A multi-sample discrete-phase BCJR algorithm for phase noise channels

A Multi-Sample Discrete-Phase BCJR Algorithm for Phase Noise Channels

B. Ripani
Politecnico di Torino
Dep. of Elec. and Telecomm. (DET)
Torino, Italy
barbara.ripani@polito.it

A. Modenini
European Space Agency
RF system division
Noordwijk, The Netherlands
andrea.modenini@esa.int

G. Montorsi
Politecnico di Torino
Dep. of Elec. and Telecomm. (DET)
Torino, Italy
guido.montorsi@polito.it

***Abstract*—In this paper, we consider channels affected by Wiener phase noise and derive a detection algorithm that exploits oversampling at the receiver for tracking better the phase noise variations.**

The resulting algorithm is a multi-sample discrete-phase BCJR that outperforms receivers using a single sample per symbol and maximizes the achievable information rate over phase noise channels. Finally, we assess the suitability of the multi-sample discrete-phase BCJR for performing iterative detection and decoding with standard codes.

I. Introduction

The majority of the communication systems suffer from phase noise, a random fluctuation in the phase of the carrier which impacts the quality of reception. One of the most critical systems is orthogonal frequency division multiplexing (OFDM). In this case, the phase noise is responsible for harming the orthogonality between sub-carriers and introducing inter-carrier interference [1]. Another example are Internet-of-Things applications [2], which are affected by high-frequency instabilities due to the adoption of low-cost local oscillators. On top of this, phase noise poses further challenges for low bit-rate transmissions (e.g., deep space satellite telecommunications [3]). In this scenario, the phase variation of the local oscillator can be several degrees over a symbol period.

In literature, the phase noise process is generally described by the Wiener model due to its simplicity, thus allowing the development of advanced synchronization algorithms while still being representative of actual oscillators [4]. However, the plurality of the studies (such as [5] and references therein) relies on a symbol-level discrete-time channel model, which can be a good approximation of the actual channel as long as the phase variation over symbol time is less than 6 degrees [4]. On the other hand, when variations are faster, the discrete-time model does not fit anymore, and receivers based on this model would experience an intrinsic information loss [6]. To tackle this issue, Ghozlan and Kramer ([7], [8]) proposed the use of oversampling to obtain a sufficient statistic.

Based on this approach, in this paper, we derive a detection technique in the Bahl-Cock-Jelinek-Raviv (BCJR, [9]) form, that is optimal for the Wiener channel (in the limit of its complexity).

The remainder of this paper is organized as follows: Section II gives the mathematical details of the channel model, while Section III provides the rationale for using oversampling as the channel model for the algorithm. Then, Section IV describes the multi-sample discrete-phase BCJR algorithm, main focus of this paper. Finally, Section V provides the numerical results and Section VI draws the conclusions.

II. Channel model

We consider the transmission of a linearly modulated signal $x(t)$ over an additive white Gaussian noise (AWGN) channel, affected by phase noise. The received signal has complex baseband expression

$$y(t) = x(t)e^{j\theta(t)} + w(t)\,, \tag{1}$$

where $w(t)$ is complex-valued white Gaussian noise with spectral density N_0 and $\theta(t)$ is the phase noise. The transmitted signal $x(t)$ reads

$$x(t) = \sum_k c_k p(t - kT)\,,$$

being $p(t)$ the shaping pulse, T the symbol time, and $\{c_k\}$ the sequence of information symbols belonging to an $M-$ary complex-valued constellation. Without loss of generality, we consider that $p(t)$ has unitary energy, satisfies the Nyquist criterion (no inter-symbol interference, ISI), and that the constellation symbols have energy $\mathbb{E}\left\{|c_k|^2\right\} = E_s$.

In this paper, we focus on the Wiener process, a simple mathematical model able to represent the phase noise in actual oscillators [4]. Namely, for a time interval τ it holds that

$$\theta(t+\tau) - \theta(t) \sim \mathcal{N}(0, S_\Delta^2 \tau)\,,$$

where $\mathcal{N}(0, S_\Delta^2\tau)$ refers to the Gaussian distribution having null average and standard deviation $S_\Delta^2\tau$, being S_Δ (measured in $[\mathrm{deg/s}]$) the speed of variation.

