POLITECNICO DI MILANO
DEPARTMENT OF ELECTRONICS, INFORMATION, AND BIOENGINEERING
DOCTORAL PROGRAMME IN INFORMATION ENGINEERING

ANALYSIS OF WIENER PHASE NOISE ISSUES IN OPTICAL TRANSMISSION SYSTEMS

Doctoral Dissertation of:
Silvio Mandelli

Supervisor:
Prof. Maurizio Magarini

Tutor:
Prof. Andrea Virgilio Monti Guarnieri

The Chair of the Doctoral Program:
Prof. Carlo Fiorini

2012/2015 – 28

Abstract

In optical communications, local oscillators and propagation introduce multiplicative phase noise that must be taken in consideration, estimated and compensated at the receiver. In the literature such channels are dealt with by considering a symbol-spaced discrete-time model where the transmitted symbol is impaired by both AWGN and a multiplicative phase noise given by a first order Wiener process. The issues given by such channels are objects of several works in the literature. The aim of this thesis is to discuss some of them and try to extend those dissertations.

First of all, in the literature this model is assumed by considering "small" phase noise but nobody has ever discussed how much "small" it must be. A statistical and mathematical analysis is derived by the author, and a threshold of validity of the so called Discrete Model is worked out, proving that the assumption is correct in almost all practical scenario and it is conservative in term of performance simulation. The analysis of phase noise channels is then deepened by studying Bayesian tracking techniques to extract all the information about the transmitted symbols. An iterative demodulation and decoding scheme is proposed and compared to others in the literature. The major gain is given by the greater spectral efficiency obtained by not transmitting Pilot Symbols and still working better than other considered similar schemes. Bayesian tracking allows also to derive the information rate of the considered Discrete Model channel and to verify that the proposed algorithm can achieve it.

The focus is then moved to analyze short reach access optical scenarios where, for spectral efficiency and receiver sensitivity, OFDM has been con-

sidered instead of single carrier systems. For cost, footprint and power consumptions requirements, Direct Detection has several advantages compared to Coherent schemes since the multiplicative phase noise introduced by the transmitting laser can be neglected. However, if dispersive compensating fibers are not used, Chromatic Dispersion impairs the signal and the phase noise cannot be canceled. The author has proposed the literature analysis of this phenomenon, enlightening its weaknesses and comparing the new results of the thesis with experimental measurements given by other authors.

Contents

Contents

List of Figures

List of Figures

List of Figures

List of Tables

List of Acronyms

AM Amplitude Modulation

ARMA AutoRegressive Moving Average

AWGN Additive White Gaussian Noise

BCJR Bahl-Cocke-Jelinek-Raviv

BER Bit Error Rate

CM Complete Model

DFB Distributed Feedback Lasers

DM Discrete-time Model

DSB Dual Side-Band

DSP Digital Signal Processing

DTFN Discrete-Time Frequency Noise

GVD Group-Velocity Dispersion

FFT Fast Fourier Transform

FSM Finite State Machine

ICI Inter-Carrier Interference

i.i.d. independent and identically distributed

ISI Inter Symbol Interference

KLD Kullback-Leibler Divergence

LB Lower Bound

MIMO Multiple-In Multiple-Out

MMSE Minimum Mean Square Error

MC Monte Carlo

MSE Mean Square Error

p.d.f. probability density function

PN Phase Noise

PCC Pearson's Correlation Coefficient

PD Power Degradation

PN Phase Noise

PRT Phase Rotation Term

PSD Power Spectral Density

QAM Quadrature Amplitude Modulation

SA Semi-Analytic

SER Symbol Error Rate

SNR Signal-to-Noise Ratio

SPN Small Phase Noise

SRRC Square-Root Raised Cosine

SSB Single Side-Band

UB Upper Bound

VCSEL Vertical Cavity Surface Emitting Lasers

ZP Zero-Padding

CHAPTER 1

Introduction

1.1 Background

Multiplicative phase noise is one of the major impairments affecting the performance of coherent optical transmission systems [Leoni et al., 2012, Magarini et al., 2012a, Goebel et al., 2011]. Phase noise is due to both laser oscillators used for up- and down-conversion [Foschini and Vannucci, 1988], and to cross-phase modulation that arises in wavelength-division-multiplexing systems [Essiambre et al., 2010]. In particular, in [Magarini et al., 2011], it is concluded that the discrete symbol-spaced Optical Channels transmission can be modeled, if the phase noise line-width is "small", as follows

$$y_i = a_i e^{j\varphi_i} + n_i \ , \tag{1.1}$$

where a_i is the transmitted complex symbol, φ_i a discretized first-order Wiener process and n_i AWGN. Several schemes have been proposed to estimate the received carrier phase for arbitrary PSK and QAM constellations in presence of phase noise. In this thesis work Bayesian Inference will be considered to analyze the features and deal with Equation (1.1) channel. In particular, Bayesian tracking is exploited to follow the hidden phase φ_i, given the measurements y_i.

Tracking the state of a dynamic system from noisy measurements is a classical problem in several fields of science. In the state-space approach to time-series modeling, the state process describes all relevant information about the system under investigation. For example, this information could be related to the kinematic characteristics of a generic target [Li and Jilkov, 2003, Blackman, 2004]. In an econometrics problem it could be related to the interest rates or the monetary flow [Gray, 1996, Duffie, 2010]. Alternatively, it could be related to the motion characteristics in video analytics applications for visual tracking, where the aim is to automatically understand the actions occurring in a monitored scene [Isard and Blake, 1998, Kwon and Lee, 2010, Zhou et al., 2004, Ross et al., 2008].

In order to make inference about the state of a dynamic system that changes over time, a model of the state evolution with time (the system model) and a model related to the noisy measurements to the state (the measurement model) are required. For dynamic state estimation, the discrete-time approach is widespread and convenient. In the Bayesian approach, system and measurement models are available in a probabilistic form. Accordingly, probabilities are used to model the state evolution and the measurement given the state, and, from the model and the measurements, inference is made on the hidden evolving state. By making inference one builds the probability of the state given all the available measurements, thus embodying all the available statistical information in the inferred distribution. Therefore it can be said that, in some sense, *Bayesian tracking* extracts the information about the state that is brought by the measurements. This provides a rigorous general framework for dynamic state estimation problems [Simon, 2006].

The most popular tool for Bayesian tracking of a system with discrete-time continuous state is the Kalman filter proposed in [Kalman, 1960] (see, e.g., [Simon, 2006, Haykin, 2004], two comprehensive books on the Kalman filter). The Kalman filter performs optimal tracking, thus leading to exact inference, when the equations that describe the system model and the measurement model are linear and the noisy processes that affect the state evolution and the measurements are additive and independent Gaussian processes. When the state transition and/or the measurement equations are non-linear and/or the noise processes are non-Gaussian, the Kalman filter is no more optimal. To face the non-optimality of the Kalman filter in case of non-linear state model and/or measurement model, the extended Kalman filter has been proposed in [Bellantoni and Dodge, 1967] and adopted in more applications: real-time traffic estimation [Wang and Papageorgiou, 2005], data-assimilation in oceanography [Pham et al., 1998], estimation

speed in induction motor [Kim et al., 1994], real-time estimation of rigid body [Marins et al., 2001], and data-fusion of Global Positioning System signals [Sasiadek et al., 2000] are some examples. This technique requires differentiable functions in the state and measurement models and it can diverge, owing to its linearization. Other techniques try to match the distribution of the state with a parametric distribution with limited number of parameters: some examples could be the parametric distributions for residential air exchange rates [Murray and Burmaster, 1995], the parametric models of geometry and illumination for the visual tracking [Hager and Belhumeur, 1998], the Tikhonov and Fourier parametrizations proposed in [Colavolpe et al., 2005] for the phase tracking problem, parametric distributions of storage time and temperature of ready-to-eat foods [Pouillot et al., 2010]. When the state-space has reduced dimensionality, an other approach is the quantization: this trivial non-parametric technique can provide satisfactory performance as in [Barletta et al., 2012a, Barletta et al., 2011] for the phase tracking problem. An other example of application of the quantization technique is the word recognition from acoustic signals [Rabiner et al., 1983]. Among the inferential techniques proposed to apply the Bayesian approach in a more generic framework, particle filter has received in the past two decades widespread interest. This technique does not need to the Gaussianity of the noise processes. The basic feature of the particle filter is to provide a non-parametric approximation to the exact distribution, thus making possible to accurately infer multi-modal distributions. Particle filtering techniques have found application in several research areas, including, to cite just a few, communication systems [Amblard et al., 2003, Punskaya et al., 2001], data fusion [Perez et al., 2004, Caron et al., 2007], non-linear control [Rigatos, 2009], target-tracking [Särkkä et al., 2007, Okuma et al., 2004] analysis of financial time series [Lopes and Tsay, 2011, Fearnhead, 2005]. The papers [Arulampalam et al., 2002, Djurić et al., 2003, Candy, 2007, Creal, 2012, Hlinka et al., 2013] take a look at the world of particle filters.

Coming back to the problem of tracking φ_i in the discrete-time Wiener phase noise channel of Equation (1.1), among the proposed methods, the blind feed-forward scheme of [Pfau et al., 2009] addresses the constraints imposed by high speed parallel processing. Pilot-aided carrier phase recovery schemes have recently gained attention as candidate phase recovery schemes for systems affected by strong phase noise. Papers [Morsy-Osman et al., 2011, Zhang et al., 2012] are based on the insertion of a pilot tone in a notch of the transmitted signal spectrum, while in papers [Magarini et al., 2012b, Spalvieri and Barletta, 2011, Barletta et al., 2013] pilot sym-

bols are inserted in time domain to demodulate the phase noise signal. Also, schemes based on time domain interleaving of robust modulation formats and less robust but more spectrally efficient modulation formats are proposed in [Barletta et al., 2012c, Le et al., 2014]. An iterative demodulation and decoding algorithm published by the author can demodulate at the information rate a phase noisy channel with low computational complexity and without pilot symbols [Pecorino et al., 2015]. The capacity of channels given by Equation (1.1) are studied in [Dauwels and Loeliger, 2008, Barletta et al., 2011, Barletta et al., 2012a, Barletta et al., 2012b].

Wiener phase noise and its issues are studied in the literature of optical transmission, where among the different strategies, OFDM has been considered in [Ma et al., 2009, Armstrong and Lowery, 2006, Shieh et al., 2008] as a good alternative to the coherent single carrier system [Beppu et al., 2015, Koizumi et al., 2012] for its good performance in terms of spectral efficiency, receiver sensitivity, and polarization dispersion resilience. The internet traffic needs strong and distributed networks, which can carry ever-growing data demand not only in long-haul and medium range network, but also in short range scenarios. For this reason, Coherent Optical OFDM (CO-OFDM) has taken an important role in long-haul transmission [Ma et al., 2009, Shieh et al., 2008], while Direct Detection Optical OFDM [Zan et al., 2008, Schmidt et al., 2009, Peng et al., 2009b, Schuster et al., 2008] (DDO-OFDM) could be very interesting in the short range scenario. However, compared with optical transport networks, the latter are more sensitive to cost, footprint and power consumptions. For this purpose, the exploitation of cost-effective and energy efficient laser sources could become mandatory in the next future. In particular, Vertical Cavity Surface Emitting Lasers (VCSEL) dominate intra-datacenter communications for low-data rate applications due to their intrinsic low cost, energy efficiency and footprint [Amann et al., 2012, Hofmann and Bimberg, 2012, Hofmann et al., 2012]. Nevertheless, Distributed Feedback Lasers (DFB) are mandatory in metro networks due to their superior performance in terms of emitted power frequency chirps and linewidth. To overcome the intrinsic bandwidth limitations, the spectral efficiency of the transmitted signal has to be increased.

Compared to CO-OFDM, in DDO-OFDM coarser lasers can be used, e.g. the already cited VCSEL and DFB. Though those lasers are costly efficient, they introduce a big phase noise in the optical channel that must be dealt with. Therefore, it is important to find in the literature some works that analyze and model phase noise, like [Wu and Bar-Ness, 2004, Liu and Bar-Ness, 2006, Mandelli et al., 2014, Mandelli et al., 2015, Magarini et al., 2011]. This thesis inserts itself in this scenario with a contribution reported

in the following Section.

1.2 Contribution and Outline

This works deepens the analysis on optical Wiener phase noise channels, aiming to bring three major contributes to the literature.

- Several works are present in the literature that exploits the "small" phase noise assumption to assume the Discrete Model of Equation (1.1). However, nobody has never considered how the phase noise should be in order to have such approximation. Obviously, if the phase noise is too "strong", some issues begin to impair the transmission, like cycle slips [Ascheid and Meyr, 1982]. If one considers infinite phase noise, i.e. the continuous-time phase noise $\varphi(t)$ being a uniform white process, it has been demonstrated that the capacity of such channel is null [Barletta and Kramer, 2014a]. The aim of Chapter 2 is to analyze the mismatch between the continuous-time phase noise introduced by oscillators [Foschini and Vannucci, 1988] and the discrete-time model assumed in the literature. A threshold of validity of the discrete model will be found out and validated through several tests.

- In Chapter 3, Bayesian tracking applied to state-based approach is presented with some examples and a scenario where this theory shows its limits. Then this framework is imported into the discrete-time Wiener phase noise channel, where phase noise is the hidden state that must be tracked. Accordingly, in Chapter 4, the Information rate of such channel is computed, finding in this way the limits of transmission over that model [MacKay, 2003]. However, Bayesian tracking allows not only to compute the theoretical optimal transmission rates, but also to design an algorithm achieving such bounds. Therefore, it is proposed in the Chapter a complete Demodulation and Decoding algorithm that can achieve the information rate and, in contrast to the other published [Barletta et al., 2013, Spalvieri and Barletta, 2011, Kamiya and Sasaki, 2013], can work without either using Pilot-Symbols or losing performance. Furthermore, the heavy computational requirements of the algorithm are reduced by smart techniques that allows practical implementation.

- As in single carrier systems, the same phase noise channel model of Equation (2) [Ma et al., 2009] can be used to evaluate the perfor-

mance of optical transmission based on OFDM. While Coherent Optical (CO) OFDM are affected by the same issues of single carrier phase noise, for Direct-Detection Optical (DDO) OFDM the carrier phase noise is canceled out by the detector [Peng et al., 2009b]. However, nobody except [Peng, 2010] has considered Single Side-Band (SSB) DDO-OFDM scenarios without CD compensation fibers. In this case CD plays an important role since, after propagation along the channel, the transmitter laser phase noise cannot be canceled by a Direct Detection receiver. The author wants to deepen the focus about this issue since the analysis developed by [Peng, 2010] is only mathematically derived and it is presented without any experimental validation. Particularly, unlike previous works, the analysis is focused onto strong laser line-widths, which are strong candidates to be used in short reach Access Networks [Alves et al., 2014]. However this is not true anymore with lasers with line-width comparable to the subcarrier spacing Δf of an OFDM modulation with a large number N_d of subcarriers. Consequently, the author analyzes the effects of Phase Noise in DDO-OFDM Transmission due to Chromatic Dispersion. Since DFB and VCSEL are typically characterized by linewidths of few MHz, the impact of phase noise after fiber propagation has to be considered. In Chapter 5 the author begins with the literature analysis of such phenomenon, with cleaner mathematical derivation. SSB transmission is considered in order to cancel the fading of Dual Side-Band (DSB) transmission. Then, the limits of [Peng, 2010, Peng et al., 2009a] are investigated with coarse lasers, like DFB and VCSEL. With this last Chapter the Monte Carlo performance of [Peng, 2010] are reproduced by a Semi-Analytical Model and compared with a proposed simulator by the author. Since the results are not the same, the limits of the Semi-Analytical derivation are investigated and the results compared with the measurements of [Schmidt et al., 2008]. Moreover, thanks to result from the simulation presented in this paper one can point out that Chromatic Dispersion limits in DDO-OFDM the transmission over medium to long tracks of fiber when laser line-widths become closer to the subcarrier spacing.

1.3 Notation

The notations written in this section are valid for the entire Thesis, unless specified differently. The uppercase bold character $\mathbf{U}$ denotes a matrix. The lowercase simple chapter u denotes either a scalar or a column vector

and the uppercase calligraphic character $\mathcal{U}$ denotes the space spanned by u. The lowercase character between brackets $\{u\}$ indicates a possibly non-stationary process, $\{u\} = U_0, U_1, \cdots$, where the uppercase indexed letter U_k denotes a random vector or a random scalar variable, whose generic realization u_k takes its values in $\mathcal{U}_k$. Also, u_i^k denotes a windowed sequence of vectors or scalars between the discrete time instant i and the discrete time instant k, that is

$$u_i^k = \begin{cases} (u_i, u_{i+1}, \cdots, u_k) & \text{if } 0 \leq i \leq k \\ \text{empty} & \text{elsewhere} \end{cases}$$

It is the same for U_i^k, a sequence of random vectors or scalar random variables. If U_i^k and u_i^k are used to indicate a sequence of scalars, random or not, respectively, they can be also interpreted as column vectors. In the following the generic vector is noted as $\mathbf{v}$, whereas a scalar is called s, when one do not want to specify its dimensionality.

Let c_k be a deterministic or random sequence. The polynomial of the complex variable z denotes as $c(z)$ identifies the z-transform of the sequence c_k:

$$c(z) = \sum_{k=-\infty}^{\infty} c_k z^{-k}$$

. For continuous random variables, $p(u_k)$ is a shorthand used to indicate the univariate probability density function $p(U_k = u_k)$, while, when using discrete random variables, the shorthand $p(u_k)$ indicates the univariate mass probability of U_k evaluated in u_k. It is the same for $p(u_i^k)$ that refers to the multivariate case or to the joint probability. In case of conditional probability as $p(u_k|q_k)$, if q_k does not exist then $p(u_k|q_k) = p(u_k)$. In the Thesis, replacing the probability $p(\cdot)$ with $q(\cdot)$ one wants to point out an approximating probability af the actual probability. For example, $q(u_k|v_k)$ is an approximation of the conditional probability density function or of the conditional probability mass function $p(u_k|v_k)$. The Table 1.1 collects all the other important notations adopted in the Thesis. The notation reported in this Section are the most used in the thesis. Other punctual definition will be reported during the dissertation.

Table 1.1: *Iterations performed by the proposed hybrid iterative demodulation and decoding algorithm.*

Notation	Description
$\|\mathcal{U}\|$	number of elements in the discrete set $\mathcal{U}$
j	imaginary unit ($j = \sqrt{-1}$)
u^T, $\mathbf{U}^T$	vector and matrix transpose, respectively
u^*	complex conjugate of the scalar u or complex conjugate of each entries of the vector u
$\mathbf{I}$	identity matrix
$\mathbf{I}_n$	$n \times n$ identity matrix
0_n	$n \times 1$ zero vector
$\mathbf{0}_n$	$n \times n$ zero matrix
$\det(\mathbf{U})$	determinant of the matrix $\mathbf{U}$
$\mathsf{E}\{U_k\}$	mean of the random variable U_k
$cov\{U_k\}$	covariance matrix of the random vector U_k or variance of the scalar random variable U_k
$\log(u)$	natural logarithm of u
$\log_2(u)$	base-2 logarithm of u
$\exp(u)$, e^u	natural exponential of u
$\mathrm{mod}(u, a)$	remainder after the division of u to a
$\mathrm{Re}(c)$	real part of the complex number c
$\mathrm{Im}(c)$	imaginary part of the complex number c
$\|c\|$	absolute value of the complex number c
$\angle c$	phase of the complex number c
$g(\mu, \sigma^2; x)$	univariate Gaussian probability density function over the real axis spanned by x with mean μ and variance σ^2
	density function over the complex plane spanned by x with mean μ and two-dimensional variance σ^2
$g(\mu, \mathbf{R}; x)$	multivariate Gaussian probability density function over the real space spanned by x with mean vector μ and covariance matrix $\mathbf{R}$
$g_c(\mu, \mathbf{R}; x)$	multivariate circular symmetric Gaussian probability density function over the complex space spanned by x with mean vector μ and covariance matrix $\mathbf{R}$
$\delta(x)$	Dirac function over the space spanned by x
$\mathcal{Z}$	set of integer numbers

CHAPTER 2

Modeling the phase Noise

This Chapter investigates the differences between the symbol-spaced discrete-time Wiener phase noise channel model, commonly assumed to represent the effect of phase noise [Mengali, 1997], and that obtained by symbol-rate sampling the filtered continuous-time received signal affected by Wiener phase-noise. All fields of interest where the phase noise is an issue that must be considered are well dealt with in the thesis Introduction and are not considered here if not necessary or examples. In the literature regarding phase noise one can find the continuous-time approach, like in [Ghozlan and Kramer, 2013a, Ghozlan and Kramer, 2013b]. Nevertheless, in most works it is usually considered a symbol-time model for the sampled signal, e.g. in [Spalvieri and Barletta, 2011, Demir et al., 2000, Magarini et al., 2011], where discrete-time phase noise after the receive filtering is considered to be a Wiener process; this is done by assuming that one has slow phase variation in one time symbol. However, nobody has studied yet how much must be slow the phase process to fit the so called Discrete Model (DM). This study is the aim of the Chapter.

The Chapter is organized as follows: in the first Section the commonly assumed Discrete-time Model (DM) is presented together with the derivation of the Complete Model (CM). Then the mismatch between the CM

obtained by sampling the continuous-time signal and the DM is found. Another Section reports the simulations that investigates the differences and the similarities of the two models. The power of this mismatch is then studied by simulations, with a particular emphasis on the phase noise of the Completed Model. In particular, for comparison, some statistical tests to check temporal and distributional properties of the two models are considered. Moreover it is observed that the receive filtering introduces memory in the phase process. The author concludes by pointing out the limits of validity of the DM, in order to validate the previous works present in the literature and to have a precise threshold to work with when assuming discrete-time Models with such scenarios.

The aim of this Chapter is simple, and can be synthesized in one question. How much "small" the phase noise should be to have the continuous-time Wiener phase noise that affects optical signal transmission should be to validate the discrete-model of the Equation (1.1)?

