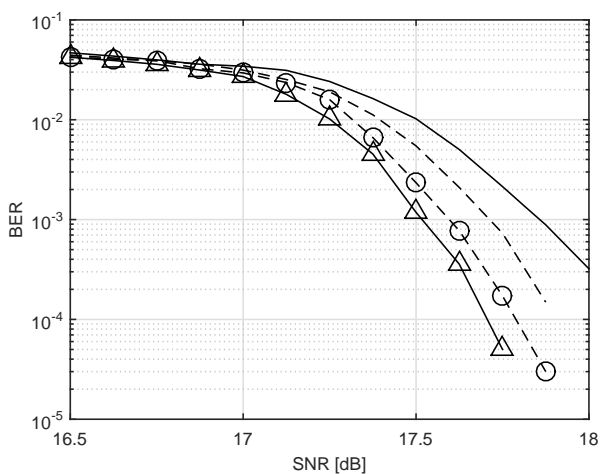
backward recursion would, at least theoretically, improve the information rate, it happens that, with strong phase noise, cycle slips affect non-data-aided backward tracking, leading to numerical problems that appear when one tries to compute the product inside the integral (14). For this reason we report only the information rate extracted by the data-aided forward recursion, which is virtually optimal because the non-data-aided recursion is much weaker than the data-aided one, therefore in practice it can be ignored. The second curve with full trellis is obtained with non-data-aided forward and backward recursions. Figures 2 and 3 show that the two demodulation
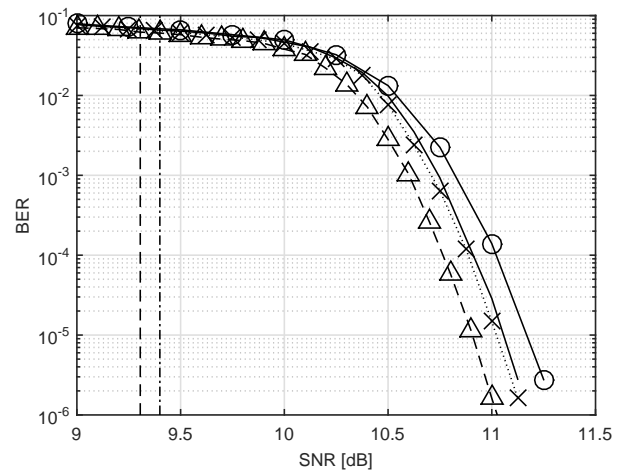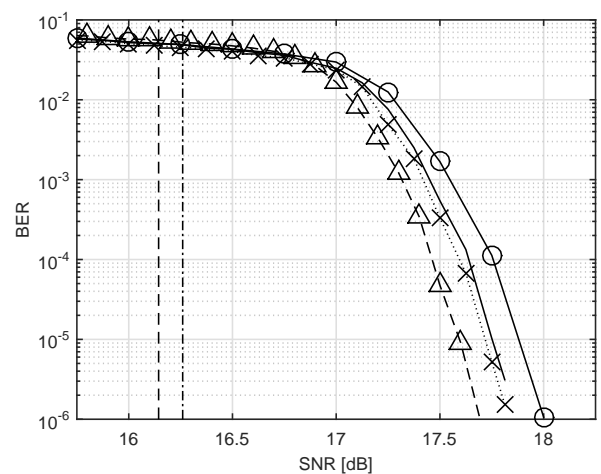
(a) 16-QAM

(a) 16-QAM

(b) 64-QAM

(b) 64-QAM

Fig. 4: BER versus SNR after the first demodulation and decoding step, no pilot symbols. (a) 16-QAM, phase range spanned by the trellis $(-\pi/2, \pi/2)$. Solid line: $|\mathcal{S}'| = 8$, $n = 73$. Dashed line: $|\mathcal{S}'| = 32$, $n = 73$. Dashed line with circles: $|\mathcal{S}'| = 16$, $n = 2$. Solid line with triangles:$|\mathcal{S}'| = 128$, $n = 2$. (b) 64-QAM, phase range spanned by the trellis $(-\pi/8, \pi/8)$, and number of nearest constellation points $N_p = 9$. Solid line: $|\mathcal{S}'| = 8$, $n = 73$. Dashed line: $|\mathcal{S}'| = 32$, $n = 73$. Dashed line with circles: $|\mathcal{S}'| = 8$, $n = 14$. Solid line with triangles:$|\mathcal{S}'| = 128$, $n = 2$.