III. On the use of an auxiliary channel with oversampling

Several papers concerning phase synchronization and detection of signals affected by phase noise (like the recent ones [5], [10]) approximate the channel model in (1) with the following discrete-time equation

$$\tilde{y}_k = c_k e^{j\theta_k} + w_k\,, \tag{2}$$

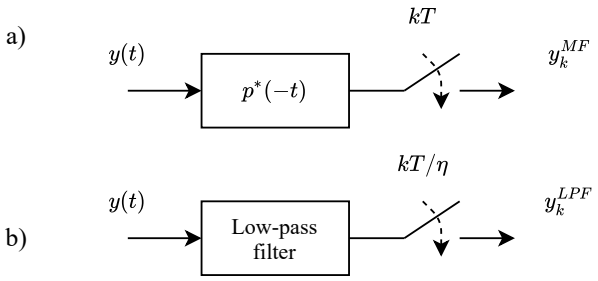

Figure 1. Receiver block diagram when a) using a matched filter and b) a low-pass filter with oversampling.

where w_k are independent identically distributed complex Gaussian random variables with variance N_0, and $\theta_k = \theta(kT)$, so that

$$\theta_{k+1} - \theta_k \sim \mathcal{N}(0, \sigma_\Delta^2)\,,$$

where $\sigma_\Delta^2 = S_\Delta^2 T$ is the phase noise variance over the reference symbol time. However, such discrete-time approximation is obtained by assuming that $\{\tilde{y}_k\}$ are the samples at the output of a matched filter (MF), as shown in Fig. 1a. In fact, if $\theta(t)$ is "slow" w.r.t. the symbol time, it holds that the observable at the MF output is such that $y_k^{MF} \approx \tilde{y}_k$. On the other hand, when the phase noise variance σ_Δ^2 is "large", this assumption holds no longer.

In this paper, we consider instead a channel model derived by means of the oversampling method described in [11]. Namely, for a given oversampling factor η, we consider that $y(t)$ passes through an ideal low-pass filter with a bandwidth equal to η/T, as shown in Fig. 1b. Its output is then sampled with an interval of T/η, resulting in the sequence $\{y_k^{LPF}\}$. Considering that the phase variation between two adjacent samples is σ_Δ^2/η, it decreases as η increases, and if the oversampling factor is chosen large enough, it holds (with a little abuse of notation) that

$$y_k^{LPF} \approx \tilde{y}_k = \sum_i c_i p_{k-i\eta} e^{j\theta_k} + w_k\,, \tag{3}$$

where $p_k = p(kT/\eta)$, $\theta_k = \theta(kT/\eta)$, and w_k has variance $N_0\eta$.

The design of receivers based on a mismatched channel is an instance of *mismatched decoding* [12], [13]. More specifically, being $\tilde{\boldsymbol{y}}, \boldsymbol{c}$ the vectors collecting the observable and transmitted symbols, if the receiver is based on an *auxiliary channel* [14] with probability density function (pdf) $q(\tilde{\boldsymbol{y}}|\boldsymbol{c})$ (different from the actual channel law), it will have an achievable information rate (AIR) equal to

$$I_{\mathrm{AIR}} = \lim_{N\to\infty} \frac{1}{N} \left[\mathbb{E}\left\{-\log_2 q(\tilde{\boldsymbol{y}})\right\} + \mathbb{E}\left\{\log_2 q(\tilde{\boldsymbol{y}}|\boldsymbol{c})\right\}\right]\,,$$

where N is the number of transmitted symbols, $q(\tilde{\boldsymbol{y}}) = \sum_{\boldsymbol{c}} q(\tilde{\boldsymbol{y}}|\boldsymbol{c})P(\boldsymbol{c})$, and the expectations are computed w.r.t. the actual channel statistics. Trivially, the AIR is upper-bounded as

$$I_{\mathrm{AIR}} \le I\left(\boldsymbol{c};\boldsymbol{y}\right)\,,$$

where $I\left(\boldsymbol{c};\boldsymbol{y}\right)$ is the information rate of the actual channel. This holds for both the information rates I_{AIR}^{MF} and I_{AIR}^{LPF}, achievable when adopting the auxiliary channels as in (2) and (3), respectively. However, when using oversampling with $\eta \to \infty$, $\tilde{y}_k^{LPF}$ tends to be a sufficient statistic for the channel output, and thus $I_{\mathrm{AIR}}^{LPF} \to I\left(\boldsymbol{c};\boldsymbol{y}\right)$. A detailed analysis of the AIR when using an oversampled auxiliary channel was done in [7], showing that I_{AIR}^{LPF} outperforms I_{AIR}^{MF}. In light of this, we derive a detection algorithm based on oversampling in the next section.