2.1 Two Models

2.1.1 Complete Model Derivation

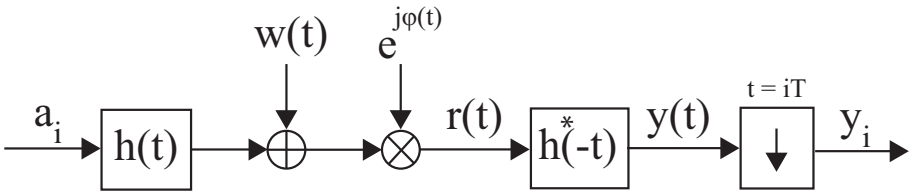


Figure 2.1: *Complex baseband representation of the transmission system with multiplicative phase noise, matched filtering, and symbol-rate sampling.*

In [Foschini and Vannucci, 1988] it is shown that phase noise introduced by laser oscillators can be modeled as a continuous-time Wiener process. Starting from the assumption above and with reference to Figure 2.1 the complex baseband model of the continuous-time signal $r(t)$ at the input of the receiver is

$$r(t) = \sum_m a_m h(t - mT) e^{j\varphi(t)} + w(t) e^{j\varphi(t)} \tag{2.1}$$

where a_i is the sequence of zero-mean complex symbols with unit variance $\sigma_a^2 = 1$ transmitted at rate $1/T$, $j = \sqrt{-1}$ is the imaginary unit, $h(t)$ is the

square-root Nyquist impulse response of the transmit shaping filter with energy E_h and $w(t)$ is the complex Additive White Gaussian Noise (AWGN) with power spectral density N_0. The signal-to-noise ratio is SNR $= E_s/N_0$, where $E_S = \sigma_a^2 E_h$ is the average energy per symbol. Just for reference the information rate between the input modulation and the continuous-time signal of Equation 2.1 is well studied in [Ghozlan and Kramer, 2013a, Ghozlan and Kramer, 2013b] while a lower bound on the capacity has been recently derived in [Barletta and Kramer, 2015]. Upper bounds on the SNR penalty due to phase noise with arbitrary discretization in time domain are given in [Barletta and Kramer, 2014b]. The random phase oscillation of a continuous-time Wiener process evolves as

$$\varphi(t) = \varphi(0) + \sigma \int_0^t \nu(\tau) d\tau \ , \tag{2.2}$$

where $\varphi(0)$ is a Uniform distributed random variable in the interval $[0, 2\pi]$, γ is a real constant, and $\nu(t)$ is a real zero-mean white Gaussian process with autocorrelation

$$E\left[\nu(\tau)\nu(\tau+t)\right] = \int_{-\infty}^{+\infty} \nu(\tau)\nu(\tau+t) d\tau = \delta(t) \ , \tag{2.3}$$

being $\delta(\cdot)$ is the Dirac delta function and $E[\cdot]$ is the expectation operator. Without loss of generality $\varphi(0)$ is set to 0. Note that, while $E[\varphi(t)] = 0$, the variance of $\varphi(t)$ is not a constant, but it linearly increases with respect to the time t

$$\text{Var}[\varphi(t)] = E[\varphi^2(t)] = \sigma^2 E\left[\left(\int_0^t \nu(\tau) d\tau\right)^2\right] dt = \sigma^2 t \tag{2.4}$$

If one recall [Foschini and Vannucci, 1988], the Power Spectral Density (PSD) of the complex exponential $e^{j\varphi(t)}$ given by a Wiener phase noise $\varphi(t)$ is known to be the Lorentzian function given by

$$\mathcal{L}(f) = \frac{4\sigma^2}{\sigma^4 + 16\pi^2 f^2} \tag{2.5}$$

with 3 dB line-width $\sigma^2/(4\pi)$ [Magarini et al., 2011]. After the description of the properties of the received signal, the author proceeds in deriving the signals processed at the receiver before any Digital Signal Processing unit, i.e. prior to the sampler.
The received signal of Equation (2.1) is filtered through the square root

Nyquist matched filter $h^*(-t)$ and sampled at the time instants $t = iT$, leading to

$$y_i = \sum_l a_{i-l} c_l^{(i)} + n_i' \ , \tag{2.6}$$

where

$$c_l^{(i)} = \int_{-\infty}^{+\infty} h(\tau - lT) h^*(\tau - iT) e^{j\varphi(t)} d\tau \ , \tag{2.7}$$

and

$$n_i' = \int_{-\infty}^{+\infty} w(\tau) e^{j\varphi(t)} h^*(\tau - iT) d\tau \ . \tag{2.8}$$

If the phase noise cannot be approximated as nearly constant within the effective duration of the impulse response of the receive filter, the Nyquist condition for Inter-Symbol Interference (ISI) free transmission is not satisfied. It is worth writing the output of the sampled matched filter as

$$y_i = a_i e^{j\varphi_i'} \cdot \rho_i' + n_i' \ . \tag{2.9}$$

Equation (2.9) defines the Complete Model (CM). In the next subsection the Discrete-time Model commonly assumed in the literature is presented.

2.1.2 Discrete-time Wiener Phase Noise Model

In the literature of digital communications, a trusted model to study the performance of phase noise and its related topics, e.g. carrier recovery, is the Discrete-time Model (DM) [Magarini et al., 2011, Mengali, 1997] and [Spalvieri and Barletta, 2011, Magarini et al., 2012b] of Equation 1.1 where

$$\begin{aligned} \varphi_i &= \varphi_{i-1} + \sigma_{\text{PN}} \nu_i \\ \sigma_{\text{PN}} &= \sigma^2 T \\ \nu_i &\sim N(0,1) \ . \end{aligned} \tag{2.10}$$

The term $\sigma_{\text{PN}} \nu_i$ can be interpreted as the instantaneous value of a white frequency noise process, being it given by the difference between two successive phase noise samples. Note that Equation (2.10) is a first-order Markov process, since its value at the time i given the one at time $(i-1)$ does not depends on the past values [Mengali, 1997].

In other words, translation from continuous to discrete-time is simply obtained by neglecting the effects of the receive filter on the multiplicative phase noise. It is trivial that DM of Equation (1.1) in an approximation of the CM of (2.9). In the next subsection the mathematical discrepancies

between the two models are pointed out together with the introduction to the statistical analysis of the mismatch developed in the following Sections.

2.1.3 Model Comparison

The model defined by (1.1) and (2.10) is commonly assumed in computer simulations for Bit Error Rate (BER) evaluation. Remarkably, the experimental results presented in [Magarini et al., 2011] show that DM can be adopted to describe carrier phase noise after nonlinear propagation in different transmission scenarios. The goal of this Chapter is to show what are the limits of applicability of the DM in the approximation of the CM. In order to do this, we perform statistical tests on temporal and distributional properties of CM for two different roll-offs that can be considered as endpoints of the range of values that are of practical interest in optical systems. Then we compare them with those performed on the DM. The main result is the proof, by simulations, that DM is a good approximation of the CM when $\sigma_{\text{PN}} < 0.1$ rad. As a further way of evaluating the accuracy provided by the approximation we present computer simulations to compare BERs of QPSK and 16-QAM and the power spectral densities of the complex exponential phase noise obtained by simulating the two models. Nevertheless a first mathematical comparison is already presented in this Subsection. Indeed the CM to the DM of Equation (2.9) and (1.1) respectively can be compared to point out that

- the additive noise n_i' process in the CM is statistically equivalent to the process n_i of the DM,
- the term $\rho_i' e^{j(\varphi_i' - \varphi_i)}$ is a distortion on the symbol a_i given by the integration of the complex exponential through the matched filter. Actually, as pointed out in [Foschini and Vannucci, 1988], since the effect of filtering is to convert phase fluctuations in amplitude variations, phase noise can have a detrimental effect not only for the case of Phase Modulations (PMs) but also for Amplitude ones (AMs). However, this PM-AM conversion is totally neglected in the DM.

The distortion term can be explained also by reasoning in frequency domain. The noiseless part of the received signal $r(t)$ in (2.1) corresponds to the multiplication of the filtered data sequence with $e^{j\varphi(t)}$. If one translates this to the frequency domain, the Power Spectral Density of the noiseless part of the received signal is the convolution between $\sigma_a^2 |H(f)|^2/T$ and the Lorentzian spectrum of the complex exponential phase noise $e^{j\varphi(t)}$ given in (2.5). Since the overall frequency response from the input of the transmit

filter to the output of the matched filter is not proportional to $|H(f)|^2$, ISI arises.
The remaining part of this Chapter is organized as follows. Sections 2.2 and 2.3 go further in depth by comparing the statistical characterizations of the discrete-time sampled-spaced output signals and by discussing the mismatch between the two models. Simulation results are presented in Section 2.4, where we compare the BER and the power spectral density of discrete-time phase noise of the CM and with its DM approximation.

2.2 Phase Noise process Properties

The aim of this Section is to check if and when the process φ'_i in Equation 2.9 is a discrete-time Wiener process, that is the same as asking if it is a good approximation of the DM phase noise of the (2.10).
In order to verify this hypothesis, one should demonstrate that

$$\eta_i = \varphi'_i - \varphi'_{i-1} \,, \tag{2.11}$$

is a white and Gaussian process. Being η_i the difference between two phases at the two successive symbol time instants, i.e. iT and $(i-1)T$, it is defined as Discrete-Time Frequency Noise (DTFN). In the following, the AWGN terms n_i and n'_i appearing in (2.9) and (1.1) will be neglected because they affect the two discrete-time models with two statistically equivalent processes. Accordingly, the noiseless part of the CM can be written as follows

$$y_i = a_i e^{j\varphi'_i} \cdot \rho'_i \,. \tag{2.12}$$

The analysis of the mismatch between CM and DM is performed by means of simulations. It is worth emphasizing that since the goal of this study is to analyze the non-linear effects introduced by phase noise, time-domain processing is implemented. In order to synthetically generate the actual signal y_i in (2.12) the continuous-time signal is sampled at a rate much higher than the symbol interval. In our simulations the oversampling factor is equal to 20. Such a value has been chosen after a preliminary analysis with the goal of providing a safe margin for aliasing free processing and, at the same time, obtaining numeric results with reasonable complexity.

The discrete-time sequence $\{\eta_i\}$ is analyzed by a Monte Carlo (MC) simulator with the block scheme as in Figure 2.2 that is described below:

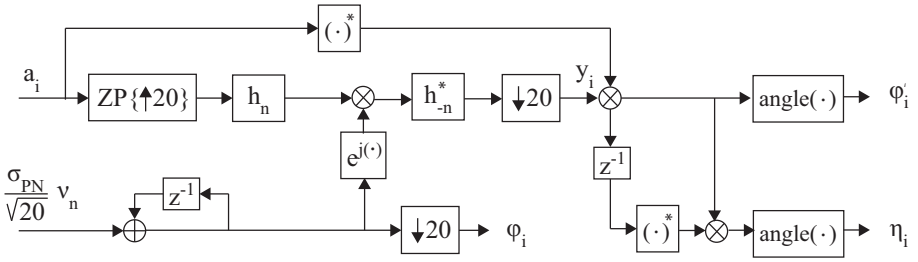


Figure 2.2: *Simulator Block Diagram for generating η_i, φ_i and φ'_i.*

- The block Zero Padding (ZP) appends 19 zeros after one input sample $\{a_i\}$, then the signal is filtered by the Square-Root Raised Cosine (SRRC) transmit filter h_n and arrives at the receiver.
- The Phase Noise (PN) process is originated through a first order filtering of the Gaussian variable $\sigma_{\text{PN}}\nu_n/\sqrt{20}$. This is done to have the cumulative sum of 20 samples have the desired variance σ^2_{PN}. Note that $\{\nu_n\} \sim N(0,1)$ is a independent identically distributed (i.i.d.) normal standard random process, whereas it is trivial that the simple down-sampling of such sequence is $\{\varphi_i\}$.
- Phase noise and receive matched filtering are applied. Then the sequence is down-sampled.
- The phase information of the actual symbol is then removed by multiplication of the complex conjugate, then the phase is the process $\{\varphi'_i\}$.
- Trivially is also obtained $\{\eta_i\}$.

In the next two Subsections statistical test about whiteness and gaussianity of the sequence η_i are performed and then the results are commented in the last Subsection of this Section.

2.2.1 Test of whiteness

In the test of whiteness we focus on the estimation of the DTFN autocorrelation. In particular we consider the Pearson's Correlation Coefficient (PCC) with time lag mT defined as [Dunn and Clark, 1986]

$$\text{PCC}_m = \frac{\text{Cov}\left[\eta_i, \eta_{i+m}\right]}{\sigma^2_\eta} = \frac{E\left[\eta_i \eta_{i+m}\right]}{E\left[\eta_i^2\right]}\,, \tag{2.13}$$

where $E[\eta_i] = 0$ can be easily assumed. In statistics, Pearson's correlation coefficient is one of the most popular tests for measuring the linear dependence between two continuous random variables [Dunn and Clark, 1986].

From (2.13) it follows that PCC_l assumes always values between -1 and $+1$. Specifically, if PCC_1 equals to 0 means that there is no correlation between the two random variables, a value of $+1$ (-1) means that there is a perfect positive (negative) relationship between them and, therefore, as one variable increases, the second one increases (decreases) in exactly the same proportion. If the DTFN sequence η_i is white it happens that

$$\text{PCC}_m = \begin{cases} 1 \text{ if } m = 0, \\ 0 \text{ otherwise.} \end{cases} \tag{2.14}$$

From the sequence of N DTFN samples $\{\eta_i\}$, generated by simulation, an estimate of the PCC of the (2.13) is obtained as

$$\hat{\text{PCC}}_1 = \frac{(N-m)^{-1}\sum_{i=1}^{N-m}\eta_i\eta_{i+m}}{(N-m)^{-1}\sum_{i=1}^{N-m}\eta_i^2} = \frac{\sum_{i=1}^{N-m}\eta_i\eta_{i+m}}{\sum_{i=1}^{N-m}\eta_i^2} \tag{2.15}$$

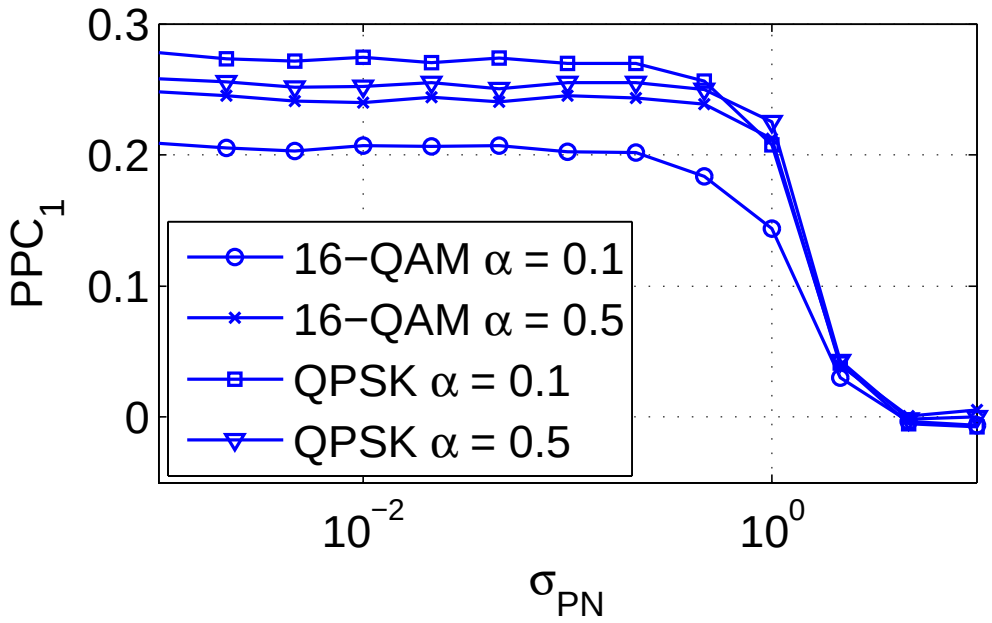


Figure 2.3: *$\hat{PCC}_1$ versus σ_{PN} for QPSK and* 16*-QAM.*

Figures 2.3 and 2.4 show $\hat{\text{PCC}}_1$ and $\hat{\text{PCC}}_2$, respectively, versus σ_{PN} for QPSK and 16-QAM with roll-off $\alpha = 0.1$ and $\alpha = 0.5$. These two roll-offs can be considered as the endpoints of the range of values that are of

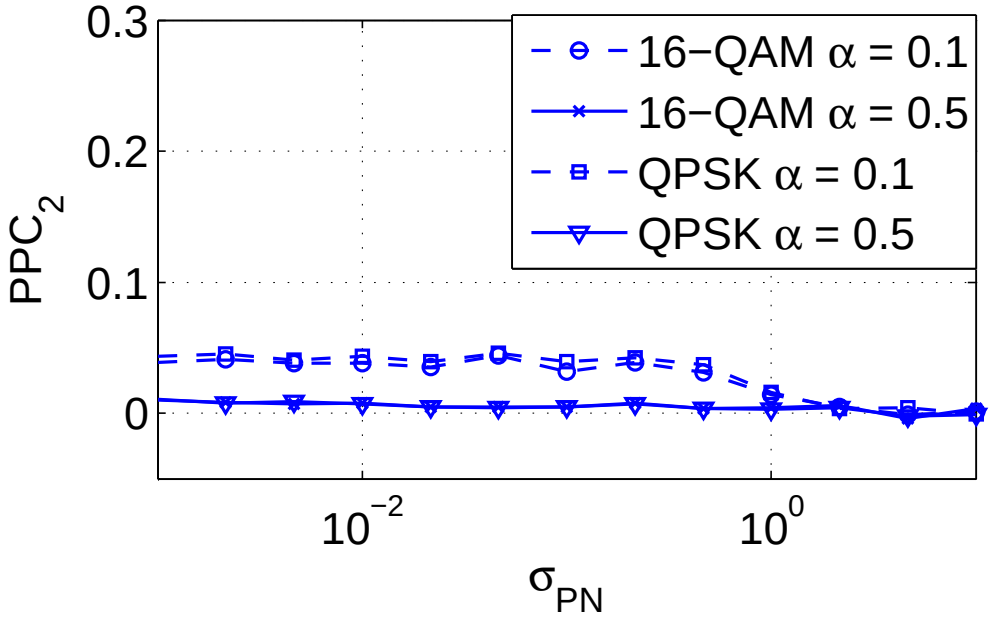


Figure 2.4: *$\hat{PCC}_2$ versus σ_{PN} for QPSK and* 16*-QAM.*

practical interest for optical systems. From Figures one can clearly distinguish between two different cases: the first where σ_{PN} is lower than 0.3 rad and the second where it is higher. In the first case one sees that while for $\alpha = 0.5$ the values of $\hat{\text{PCC}}_1$ and $\hat{\text{PCC}}_2$ are virtually not affected by the modulation format, for $\alpha = 0.1$ this property is satisfied only by $\hat{\text{PCC}}_2$. Concerning with $\hat{\text{PCC}}_1$, a higher value can be observed for QPSK than for 16-QAM. A possible explanation of this behavior resides in the combined effect of different amplitude levels of 16-QAM and slow-decaying tails of the Nyquist impulse responses with small roll-off values. The fast amplitude variations within a symbol interval induced by higher tails and amplitude levels of 16-QAM interfere in a stronger way thus reducing the observed correlation between successive samples of η_i. In the second case, where σ_{PN} has values higher than 0.3 rad, we can see that the large phase change occurring between successive samples of the Wiener phase noise process totally decorrelates the sequence of samples η_i.
Moreover, independently on the roll-off value and modulation type, $\hat{\text{PCC}}_2$ can be considered negligible with respect to $\hat{\text{PCC}}_1$. Other values $\hat{\text{PCC}}_m$, with $m > 2$, are not reported, since $\hat{\text{PCC}}_m \approx 0$.

2.2.2 Test of Gaussianity

The sequence of N DTFN samples $\{\eta_i\}$ is used to build a histogram $p_\eta(x)$ of the samples distribution. In order to test the gaussianity of the DTFN, we compute the Kullback-Leibler Divergence (KLD) between $p_\eta(x)$ and the Gaussian distribution

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ , \tag{2.16}$$

where μ and σ^2 are the mean and the variance of the distribution respectively. The KLD is defined in [Arizono and Ohta, 1989] as

$$\text{KLD}(p_\eta(x)||g(x) = \int_{-\infty}^{+\infty} p_\eta(x) \ln \left[\frac{p_\eta(x)|}{g(x)} \right] dx \ . \tag{2.17}$$

and provides a measure of the discrepancy between two probability distributions. With some easy mathematical derivations one can write

$$\begin{aligned}
\text{KLD}(p_\eta(x)||g(x) &= \int_{-\infty}^{+\infty} p_\eta(x) \ln \left[\frac{p_\eta(x)|}{g(x)} \right] dx = \\
&= -H(p_\eta(x)) - \int_{-\infty}^{+\infty} p_\eta(x) \left[\frac{1}{2} \ln(2\pi\hat{\sigma}_\eta^2) - \frac{x^2}{2\hat{\sigma}_\eta^2} \right] dx = \\
&= -H(p_\eta(x)) + \frac{1}{2} \ln(2\pi\hat{\sigma}_\eta^2) + \frac{1}{2} \frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\eta^2} = \\
&= -H(p_\eta(x)) + \frac{1}{2} \ln(2\pi e\hat{\sigma}_\eta^2) = \\
&= H(g) - H(p_\eta(x)) \ ,
\end{aligned} \tag{2.18}$$

where the entropy $H(p_z)$ of the random process z with probability density function (p.d.f.) $p_z(x)$ is defined in bit as [MacKay, 2003]

$$H(p_z) = E_{p_z}[-\log_2 p_z(x)] = -\int_x p_z(x) \log_2 p_z(x) dx \ . \tag{2.19}$$

It is worth to point out that one can easily derive the entropy of a Gaussian variable with variance σ^2, i.e. $H(g) = 0.5 \ln(2\pi e\sigma^2)$.
Simulations were carried out for QPSK and 16-QAM constellations. Figure 2.5 reports the KL divergence in nats versus σ_{PN} for roll-off $\alpha = 0.1$. For each point in the Figure, the histogram that approximates the p.d.f. is built with 10^3 bins and $N = 2 \cdot 10^5$.

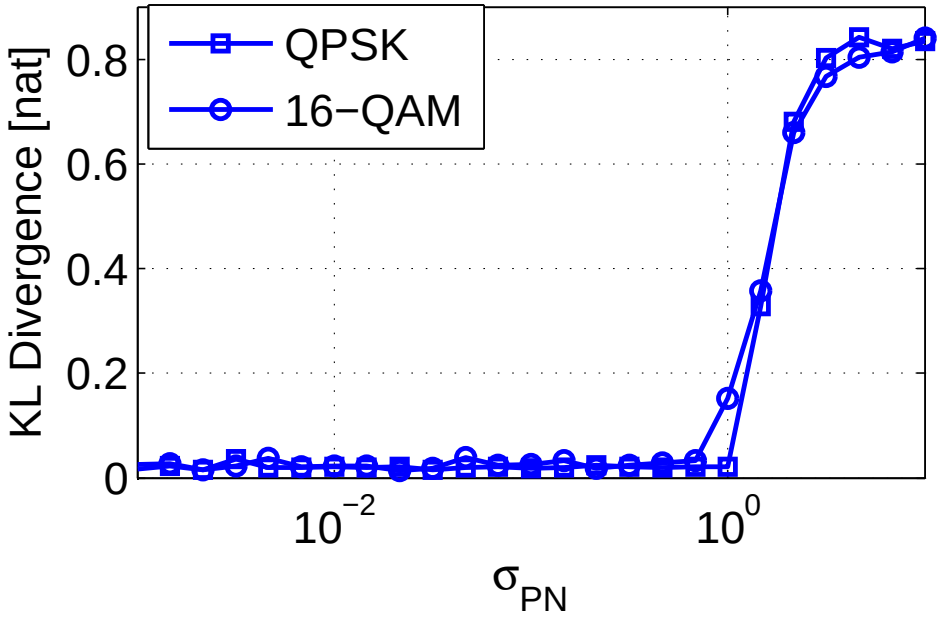


Figure 2.5: *Kullback-Leibler Divergence versus σ_{PN} with QPSK and 16-QAM.*

2.2.3 Discussion of the statistical tests

Values of $\mathrm{KLD}(p_\eta(x)||g(x)) \approx 0$ mean that $\{\eta_i\}$ is virtually Gaussian. From Figure 2.5 we observe that this happens for values of σ_{PN} lower than the threshold value $\overline{\sigma}_{\mathrm{PN}} \approx 0.3$ rad. From the Figure it is clear that the KL divergence measured for σ_{PN} below the threshold σ_{PN} is never greater than 0.04. Above $\overline{\sigma}_{\mathrm{PN}}$ the discrete-time frequency noise cannot be considered Gaussian. This means that if the phase noise is too strong then the approximation of the phase noise in the CM with the one of the DM does not hold anymore.
A completely different behavior can be observed in Figure 2.3 for $\hat{\mathrm{PCC}}_1$. Indeed when $\overline{\sigma}_{\mathrm{PN}} > \sigma_{\mathrm{PN}}$ $\hat{\mathrm{PCC}}_1$ tends to 0. This would lead us to the conclusion that the phase noise of the CM cannot be approximated to a discrete-time Wiener process, at least for small values of $\hat{\mathrm{PCC}}_1$. However, Sections 2.3 and 2.4 will enlighten that the difference between the non-white discrete-time frequency noise η_i, obtained from simulations, and the discrete-time white frequency process, defining the random increment of the Wiener phase noise in (2.10), does not have any significant impact on the power of the error associated with the mismatch due to the use of the

two models and on the associated measured BERs. As a consequence, the non-whiteness of the discrete-time phase noise can be neglected in practical cases.