Fig. 5: BER versus SNR. Dashed line: performance limit of AWGN channel. Dash-dotted line: performance limit of AWGN and phase noise channel. Dashed line with triangles: pure AWGN. Solid line: hybrid iterative demodulation and decoding without pilot symbols. Dotted line with crosses: full trellis with data-aided forward recursion and non-data-aided backward recursion. Solid line with circles: iterative demodulation and decoding of [3] with pilot rate 1/25.

complexity trellis with non-data-aided forward and backward recursion and 5 iterations of the LDPC decoder. The BER after the first demodulation and decoding step is reported in Fig. 4 for different parameters of the reduced-complexity trellis. A trellis with complexity greater than the one used to get the results marked with the solid line with triangles does not improve performance. From the results reported in Fig. 4 one observes that 64-QAM is much more sensitive than 16-QAM to n, the number of merged steps. After the first demodulation and decoding step, parametric demodulation and soft-output LDPC decoding are iterated as depicted in the block diagram of Fig. 1. The results of Fig. 5 are obtained with trellis parameters reported in Table I. In Fig. 5, the limit SNR with the given constellation with the pure AWGN channel and

the phase noise channel is drawn. Also, Fig. 5 reports the BER versus SNR for different demodulation methods with phase noise and the reference curve for the pure AWGN channel. For the phase noise channel, Fig. 5 reports the BER with ideal demodulation based on a finely discretized version of (14), the proposed hybrid iterative method of Fig. 1 with reduced-complexity trellis at the first demodulation stage with the complexity reduction of Table I, and the competitor method of [3] with pilot rate 1/25 initialized by the unconstrained Wiener filter studied in [13]. The parameter of iterative demodulation and decoding of the proposed method and of [3] are reported in Table II. It is apparent from Fig. 5 that, although the proposed method does not make use of pilot symbols, it outperforms the adversary method based on pilot symbols. Also, we observe that, despite the dramatic complexity reduction, the performance with reduced complexity trellis is virtually

TABLE II: Iterations performed by the proposed hybrid iterative demodulation and decoding algorithm.

	16-QAM	64-QAM
Number of demodulation and decoding iterations	3	2
Number of iterations of the LDPC decoder for each iteration of demodulation and decoding	5	5
Number of iterations of the LDPC decoder after the last demodulation	50	50

optimal. Actually, we have found that trellis complexity can be reduced without any appreciable performance degradation until cycle slips stars appearing. Moreover, the dramatic complexity reduction applied on the trellis makes the computation of our proposed method affordable also in high-speed data transmission.

VII. CONCLUSION

The paper has shown that there exists a threshold that divides the operation of iterative demodulation and decoding in two regions: low-to-intermediate phase noise and strong phase noise. Focusing on the region of low-to-intermediate phase noise, the paper has shown that the iterative process can bootstrap also without pilot symbols, provided that nonparametric demodulation, such as trellis-based demodulation, is adopted before the first decoding iteration. Complexity reduction techniques are proposed to mitigate the complexity of the forward/backward recursions on the trellis. The performance of iterative demodulation and decoding turns out to be virtually unaffected by complexity reduction, provided that it is adequate to guarantee cycle-slip-free operation of the first demodulation stage.

REFERENCES

[1] G. Colavolpe, "Communications over phase noise channels: a tutorial review," *Int. J. Satell. Commun. Network*, Wiley Online Library, vol. 32, no. 3, pp. 167-185, May/June 2014.

[2] G. Durisi, A. Tarable, C. Camarda, R. Devassy, and G. Montorsi, "Capacity bounds for MIMO microwave backhaul links affected by phase noise phase noise," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 920-929, March 2014.

[3] N. Kamiya and E. Sasaki, "Pilot-symbols assisted and code-aided phase error estimation for high-order QAM transmission," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4369-4380, Oct. 2013.

[4] P. Leoni, V. Sleiffer, S. Calabrò, M. Kuschnerov, S. L. Jansen, B. Spinnler, and B. Lankl, "On the performance of a soft decision FEC scheme operating in highly non-linear regime," in *Advanced Photonics Congress*, OSA Technical Digest (online) (Optical Society of America, 2012), paper SpTu3A.6.

[5] M. Magarini, L. Barletta, A. Spalvieri, A. Leven, M. Pepe, and G. Gavioli, "Impact of non-ideal phase reference on soft decoding of differentially encoded modulation," *IEEE Photon. Technol. Lett.*, vol. 24, no. 23, pp. 2179–2182, Dec. 1, 2012.