IV. A multi-sample dp-BCJR algorithm

For the discrete-time auxiliary channel in (2), Colavolpe et al. derived in [15] a discrete-phase BCJR (dp-BCJR) algorithm. Here, we adopt a similar approach by defining a new dp-BCJR algorithm working on the oversampled observable.

Unfortunately, its derivation for the channel model in (3), in its most general form, is not trivial. It would require the definition of a BCJR algorithm that can account for both ISI and the phase noise model. However, when adopting time-limited shaping pulses, the problem becomes tractable. Without loss of generality, if we consider a squared shaping pulse[1], the channel equation can be simplified as

$$\tilde{y}_k = c_{\lfloor k/\eta \rfloor} e^{j\theta_k} + w_k\,, \tag{4}$$

and we can derive a BCJR algorithm by means of factor graphs (FGs) and the sum-product algorithm [16].

Namely, the pdf $q(\boldsymbol{c},\boldsymbol{\theta}|\tilde{\boldsymbol{y}})$ can be factorized as

$$\begin{aligned} q(\boldsymbol{c},\boldsymbol{\theta}|\tilde{\boldsymbol{y}}) &\propto q(\tilde{\boldsymbol{y}}|\boldsymbol{c},\boldsymbol{\theta})P(\boldsymbol{c})p(\boldsymbol{\theta}) \\ &= \left(\prod_{k=0}^{N\eta-1} q(\tilde{y}_k|c_{\lfloor k/\eta\rfloor},\theta_k)\mathcal{P}(c_{\lfloor k/\eta\rfloor})\right) p(\boldsymbol{\theta})\,, \end{aligned} \tag{5}$$

where $q(\tilde{y}_k|c_{\lfloor k/\eta\rfloor},\theta_k)$ is a Gaussian pdf with average $c_{\lfloor k/\eta\rfloor}e^{j\theta_k}$ and variance $N_0\eta$. The term $\mathcal{P}(c_{\lfloor k/\eta\rfloor})$ is instead an indicator function that takes values

$$\mathcal{P}(c_{\lfloor k/\eta\rfloor}) = \begin{cases} P(c_{\lfloor k/\eta\rfloor}) & \text{if } (k+1) \text{ is multiple of } \eta \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

being $P(c_{\lfloor k/\eta\rfloor})$ the probability of the information symbol $c_{\lfloor k/\eta\rfloor}$. In turn, the probability of the phase is factorized as

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\theta_0)\prod_{k=1}^{N\eta-1} p(\theta_k|\theta_{k-1}) \\ &= \frac{1}{2\pi}\prod_{k=1}^{N\eta-1} p_\Delta(\theta_k - \theta_{k-1})\,, \end{aligned} \tag{7}$$

where $p_\Delta(\theta)$ is a wrapped Gaussian pdf in $[0, 2\pi)$ as

$$p_\Delta(\theta) \propto \sum_i e^{\frac{(\theta-2\pi i)^2}{2\sigma_\Delta^2}}\,.$$

By plugging (7) into (5), the factorization can be visualized in the FG of Figure 2.

[1]It is pointed out that the algorithm can easily take into account any shaping pulse with duration $\le T$ by including the coefficients $p_k = p(kT/\eta)$ into (4).

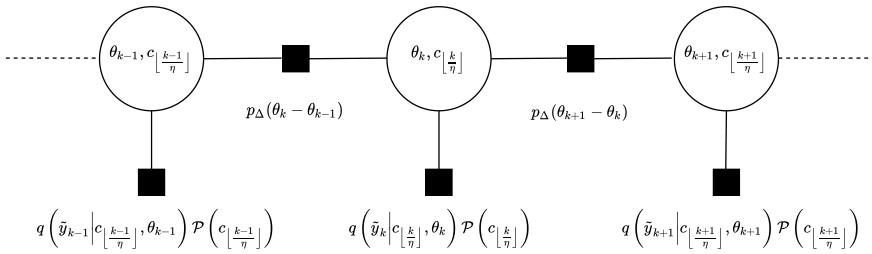

Figure 2. Factor graph representing the factorization in (5).