2.3 Mismatch Power Analysis

In Section 2.2 the mismatch of the phase noise process between the CM and DM is studied. Still the issue about the $\hat{\text{PCC}}_1 \neq 0$ must be solved, but one can trivially understand the presence of some memory after the matched filter.
Rather than focusing on modeling processes and check their statistical agreement in this Section the author treats the model mismatch in rawer, but more efficient way. Therefore, in what follows, the power of the model mismatch is simulated and discussed.
Consider the Mean Square Error of the model mismatch, defined as

$$P = E\left[\left|x_i - a_i e^{j\varphi_i}\right|^2\right] , \tag{2.20}$$

where the processes $\{a_i\}$, $\{x_i\}$, and $\{\varphi_i\}$ are simulated from Figure 2.2. P is defined as the mismatch power.
In the previous Section it is pointed out that the CM innovation process, i.e. η_i, is not white. Particularly η_i has a non-zero correlation at $m = 1$-step.
One can be also interested in the analysis of the mismatch power P_F between the CM and a hybrid version of the DM

$$P_F = E\left[\left|x_i - a_i e^{j\varphi_{F,i}}\right|^2\right] , \tag{2.21}$$

with

$$\varphi_{F,i} = \angle\left(z_1\varphi_{i-1} + \varphi_i + z_1\varphi_{i+1}\right) , \tag{2.22}$$

where $z_1 = \hat{\text{PCC}}_1$.

From Figures 2.6 and 2.7 one can notice that for both the two considered roll-off values P and P_F scale with respect to σ_{PN} with a 20 dB/decade slope up to about 0.3 rad. By comparing Figures 2.6 and 2.7 one realizes that $10\log_1 0(P/P_F) \approx 1$ dB. This small difference means that the memory in the DTFN does not dominate the quality of the approximation. Even more important to note is that, with σ_{PN} smaller than 0.1 rad, the powers of the errors P and P_F are still very low, being more than 20 dB below the signal power. It should be said that $\sigma_{\text{PN}} = 0.1$ rad is really strong phase

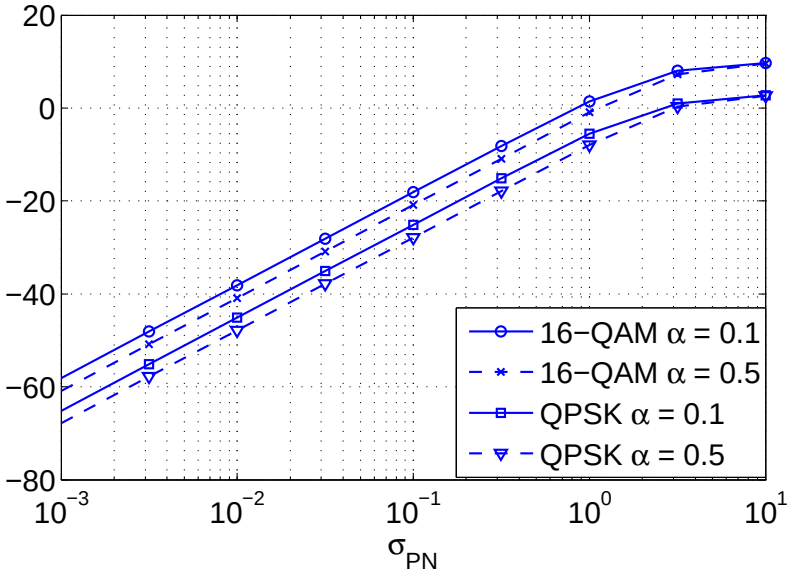


Figure 2.6: *Model Mismatch Power P between CM and DM versus σ_{PN} with QPSK and* 16*-QAM.*

noise, which can be tolerated only by robust systems as coded BPSK or coded QPSK [Pattan, 2000]. Since the threshold SNR for these systems is typically below 10 dB, we can conclude that for both the two models the level of the error power due to phase noise mismatch is much lower than the AWGN. This makes negligible the impact of the mismatch introduced by DM on system performance.

2.4 BER Performance Comparison

While in Section 2.3 the Model Mismatch is analyzed by deepening in its features, i.e. statistical test on DTFN η_i and the mismatch powers P and P_F, in this Section the author analyzes the agreement of DM with CM in terms of BER.
Particularly one expects from the results of Mismatch Power that the BER of DM and CM agrees, and starts to diverge with $\sigma_{\text{PN}} \to \overline{\sigma}_{\text{PN}}$.
Accordingly MC Simulation are run in order to have an estimate of the BER of a QPSK and a 16-QAM. The DM signals are generated according to Equations (1.1) and (2.10), while CM signal are generated according to the scheme reported in Fig. 2.2. The CM BER is measured with the roll-off values of the Square-Root Raised Cosine (SRRC) considered in previous

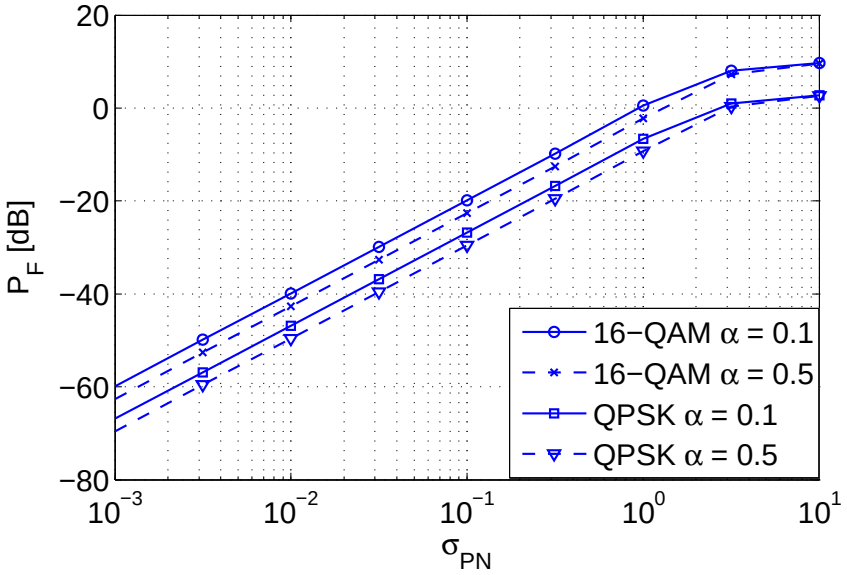


Figure 2.7: *Model Mismatch Power P_F between CM and the Filtered DM versus σ_{PN} with QPSK and* 16*-QAM.*

Sections, i.e. for $\alpha = 0.1$ and $\alpha = 0.5$. As standard deviation are considered values for σ_{PN} equal to the set $\mathcal{S} = \{3 \cdot 10^{-2}, 6.6 \cdot 10^{-2}, 0.135\}$ rad. These values of σ_{PN} represents optical systems where strong phase noise values already considered in the literature to be of practical interest [Magarini et al., 2011, Bisplinghoff et al., 2011]. The reader can consider that above those values, the transmission is so prohibitive that common constellation, e.g. QAMs, can not be considered. Hence, in these Subsections, one has to exploit strong codes and robust modulations [Pattan, 2000].

In the three Figures 2.8, 2.9 and 2.10 the results for the three values of $\sigma_{\text{PN}} \in \mathcal{S}$ are reported. It is worth noting that, only $\sigma_{\text{PN}} = 3 \cdot 10^{-2}$ and $\sigma_{\text{PN}} = 6.6 \cdot 10^{-2}$ are below the threshold of $\overline{\sigma}_{\text{PN}} \approx 0.1$ rad that defines the maximum standard deviation for which there is a good agreement between the two models. That threshold has already been observed in the previous Sections where dealing with the model properties, i.e. statistic of the phase noise, and mismatch power. In this Section that threshold is also validated through BER MC simulation.

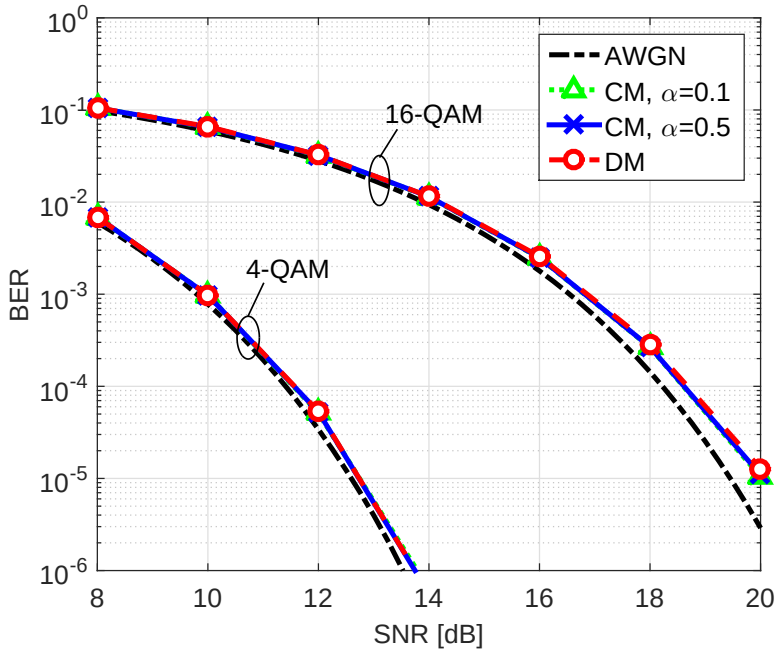


Figure 2.8: *BER versus SNR with QPSK and* 16*-QAM with* $\sigma_{PN} = 3 \cdot 10^{-2}$ *with QPSK and 16-QAM.*

Coherent demodulation of the discrete-time sequence is realized by pilot-aided trellis based demodulation scheme proposed in [Barletta et al., 2013]. Such a method is able to provide good tolerance to phase noise because it implements virtually optimal Bayesian tracking of the unknown phase. It relies on the insertion of known pilot symbols that are time-division multiplexed with the information-bearing symbols. In the results shown in this Section a pilot overhead of 5% is used. Bayesian tracking [MacKay, 2003] is an important topic in this thesis and it will be introduced in Chapters 3 and 4.

Figure 2.8 reports the BER versus SNR for $\sigma_{\text{PN}} = 3 \cdot 10^{-2}$. An excellent fit is found between the BER curves of the two models. The AWGN performance is also reported as a reference in all Figures of this Section. In the case of $\sigma_{\text{PN}} = 6.6 \cdot 10^{-2}$ the simulated BERs are shown in Figure 2.9. One can observe that for QPSK we still have a good agreement between the two models, while, in contrast, with 16-QAM a small deviation appears at BER values lower than 10^{-3}, being the performance achieved by DM slightly worse than that achieved by CM. Figure 2.10 shows results for $\sigma_{\text{PN}} = 0.135$. Due to strong phase noise, for both the two models a

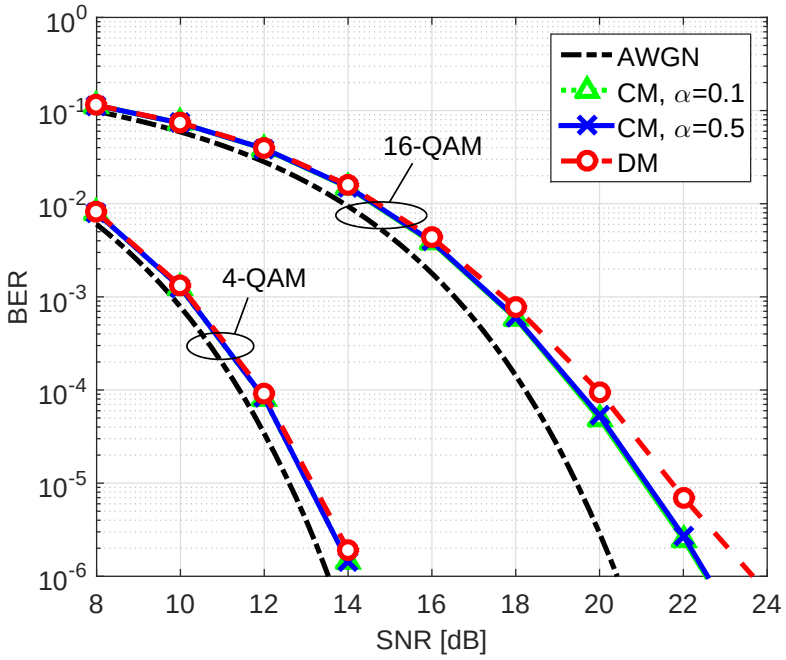


Figure 2.9: *BER versus SNR with QPSK and* 16*-QAM with* $\sigma_{PN} = 6.6 \cdot 10^{-2}$ *with QPSK and 16-QAM.*

BER floor is observed at high SNR with 16-QAM. The DM channel model exhibits a BER floor that is one order of magnitude lower than that of DM. From these results one concludes that when the DM channel model is used in computer simulations the resulting BER measure is always conservative. This gives to previous works exploiting DM rather than considering a CM approach an even more secure validation. Not only the DM is a good approximation under the condition $\sigma_{\text{PN}} < \overline{\sigma}_{\text{PN}}$, but it is even conservative in its BER analysis. The true system can work even better because of the memory introduced by all the filtering steps. Finally, in all the considered cases the roll-off factor has negligible impact on the BER performance achieved by the CM.

2.5 Phase Noise PSD

In this brief Section the author gives a frequency-domain analysis about the DM and CM mismatch, giving also a possible explanation of the better behavior of the CM compared to the DM.
The PSD of a sampled signal can be demonstrated to be equal to expected

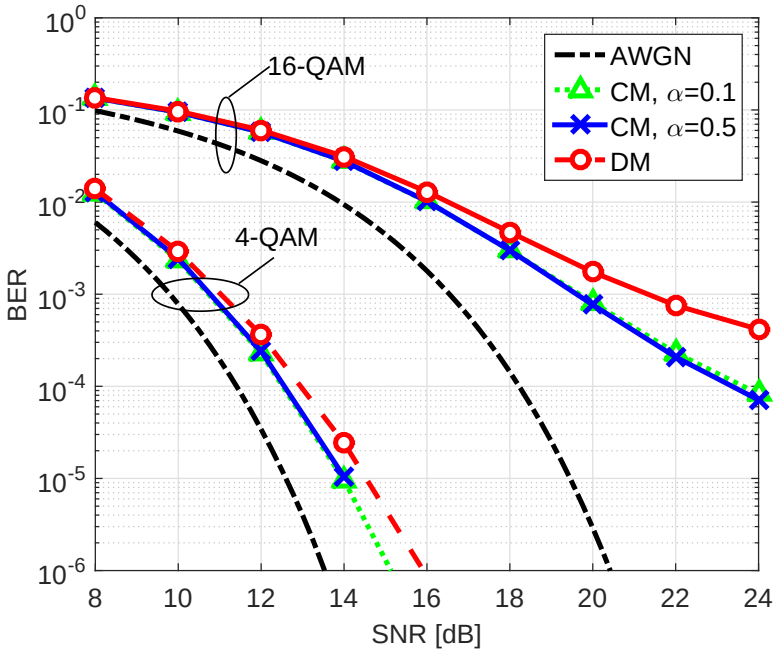


Figure 2.10: *BER versus SNR with QPSK and* 16*-QAM with* $\sigma_{PN} = 0.135$ *with QPSK and 16-QAM.*

value of the its periodogram

$$|\mathcal{P}_x(f)|^2 = \frac{T \cdot E\left[\left|\sum_{n=1}^{N} x_n e^{j2\pi nf}\right|^2\right]}{N}, \tag{2.23}$$

where T is the sampling time, i.e. symbol interval, and N the number of samples [Rabiner and Gold, 1975].
Recalling the simulation with block scheme as in Figure 2.2, the Power Spectral Density (PSD) of the two processes are simulated as the mean of the periodograms

$$|\hat{\mathcal{P}}_x(f)|^2 = \frac{T}{NM} \sum_{m=1}^{M} \left| \sum_{n=1}^{N} x_n e^{j2\pi nf} \right|^2, \tag{2.24}$$

where M is the number of averages of the periodogram taken by the simulator.

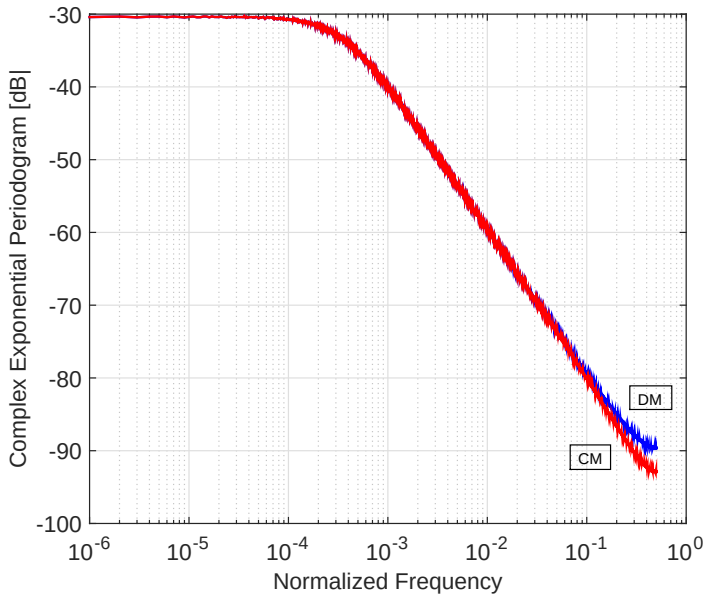


Figure 2.11: *PSD of φ_i and φ'_i of DM and CM respectively with $\sigma_{PN} = 6.6 \cdot 10^{-2}$.*

In Figures 2.11 and 2.12 the estimated PSDs of the DM and the CM phase noise complex exponential, i.e. $e^{j\varphi}$ and $e^{j\varphi'}$, are plotted. In both scenarios, the CM phase noise PSD is narrower than the DM one. The difference between the two is apparent for normalized frequency greater than 10^{-1}. This difference can be explained by observing that CM has been sampled after having been filtered through the matched filter, which increases the duration of the continuous-time phase noise memory thus narrowing the spectrum.
It is strongly intuitive that if one wants to track the phase noise in a carrier recovery system, the one with broader spectrum is worse tracked
[Rabiner and Gold, 1975]. The are several techniques to perform carrier recovery, like the Pilot-Aided trellis demodulation already presented
[Barletta et al., 2013] or blind techniques, e.g. decision-directed-mode
[Fatadin et al., 2009].
Finalizing the current dissertation, if one compares the results of this Section with the previous one about BER of the two models, he can find a sort of agreement in the behavior of the model mismatch. The better CM BER performances of the Bayesian tracking system in Section 2.4 find an explanation when comparing the two different PSD of the processes φ and φ'. Particularly, since the power of φ' is lower with higher normalized fre-

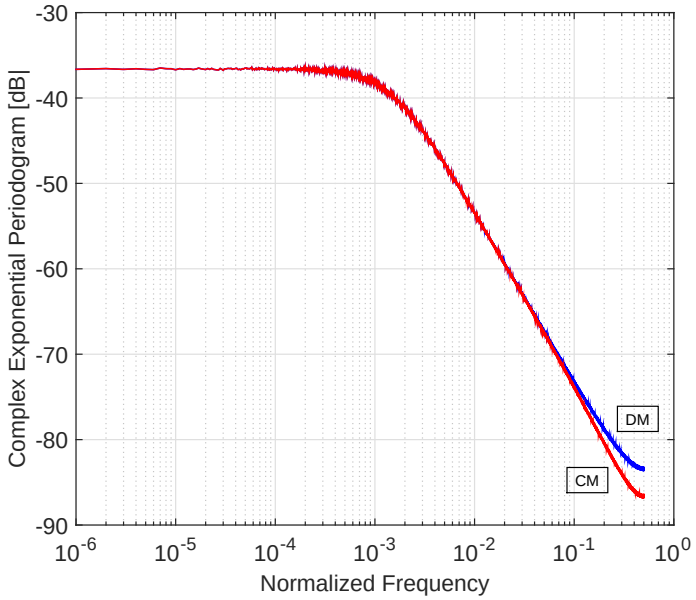


Figure 2.12: *PSD of φ_i and φ'_i of DM and CM respectively with $\sigma_{PN} = 0.135$.*

quencies can be better tracked by the low-pass nature of the Bayesian filter used to demodulate the sequence. This is another strong point to validate the results presented in the Chapter.

2.6 Conclusion

In this Chapter the limits of the symbol-spaced discrete-time channel model of Equations (1.1) that is commonly adopted to evaluate performance degradation introduced by discrete-time multiplicative Wiener phase noise have been analyzed. This so called Discrete Model (DM) is commonly assumed in the literature cited in this Chapter under the condition of "small" phase noise. However, nobody have yet studied how much "small" the phase noise must be.

Thus the author has derived the more accurate model obtained by filtering and sampling at symbol-rate the continuous-time received signal affected by multiplicative continuous-time Wiener phase noise reported in Equation (2.9). In the Chapter this model is defined as Complete Model (CM). Particularly, CM takes into account the effects that affect a received signal before sampling and any Digital Signal Processing. The fit between the two models has been analyzed by means of statistical tests aiming to verify

temporal and distributional properties of both the DM and the CM phase noise processes. The power of the error resulting from the mismatch of the noiseless signals between the two models have also been simulated. One can conclude that, when the standard deviation of the discrete-time Wiener phase-noise increment is below the threshold of DM validity $\overline{\sigma}_{\text{PN}} = 0.1$ rad, the discrete-time model provides a good approximation to the sampled filtered model with continuous-time phase noise with the same width of the spectral line. This can be concluded from the results of Section 2.3, where it is shown by MC simulation that the Power of the mismatch error is 20 dB below the power of the signals in the worst case, i.e. $\sigma_{\text{PN}} \to \overline{\sigma}_{\text{PN}}$. Note that in most practical scenarios the values of σ_{PN} are way lower than the threshold of validity of the DM.

The only notable difference between CM and DM phase noise process is the non-null correlation between the innovation of the phase noise η_i and itself delayed by one sample. The little memory introduced however is negligible in practical application because the difference produced by exploiting this information has brought only a gain of 1 dB in terms of power of mismatch, that is already very low.

The good quality of the approximation is also demonstrated by analysis of Bit Error Rate performance, showing that BERs of QPSK and 16-QAM are close to each other for the two models. Moreover, from results coming from Sections 2.4 and 2.5, one can conclude that, not only DM is a good approximation of the CM when $\sigma_{\text{PN}} < \overline{\sigma}_{\text{PN}}$, but the approximation is conservative. Indeed as one can see from Sections 2.4 and 2.5 DM's BER is always a little bit worse than the CM one and the DM phase noise process is a bit more difficult to track with a carrier recovery system for its higher PSD components with high normalized frequencies.