[6] B. Goebel, R.-J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5720-5736, Sept. 2011.

[7] S. Zafra, X. Pang, G. Jacobsen, S. Popov, and S. Sergeyev, "Phase noise tolerance study in coherent optical circular QAM transmissions with Viterbi-Viterbi carrier phase estimation," *Opt. Express*, vol. 22, no. 25, pp. 30579-30585, 2014.

[8] G. J. Foschini and G. Vannucci, "Characterizing filtered light waves corrupted by phase noise," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1437–1448, Nov. 1988.

[9] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 15, 2010.

[10] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989-999, Apr. 15, 2009.

[11] M. Morsy-Osman, Q. Zhuge, L. R. Chen, and D. V. Plant, "Feedforward carrier recovery via pilot-aided transmission for single-carrier systems with arbitrary M-QAM constellations," *Opt. Express*, vol. 19, pp. 24331–24343, Nov. 2011.

[12] F. Zhang, Y. Li, J. Wu, W. Li, X. Hong, and J. Lin, "Improved pilot-aided optical carrier phase recovery for coherent M-QAM," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 1577–1580, Sept. 2012.

[13] A. Spalvieri and L. Barletta, "Pilot-aided carrier recovery in the presence of phase noise," *IEEE Trans. Commun.*, vol. 59, pp. 1966-1974, July 2011.

[14] M. Magarini, L. Barletta, A. Spalvieri, F. Vacondio, T. Pfau, M. Pepe, M. Bertolini, and G. Gavioli, "Pilot-symbols-aided carrier-phase recovery for 100-G PM-QPSK digital coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 739-741, May 2012.

[15] L. Barletta, F. Bergamelli, M. Magarini, N. Carapellese, and A. Spalvieri, "Pilot-aided trellis-based demodulation," *IEEE Photon. Technol. Lett.*, vol. 25, pp. 1234-1237, July 2013.

[16] T. Koike-Akino, K. Kojima, D. Millar, K. Parsons, Y. Miyata, W. Matsumoto, T. Sugihara, and T. Mizuochi, "Cycle slip-mitigating turbo demodulation in LDPC-coded coherent optical communications," in *Opt. Fiber Commun. Conf. (OFC)*, 2014, paper M3A.3.

[17] T. Yoshida, T. Sugihara, K. Ishida, and T. Mizuochi, "Cycle slip compensation with polarization block coding for coherent optical transmission: Two-dimensional phases constellation corresponds to a slip state," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 57–69, Mar. 2014.

[18] L. Liu and L. Li, "Cycle-slip correction in 100 Gb/s PM-QPSK systems," in *Opt. Fiber Commun. Conf. (OFC)*, 2014, paper Th4D.2.

[19] H. Cheng, Y. Li, M. Yu, J. Zang, J. Wu, and J. Lin, "Experimental demonstration of pilot-symbols-aided cycle slip mitigation for QPSK modulation format,"in *Opt. Fiber Commun. Conf. (OFC)*, 2014, paper Th4D.1.

[20] H. Cheng, Y. Li, F. Zhang, J. Wu, J. Lu, G. Zhang, J. Xu, and J. Lin, "Pilot-symbols-aided cycle slip mitigation for DP-16QAM optical communication systems," *Opt. Express*, vol. 21, no. 19, pp. 22166–22172, Sep. 2013.

[21] M. Leong, K. Larsen, G. Jacobsen, S. Popov, D. Zibar, and S. Sergeyev, "Dimensioning BCH codes for coherent DQPSK systems with laser phase noise and cycle slips," *IEEE J. Lightw. Technol.*, vol. 32, no. 21, pp. 3446–3450, Nov. 2014.

[22] L. Barletta, M. Magarini, and A. Spalvieri, "Staged demodulation and decoding," *Opt. Express*, vol. 20, no. 21, pp. 23728–23734, Oct. 8, 2012.

[23] S. T. Le,T. Kanesan, E. Giacoumidis, N. J. Doran, and A. D. Ellis, "Quasi-pilot aided phase noise estimation for coherent optical OFDM systems," *IEEE Photon. Technol. Lett.*, vol. 26, pp. 504-507, Mar. 2014.

[24] 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, Sept. 2005.

[25] N. Noels, J. Bhatti, H. Bruneel, and M. Moeneclaey, "Block-processing soft-input soft-ouput demodulator for coded PSK using DCT-based phase noise estimation," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2939–2950, Aug. 2014.