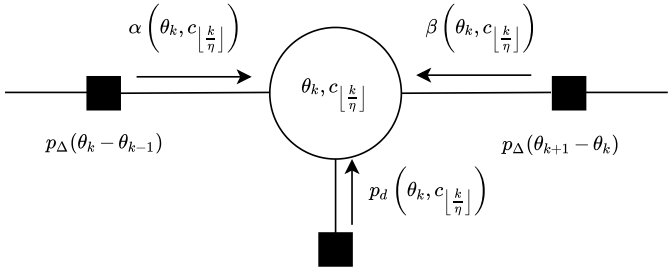

Figure 3. In-depth view of the messages passing through a single variable node.

If we define $\alpha(\cdot)$, $\beta(\cdot)$, and $p_d(\cdot)$ as the messages shown in Figure 3, we can derive the forward recursion as

$$\alpha(\theta_k, c_{\lfloor k/\eta \rfloor}) = \sum_{c_{\lfloor (k-1)/\eta \rfloor}} \int_0^{2\pi} \alpha(\theta_{k-1}, c_{\lfloor (k-1)/\eta \rfloor}) \, p_d(\theta_{k-1}, c_{\lfloor (k-1)/\eta \rfloor}) p_\Delta(\theta_k - \theta_{k-1}) \, \mathrm{d}\theta_{k-1} \quad (8)$$

when k is multiple of η, and,

$$\alpha(\theta_k, c_{\lfloor k/\eta \rfloor}) = \int_0^{2\pi} \alpha(\theta_{k-1}, c_{\lfloor (k-1)/\eta \rfloor}) \, p_d(\theta_{k-1}, c_{\lfloor (k-1)/\eta \rfloor}) p_\Delta(\theta_k - \theta_{k-1}) \, \mathrm{d}\theta_{k-1} \quad (9)$$

otherwise, where

$$p_d(\theta_k, c_{\lfloor k/\eta \rfloor}) = q(\tilde{y}_k | c_{\lfloor k/\eta \rfloor}, \theta_k) \mathcal{P}(c_{\lfloor k/\eta \rfloor}) \, .$$

Similarly, the backward recursion can be derived as

$$\beta(\theta_k, c_{\lfloor k/\eta \rfloor}) = \sum_{c_{\lfloor (k+1)/\eta \rfloor}} \int_0^{2\pi} \beta(\theta_{k+1}, c_{\lfloor (k+1)/\eta \rfloor}) \, p_d(\theta_{k+1}, c_{\lfloor (k+1)/\eta \rfloor}) p_\Delta(\theta_{k+1} - \theta_k) \mathrm{d}\theta_{k+1} \quad (10)$$

when $k+1$ is multiple of η, and

$$\beta(\theta_k, c_{\lfloor k/\eta \rfloor}) = \int_0^{2\pi} \beta(\theta_{k+1}, c_{\lfloor (k+1)/\eta \rfloor}) \, p_d(\theta_{k+1}, c_{\lfloor (k+1)/\eta \rfloor}) p_\Delta(\theta_{k+1} - \theta_k) \mathrm{d}\theta_{k+1} \quad (11)$$

otherwise. Finally, the maximum a posteriori probability on the information symbols can be derived as $q(\tilde{\boldsymbol{y}}|c_k)P(c_k)$, being $q(\tilde{\boldsymbol{y}}|c_k)$ the *extrinsic information* computed as

$$q(\tilde{\boldsymbol{y}}|c_k) = \int_0^{2\pi} \alpha(\theta, c_k) q(\tilde{y}_k | c_k, \theta) \beta(\theta, c_k) \mathrm{d}\theta \, . \quad (12)$$

For the sake of clarity, in (12), we defined $\theta = \theta_{k\eta-1}$, and $k = 0, ..., N-1$.

Clearly, the exact computation of (8)–(12) is impractical, since it involves continuous pdfs. However, as done in [15], the integral can be solved by considering a discrete phase in L values between $[0, 2\pi)$, thus obtaining a multi-sample dp-BCJR algorithm. With this approach, it is easy to see that the complexity scales as $\mathcal{O}\left(NML^2\eta\right)$. Although the complexity is η times larger than the one of the classical dp-BCJR, it can be still tractable with a standard workstation for practical cases (like those shown in the numerical results).

The multi-sample dp-BCJR algorithm we derived can be easily adopted for iterative detection and decoding: it is sufficient to set the probabilities in (6) with the *a priori* information on the symbols, and to use the extrinsic information in (12) as input to the decoder.