CHAPTER 3

Bayesian Inference in State-based Problems

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as prediction for new observation.
After an introductory Section about Bayesian Inference, the author focuses the attention on the tracking of unknown processes, e.g. phase noise, to better analyze Channel performances and its models [Gelman et al., 2014]. Bayesian tracking is a really powerful tool, but it has its own limits when experiencing complex state models, e.g. physical transmission over an optic dispersive channel. This last one will be used as an example of the limits of such approach in the last Chapter Section.

3.1 Bayesian Tracking of Channel State

As it is said in the Introduction, Bayesian Tracking can be exploited in many applications. The main objective of a communication system is the transfer of information over a channel [Nguyen and Shwedyk, 2009]. When

a signal is transmitted through a communication channel, there are two types of imperfections that cause the received signal to be different from the transmitted signal. One type of imperfection is deterministic in nature, such as linear and nonlinear distortions, ISI, etc. The second type is non-deterministic, such as addition of noise, interference, phase noise, etc. For a quantitative study of these nondeterministic phenomena, a random model is required. In this Section it is explained how one can recur to Bayesian tracking in order to study channel models. This powerful theory allows to improve communication systems in two ways. First of all, it permits to theoretically study such channels and to design better communication systems that exploit the information given by Bayesian tracking and have better performances.

If one wants to track an evolving channel in digital communication, he can deal with the so called "State-space approach", that is the focus on time-series modeling on the evolving state of a system. The state contains all relevant information required about the system under investigation. It is important to distinguish between the state, which is usually hidden, and the measurements, which are function of the state and random variables. For the purposes of this thesis work, the discrete-time formulation of the problem is considered. A complete book on the state-space approach is [Simon, 2006].

3.1.1 State-based Approach

The state process is denoted with $\{S\}$, that is composed by the random variables $\{S_0, S_1, \ldots, S_n\}$, and the measurement process with Y, that is composed by $\{Y_1, Y_2, \ldots, Y_n\}$ where n is the duration of the state and measurements sequences. As assumption, the state process starts at time index $k = 0$, while the measurements are available from the time index $k = 1$. The state-space approach is based on a *state transition equation*

$$S_k = f_{k-1}(S_{k-1}, V_{k-1}), \tag{3.1}$$

that defines the *system model*, and on a *measurement equation*

$$Y_k = h_k(S_k, N_k), \tag{3.2}$$

that defines the *measurement model*. It is important to note that in what follows, we let $k = 1, 2, \cdots$. In Equations 3.1 and 3.2 the process $\{V\}$ consists of independent variables called *process noise* and the process $\{N\}$ consists of independent variables called *measurement noise* and it is inde-

pendent of V.

$$p(v_1^n) = \prod_{k=1}^{n} p(v_k), \tag{3.3}$$

$$p(n_1^n) = \prod_{k=1}^{n} p(n_k), \tag{3.4}$$

where $p(v_k)$ and $p(n_k)$ are known probabilities, $\forall k$. From (3.2), one can point out that the evolution of the state is independent of the measurement process Y. No assumptions about the dimension of the random variables are done, i.e. according to the investigated problem, they can be random vectors or scalar random variables. Moreover, the variables can be discrete or continuous. In this way, both with continuous and discrete state system models are taken into account.

In this thesis the author considers dynamical systems that can be mapped onto the framework of first-order Markov hidden state processes. If one wants to define a first-order discrete-time Markovian state process $\{S\}$, e.g. like the one in Equation 2.10 from its joint probability, he can write

$$p(s_0^n) = p(s_0) \prod_{k=1}^{n} p(s_k|s_{k-1}), \tag{3.5}$$

where $p(s_k|s_{k-1})$ is called *state transition probability* and it is known by derivation from Equation (3.1). The measurement y_k is memoryless given the state s_k and it is characterized by the conditional distribution

$$p(y_1^n|s_0^n) = \prod_{k=1}^{n} p(y_k|s_k), \tag{3.6}$$

where $p(y_k|s_k)$ is called *measurement probability*. Note that Equation 3.6 can be easily derived from the (3.2). From Equations (3.5) and (3.6), one can write the joint distribution

$$\begin{aligned} p(s_0^n, y_1^n) &= p(y_1^n|s_0^n)p(s_0^n) \\ &= p(s_0) \prod_{k=1}^{n} p(y_k|s_k)p(s_k|s_{k-1}), \end{aligned} \tag{3.7}$$

that is all he needs to characterize the system model in a probabilistic way. After some straightforward passages, one gets

$$p(s_k|s_0^{k-1}, y_1^{k-1}) = p(s_k|s_{k-1}), \tag{3.8}$$

that is the well-known *Markovian property*.

Proof. Matching the equation (3.7) with

$$p(s_0^n, y_1^n) = p(s_0) \prod_{k=1}^{n} p(s_k, y_k | s_0^{k-1}, y_1^{k-1}),$$

one writes

$$p(s_k, y_k | s_0^{k-1}, y_1^{k-1}) = p(y_k|s_k) p(s_k|s_{k-1}). \tag{3.9}$$

Let us write

$$p(s_k, y_k | s_0^{k-1}, y_1^{k-1}) = p(y_k | s_0^k, y_1^{k-1}) p(s_k | s_0^{k-1}, y_1^{k-1})$$

henceforth by (3.6)

$$p(s_k, y_k | s_0^{k-1}, y_1^{k-1}) = p(y_k|s_k) p(s_k | s_0^{k-1}, y_1^{k-1}). \tag{3.10}$$

Equating the right sides of (3.9) and (3.10) one gets

$$p(s_k | s_0^{k-1}, y_1^{k-1}) = p(s_k|s_{k-1}).$$

□

Note that, unlike the expressions of $p(s_0^n)$, $p(y_1^n|s_0^n)$, and $p(s_0^n, y_1^n)$, reported in (3.5), (3.6), and (3.7), respectively, the a posteriori probability of the state given the measurements $p(s_0^n|y_1^n)$ is intractable. Using (3.7) it results

$$p(s_0^n|y_1^n) = \frac{p(s_0) \prod_{k=1}^{n} p(y_k|s_k) p(s_k|s_{k-1})}{p(y_1^n)},$$

where the denominator can not be written as product of marginal distributions, if not by chain rule as

$$p(y_1^n) = \prod_{k=2}^{n} p(y_k | y_1^{k-1}),$$

where $p(y_k|y_1^{k-1})$ is still not easily definable. In the following an example of the state-space approach is reported by describing the tracking of a noisy sinusoid's phase.

Esample: Sinusoid embedded in noise in the state-space approach

Estimating the phase of a sinusoid embedded in noise is a fundamental synchronization problem in many telecommunication systems. When a sinusoid is affected by phase noise, one wants to track the phase sequence,

not only to estimate a parameter. The state-based approach is used to bring the knowledge of the spectrum of the sinusoid inside the tracking problem.

A generic model for the phase noise Φ is the AutoRegressive Moving Avarage (ARMA) model. With this model different spectra of sinusoid embedded in noise can be fitted. Process Φ can be obtained as accumulation of frequency noise (like the DTFN η_i of Chapter 2 that produced φ_i), that is

$$\Phi(z) = \frac{z^{-1}}{1 - z^{-1}} \Lambda(z), \tag{3.11}$$

where the frequency noise process $\{\Lambda\}$ is the sequence of coefficients of the polynomial of complex variable z

$$\Lambda(z) = c(z)V(z), \tag{3.12}$$

where $\{V\}$ is white scalar Gaussian process noise with zero mean and variance σ^2, and

$$c(z) = \frac{\prod_{i=1}^{m}(1 - \beta_i z^{-1})}{\prod_{i=1}^{m}(1 - \alpha_i z^{-1})} = \frac{1 + \sum_{i=1}^{m} b_i z^{-i}}{1 - \sum_{i=1}^{m} a_i z^{-i}}, \tag{3.13}$$

where m is the order of the ARMA model, $|\alpha_i| < 1,\ |\beta_i| \leq 1$, therefore the transfer function $c(z)$ is causal, monic, and minimum phase [Rabiner and Gold, 1975]. Since the phase is observed through the complex exponential, to prevent the overflow in the accumulation one can periodically reduce it modulo 2π. Note that in this Chapter the simple notation σ is used instead of the longer σ_{PN} of the previous one to point at the standard deviation of the discrete-time innovation process of the phase noise.
The ARMA phase noise can be cast in the framework of the state-based approach by defining the state at time k as the column vector with dimension $d = m + 1$

$$S_k = \begin{bmatrix} \Phi_k \\ \Omega_{k-1} \\ \Omega_{k-2} \\ \vdots \\ \Omega_{k-m} \end{bmatrix}, \tag{3.14}$$

where, modeling the filter with transfer function (3.13) as a shift register with feedback taps a_1^m and forward taps b_1^m, Ω_{k-m}^{k-1} is the content of the shift register at the k-th channel use, that is

$$\Omega(z) = \frac{V(z)}{1 - a(z)}.$$

The relation between S_k and S_{k-1} results

$$S_k = \mathbf{F} S_{k-1} + \begin{bmatrix} V_{k-1} \\ V_{k-1} \\ 0_{m-1} \end{bmatrix}, \tag{3.15}$$

where the state transition matrix is

$$\begin{aligned} \mathbf{F} &= \begin{bmatrix} 1 & (a_1^m + b_1^m)^T \\ 0 & (a_1^m)^T \\ 0_{m-1} & \mathbf{I}_{m-1} \quad 0_{m-1} \end{bmatrix} \\ &= \begin{bmatrix} 1 & a_1+b_1 & a_2+b_2 & \cdots & a_{m-2}+b_{m-2} & a_{m-1}+b_{m-1} & a_m+b_m \\ 0 & a_1 & a_2 & \cdots & a_{m-2} & a_{m-1} & a_m \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}. \end{aligned} \tag{3.16}$$

The equation (3.15) is the required transition equation (3.1) of the state-based approach, where $f_{k-1}(\cdot)$ is independent of time index k.

The k-th complex measurement is

$$\begin{aligned} Y_k &= e^{jS_{k,1}} + N_k, \\ &= e^{j\Phi_k} + N_k, \end{aligned} \tag{3.17}$$

where j is the imaginary unit, $S_{k,1}$ is equal to the first entry of the state random vector S_k, and N is a complex AWGN process independent of Φ with zero mean and variance SNR^{-1}. The variance of the measurement noise N is usually expressed as the inverse of the Signal-to-Noise Ratio (SNR) when the power of the sinusoid is unitary, as in (3.17). Note that the random variables in the measurement equation are all scalar.

Figure 3.1 shows the block diagram of the model given by equations (3.14) to (3.17) with $m = 1$.

If we consider the unwrapped version of Φ_k, aware of the possibility to include Φ_k modulo 2π in the state definition if needed, the state transition probability and the measurement probability are equal to

$$\begin{aligned} p(y_k|s_k) &= g_c(e^{js_{k,1}}, \text{SNR}^{-1}; y_k) \\ &= g_c(e^{j\phi_k}, \text{SNR}^{-1}; y_k) \end{aligned} \tag{3.18}$$

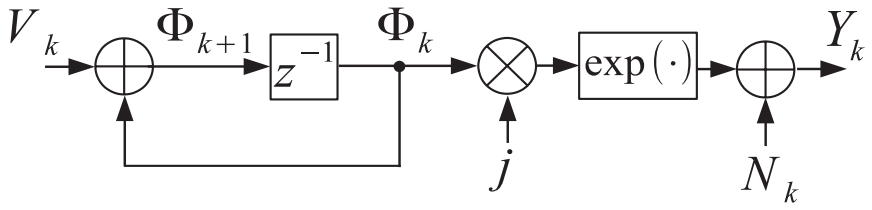


Figure 3.1: *Block diagram of the sinusoid embedded in noise and affected by ARMA phase noise with order $m = 1$, i.e. Wiener Phase Noise.*

and

$$p(s_k|s_{k-1}) = g(\mathbf{F}s_{k-1}, \mathbf{\Sigma}_v; s_k), \tag{3.19}$$

respectively, with

$$\mathbf{\Sigma}_v = \begin{bmatrix} \sigma^2 & \sigma^2 & 0_{m-1}^T \\ \sigma^2 & \sigma^2 & 0_{m-1}^T \\ 0_{m-1} & 0_{m-1} & \mathbf{0}_{m-1} \end{bmatrix}, \tag{3.20}$$

where $\mathbf{0}_m$ is an all-zero $m \times m$ matrix. Given S_{k-1}, S_k is determined if also V_{k-1} is known, hence the covariance matrix of the state transition probability has unit rank. All the probability distributions needed in the state-space approach have been defined, except the probability of the initial state $p(s_0)$ required in the equation (3.5), that is usually set with an uniform distribution in the range $[0, 2\pi)$ if there is no information about the initial phase of the sinusoid, or it is set with a Dirac delta function as distribution if the initial phase is precisely known. Note that both the state model and the measurement model are Gaussian, but the state model is also linear, while the measurement model is non-linear.

As a representative case of a class of frequency noise spectra with we take

$$c(z) = \prod_{i=1}^{m} \frac{1 - (1 - 3 \cdot 4^{-2i+1})z^{-1}}{1 - (1 - 3 \cdot 4^{-2i})z^{-1}}, \tag{3.21}$$

where the coefficients in the brackets match α_i and β_i of (3.13). The m poles and m zeros in the right side of (3.21) are interleaved and spectrally spaced of two octaves from each other. Starting from low frequency, one finds for $i = m$ the pole at $z = 1 - 3 \cdot 4^{-2m}$. This pole is followed by pairs of the type zero-pole, and the sequence of zeros and poles terminates when $i = 1$ with the zero at $z = 0.25$. Denoting by T the time delay represented

by z^{-1}, the transfer function (3.21) is that of a low-pass filter with -3 dB normalized frequency

$$f_{-3}T \approx \frac{3 \cdot 4^{-2m}}{2\pi}$$

determined by the pole at $z = 1 - 3 \cdot 4^{-2m}$.

A particular case of the ARMA phase noise is the well-know Wiener's phase noise. It can be derived setting $m = 0$ in the right side of the equation (3.13), that produces $c(z) = 1$ and $\Lambda(z) = V(z)$. In other words, the frequency noise V is only accumulated. In this case the state process coincides with the phase process ($S_k = \Phi_k, \forall k$) and the system and measurement model are defined by

$$S_k = S_{k-1} + V_{k-1}, \tag{3.22}$$

$$Y_k = e^{jS_k} + N_k, \tag{3.23}$$

where al the random variables are scalar, and the related probability distribution becomes

$$p(s_k|s_{k-1}) = g(s_{k-1}, \sigma^2; s_k), \tag{3.24}$$

$$p(y_k|s_k) = g_c(e^{js_k}, \text{SNR}^{-1}; y_k). \tag{3.25}$$

Figure 3.1 shows the block diagram of the Wiener's phase noise model given by equations (3.22) to (3.23).

The phase process (3.22) 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 reported below and is the same of Equation (2.5) Figure 3.2 reports the PSD of the ARMA phase noise obtained by the filter in (3.21) and $m = 4$, and of the Wiener's phase noise Lorentzian spectrum. For both σ is set equal to 1. As comparison, the white phase noise and a phase noise characterized by second-order model have been reported in Fig. 3.2. The second-order model is often adopted to characterize the phase noise of the free-running oscillator in synchronization problems, as in [Driessen, 1994, Patapoutian, 1999, Patapoutian, 2002, Christiansen, 1994, Spalvieri, 2006, Spalvieri and Magarini, 2008]. Note that the white phase noise does not need to bo defined in the framework of the state-based approach because it has no state.

From Fig. 3.2 one appreciates that the spectrum of phase noise obtained by frequency noise generated by (3.21) closely fits the slope of -30 dB/decade at normalized frequency higher than $f_{-3}T$, a slope that is often encountered in real world oscillators. The frequency noise that generates a phase

noise whose spectrum has a slope of -30 dB/decade is called Flicker frequency noise, or pink frequency noise, and its spectrum shows a slope of -10 dB/decade. In Figures 3.3 and 3.4 two examples of realizations of phase noise are shown, one related to the Wiener's phase noise and the other related to tha ARMA phase noise. For a detailed discussion about models of phase noise than can affect a sinusoid embedded in noise one can refer to [Meyr and Ascheid, 1990].

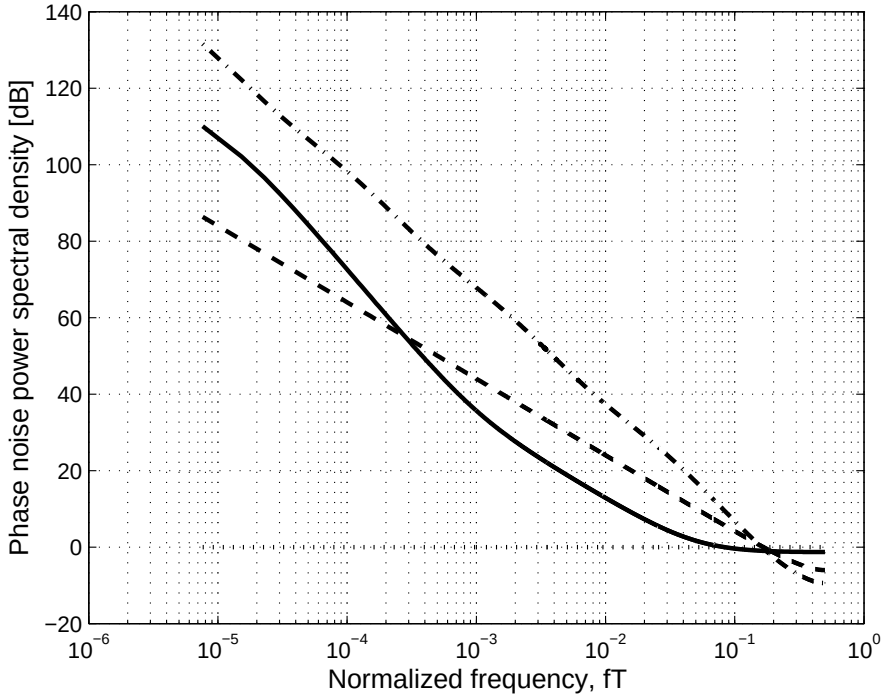


Figure 3.2: *Power spectral density of phase noise generated by accumulating white Gaussian noise with zero mean and unit variance ($\sigma = 1$) filtered through a causal, monic, and minimum phase transfer function. Solid line: phase noise model of the free-running oscillator in synchronization problems, reported in [Spalvieri and Magarini, 2008]. Dash-dotted line: phase noise generated by (3.21) with $m = 4$ followed by accumulation. Dashed line: Wiener's phase noise. Dotted line: white phase noise.*

3.1.2 Bayesian Tracking

The Bayesian tracking of an hidden Markovian state is a general probabilistic approach to estimate the unknown a-posteriori probability distribution of the evolving state recursively over time. This is done using incoming

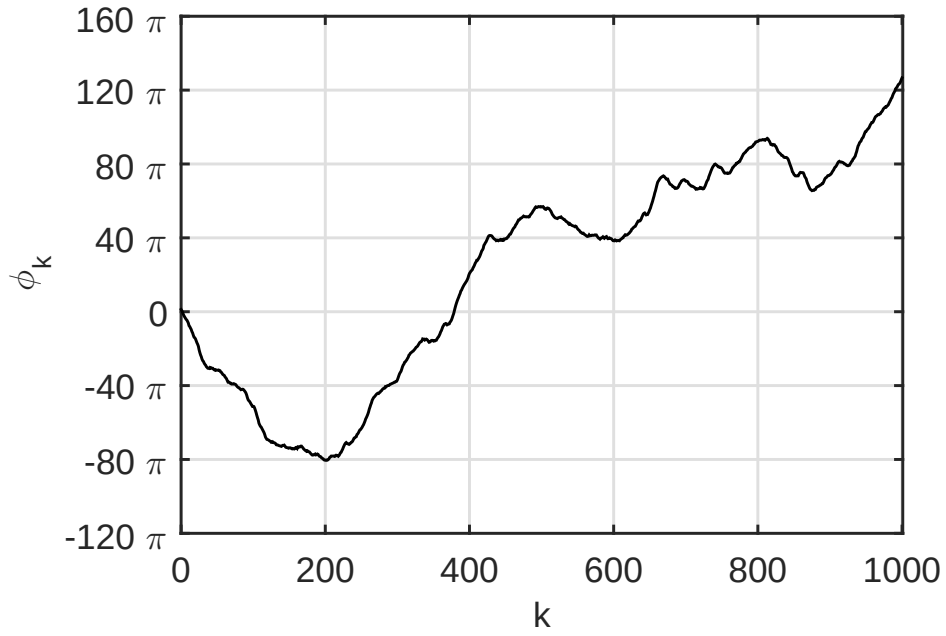


Figure 3.3: *ARMA phase noise example generated by (3.21) with* $m = 4$ *and* $\sigma = 1$.

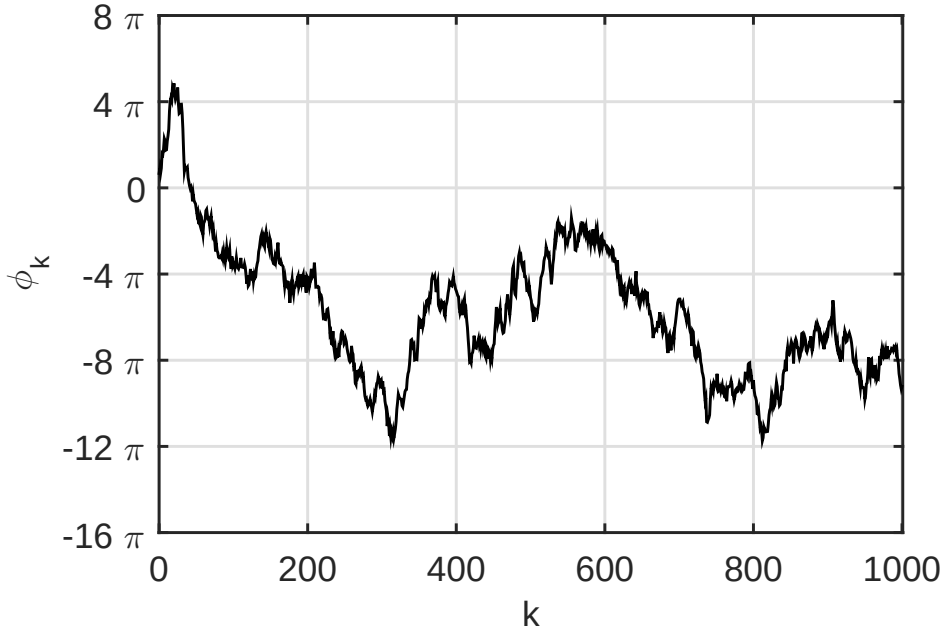


Figure 3.4: *Wiener's phase noise example with* $\sigma = 1$.

measurements and the system and measurement model in (3.1) and (3.2). Bayesian tracking allows to track the state when the state transition probability and the measurement probability are know and treatable. Without loss of generality, we consider the Markovian state continuous. Therefore one wants to track the probability density function of the state given the measurements. The case of discrete state can be directly derived, taking into account that the tool tracks the probability mass function of the state.