[26] J. Dauwels and H.-A. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 406–409, Jan. 2008.

[27] L. Barletta, M. Magarini, and A. Spalvieri, "Estimate of information rates of discrete-time first-order Markov phase noise channels," *IEEE Photon. Technol. Lett.*, vol. 23, no. 21, pp 1582–1584, Nov. 1, 2011.

[28] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3223–3228, Dec. 2011.

[29] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *IEEE J. Lightw. Technol.*, vol. 30, no. 10, pp. 1480–1486, May 15, 2012.

[30] L. Barletta, M. Magarini, and A. Spalvieri, "A new lower bound below the information rate of Wiener phase noise channel based on Kalman carrier recovery," *Opt. Express*, vol. 20, no. 23, pp. 25471–25477, Nov. 5, 2012.

[31] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," *IEEE Intern. Symposium on Inf. Theory (ISIT),* pp. 2279–2283, July 2013, Istanbul, Turkey.

[32] M. Martaló, C. Tripodi, and R. Raheli, "On the information rate of phase-noise limited communications," *Proc. Inf. Theory and Applications Workshop (ITA),* pp. 1–7, Feb. 2013, San Diego, CA, USA.

[33] H. Ghozlan and G. Kramer, "Multi-sample receivers increase information rates for Wiener phase noise channels," *Proc. Globecom,* pp. 1897–1902, Dec. 2013, Atlanta, GA, USA.

[34] L. Barletta and G. Kramer, "Upper bound on the capacity of discrete-time Wiener phase noise channels," *IEEE Inf. Theory Workshop (ITW),* Apr. 2015, Jerusalem, Israel. Available online at http://arxiv.org/pdf/1411.0390v2.pdf.

[35] L. Barletta and G. Kramer, "Signal-to-noise ratio penalties for continuous-time phase noise channels," *Int. Conf. on Cognitive Radio Oriented Wirel. Networks (CROWNCOM'14),* June 2014, Oulu, Finland.

[36] L. Barletta and G. Kramer, "Lower bound on the capacity of continuous-time Wiener phase noise channels," *Submitted to IEEE Intern. Symposium on Inf. Theory (ISIT),* 2015. Available online at http://arxiv.org.

[37] L. Barletta, M. Magarini, and A. Spalvieri, "Tight upper and lower bounds to the information rate of the phase noise channel," *IEEE Intern. Symposium on Inf. Theory (ISIT),* pp. 2284–2288, Jul. 2013, Istanbul, Turkey.

[38] L. Barletta, M. Magarini, S. Pecorino, and A. Spalvieri, "Upper and lower bounds to the information rate transferred through first-order markov channels with free-running continuous state," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3834–3844, Jul. 2014.

[39] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," *IEEE Intern. Symposium on Inf. Theory (ISIT),* pp. 2426–2429, June 2014, Honolulu, HI, USA.

[40] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," *Information Theory and Applications Workshop (ITA),* pp.1–7, Feb. 2013, San Diego, CA, USA.

[41] M. R. Khanzadi, G. Durisi, T. Eriksson, "Capacity of SIMO and MISO phase-noise channels with common/separate oscillators," to appear on *IEEE Trans. Commun.*

[42] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear block codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[43] M. V. Eyuboglu "Detection of coded modulation signals on linear severely distorted channels using decision-feedback noise prediction and interleaving," *IEEE Trans. Commun.* vol. 26, no. 4, pp. 401–409, Apr. 1988.

[44] U. Wachsmann, R. F. H. Fischer, and J B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[45] M. Ferrari, S. Bellini, and A. Spalvieri, "Coded modulation schemes based on partial Gray mapping and unpunctured high rate turbo codes," *IEE Proceedings on Communications*, vol. 153, no. 2, pp. 288–294, Feb. 2006.

[46] G.D. Forney Jr., M. D. Trott, and S.-Y. Chung, "Sphere-Bound-Achieving Coset Codes and Multilevel Coset Codes," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[47] "GSFC-STD-9100, Low density parity check code for rate 7/8." Available online: https://standards.nasa.gov/documents/gsfc.

[48] M. Beerman, T. Breddermann, and P. Vary, "Rate-compatible LDPC codes using optimized dummy bit insertion," *Intern. Symposium Wirel. Commun. Systems,* pp.447–451, Nov. 2011, Aachen, Germany.