V. Numerical results

We present the performance of the multi-sample dp-BCJR detection algorithm introduced in Section IV. We simulated the algorithm over the channel (1), using a squared shaping pulse $p(t)$, for different modulation formats and high phase noise levels ($\sigma_\Delta > 10°$, for which oversampling has advantages). For comparison, we adopted the classical dp-BCJR in [15] that is optimal (to the extent of the maximum number of discrete levels adopted) for the discrete-time auxiliary channel, hence representative of the best performance that can be obtained by a general algorithm based on (2).

By means of the Monte Carlo method shown in [14], we first derived the AIR I_{AIR}^{LPF} as a function of the signal-to-noise ratio E_s/N_0, which is representative of the AIR of the multi-sample dp-BCJR. Figure 4 shows the example of a quaternary phase-shift keying (QPSK) and 16-PSK, over a phase noise channel with $\sigma_\Delta = 28°$ when using $L = 32$ and $\eta = 4$. In all cases, we considered 1 pilot for every 20 symbols. The same figure reports I_{AIR}^{MF} for $L = 32$, representative of the AIR of the classical dp-BCJR. Similar to what has been proven in [7], the multi-sample dp-BCJR has a higher AIR as the E_s/N_0 increases, i.e., when the phase noise becomes the main limiting factor. The gain improves as the cardinality of the constellation M grows, especially at high values of σ_Δ, such as the one previously proposed. This holds particularly for M-PSK modulation. On the other hand, when using constellations more robust to the phase noise, like quadrature amplitude

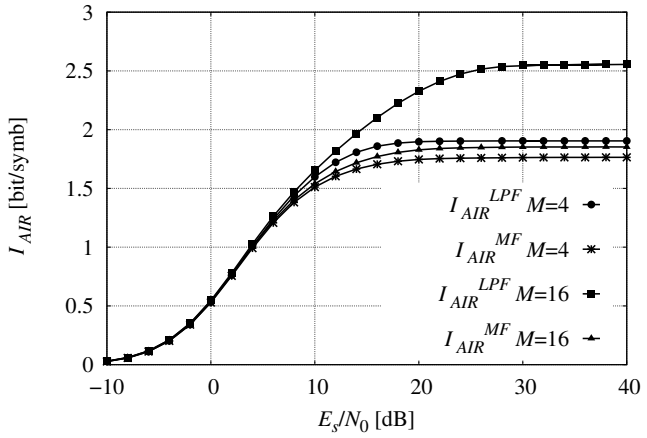

Figure 4. Achievable information rates for QPSK (M=4) and 16-PSK (M=16) constellation, with $\sigma_\Delta = 28°$, and $L = 32$.

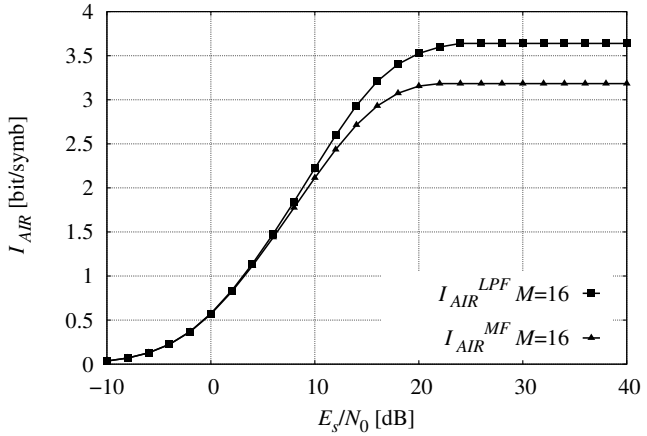

Figure 5. Achievable information rates for 16-QAM (M=16) constellation with $\sigma_\Delta = 28°$, and $L = 32$.

modulation (QAM), the benefit of the multi-sample dp-BCJR tends to diminish. For instance, Figure 5 shows the case of a 16-QAM over a phase noise channel with $\sigma_\Delta = 28°$, where the gain of the multi-sample dp-BCJR (over the classical one) is just $\sim 14\%$ in the saturation region versus the $\sim 38\%$ that was achieved for 16-PSK. Thus, the use of the multi-sample dp-BCJR should always be properly traded off between complexity and target information rate.