The well-known *two-step Bayesian recursion* allows to track the desired distribution of the state alternating predict and update step. For $k = 1, 2, \cdots$, two-step recursion is composed by the predict step, that is

$$p(s_k|y_1^{k-1}) = \int_{\mathcal{S}_{k-1}} p(s_k|s_{k-1})p(s_{k-1}|y_1^{k-1})\,\mathrm{d}s_{k-1}, \tag{3.26}$$

and the update step, that is

$$p(s_k|y_1^k) = \frac{p(s_k|y_1^{k-1})p(y_k|s_k)}{p(y_k|y_1^{k-1})}, \tag{3.27}$$

where $p(s_k|y_1^{k-1})$ is the *predictive* distribution and $p(s_k|y_1^k)$ is the *posterior* distribution. The term $p(y_k|y_1^{k-1})$ is a normalization factor independent of the state such that the left-hand side of (3.27) is a probability, i.e. it sums to 1 [Walpole et al., 1993]. This last factor can be computed by the Chapman-Kolmogorov equation

$$p(y_k|y_1^{k-1}) = \int_{\mathcal{S}_k} p(s_k|y_1^{k-1})p(y_k|s_k)\,\mathrm{d}s_k, \tag{3.28}$$

The predictive distribution is the result of the total probability law with the state transition probability $p(s_k|s_{k-1})$ in place of $p(s_k|s_{k-1}, y_1^{k-1})$ thanks to the Markovian property in (3.8). The update distribution derives from the Bayes rule and the substitution of $p(y_k|s_k, y_1^{k-1})$ with $p(y_k|s_k)$ thanks to (3.6). For $k = 1$ the predictive distribution is

$$p(s_1) = \int_{\mathcal{S}_0} p(s_1|s_0)p(s_0)\,\mathrm{d}s_0. \tag{3.29}$$

When one knows the initial state $s_0 = s_0^\dagger$, $p(s_0)$ is a Delta function and $p(s_1)$ is equal to the state transition probability $p(s_1|s_0 = s_0^\dagger)$. When one does not have any information about the initial state, $p(s_0)$ is uniform on the domain $\mathcal{S}_0$ and $p(s_1)$ often remains uniform on its domain. When the state is discrete we can easily replace the integration operator in (3.26), (3.27),

and (3.28) with a sum. The two steps in (3.26) and (3.27) can be merged getting the *one-step Bayesian recursion* for the posterior distribution

$$p(s_k|y_1^k) = \frac{p(y_k|s_k)}{p(y_k|y_1^{k-1})} \int_{\mathcal{S}_{k-1}} p(s_k|s_{k-1}) p(s_{k-1}|y_1^{k-1}) \, \mathrm{d}s_{k-1}, \tag{3.30}$$

where the normalization factor can be computed as

$$p(y_k|y_1^{k-1}) = \int_{\mathcal{S}_{k-1}^k} p(y_k|s_k) p(s_k|s_{k-1}) p(s_{k-1}|y_1^{k-1}) \, \mathrm{d}s_{k-1}^k. \tag{3.31}$$

According to the application one can choose to employ the two-step or the one-step Bayesian recursion. In those applications where only the ratio of the probability of one state to the probability of the other states is required, computing $p(y_k|y_1^{k-1})$, which is common to the entire state space, is not necessary, therefore one si satisfied with

$$\begin{aligned} p(s_k|y_1^k) &\propto p(s_k|y_1^{k-1}) p(y_k|s_k) \\ &\propto p(y_k|s_k) \int_{\mathcal{S}_{k-1}} p(s_k|s_{k-1}) p(s_{k-1}|y_1^{k-1}) \, \mathrm{d}s_{k-1}, \end{aligned} \tag{3.32}$$

where the first and the second lines are the proportional version of (3.27) and (3.30) respectively.

Example: Phase tracking of a sinusoid embedded in noise affected by Wiener's phase noise

This example recalls the previous Subsection related to the sinusoid embedded in noise affected by Wiener's phase noise. We assume that the initial state is know and it is equal to 0, which leads to $p(s_0) = \delta(s_0)$. In Fig. 3.5 one can see an example of evolution of the two-step Bayesian recursion for SNR $= 1$ dB and $\sigma = 1$. The first predictive distribution, that bootstraps the algorithm, coincides with the state transition probabilities with conditioning term s_0 equal to 0. After this, each predictions make smoother the previous update distribution (this is the effect of the integral in (3.26)), while each update makes sharper the previous predictive distribution using the actual measurement (this is the effect of the product in (3.27)). In this case study, we can see as the predictive and posterior distributions are multimodal. This happens when the phase noise impairment is consistent (in other words σ is high) and the SNR is sufficiently low.

Figure 3.6 shows a realization of the Maximum A-posteriori Probability (MAP) estimation obtained by maximization of the posterior distribution

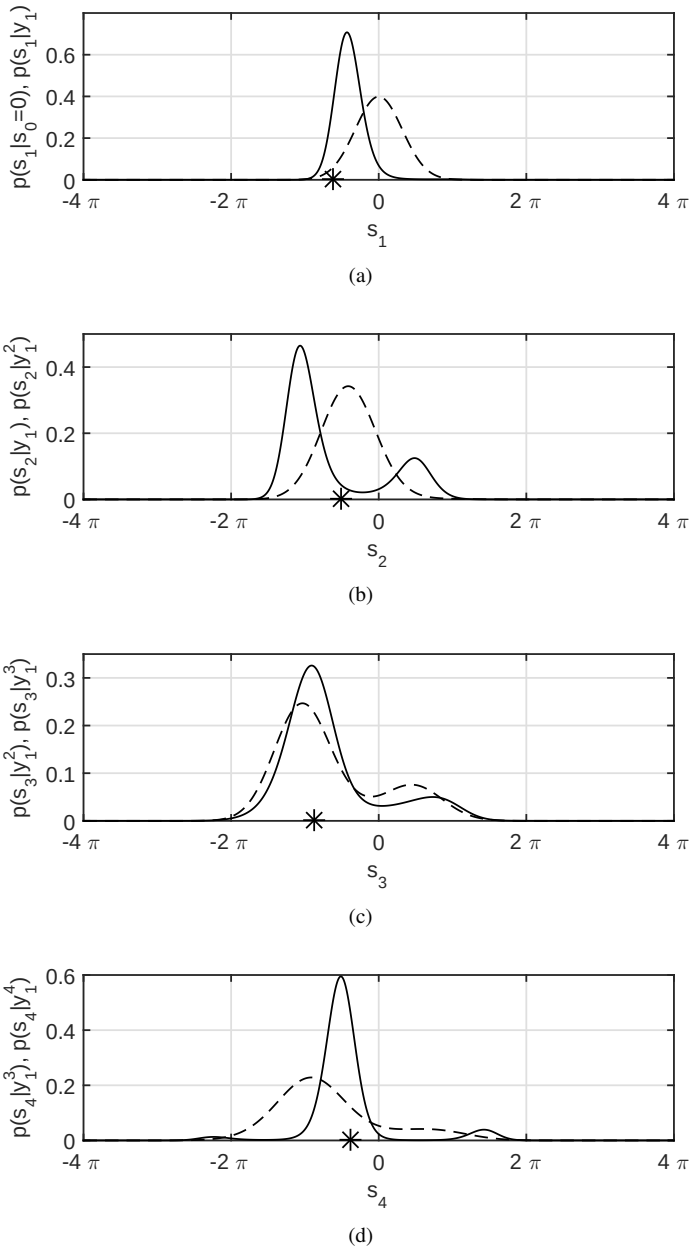


Figure 3.5: *Example of two-step Bayesian recursion applied to the tracking of the phase of the sinusoid embedded in noise affected by Wiener's phase noise. SNR* $= 1dB$*,* $\sigma = 1$*, and* $p(s_0) = \delta(s_0)$*. Dashed line: predictive distribution* $p(s_k|y_1^{k-1})$*. Solid line: posterior distribution* $p(s_k|y_1^k)$*. Asterisk: actual value of the phase. (a)* $k = 1$*. (b)* $k = 2$*. (c)* $k = 3$*. (d)* $k = 4$*.*

$p(s_k|y_1^k)$ recursion. The figure reports the case of sinusoid embedded in Wiener's phase noise with SNR $= 5$ dB and $\sigma = 2$. In Fig. 3.6 the cycle slip phenomenon is present: the tricky problem to track the unwrapped phase of a sinusoid is to not perform phase slips multiple of 2π. This occurs because the measurements are function of the phase modulo 2π. When the phase noise impairment is consistent and the SNR is not sufficiently high, the phase tracking with an error of 2π or its multiples can occur. These errors are not a problem, since the phase is defined modulo 2π, but the transient in its estimate given by a cycle slip can heavily affect the system performance.

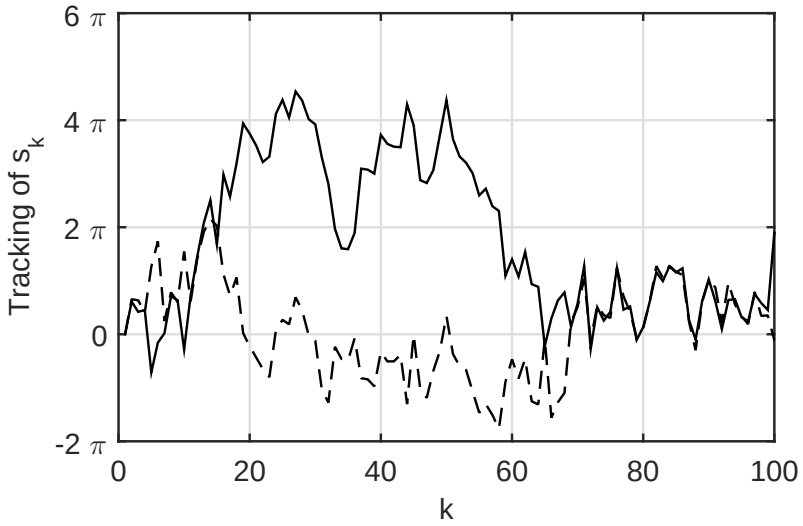


Figure 3.6: *Example of MAP estimation applied to the phase tracking of a sinusoid embedded in noise affected by Wiener's phase noise. Specifically, the estimation is obtained by the maximization of the posterior probability of the Bayesian recursion. SNR* $= 5dB$*,* $\sigma = 2$*, and* $p(s_0) = \delta(s_0)$*. Dashed line: MAP estimation of the phase* s_k*. Solid line: actual phase* s_k*.*

Until now, the Bayesian recursion has been presented following the natural evolution of the time. Specifically the forward version of the Bayesian recursion have been described. It is possible to see Equations (3.26), (3.27),

and (3.28) in the backward direction as follows

$$p(s_k|y_{k+1}^n) = \int_{\mathcal{S}_{k+1}} p(s_k|s_{k+1})p(s_{k+1}|y_{k+1}^n)\,\mathrm{d}s_{k+1}, \tag{3.33}$$

$$p(s_k|y_k^n) = \frac{p(s_k|y_{k+1}^n)p(y_k|s_k)}{p(y_k|y_{k+1}^n)}, \tag{3.34}$$

$$p(y_k|y_{k+1}^n) = \int_{\mathcal{S}_k} p(s_k|y_{k+1}^n)p(y_k|s_k)\,\mathrm{d}s_k, \tag{3.35}$$

where the recursion starts from the probability distribution of the final state $p(s_{n+1})$ and it goes on in the past alternating prediction and update steps.

If one wants to make inference on S_k using the entire measurement Y_1^n then one can exploit the *forward-backward Bayesian recursion*, that is

$$p(s_k|y_1^n) = \frac{p(s_k, y_1^k)p(y_{k+1}^n|s_k)}{p(y_1^n)} \tag{3.36}$$

$$= \frac{p(s_k|y_1^k)p(s_k|y_{k+1}^n)}{p(s_k)}\frac{p(y_1^k)p(y_{k+1}^n)}{p(y_1^n)} \tag{3.37}$$

$$\propto \frac{p(s_k|y_1^k)p(s_k|y_{k+1}^n)}{p(s_k)}. \tag{3.38}$$

Proof. Starting from the definition of conditional probability and splitting the sequence of the measurements, one has

$$\begin{aligned} p(s_k|y_1^n) &= \frac{p(s_k, y_1^n)}{p(y_1^n)} \\ &= \frac{p(s_k, y_1^k, y_{k+1}^n)}{p(y_1^n)} \\ &= \frac{p(s_k, y_1^k)p(y_{k+1}^n|s_k)}{p(y_1^n)}, \end{aligned}$$

where the last step exploits the equality $p(y_{k+1}^n|s_k, y_1^k) = p(y_{k+1}^n|s_k)$. This last equation is equal to (3.36). Substituting the first factor of the numerator with $p(s_k|y_1^k)p(y_1^k)$ and exploiting the Bayes rule for the second factor, one writes

$$p(s_k|y_1^n) = \frac{p(s_k|y_1^k)p(s_k|y_{k+1}^n)}{p(s_k)}\frac{p(y_1^k)p(y_{k+1}^n)}{p(y_1^n)},$$

that is equal to (3.37). Taking into account that the second ratio is independent to s_k, the equation (3.38) can be written. □

The forward-backward Bayesian recursion is the result of the product of the forward posterior distribution $p(s_k|y_1^k)$ and the backward predictive distribution $p(s_k|y_{k+1}^n)$, as described in (3.37). The $p(s_k)$ distribution can be usually derived from the system model. It often happens that the distribution of the state S_k without other conditioning terms is uniform and then $p(s_k)$ becomes a normalization factor together with the second ratio of the (3.37). Also in this case, when the application requires the ratio of the probability $p(s_k|y_1^n)$ of one state to the probability of the other states, one can use the Equation (3.38). It is easy to demonstrate that if one implements the product between the forward predictive distribution and the backward posterior distribution in place of the product between the forward posterior distribution and the backward predictive distribution, the result does not change:

$$p(s_k|y_1^n) = \frac{p(s_k|y_1^{k-1})p(s_k|y_k^n)}{p(s_k)} \frac{p(y_1^{k-1})p(y_k^n)}{p(y_1^n)}.$$

Proof. By Bayes rule one has

$$\begin{aligned} p(s_k|y_1^k) &= \frac{p(y_k|s_k)p(s_k|y_1^{k-1})}{p(y_k|y_1^{k-1})}, \\ p(s_k|y_k^n) &= \frac{p(y_k|s_k)p(s_k|y_{k+1}^n)}{p(y_k|y_{k+1}^n)}, \end{aligned} \tag{3.39}$$

where the second equation leads to

$$p(s_k|y_{k+1}^n) = \frac{p(s_k|y_k^n)p(y_k|y_{k+1}^n)}{p(y_k|s_k)} \tag{3.40}$$

by inversion. In addition one writes:

$$\begin{aligned} p(y_1^k) &= p(y_k|y_1^{k-1})p(y_1^{k-1}), \\ p(y_k^n) &= p(y_k|y_{k+1}^n)p(y_{k+1}^n), \end{aligned} \tag{3.41}$$

where the second equation leads to

$$p(y_{k+1}^n) = \frac{p(y_k^n)}{p(y_k|y_{k+1}^n)} \tag{3.42}$$

by inversion. Replacing (3.39), (3.40), (3.41), and (3.42) in (3.37) one has

$$\begin{aligned} p(s_k|y_1^n) &= \frac{p(y_k|s_k)p(s_k|y_1^{k-1})p(s_k|y_k^n)p(y_k|y_{k+1}^n)}{p(s_k)p(y_k|y_1^{k-1})p(y_k|s_k)} \\ &\quad \frac{p(y_k|y_1^{k-1})p(y_1^{k-1})p(y_k^n)}{p(y_k|y_{k+1}^n)p(y_1^n)} \\ &= \frac{p(s_k|y_1^{k-1})p(s_k|y_k^n)}{p(s_k)} \frac{p(y_1^{k-1})p(y_k^n)}{p(y_1^n)}. \end{aligned}$$

□

The two factors in the numerator of (3.36) can be obtained, by the forward and backward recursions of the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm described in [Bahl et al., 1974], where it has been applied to decode the convolutional and block codes with the the goal of to minimize the Symbol Error Rate (SER). Therefore, it has been presented for discrete-state case, but it is employable also to track continuous-state. The main equations of BCJR are

$$p(s_k, y_1^n) = p(s_k, y_1^k)p(y_{k+1}^n|s_k), \tag{3.43}$$

$$p(s_k, y_1^k) = \sum_{\mathcal{S}_{k-1}} p(s_{k-1}|y_1^{k-1})p(s_k, y_k|s_{k-1}), \tag{3.44}$$

$$p(y_{k+1}^n|s_k) = \sum_{\mathcal{S}_{k+1}} p(y_{k+2}^n|s_{k+1})p(s_{k+1}, y_{k+1}|s_k), \tag{3.45}$$

where (3.44) is the BCJR forward recursion, (3.45) is the BCJR backward recursion, and the product of (3.44) and (3.45) leads to (3.43). In the BCJR algorithm the forward recursion (3.44) bootstraps with $p(s_0)$ in place of $p(s_{k-1}|y_1^{k-1})$ when $k = 1$, while the backward recursion (3.45) starts with $p(y_{k+2}^n|s_{k+1}) = 1$ when $k = n - 1$. Contrary to the forward-backward Bayesian tracking in (3.37), the BCJR forward-backward recursion tracks the joint probability $p(s_k, y_1^n)$. It can be used when we are interested to a proportional value of $p(s_k|y_1^n)$, as in (3.38).

Example: Forward-backward phase tracking of a sinusoid embedded in noise affected by Wiener's phase noise

In this example the sinusoid embedded in noise affected by Wiener's phase noise in still considered, but now taking into account the wrapped version of the phase. The state transition equation becomes non-linear, while the

measurement model does not change:

$$S_k = mod(S_{k-1} + V_{k-1}, 2\pi), \tag{3.46}$$
$$Y_k = e^{jS_k} + N_k. \tag{3.47}$$

The state domain is bounded in $[0, 2\pi)$. In this case the state transition probability becomes

$$p(s_k|s_{k-1}) = \sum_{i=-\infty}^{\infty} g(s_{k-1} + 2i\pi, \sigma^2; s_k). \tag{3.48}$$

In Fig. 3.7 one can see an example of evolution of the forward-backward Bayesian recursion for SNR $= 8dB$ and $\sigma = 0.5$, applied to a measurement frame with length $n = 100$. In the example the initial and final states s_0 and s_{101}, respectively, have been assumed know. The predict and update steps of the two-step Bayesian recursion in both forward and backward direction have been numerically computed. As well as in previous examples, one can say that the predictive probabilities are smoother then the posterior ones, both in forward and backward direction.

3.1.3 Parametric Bayesian tracking

Taking into account continuous state process, when the measurement and the state evolution are expressed by a linear and additive noise model with Gaussian measurement noise and process noise, tracking of the state is feasible by the Kalman filter [Kalman, 1960]. Suppose that the processes S and Y are composed by random vectors. The case of processes consisting of scalar random variables is easily derivable. The state transition equation and the measurement equation are

$$S_k = \mathbf{F}_{k-1}\, S_{k-1} + V_{k-1}, \tag{3.49}$$
$$Y_k = \mathbf{H}_k\, S_k + N_k, \tag{3.50}$$

where $\mathbf{H}_k$ and $\mathbf{F}_{k-1}$ are known. The vectors V_{k-1} and N_k are jointly independent and white real Gaussian random vectors with zero mean vectors and covariance matrices $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$, respectively. The case of white circular complex Gaussian noises can be derived from the studied real case replacing in the following equations the transpose operator with the hermitian operator and the real Gaussian distribution with circular complex Gaussian distribution.

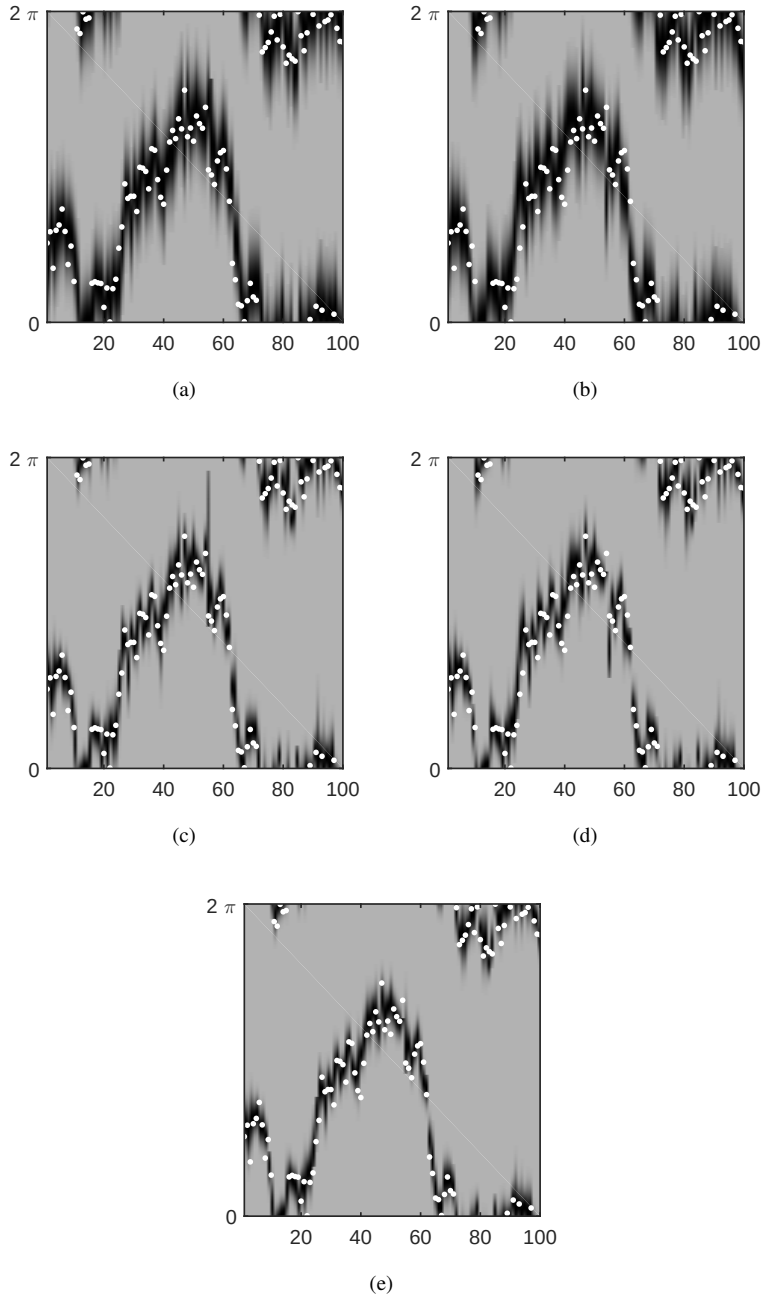


Figure 3.7: *Example of forward-backward Bayesian recursion applied to the tracking of the wrapped phase of the sinusoid embedded in noise affected by Wiener's phase noise. SNR* $= 8dB$*,* $\sigma = 0.5$*, and* $n = 100$*.* $p(s_0)$ *and* $p(s_{101})$ *are Dirac function. Grayscale image: probabilities distribution for each* k *time indexes in the* x*-axis. White dots: actual value of the phase. (a) forward predictive distribution* $p(s_k|y_1^{k-1})$*. (b) backward predictive distribution* $p(s_k|y_{k+1}^n)$*. (c) forward posterior distribution* $p(s_k|y_1^k)$*. (d) backward posterior distribution* $p(s_k|y_k^n)$*. (e) forward-backward distribution* $p(s_k|y_1^n)$*.*

From the models (3.49) and (3.50) the state transition and the measurement probabilities are Gaussian

$$p(s_k|s_{k-1}) = g(\mathbf{F}_{k-1}\ s_{k-1}, \mathbf{Q}_{k-1}; s_k)$$
$$p(y_k|s_k) = g(\mathbf{H}_k\ s_k, \mathbf{R}_k; s_k)$$

By the assumptions made, inferring the state by the two-step Bayesian recursion given by Equations (3.26) and (3.27), the predictive and the posterior distributions still remain multivariate Gaussian distributions. Specifically, one has the predictive distribution equal to

$$p(s_k|y_1^{k-1}) = g(\overline{m}_k, \overline{\mathbf{\Sigma}}_k; s_k)$$

with parameters

$$\begin{aligned}\overline{m}_k &= \mathsf{E}\left\{S_k|Y_1^{k-1}\right\} \\ &= \mathbf{F}_{k-1}m_{k-1}, \qquad (3.51)\\ \overline{\mathbf{\Sigma}}_k &= cov\left\{S_k|Y_1^{k-1}\right\} \\ &= \mathbf{F}_{k-1}\mathbf{\Sigma}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \qquad (3.52)\end{aligned}$$

that is the result of a convolution between two Gaussian functions, and the posterior distribution equal to

$$p(s_k|y_1^k) = g(m_k, \mathbf{\Sigma}_k; s_k)$$

with parameters

$$\begin{aligned}m_k &= \mathsf{E}\left\{S_k|Y_1^k\right\} \\ &= \overline{m}_k + \mathbf{G}_k u_k, \qquad (3.53)\\ \mathbf{\Sigma}_k &= cov\left\{S_k|Y_1^k\right\} \\ &= \overline{\mathbf{\Sigma}}_k - \mathbf{G}_k\mathbf{H}_k\overline{\mathbf{\Sigma}}_k, \qquad (3.54)\end{aligned}$$

and

$$u_k = y_k - \mathbf{H}_k\overline{m}_k, \qquad (3.55)$$
$$\mathbf{G}_k = \overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T\left(\mathbf{H}_k\overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}. \qquad (3.56)$$

The random vector U_k is the k-th element of the *innovation process* $\{U\}$. The matrix $\mathbf{G}_k$ is known as *Kalman gain*. Following these last six equations, one can say that the Kalman filter is a specific parametric tracking of the two-step Bayesian recursion composed by the prediction and the update

of two parameters, that are the mean and the covariance of the predictive and update Gaussian distributions.