The rates promised by the AIR can be usually achieved by means of iterative detection and decoding. For checking its feasibility with standard error correction codes (designed for the AWGN channel) at high values of σ_Δ, we considered the low-density parity-check (LDPC) codes (block length 4096 and 64800 bits) with row-by-column interleaving and used the extrinsic information transfer (EXIT) charts [17]. Focusing on the QPSK scenario previously presented, Figure 6 shows the EXIT chart of the two dp-BCJR algorithms, for $E_s/N_0 = 10$ dB. Based on the AIR, at $E_s/N_0 = 10$ dB it

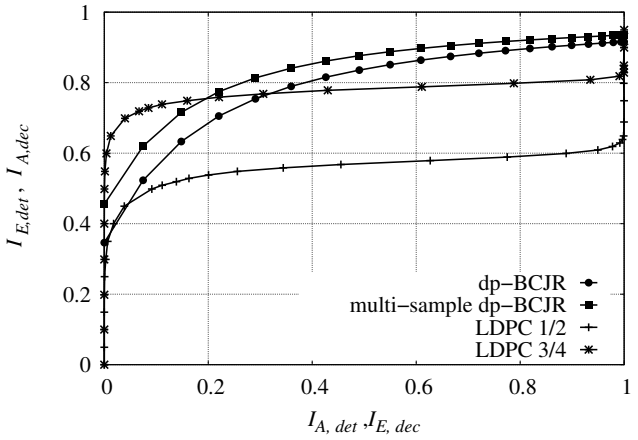

Figure 6. EXIT chart for $E_s/N_0 = 10$ dB, for the multi-sample and classical dp-BCJR, QPSK modulation, $L = 32$, and for LDPC codes with rate 1/2 and 3/4.

should be feasible to adopt a code with a rate $3/4$. However, when checking the EXIT of the LDPC code, we see that the curves intersect each other, for both the case of the multi-sample and the classical dp-BCJR. The highest code rate that allows the iterative process to start is $1/2$. Thus, codes performing well on the AWGN will have a large loss when adopted on channels with strong phase noise, independently of the algorithm adopted.

This behavior is confirmed by the frame error rate (FER): we analyzed the FER for a QPSK modulation over a channel with $\sigma_\Delta = 28°$ when using the LDPC $(1/2,\ 4096)$ for the multi-sample and classical dp-BCJR. Figure 7 illustrates the FER curves, obtained with 5 global iterations, each with 10 inner decoder iterations, and each simulated point counts 50 frame errors. For complexity reasons, we simulated FER only down to 10^{-4}. As predicted by the EXIT, the multi-sample dp-BCJR FER curve starts the waterfall region at $E_s/N_0 \sim$10 dB, about 3 dB lower than the dp-BCJR. On the other hand, both cases show a poor slope in the waterfall region, a sign of the presence of correlations among the samples. A steeper slope of the curve can be achieved by adopting "longer" codes, as the one in Figure 8, obtained with the same number of iterations of Figure 7. In fact, the EXIT chart predictions become more reliable as the block length tends to infinity. However, a re-design of the error correction code is mandatory.

As additional proof, we examined the decoder and detector (dp-BCJR and multi-sample dp-BCJR, respectively) trajectories, obtained by simulating fifteen random codewords. The result, depicted in Figure 9 and 10 is in agreement with the EXIT charts (depicted on top) and in line with the FER curves.

VI. Conclusions and future work

In this paper, we presented a multi-sample dp-BCJR, that uses oversampling as an observable for better tracking the phase noise variations. An instance of the algorithm was derived for the case of shaping pulses that have a duration

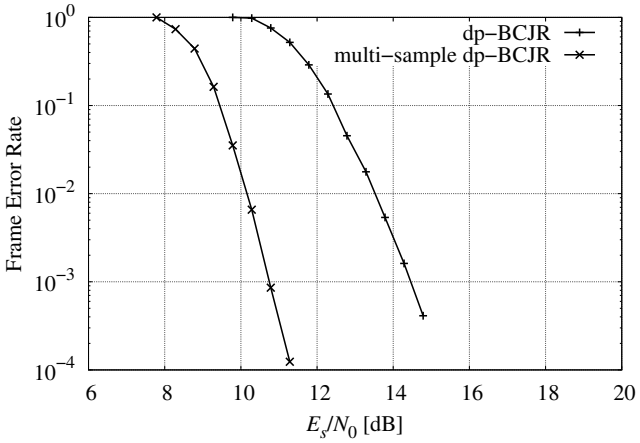

Figure 7. FER as function of E_s/N_0, for the multi-sample BCJR and the classical dp-BCJR with QPSK, when $\sigma_\Delta = 28°$, $L = 32$, with a LDPC code having rate 1/2, block length 4096.

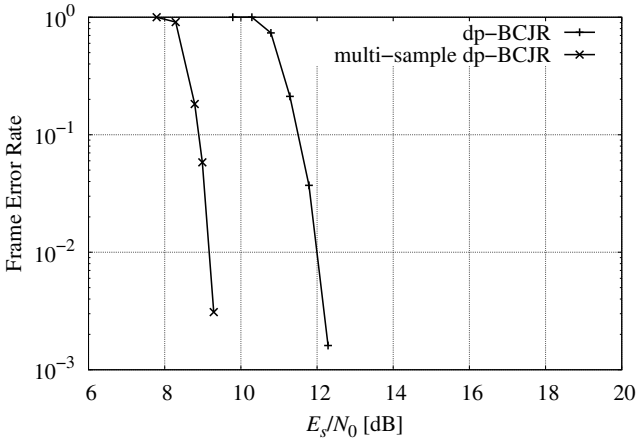

Figure 8. FER as function of E_s/N_0, for the multi-sample dp-BCJR and the classical dp-BCJR with QPSK, when $\sigma_\Delta = 28°$, $L = 32$, with a LDPC code having rate 1/2, block length 64800.

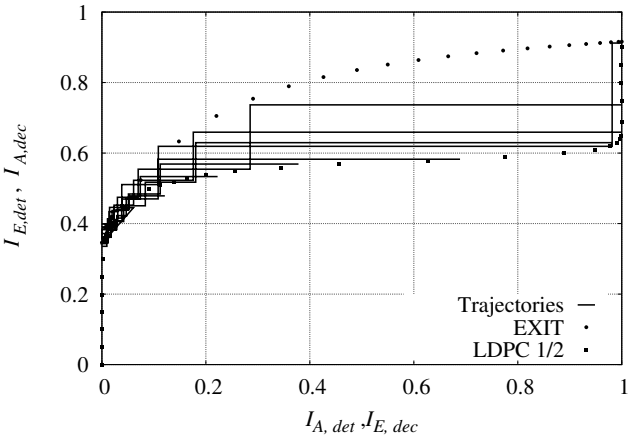

Figure 9. dp-BCJR detector and LDPC (1/2, 4096) decoder trajectories compared to the respective EXIT charts with QPSK, when $\sigma_\Delta = 28°$, $L = 32$, and $E_s/N_0 = 10$ dB.

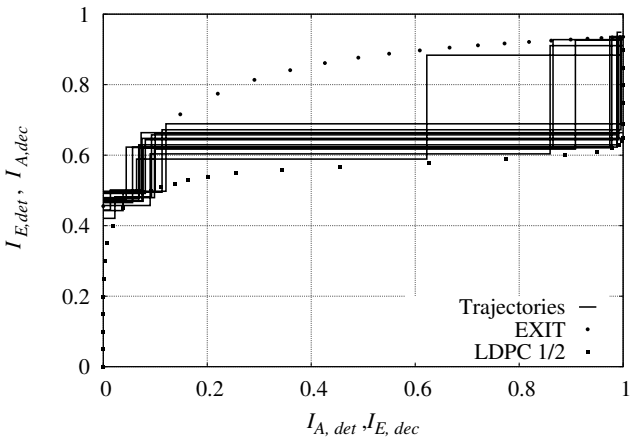

Figure 10. Multi-sample dp-BCJR detector and LDPC (1/2, 4096) decoder trajectories compared to the respective EXIT charts with QPSK, when $\sigma_\Delta = 28°$, $L = 32$, and $E_s/N_0 = 10$ dB.

equal to the symbol time. We showed that the proposed multi-sample dp-BCJR outperforms the classical dp-BCJR and, more in general, algorithms using a single sample per symbol.

However, although the here-presented algorithm shows promising results, we identified three major aspects that require additional research: complexity reduction, the generalization to other shaping pulses, and the design of forward error correction codes.

For the first, we have shown that the multi-sample dp-BCJR has a complexity that scales as $\mathcal{O}\left(NML^2\eta\right)$. That is acceptable for the here-analyzed scenarios but can become impractical for large constellations (e.g., $M > 64$). For making the problem tractable again, a possibility is the use of canonical distributions as the Tikhonov pdf [15], that can be propagated along the FG by means of a Kullback-Leibler minimization [18], expectation propagation [19], or other message-passing techniques. Another possibility could be the optimization (from an AIR perspective) of $p_\Delta(\theta)$ for a fixed number of phase level L, similarly to what is done for ISI channels when using the channel shortening technique [20].