When the model is that in (3.49) and (3.50) with Gaussian state noise and measurement noise, the Kalman filter performs the *exact* Bayesian tracking. Conversely, when the noise sequences are not Gaussian and/or the system and measurement models are non-linear, the Kalman filter computes only an *approximated* Bayesian tracking. For jointly Gaussian noise sequences, the predictive mean $\overline{m}_k$ and the posterior mean m_k are the optimal estimations under the MAP and Minimum Mean Square Error (MMSE) criteria:

$$\begin{aligned}
\overline{m}_k &= \arg\max_{s_k \in \mathcal{S}_k} p(s_k|y_1^{k-1}) \\
&= \arg\min_{\widehat{s}_k \in \mathcal{S}_k} \mathsf{E}\left\{(S_k - \widehat{s}_k)^T (S_k - \widehat{s}_k)\right\}, \\
m_k &= \arg\max_{s_k \in \mathcal{S}_k} p(s_k|y_1^{k}) \\
&= \arg\min_{\widetilde{s}_k \in \mathcal{S}_k} \mathsf{E}\left\{(S_k - \widetilde{s}_k)^T (S_k - \widetilde{s}_k)\right\},
\end{aligned}$$

where $\widehat{s}_k$ and $\widetilde{s}_k$ are functions of y_1^{k-1} and y_1^k, respectively, and are the generic estimator of the state s_k in the prediction and update step. More in general, for Gaussian noise processes, the Kalman filter is the best predictor in the following sense:

$$\overline{m}_k = \arg\min_{\widehat{s}_k \in \mathcal{S}_k} \mathsf{E}\left\{(S_k - \widehat{s}_k)^T \mathbf{M}_k (S_k - \widehat{s}_k)\right\}, \tag{3.57}$$

where $\mathbf{M}_k$ is an user-defined positive definite matrix. When the measurement noise $\{N\}$ and the process noise $\{V\}$ are not Gaussian, the prediction made by the Kalman filter is not optimal, but it is the best linear solution to the minimization problem in (3.57). Following the equations (3.51) and (3.53), the Kalman filter can be seen as a 1-casual predictive filter, as depicted in Fig. 3.8.

The innovation process $\{U\}$ contains the novelty carried by the measurement about the state after the prediction step, as one can see in the equation (3.55). The covariance of the random vector U_k is

$$cov\left\{U_k|Y_1^{k-1}\right\} = \mathbf{H}_k \overline{\mathbf{\Sigma}}_k \mathbf{H}_k^T + \mathbf{R}_k$$

The innovation process is white and, with Gaussian noise processes, this means that all the information that is present in the past about the present has been extracted. One way to check that the filter is working is to verify that the innovation sequence is white.

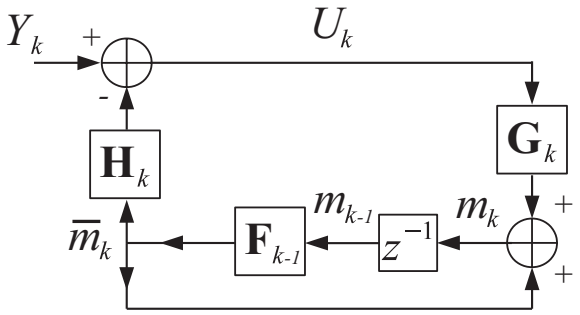


Figure 3.8: *Block diagram of the Kalman filter.*

From an other point of view the Kalman prediction is the best linear solution linear solution of the following criterion

$$\overline{m}_k = \arg \min_{\hat{s}_k \in \mathcal{S}_k} \mathsf{E}\left\{U_k^T \mathbf{M}_k U_k\right\},$$

where $U_k = Y_k - \mathbf{H}_k \hat{s}_k$, also when the noise processes are non-Gaussian.

In practical problem, the state transition and the measurement Equations are often non-linear. In this case performing exact Bayesian tracking is often infeasible. One can linearize the nonlinear equation around the Kalman filter estimate. This is the idea of the *extended Kalman filter*, which was originally proposed by Stanley Schmidt so that the Kalman filter could be applied to nonlinear spacecraft navigation problems [Bellantoni and Dodge, 1967]. The system and measurement models need not be linear functions of the state but may instead be differentiable functions. We refer to the following model:

$$S_k = f_{k-1}(S_{k-1}) + V_{k-1}, \tag{3.58}$$
$$Y_k = h_k(S_k) + N_k, \tag{3.59}$$

where V_{k-1} and N_k are jointly independent random vectors with zero mean vectors and covariance matrices $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$, respectively. The differentiable functions $f_{k-1}(\cdot)$ and $h_k(\cdot)$ can be linearized by partial derivation around an estimation. Let

$$\mathbf{F}_{k-1} = \left.\frac{\partial f_{k-1}(s_k)}{\partial s_k}\right|_{s_k = m_{k-1}} \tag{3.60}$$

$$\mathbf{H}_k = \left.\frac{\partial h_k(s_k)}{\partial s_k}\right|_{s_k = \overline{m}_k} \tag{3.61}$$

the Jacobians matrices calculated in

$$\overline{m}_k = f_{k-1}(m_{k-1})$$
$$m_k = h_k(\overline{m}_k),$$

that are the means of the predictive and posterior inferred distribution, respectively, used as estimation. The covariance of these distributions can be recursively calculated with (3.52) and (3.54). Following this steps one can approximate the predictive and posterior distributions as Gaussian:

$$p(s_k|y_1^{k-1}) \simeq g(\overline{m}_k, \overline{\mathbf{\Sigma}}_k; s_k),$$
$$p(s_k|y_1^{k}) \simeq g(m_k, \mathbf{\Sigma}_k; s_k).$$

The extended Kalman filter can be generalized also in case of model in (3.1) and (3.2), as explained in [Simon, 2006].

3.1.4 Non-Gaussian parametrization: the Tikhonov distribution

In the general case, when the noise processes are non-Gaussian and/or the model is non-linear, exact Bayesian tracking can be difficult and one usually makes use of the approximated tracking. The quality of the approximation depends on the fit between the model and the actual probabilities. To impose a parametric form to the predictive and posterior distributions of the two step Bayesian recursion means to implement a parametric tracking. It is shown in the previous Subsection that, when the predictive and posterior distribution are or are approximated Gaussian, the parametric tracking can be performed by Kalman filtering or extended Kalman filtering, respectively. Conversely, when the predictive and posterior are and are imposed non-Gaussian, in some way we need to a non-Gaussian version of the Kalman filter. There are a lot of possible non-Gaussian parametric parametrizations and they are usually tighly connected to the specifific tracking problem. In the following the Tikhonov parametrization is described. This parametric tracking technique is usually employed for phase tracking problems.

The idea is to model the predictive and posterior distributions of the continuous state as Tikhonov distributions (also known as Von Mises distributions) with probability density function equal to

$$t(z, a) = \frac{e^{\mathrm{Re}\left(a\, e^{-jz}\right)}}{2\pi I_0(|a|)}$$

where a is the complex parameter, $z \in [0, 2\pi)$ is the domain, and $I_0(\cdot)$ is the zero-th order modified Bessel function of first kind. $\angle a$ is the mean,

while $|a|$ is a measure of concentration (a reciprocal measure of the dispersion). Some example of this distribution are drawn in Fig. 3.9. In this case,

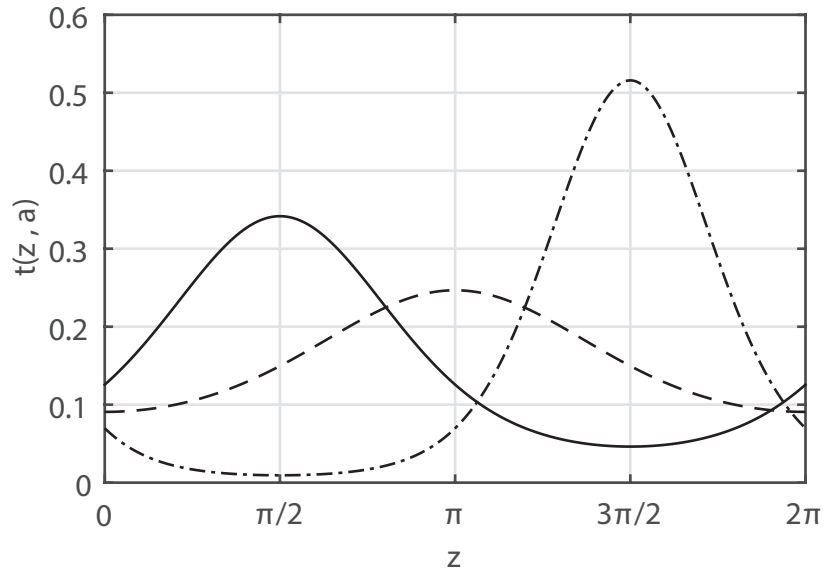


Figure 3.9: *Tikhonov distribution for different values of the parameter a. Solid line:* $a = \exp(j\pi/2)$. *Dashed line:* $a = 0.5\exp(j\pi)$. *Dash-dotted line:* $a = 2\exp(j3\pi/2)$

the two-step Bayesian recursion becomes the succession of prediction and update operations of a evolving complex parameter a. Defining with

$$p(s_k|y_1^{k-1}) = t(s_k, \overline{a}_k) \tag{3.62}$$

$$p(s_k|y_1^{k}) = t(s_k, a_k) \tag{3.63}$$

the predictive and posterior distributions, respectively, the goal is to find a recursive law between $\overline{a}_k$ and a_k. This law depends on the system and measurement model. If the real predictive and posterior distributions are Tikhonov distributions, the parametric Bayesian tracking is exact, otherwise is approximated. The following example shows a possible application.

Example: Tikhonov's phase tracking of a sinusoid embedded in noise affected by Wiener's phase noise

We refer to the usual example of the sinusoid embedded in noise affected by Wiener's phase noise, where the state is equal to the wrapped phase. A similar problem has been addressed in [Pecorino et al., 2015], where the Tikhonov approximation has been employed to track the phase in a

Wiener's phase noise channel, inside an iterative demodulation and decoding scheme. The non-linear state transition equation and the linear measurement equation are reported in (3.46) and (3.47). The measurement probability can be written as

$$\begin{aligned} p(y_k|s_k) &= g_c(e^{js_k}, \text{SNR}^{-1}; y_k) \\ &\propto \exp\left\{2\ \text{SNR}\ \text{Re}\{y_k e^{-js_k}\}\right\}. \end{aligned} \tag{3.64}$$

By the same approximation reported in [Colavolpe et al., 2005], when the phase of the sinusoid is slowing variant ($\sigma \to 0$), the state transition probability can be approximated by

$$p(s_k|s_{k-1}) = \delta(s_k - s_{k-1}).$$

Substituting this transition probability and the posterior distribution (3.63) at time $k-1$ in the predict step (3.26) one has

$$\overline{a}_k = a_{k-1}. \tag{3.65}$$

Then, the predictive distribution is equal to the previous posterior distribution. When the approximation of slowing variant phase does not hold, the authors of [Colavolpe et al., 2005] have proposed only a scaling of the posterior parameter:

$$\overline{a}_k = \frac{a_{k-1}}{1 + \sigma^2|a_{k-1}|}. \tag{3.66}$$

Replacing the predictive distribution (3.62) and the measurement probability (3.64) in the update step (3.27), one can write

$$a_k = \overline{a}_k + 2\ \text{SNR} y_k.$$

Here parametric tracking as two-step Bayesian recursion has been presented, while in [Colavolpe et al., 2005] the authors reports an one-step formulation, that results equal to

$$\overline{a}_k = \frac{\overline{a}_{k-1} + 2\ \text{SNR} y_{k-1}}{1 + \sigma^2|\overline{a}_{k-1} + 2\ \text{SNR} y_{k-1}|}.$$

Figure 3.10 shows an example parametric Bayesian tracking by Tikhonov approximation with the employment in the predictive step of both the slowing variant phase assumption (3.65) and the scaling operation in (3.66). The initial state is unknown, then $p(s_0) = (2\pi)^{-1}$, that corresponds to bootstrap the recursion with $a_0 = \overline{a}_1 = 0$. Therefore, the predictive distribution of

s_1 remains uniform as one can see from the Figure. Moreover, the tracking with the slowing variant phase assumption has been outperformed by the tracking with the scaling operation in the predictive step. As in this example, this is observable when the phase noise is consistent, i.e. σ is high. When σ is low, the two algorithms have the same performance. Particularly, if $\sigma \to 0$, Equation (3.66) approaches the (3.65). One can conclude that the Tikhonov parametrization allows an approximated Bayesian tracking close to the exact Bayesian tracking for the problem of a sinusoid embedded in noise.

As for the application studied in [Colavolpe et al., 2005], if one is interested to the forward-backward Bayesian tracking, the merging of the forward and the backward distribution is trivial thanks to the Tikhonov parametrization. Using Equation (3.38) and denoting with $\overline{a}_k^f$ and a_k^f the predictive and posterior parameters in the forward direction and with $\overline{a}_k^b$ and a_k^b the predictive and posterior parameters in the backward direction, one can get the parametric forward-backward tracking algorithm by Tikhonov approximation:

$$\begin{aligned}
\overline{a}_k^f &= \frac{a_{k-1}^f}{1+\sigma^2|a_{k-1}^f|} \\
a_k^f &= \overline{a}_k^f + 2\ \text{SNR} y_k \\
\overline{a}_k^b &= \frac{a_{k+1}^b}{1+\sigma^2|a_{k+1}^b|} \\
a_k^b &= \overline{a}_k^b + 2\ \text{SNR} y_k \\
p(s_k|y_1^n) &\simeq t(s_k, \overline{a}_k^f + a_k^b)
\end{aligned}$$

It is sufficient to sum the forward predictive parameter with the backward posterior one or the forward posterior parameter with the backward predictive one, to get the parameter of the Tikhonov approximation of $p(s_k|y_1^n)$.

3.1.5 Non-parametric tracking: State-space quantization

Taking into account the case of continuous state and non-Gaussian noise processes and/or non-linear model, when one has no idea about the shape of the predictive and posterior distributions, the state-space quantization techniques is a practical solution to the state tracking problem. The same algorithm shown in this Subsection, is employable if the state process is discrete. In this case, the Bayesian tracking can be exact, while in the continuous state case the Bayesian tracking is approximated because it certainly

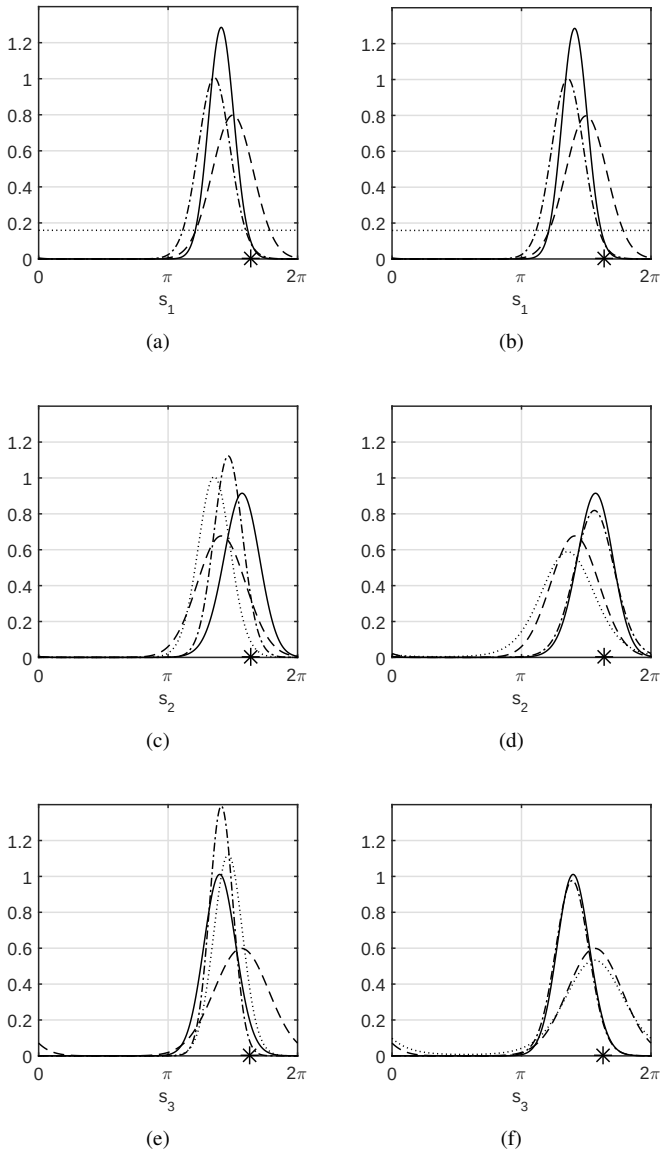


Figure 3.10: *Examples of parametric Bayesian tracking by Tikhonov approximation for the phase tracking problem of a sinusoid affected by Wiener's phase noise.* $\sigma = 0.5$*,* $SNR = 1dB$*,* $p(s_0) = (2\pi)^{-1}$*. Solid line: actual posterior distribution* $p(s_k|y_1^k)$*. Dashed line: actual predictive distribution* $p(s_k|y_1^{k-1})$*. Dash-dotted line: approximated posterior distribution* $t(s_k, a_k)$ *by Tikhonov parametrization. Dotted line: approximated predictive* $t(s_k, \overline{a}_k)$ *distribution by Tikhonov parametrization. Asterisk: actual value of the phase* s_k*. (a)* $k = 1$*, Tikhonov tracking with prediction rule in (3.65). (b)* $k = 1$*, Tikhonov tracking with prediction rule in (3.66). (c)* $k = 2$*, Tikhonov tracking with prediction rule in (3.65). (d)* $k = 2$*, Tikhonov tracking with prediction rule in (3.66). (e)* $k = 3$*, Tikhonov tracking with prediction rule in (3.65). (f)* $k = 3$*, Tikhonov tracking with prediction rule in (3.66).*

exploits the quantization approximation of the state. Focusing on continuous state processes, the main idea is to have the state-space divided into a limited number of bins with equal or different dimensions. These bins are the centroids of the histogram approximation carried out on the state probabilities density functions. The discrete or discretized space allows to define the finite state model, that is mapped onto a Finite State Machine (FSM), whose time evolution is described by the so-called *trellis* diagram. A comprehensive book about the FSMs is [Minsky, 1967]. In Figure 3.11 an example of FSM with four states and related trellis diagram is shown.

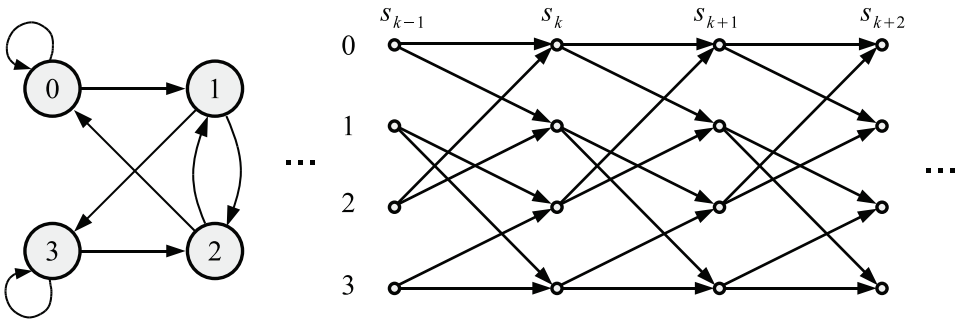


Figure 3.11: *Example of state diagram (left) and trellis digram (right) of a finite state machine.*

In these subsection $\{T\}$ is denoted as the discretized state process. The same equations can be adopted when the state process $\{S\}$ is discrete, imposing $\{S\} = \{T\}$. The trellis structure generally tracks the posterior probability mass function $p(t_k|y_1^k)$. In this context the one-step Bayesian recursion in (3.32) is commonly performed. The probabilities $p(t_k|y_1^k)$ for each possible value of t_k are associated to each state of the trellis at time k and are called *state metric*. The joint conditional probabilities

$$p(t_k, y_k|t_{k-1}) = p(y_k|t_k)p(t_k|t_{k-1})$$

are associated to the transitions between the states at time $k-1$ and states at time k in the trellis and are called *transition metrics*. Starting from the discrete version of (3.32), the state metric for each possible t_k is computed as

$$p(t_k|y_1^k) \propto \sum_{t_{k-1} \in \mathcal{S}_{k-1}} p(t_k, y_k|t_{k-1})p(t_{k-1}|y_1^{k-1}), \tag{3.67}$$

that is the sum of the transition metrics of each incoming arcs multiplied by the state metrics associated to the previous states connected to arcs' tails.

In addition to the one-step formulation, the trellis can separately tracks the predictive and posterior distributions of the states (the discrete version of (3.26) and (3.27)) and also it can perform the forward and backward recursion (the discrete version of (3.38)). In the classical applications, since the trellis structure tracks the probabilities distribution for all the possible state values, it is sufficient to calculates a proportional values of the wanted probabilities.