Concerning instead the generalization to other shaping pulses, as previously mentioned, the main challenge appears to derive an FG that can take into account the ISI in the observable in (3) and the phase noise model. At the time of writing, a formulation of this problem (to the best of the authors' knowledge) still does not exist.

Finally, with reference to error correction codes, we have shown in the previous section that when the phase noise is too strong, standards codes (for the AWGN channel) do not perform well with both the multi-sample and classical dp-BCJR. It means that one shall design a code (specific for the scenario) to achieve the rates predicted by the AIR. This can be carried out, for instance, by means of the techniques presented in [21] and [22].

DISCLAIMER

The view expressed herein can in no way be taken to reflect the official opinion of the European Space Agency.

REFERENCES

[1] M. R. Khanzadi, D. Kuylenstierna, A. Panahi, T. Eriksson, and H. Zirath, "Calculation of the performance of communication systems from measured oscillator phase noise," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 5, pp. 1553–1565, 2014.

[2] C. A. Hofmann, K.-U. Storek, and A. Knopp, "Impact of phase noise and oscillator stability on ultra-narrow-band-IoT waveforms for satellite," in *Proc. IEEE Intern. Conf. Commun.*, 2021, pp. 1–6.

[3] M. Baldi, M. Bertinelli, F. Chiaraluce, P. Closas, R. Garello, N. Maturo, M. Navarro, J. M. Palomo, E. Paolini, S. Pfletschinger *et al.*, "NEXCODE: Next generation uplink coding techniques," in *International Workshop on Tracking, Telemetry and Command Systems for Space Applications (TTC)*, Noordwijk, The Netherlands, 2016.

[4] A. Piemontese, G. Colavolpe, and T. Eriksson, "Phase noise in communication systems: from measures to models," *Available on Arxiv*, Apr. 2021.

[5] A. Kreimer and D. Raphaeli, "Efficient low-complexity phase noise resistant iterative joint phase estimation and decoding algorithm," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 4199–4210, 2018.

[6] M. Martalò, C. Tripodi, and R. Raheli, "On the information rate of phase noise-limited communications," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, USA, 2013, pp. 1–7.

[7] H. Ghozlan and G. Kramer, "Multi-sample receivers increase information rates for Wiener phase noise channels," in *Proc. IEEE Global Telecommun. Conf.*, Atlanta, GA, U.S.A., Dec. 2013, pp. 1897–1902.

[8] ——, "Phase modulation for discrete-time Wiener phase noise channels with oversampling at high SNR," in *Proc. IEEE International Symposium on Information Theory*, Honolulu, HI, USA, 2014, pp. 1554–1557.

[9] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[10] L. Gaudio, B. Matuz, T. Ninacs, G. Colavolpe, and A. Vannucci, "Approximate ML decoding of short convolutional codes over phase noise channels," *IEEE Commun. Letters*, vol. 24, no. 2, pp. 325–329, 2020.

[11] H. Meyr, M. Oerder, and A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers," *IEEE Trans. Commun.*, vol. 42, pp. 3208–3214, Dec. 1994.

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

[13] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[14] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[15] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Select. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.

[16] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[17] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[18] S. Shayovitz and D. Raphaeli, "Message passing algorithms for phase noise tracking using tikhonov mixtures," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 387–401, 2016.

[19] G. Colavolpe and A. Modenini, "Iterative carrier synchronization in the absence of distributed pilots for low SNR applications," in *International Workshop on Tracking, Telemetry and Command Systems for Space Applications (TTC)*, Darmstadt, Germany, 2012.

[20] A. Modenini, F. Rusek, and G. Colavolpe, "Optimal transmit filters for constrained complexity channel shortening detectors," in *Proc. IEEE Intern. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 1688–1693.

[21] G. Liva, S. Song, L. Lan, Y. Zhang, S. Lin, and W. E. Ryan, "Design of LDPC codes: A survey and new results," *Journal of Communications Software and Systems*, vol. 2, no. 3, pp. 191–211, Sep. 2006.

[22] G. Montorsi, "Design of LDPC codes with tunable slope of their EXIT charts," in *Proc. Intern. Symp. on Turbo Codes & Relat. Topics*. IEEE, 2016, pp. 126–130.