The main limit of this approach, either if the state is discrete by nature or it is quantized, is the number of state bins. Many bins increase the computational complexity. However, one cannot reduce at will the bins number. When the state is continuous the bins may be insufficient to have a good approximation of the actual probabilities, specially when the state is multi-dimensional.

Example: State-space quantization technique applied to the phase tracking of a sinusoid embedded in noise affected by Wiener's phase noise

Referring to the already known example of the sinusoid embedded in noise affected by Wiener's phase noise, the wrapped phase is considered as state. The non-linear state transition equation and the linear measurement Equations are reported in (3.46) and (3.47). To cope with the continuous-state model of the Wiener's phase noise, 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 phase evolution in the discrete-state space, where trellis' states are the centroids of the bins, as done in [Barletta et al., 2012a, Spalvieri and Barletta, 2011, Pecorino et al., 2015, Barletta et al., 2011].

The continuous phase S is discretized into $|\mathcal{S}|$ bins. Then the discrete state T, takes its values in the set of bins' centroids:

$$\mathcal{S} = \{\Delta, 3\Delta, 5\Delta \cdots, (2|\mathcal{S}| - 1)\Delta\}, \tag{3.68}$$

where

$$\Delta = \frac{\pi}{|\mathcal{S}|} \tag{3.69}$$

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

approximated by

$$p(s_k|s_{k-1}) \simeq p(t_k|t_{k-1}) = \int_{\mathcal{R}(t_{k-1}^k)} \frac{p(\phi_k|\phi_{k-1})}{\Delta} \,\mathrm{d}\phi_s \,\mathrm{d}\phi_{s-1}, \tag{3.70}$$

where $\mathcal{R}(t_{k-1}^k)$ indicates the two-dimensional quantization region whose centroid is t_{k-1}^k. Since $p(t_k|t_{k-1})$ depends only on the difference $t_k - t_{k-1}$, it takes its values in a set of $|\mathcal{S}|$ numbers that can be pre-computed by Equation (3.70). The probability $p(y_k|t_k)$ is similar to (3.25) with the centroid t_k in place of the continuous phase. All the needed information to compute the phase tracking in (3.67) by mean of the trellis structure is available. An example of application is reported in Figure 3.12, where the wrapped phase is discretized in $|T| = 8$ bins, the initial phase is supposed know, SNR $= 1dB$, and $\sigma = 0.3$. In the Figure one can see an example of the forward posterior distribution approximated by the trellis computation. In addition the related broken line, that depicts the MAP estimation, is drawn.

3.1.6 Particle filtering

Particle filtering is among the most popular and effective techniques to evaluate non-parametric approximations to the posterior and predictive probabilities density functions of a continuous state. It can be also applied to the case of discrete state: when the complete analysis of the state space by a trellis structure is infeasible because the number of possible state is too high, the particle filter could be a solution to track the state evolution. This technique is well resumed in [Arulampalam et al., 2002]. Other resources are the papers [Carpenter et al., 1999, Pitt and Shephard, 1999, Nummiaro et al., 2003, Van Der Merwe et al., 2000] and the comprehensive book [Smith et al., 2013].
The term *particle filter* was first coined in 1996 by Del Moral in the paper [Del Moral, 1996] in reference to mean field interacting particle methods used in fluid mechanics since the beginning of the 1960s. This technique is a sequential importance sampling algorithm based on a Monte-Carlo method. It merges the Monte-Carlo sampling concept with the importance sampling one, applying them to the Bayesian tracking framework.

The basic concept of the *Monte-Carlo sampling* is to approximate a distribution by considering it as sum of Dirac delta functions

$$p(x) \simeq \sum_{i=1}^{P} w^{(i)} \delta(x - x^{(i)}), \tag{3.71}$$

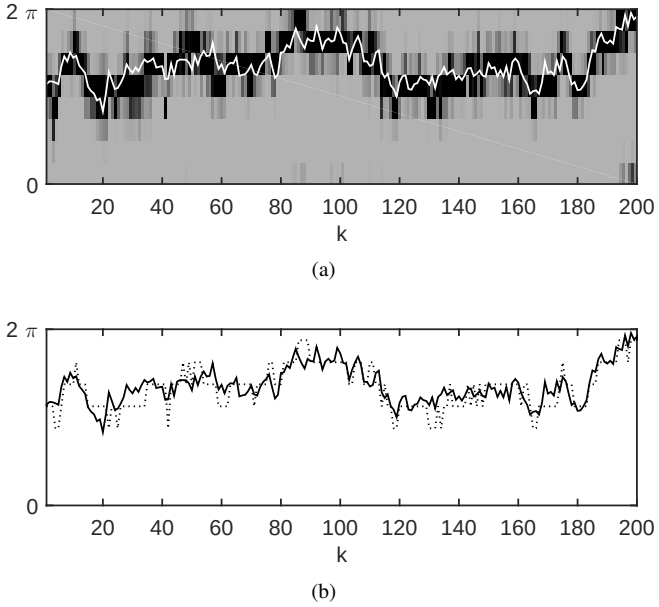


Figure 3.12: *Example of forward Bayesian recursion by state-space quantization applied to the tracking of the wrapped phase of the sinusoid embedded in noise affected by Wiener's phase noise. SNR* $= 1dB$*,* $\sigma = 0.3$*, and* $|\mathcal{S}| = 8$*.* $p(s_0)$ *is a Dirac function. (a) Grayscale image: forward posterior distribution* $p(s_k|y_1^k)$ *approximated by state-space quantization; solid white line: actual phase evolution. (b) Dotted line: MAP estimation from the forward posterior distribution* $p(s_k|y_1^k)$ *approximated by state-space quantization; solid black line: actual phase evolution.*

where $w^{(i)} = P^{-1}$, and $\{x^{(i)}\}$, with

$$x^{(i)} \sim p(x), \qquad i = 1, 2, \cdots P,$$

is the set of samples taken with probability $p(x)$. The main property of (3.71) is that it leads to straightforward computation of expected values:

$$\begin{aligned} E_p\{f(x)\} &= \int_{\mathcal{X}} f(x)p(x)dx \\ &\simeq \sum_{i=1}^{n} w^{(i)} f(x^{(i)}). \end{aligned} \tag{3.72}$$

Importance sampling is based on the idea of generating samples from the *importance* distribution $q(x)$, with

$$x^{(i)} \sim q(x) \qquad i = 1, 2, \cdots P.$$

We can assign weights according to

$$w^{(i)} \propto \frac{p(x^{(i)})}{q(x^{(i)})}, \qquad i = 1, 2, \cdots P,$$

and, after an operation of normalization to get

$$\sum_{i=1}^{P} w^{(i)} = 1,$$

we can use the same equation (3.71) and (3.72) to approximate the distribution $p(x)$ and theirs expected values, respectively. When $q(x) = p(x)$ one has the basic Monte-Carlo sampling, where all the weights are equal.

In the *sequential important sampling* the samples evolve in time. The idea is to track the joint posterior distribution of the state sequence $p(s_0^k|y_1^k)$, updating weights and samples. In this context we are talking about the particle filtering method, where the samples are called *particles*. Let P be the number of particles, $s_k^{(i)}$ the state visited by the i-th particle at time k, and $w_k^{(i)}$ the weight of the i-th particle at time k. The particle $s_k^{(i)}$ is generated from $s_{k-1}^{(i)}$, which is generated from $s_{k-2}^{(i)}$, and so on. The particles sequences give a Monte-Carlo importance sampling of the joint posterior state distribution, as

$$p(s_0^k|y_1^k) \simeq \sum_{i=1}^{P} w_k^{(i)} \delta(s_0^k - s_0^{k,(i)}),$$

where the weight $w_k^{(i)}$ carries the entire history of the i-th particle

$$w_k^{(i)} \propto \frac{p(s_0^{k,(i)}|y_1^k)}{q(s_0^{k,(i)}|y_1^k)} \qquad i = 1, 2, \cdots P. \tag{3.73}$$

The importance density $q(s_0^k|y_1^k)$ is up to the user. If it is chosen as product of two importance densities

$$q(s_0^k|y_1^k) = q(s_k|s_0^{k-1}, y_1^k)q(s_0^{k-1}|y_1^{k-1}), \tag{3.74}$$

one can get the samples of $s_0^{k,(i)} \sim q(s_0^k|y_1^k)$ by augmenting each of existing samples of $s_0^{k-1,(i)} \sim q(s_0^{k-1}|y_1^{k-1})$ with the new samples $s_k^{(i)} \sim q(s_k|s_0^{k-1}, y_1^k)$. Using (3.74) and

$$\begin{aligned} p(s_0^k|y_1^k) &\propto p(y_k|s_0^k, y_1^{k-1})p(s_0^k|y_1^{k-1}) \\ &= p(y_k|s_0^k, y_1^{k-1})p(s_k|s_{k-1})p(s_0^{k-1}|y_1^{k-1}), \end{aligned} \tag{3.75}$$

where the Bayes rule was exploited, also the weights of the particles in (3.73) can be calculated by iteration

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(y_k|s_0^{k,(i)}, y_1^{k-1})p(s_k^{(i)}|s_{k-1}^{(i)})p(s_0^{k-1,(i)}|y_1^{k-1}))}{q(s_k^{(i)}|s_0^{k-1,(i)}, y_1^k)q(s_0^{k-1,(i)}|y_1^{k-1})} \\ &= \frac{p(y_k|s_k^{(i)})p(s_k^{(i)}|s_{k-1}^{(i)})}{q(s_k^{(i)}|s_0^{k-1,(i)}, y_1^k)} w_{k-1}^{(i)} \qquad i = 1, 2, \cdots P, \end{aligned} \tag{3.76}$$

where the Markovian property in (3.8) is employed. In practical case the importance density $q(s_k|s_0^{k-1}, y_1^k)$ is often chosen only dependent on the current measurement and the previous state, as

$$q(s_k|s_0^{k-1}, y_1^k) = q(s_k|s_{k-1}, y_k)$$

In this way there is no need to save the history of the particles and the measurements. If the importance density is chosen equal to the state transition probabilities as

$$q(s_k|s_0^{k-1}, y_1^k) = p(s_k|s_{k-1}),$$

the weights update is simplified in

$$w_k^{(i)} = \alpha_k p(y_k|s_k^{(i)})w_{k-1}^{(i)} \qquad i = 1, 2, \cdots P, \tag{3.77}$$

where

$$\alpha_k = \left(\sum_{j=1}^{P} p(y_k|s_k^{(j)}) w_{k-1}^{(j)} \right)^{-1} \tag{3.78}$$

is the normalization factor. This choice is often adopted. Starting from an initial set of weights $\{w_0^{(i)},\ i = 1, 2, \cdots P\}$ and from an initial set of particles $\{s_0^{(i)},\ i = 1, 2, \cdots P\}$ the particle filtering alternates the generation of the new particles and the update of theirs weights.

The two-step Bayesian recursion tracks the predictive $p(s_k|y_1^{k-1})$ and posterior $p(s_k|y_1^k)$ distributions, while the particle filtering describes the evolution of the joint distribution $p(s_0^k|y_1^k)$. The posterior distribution is trivially approximated as

$$p(s_k|y_1^k) \approx \sum_{i=1}^{P} w_k^{(i)} \delta(s_k - s_k^{(i)}), \tag{3.79}$$

therefore the weights for the joint distribution $p(s_0^k|y_1^k)$ remains the weights of $p(s_k|y_1^k)$.

Proof. By integration one has

$$\begin{aligned} p(s_k|y_1^k) &= \int_{\mathcal{S}_0^{k-1}} p(s_0^k|y_1^k)\, \mathrm{d}s_0^{k-1} \\ &\approx \int_{\mathcal{S}_0^{k-1}} \sum_{i=1}^{P} w_k^{(i)} \delta(s_0^k - s_0^{k,(i)})\, \mathrm{d}s_0^{k-1} \end{aligned}$$

Using the definition

$$\delta(s_0^k - s_0^{k,(i)}) = \prod_{j=0}^{k} \delta(s_j - s_j^{(i)})$$

one can write

$$\begin{aligned} p(s_k|y_1^k) &\approx \int_{\mathcal{S}_0^{k-1}} \left[\sum_{i=1}^{P} w_k^{(i)} \prod_{j=0}^{k} \delta(s_j - s_j^{(i)}) \right] \mathrm{d}s_0^{k-1} \\ &= \sum_{i=1}^{P} w_k^{(i)} \delta(s_k - s_k^{(i)}) \prod_{j=0}^{k-1} \int_{\mathcal{S}_j} \delta(s_j - s_j^{(i)})\, \mathrm{d}s_j \\ &= \sum_{i=1}^{P} w_k^{(i)} \delta(s_k - s_k^{(i)}), \end{aligned}$$

where one exploit the property

$$\prod_{j=0}^{k-1} \int_{\mathcal{S}_j} \delta(s_j - s_j^{(i)})\, \mathrm{d}s_j = 1.$$

□

Using the general definition in (3.76), the predictive distribution can be obtained as

$$p(s_k|y_1^{k-1}) \propto \sum_{i=1}^{P} \frac{p(s_k^{(i)}|s_{k-1}^{(i)})}{q(s_k^{(i)}|s_0^{k-1,(i)}, y_1^k)} w_{k-1}^{(i)} \delta(s_k - s_k^{(i)}). \tag{3.80}$$

Proof. By Bayes rule one has

$$p(s_k|y_1^k) \propto p(y_k|s_k) p(s_k|y_1^{k-1})$$

where the Markovian property (3.8) is employed. Making the inverse of the previous equation and using (3.76) and (3.79), the predictive distribution becomes

$$\begin{aligned} p(s_k|y_1^{k-1}) &\propto \frac{p(s_k|y_1^k)}{p(y_k|s_k)} \\ &\simeq \sum_{i=1}^{P} \frac{w_k^{(i)}}{p(y_k|s_k^{(i)})} \delta(s_k - s_k^{(i)}) \\ &\propto \sum_{i=1}^{P} \frac{p(s_k^{(i)}|s_{k-1}^{(i)})}{q(s_k^{(i)}|s_0^{k-1,(i)}, y_1^k)} w_{k-1}^{(i)} \delta(s_k - s_k^{(i)}), \end{aligned}$$

□

In case of importance density equal to the state transition probabilities and then adopting the weights update rule in (3.77), the equation (3.80) becomes

$$p(s_k|y_1^{k-1}) \simeq \sum_{i=1}^{P} w_{k-1}^{(i)} \delta(s_k - s_k^{(i)}), \tag{3.81}$$

In this case, the predictive step of the Bayesian recursion is computed by only the Monte-Carlo update of the particles as $s_k^{(i)} \sim p(s_k|s_{k-1}^{(i)})$, while the update step of the Bayesian recursion is performed updating the weights with (3.77). Taking into account the entire particles sequence, the samples

are generates by $s_0^{k,(i)} \sim p(s_0^k)$, while the weight carries the entire history of the particles as

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(s_0^{k,(i)}|y_1^k)}{p(s_0^{k,(i)})} \\ &\propto p(y_1^k|s_0^{k,(i)}). \end{aligned}$$

In this thesis work, the state transition equation will be always exploited to generate the new particles. Making explicit the system model in (3.1), the particles are generated with

$$s_k^{(i)} = f_{k-1}(s_{k-1}^{(i)}, v_{k-1}^{(i)}), \tag{3.82}$$

where $\{v_{k-1}^{(i)},\ i = 1, 2, \cdots P\}$ is a set of independent samples of process noise.
After updating the particles with (3.82), a *resampling* procedure may be necessary to prevent particles from collapsing onto one particle of unitary weight. For every particle filtering implementation in the thesis, the classical resampling algorithm is adopted, as described in [Arulampalam et al., 2002]. The minimum *effective sample size* has been chosen, as called in [Arulampalam et al., 2002], equal to the typical value $0.3P$. The effective sample size equal to $\sum_{i=1}^{P}(w_k^{(i)})^2$ is an estimation of particles number with weight not negligible. The algorithm 1 summarizes the steps of the particle filtering, with the choice of importance function equal to the state transition probability and resempling procedure as in [Arulampalam et al., 2002].

Example: Particle phase tracking of a sinusoid embedded in noise affected by ARMA phase noise

This example considers the sinusoid embedded in noise in case of ARMA phase noise presented in Subsection 3.1.1. The non-linear state transitionn and the linear measurement Equations are reported in (3.15) and (3.17). Here the problem of tracking the phase of the sinusoid cannot worked out with the state-space quantization techniques, because the state-space is not confined and its dimension $m + 1$ is high. Specifically, working out the phase tracking of the sinusoid affecting by ARMA phase noise is a challenging problem, because

- the state space is not finite and it is multidimensional, therefore it cannot be approached by trellis-based techniques shown in the Subsection

Algorithm 1 Particle filtering

Generate $(s_0^n, y_1^n) \sim p(s_0^n, y_1^n) = \delta(s_0) \prod_{k=1}^{n} p(s_k|s_{k-1})p(y_k|s_k)$
Generate $s_0^{(i)} \sim p(s_0)$ for $i = 1, \ldots, P$
$w_0^{(i)} \leftarrow P^{-1}$ for $i = 1, \ldots, P$
for $k = 1, \ldots, n$ **do**
  Generate $v_{k-1}^{(i)} \sim p(v_{k-1})$ for $i = 1, \ldots, P$
  $s_k^{(i)} \leftarrow f_k(s_{k-1}^{(i)}, v_{k-1}^{(i)})$ for $i = 1, \ldots, P$
  $w_k^{(i)} \leftarrow w_{k-1}^{(i)} p(y_k|s_k^{(i)})$ for $i = 1, \ldots, P$
  $\alpha_k \leftarrow \sum_{i=1}^{P} w_k^{(i)}$
  $w_k^{(i)} \leftarrow w_k^{(i)} / \alpha_k$ for $i = 1, \ldots, P$
  if $\sum_{i=1}^{P} (w_k^{(i)})^2 > (0.3P)^{-1}$ **then**
    $\left(\{s_k^{(i)}\}, \{w_k^{(i)}\}\right) \leftarrow$ resample $\left(\{s_k^{(i)}\}, \{w_k^{(i)}\}\right)$
  end if
end for

3.1.5 based on quantization of the state space, because the number of states of the trellis would be enormous, and

- the measurement is a nonlinear function of the state, therefore the Kalman filter shown in Subsection 3.1.3 can not be used and the extended Kalman filter can be far from being optimal.

The sampling by particle filtering can be a possible solution: in Figure 3.13 an example of application is presented, with SNR $= 0$ dB and ARMA phase noise with $\sigma = 1$ generated by (3.21) and $m = 4$. The number of particles is $P = 1000$. The forward Bayesian recursion for four time steps is shown. When the state is multidimensional (as in this case that is 5-dimensional) the predictive and posterior distribution can not be graphically pictured. In this case is interesting to look the evolving distribution of the phase ϕ_k that is the first entry of the state. The particle filter approximates the probability density functions as sum of Dirac functions, which are not drawable. In this case one can use the *Parzen method* for a kernel density estimation of the probability density functions [Silverman, 1986]. Easily, the Dirac function can be substituted with a chosen kernel function centered in the position of Dirac pulse. The use of a Gaussian function as kernel is widely employed. Therefore the posterior distribution of the first entry of the state

can be approximated as

$$p(\phi_k|y_1^k) \simeq \sum_{i=1}^{P} w_k^{(i)} g(\phi_k^{(i)}, \sigma_g^2; s_k), \tag{3.83}$$

where $\phi_k^{(i)}$ is the first entry of the i-th particle and σ_g^2 is up to the user. In the example of Figure 3.13 σ_g^2 is equal to 0.1. The Parzen method can be also used in the applications where the evaluation of the posterior and/or predictive distribution in all the state domain is required.

As it is shown in Figure 3.13, the particle method is able to track the sinusoid phase also in presence of strong phase noise and low SNR. Iteration by iteration the particles increase the explored region allowing the appearance of secondary lobes (that are far multiples of 2π from the main lobe): this is the effect of the cycle slip, observable at high σ.

3.2 Non linear Optical Channel - An example of Bayesian Tracking Limits

In Section 3.1, the State-based Approach is presented together with all techniques to track an hidden state channel. However, this powerful technique has its own weaknesses. Several parameters make the problem dramatically scale. If one uses particle filtering to track the probabilities of the Channel-state, the parameter that he must take into account if he wants to analyze the complexity is the order of the Markovian state. In this Section the theoretical formulation of Bayesian Tracking is presented in order to model the propagation of a signal over an optical fiber. All possible approximations are worked out in order to cast the problem onto the State-based Approach defined previously.
In the end the limits of the described method are analyzed in an inductive way, taking this Section as a practical example of the State-based Approach leaks.

3.2.1 Non Linear Optical Channel

In this Section it is introduced also the Kerr non-linear effect of optic propagation given by the cubic electrical field term in the Equation below [Essiambre et al., 2010]

$$\frac{\partial E}{\partial L} + \frac{\alpha}{2}E + j\beta_2\frac{\partial^2 E}{\partial t^2} - j\gamma|E|^2E = iN(L,t)\ , \tag{3.84}$$

where the terms

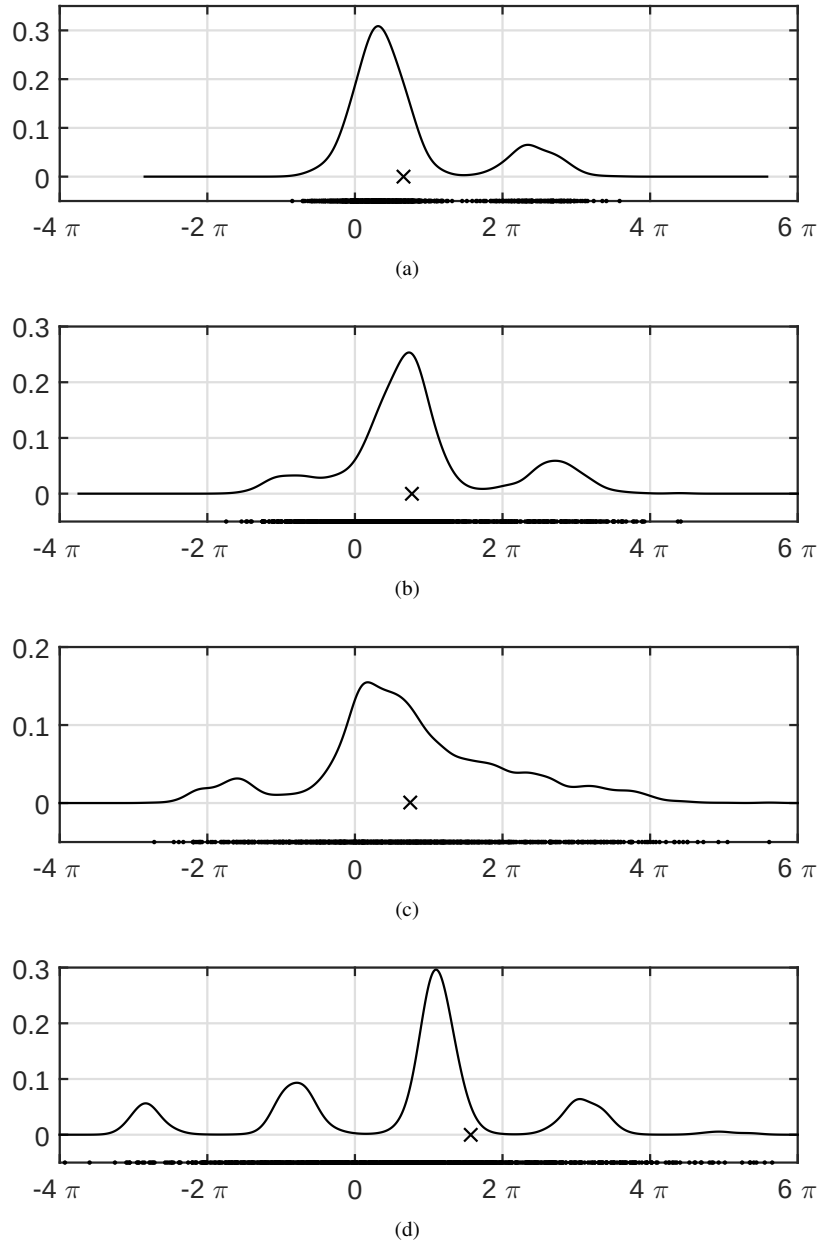


Figure 3.13: *Example of forward Bayesian recursion by particle filtering applied to the tracking of the unwrapped phase of the sinusoid embedded in noise affected by ARMA phase noise. SNR* $= 0dB$*, ARMA phase noise generated by (3.21) with* $m = 4$ *and* $\sigma = 1$*, and* $p(\phi_0)$ *uniform between* $[0, 2\pi)$*. Number of particles* $P = 1000$*. Solid line: posterior distribution* $p(\phi_k|y_1^k)$ *by formula (3.83) with* $\sigma_p^2 = 0.1$*. Crosses: actual phase. Dots on the* x*-axis: values of the first entry of the particles* ($\phi_k^{(i)}$) *(a)* $k = 1$*. (b)* $k = 2$*. (c)* $k = 3$*. (d)* $k = 4$*.*

- $E(L,t)$ is the electrical field that propagates over the fiber L kilometer length with time t,
- α is the fiber loss coefficient net of the fiber distributed gain,
- β_2 is the Group-Velocity Dispersion (GVD). Remember that the already presented Chromatic Dispersion parameter D is related to GVD as follows
$$D = -\frac{2\pi c}{\lambda^2}\beta_2 \ ,$$
- the fast change in the fiber medium refractive index that occurs in the presence of an intense electric field $E(L,t)$ is referred to as the instantaneous Kerr nonlinearity. Propagation of a signal field in the presence of loss, gain and instantaneous Kerr nonlinearity, with non linear parameter
$$\gamma = \frac{n_2\omega_s}{cA_{\text{eff}}} \ ,$$
where n_2 is fiber nonlinear refractive index, $\omega_s = 2\pi\nu_s$ is the angular frequency and A_{eff} the fiber effective area [Agrawal, 2007].
- $N(L,t)$ is the fiber noise generation phenomenon.

If one wants only to take into account the non-linearities, he can consider only a simplified version of the (3.84)

$$\frac{\partial E}{\partial L} + \frac{\alpha}{2}E - j\gamma|E|^2E = 0 \ , \tag{3.85}$$

that has exact solution in

$$E(L,t) = E(0,t)e^{-\alpha L/2}e^{j\phi_{NL}(z)} \ , \tag{3.86}$$

where the integrated nonlinear phase is defined as

$$\phi_{NL}(L) = \gamma|E(L,t)|^2L_{\text{eff}}(L) \text{ , with} \tag{3.87}$$

$$L_{\text{eff}}(L) = \frac{1-e^{-\alpha L}}{\alpha} \ . \tag{3.88}$$

However, the integrated nonlinear phase for an arbitrary signal power evolution is

$$\phi_{NL}(L) = \gamma\int_0^L P(L')dL' \ , \tag{3.89}$$

where the power is defined as $P(z) = |E(z)|^2$ [Agrawal, 2007]. This type of nonlinearity is called *Instantaneous Kerr Nonlinearity* and it is due to the

fast change in the fiber medium refractive index that occurs in the presence of an intense electric field [Essiambre et al., 2010]. The other phenomenon modeled in this Section is the Chromatic Dispersion (CD), which is due to the term with β_2 in Equation (3.85). Particularly, Optical fibers are made of fused silica, a material that exhibits inherent CD. Standard single-mode fibers (SSMFs) have a waveguide dispersion smaller than the material dispersion with a combined dispersion of $D \approx 17$ ps/(nm $\cdot$ Km). Independent of the origins of dispersion, the equation describing dispersive propagation in fibers, neglecting all other terms in (3.85), can be written as

$$\frac{\partial E}{\partial L} + j\beta_2 \frac{\partial^2 E}{\partial t^2} = 0 \ , \tag{3.90}$$

that has exact solution in the spectral domain

$$\tilde{E}(L,\omega) = \tilde{E}(L,0)e^{j\beta_2\omega^2 L} = \tilde{E}(L,0)e^{jD\lambda^2 f^2 L/c} = \tilde{E}(L,0)H_{\text{CD}}(L,f) \ , \tag{3.91}$$

where $\omega = 2\pi f$ and λ the operating wavelength in micron. As its name suggests, CD produces a spread in time of the various frequency components of a signal due to the difference in group velocity experienced by each frequency component. As CD accumulates, neighboring symbols start to overlap in time, with the number of symbols overlapping increasing with the accumulation of CD, see Chapter XXXX. Moreover, in terms of information theory, CD introduces memory to the channel [Essiambre et al., 2010].

3.2.2 State-based Approach

If one wants to model with the State-based Approach the effects of the optic channel reported in the previous Subsection, i.e. Cromatic Dispersion and the Nonlinear Phase, he can solve the considered fiber propagating Equation

$$\frac{\partial E}{\partial L} + j\beta_2 \frac{\partial^2 E}{\partial t^2} + \frac{\alpha}{2}E - j\gamma|E|^2 E = 0 \tag{3.92}$$

by dividing the problem into small propagation sections. If one recalls the generic State transition and measurement Equations (3.1) and (3.2)

$$\begin{aligned} S_k &= f_{k-1}(S_{k-1}, V_{k-1}), \\ Y_k &= h_k(S_k, N_k), \end{aligned}$$

he can fit up the model as it is done in Figure 3.14 for two fiber sections and explained for the general case below.

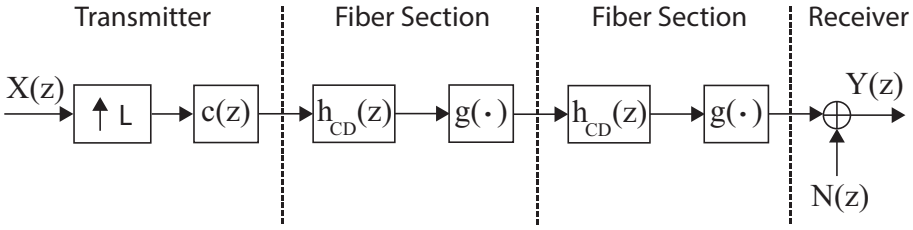


Figure 3.14: *Block Scheme of the State-based Approach of a two sections optic channel.*

- the time index $k = 1, 2, \cdots, KL$ can simulate the transmission of K symbols, padded with $L-1$ other samples. This is done to have a good approximation of the continuous-time when the signals' evolution are analyzed in a discrete-time manner.
- The two transition functions $f(\cdot)$ and $h(\cdot)$ are time-invariant.
- The process noise $\{V\}$ is modeled here as cyclostationary process built by zero-padding the modulation one $\{X\}$, leading to the Z-transform

$$V(z) = \sum_{i=1}^{K} X_i z^{-iL} \,, \tag{3.93}$$

  where the random variables X_i are the i.i.d. modulation symbols.
- The state transition equation $f(\cdot)$ is the cascade of a shaping transmit filter, e.g. square root raised cosine, and N sections that model nonlinear propagation in optical fiber.
- The shaping filter is characterized by the transfer function

$$c(z) = \frac{\prod_{l=1}^{m}(1-\beta_l z^{-1})}{\prod_{l=1}^{m}(1-\alpha_l z^{-1})} = \frac{1+\sum_{l=1}^{m} b_l z^{-l}}{1-\sum_{l=1}^{m} a_l z^{-l}} \,, \tag{3.94}$$

  where $|\alpha_l| < 1$ and $|\beta_l| \leq 1 \; \forall l$. Therefore the transfer function $c(z)$ is causal, monic, and minimum phase.
- The i-th fiber's section is the cascade of a filter followed by a nonlinear memoryless transformation whose output is the input to the next section. Given the theory of the previous Subsection 3.2.1, the frequency response of the filter taking into account the CD is

$$H_{\text{CD}}(L, f) = e^{-\alpha L/2} e^{-j2\pi D\lambda^2 f^2 L/c} \,, \tag{3.95}$$

where the two distinct contributes of the signal attenuation and CD are split. In order to better focus on fiber nonlinearities and CD, fiber attenuation α is set to zero. The next aim is to approximate Equation (3.95) with a linear transfer function in order to fit the state-based framework, i.e. have a Markovian state. Taking into account the dimensionality problem discussed in the previous Section, one limits the section transfer function to a 2-order one, in the following way

$$h_{\text{CD}}(z) = \frac{1 - \left(z_0^{-1}\right)^* z^{-1}}{1 - z_0 z^{-1}} \cdot \frac{1 - z_0^{-1} z^{-1}}{1 - z_0^* z^{-1}} \,. \tag{3.96}$$

Note that $h_{\text{CD}}(z)$ is all-pass with complex parameter z_0. Depending on $|z_0|$ and $\angle z_0$ one can control the phase transition slope and position respectively in the frequency domain.

- The memoryless nonlinearity applied after each fiber section transfer function is easily derived from Equations (3.86) and (3.87)

$$g(E(L,t)) = E(L,t) e^{j\gamma |E(L,t)|^2 L} \,, \tag{3.97}$$

where $L_{\text{eff}} = L$ given that $\alpha = 0$. Note that one must also determine L, that is the fiber sections' length.

3.2.3 Model Fitting

The remaining work is to find the Z-transfer functions $c(z)$, $h_{\text{CD}}(z)$. The method is explained below, while the discussion of weak points is reported later.

The transmit transfer function $c(z)$ approximates a SRRC filter with an Infinite Impulse Response (IIR) one. From Equation (3.94), $c(z)$ has order m. In Figure 3.15 are reported the transfer functions of an m-th order, $m = \{4, 5, 6\}$, that limit the state order to an acceptable value. Note that the Up-Sampling Factor USF $= 32$ let the sampling time being $T_S = T/32$. This expands the normalized frequency axes, such that the frequency cut in [π rad/sample] is $f_c = 1/32$. All the results of Figure 3.15 are obtained with apposite Matlab library (rcosine) that gives the best 2-norm IIR filter approximation for a SRRC filter, given its order m.

The second-order Chromatic Dispersion transfer function $h_{\text{CD}}(z)$ is reported in Figure 3.16. Note that $z_0 = 0.993 e^{j2\pi \cdot 0.5/32}$ is chosen such that the phase 2π slope is centered around the normalized frequency $f_c = 0.5/32$, where USF = 32. Note that the phase response in the frequency domain and its quadratic MMSE approximation only fits in the central frequencies, that

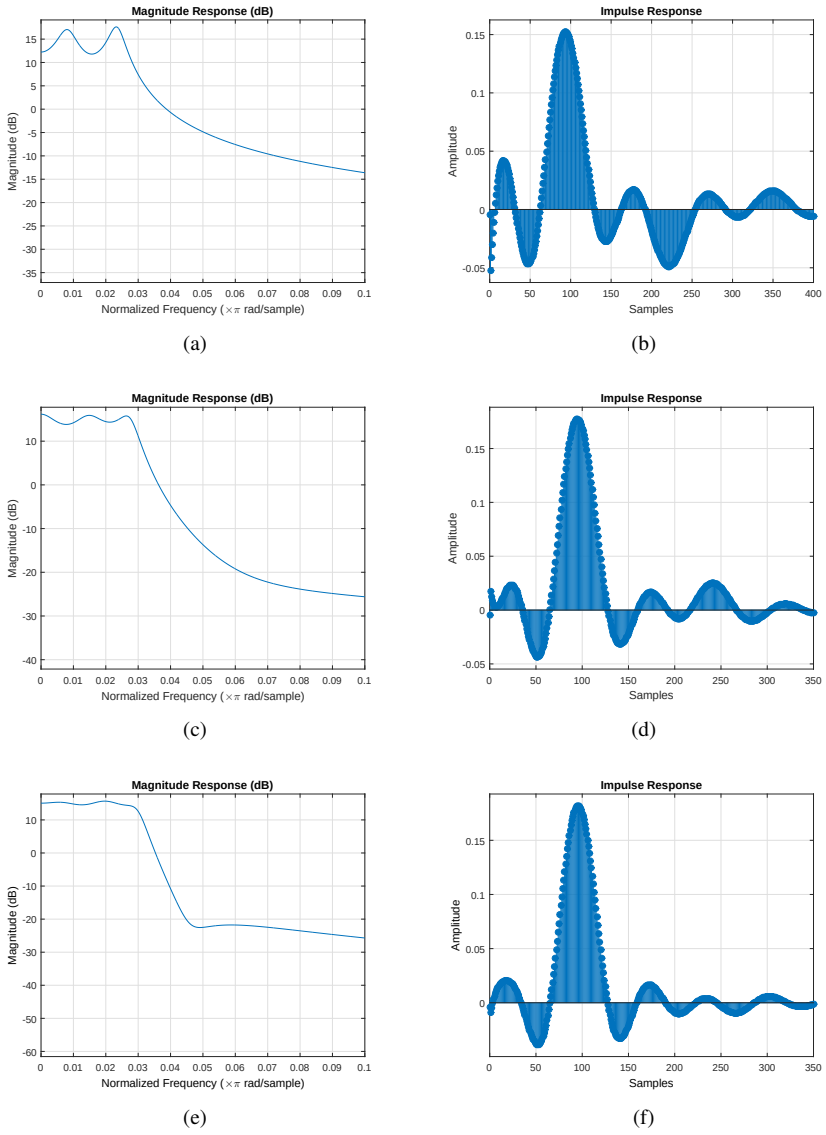


Figure 3.15: *Frequency and Impulse Responses of approximated Square-Root Raised Cosine by m-th order transfer functions.*
(a) $m = 4$ Impulse Response. (b) $m = 4$ Frequency Response.
(c) $m = 5$ Impulse Response. (d) $m = 5$ Frequency Response.
(e) $m = 6$ Impulse Response. (f) $m = 6$ Frequency Response.

are the ones spanned by the transmit filters. However, there is not a perfect fit of the approximated transfer function. From the quadratic coefficient of

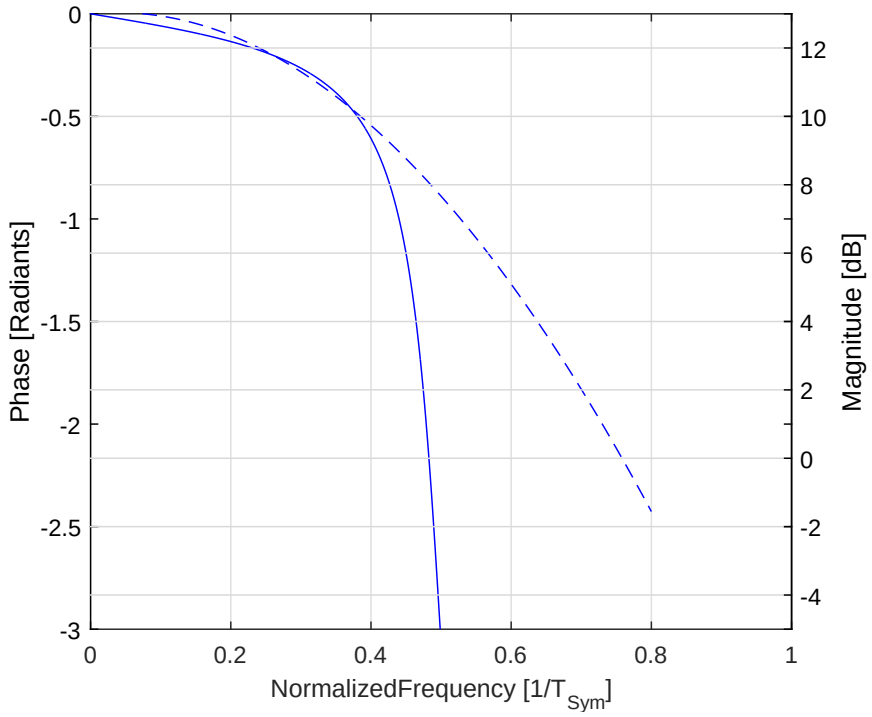


Figure 3.16: *Phase Response of $h_{CD}(z)$, with $z_0 = 0.993e^{j2\pi\cdot 0.5/32}$ (solid line) and its quadratic MMSE approximation (dashed).*

the parabolic approximation of the phase response in Figure 3.16 one can compute the modeled fiber sections' length L. Recalling the (3.95), and given D, λ, a fixed frequency of interest f, and the approximated quadratic coefficient $\overline{a}$

$$L = \frac{\overline{a}\, c}{2\pi D\lambda^2}\ . \tag{3.98}$$

Now, one must choose the number of sections and the order m of the transmit transfer function to set up the state-based approximated optical channel, therefore he can apply the particle filtering techniques described in this Chapter to derive information bounds and demodulating techniques that exploits all the available data to have an optimal estimate of both the channel state and the transmitted sequence. One should ask at this point why the actual Section has the second part of its name "An example of Bayesian Tracking Limits". This is due to the fact that Bayesian Tracking

applied to state-based problems works when one is simulating phenomenon with limited order transfer functions. However, nonlinear optic propagation is not so easy to be cast in such framework. Remember that the state-dimension plays a strong limiting role of application of such techniques. The simulating platform is given by a 2000 particles filter programmed with Matlab and run on an Intel Core i7-3770 Processor (8M Cache, up to 3.90 GHz) and a 16 GB memory. With such model, the state order that can be simulated in an acceptable time length is limited to be 11. Two parameters are linked to the state dimension and are reported below.

- In Figure 3.15 one can observe that the transmit filter is approximated by the m-th order transfer functions. Obviously m let the state increase accordingly.
- The number of fiber Sections F increases the state by $2F$ since at each fiber section the second-order transfer function $h_{\text{CD}}(z)$ is applied.

If one wants to resume this concept, he can write that the state dimension $|S|$ is given and limited by

$$|S| = m + 2F \leq 11 \ , \tag{3.99}$$

meaning that, if $m = 5$, the maximum number of fibers sections F is 3. If one wants to add also additive independent noise at each fiber Section

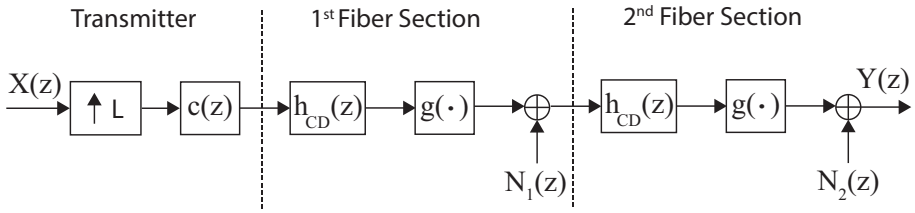


Figure 3.17: *Block Scheme of the State-based Approach of a two sections optic channel with additive noise at each fiber section.*

as schemed in Figure 3.17 he has to model the state including also the values of all the noise added except the last one, that does not limit the Markovian properties of the state. However all other noises $N_i(z)$, with $i = 1, \cdots, F-1$, since they pass through filters and they affect the channel state must be included in the state, modifying the constraint Equation (3.99) into

$$|S| = m + 3F - 1 \leq 11 \ . \tag{3.100}$$

Given Equation (3.100) and $m = 6$, one can only deal with $F = 2$ fiber sections.

The limits of the state-based approach applied to this difficult continuous-time channel are easily given by considering the trade-off between model complexity and fitting on one side and practical implementation on the other.

- With the proposed model without noise, only 3 fiber sections can be simulated, leading to a total fiber length less than one kilometer. The number becomes 2 if one wants also to take into account the optical noise introduced during the propagation in the fiber.
- With computational heavy techniques one can track the channel evolution and have optimal demodulators [Pecorino et al., 2015] that can transmit at the theoretical channel information rate with minimal rate loss. However, this loss increases as more the state-based channel is different from the true one. Consider the scenario of nonlinear optical channels presented in this Section. From Figures 3.15 and 3.16 it is trivial that the proposed approximation has not a perfect fit with the true optical channel and a SRRC transmit filter. From the first results the state-based simulated transmission behavior is so far from a true Split-Step Fourier Method [Sinkin et al., 2003] propagation that it is not worth to proceed in this work.
- Even if the model fitting to the physical behavior of the optical channel were perfect, there would still remain the problem of implementation of such method in practical optical demodulators, where Digital Signal Processing system must work at much high rates [Tamir et al., 2009]. Such Baud Rates can be achieved by expensive and big systems, like backbone and long haul optic transmission, where the fiber length consists in hundreds of kilometers, that is not our case. On the other hand, for short reach scenarios, it is not worthed to charge such systems with heavy computational capacity. Indeed, the gain of such techniques, as can be seen in the following Section, is still not high enough to let the performance gain over the economic, and dimension, loss given by such complexity.

CHAPTER 4

Bayesian Tracking in Wiener Phase Noise Channels

In Chapter 3 Bayesian Tracking applied to state-based approaches is explained. Both parametric and non-parametric techniques are proposed to track the hidden state of a channel, with focus onto phase processes. In this Chapter the knowledge given by this theory are exploited to better analyze the discrete Wiener phase noise channel, whose model, the DM, has been validated in Chapter 2. After a Section where Bayesian Tracking is exploited to compute the Information rate that can be transmitted over a state-based channel, the DM Information rate is computed in Section 4.2. In order to achieve such rate, one must perform Bayesian tracking of the channel state, i.e. phase noise. However, this is often computational expensive and requires Pilot Symbols, like the already cited [Barletta et al., 2013, Spalvieri and Barletta, 2011] and [Kamiya and Sasaki, 2013]. In the last Section of the Chapter it is proposed an Iterative demodulation and decoding algorithm that can bootstrap without Pilot Symbols and demodulate a signal transmitted over the DM channel. This algorithm has far less computational load than a conventional and complete tracking algorithm, but it can still reach the Information Rate and has negligible losses compared to

the competitors.

4.1 Information rate

4.1.1 Exact Information Rate

In this Section the concept of information rate is firstly introduced. Then it is described how Bayesian inference developed on auxiliary, and easier, channels can be exploited to derive information rate bounds and how to compute them by numerical simulation with particle filtering. The information between the state and the measurements can be calculated starting from the definition of the Shannon mutual information, that is

$$I(S;Y) = h(Y) - h(Y|S), \tag{4.1}$$
$$= h(S) - h(S|Y), \tag{4.2}$$

where $h(X) = -\mathsf{E}\left\{\log_2(p(x))\right\}$ is the entropy of the continuous or discrete random variable X, as applicable. Since in the state-based approach the process S and Y have memory, the information rate can be written as

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} I(S_1^n; Y_1^n), \tag{4.3}$$

that, starting from (4.1), becomes

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left\{\log_2\left(\frac{p(y_k|s_k)}{p(y_k|y_1^{k-1})}\right)\right\} \tag{4.4}$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(S_k; Y_k|Y_1^{k-1}). \tag{4.5}$$

Proof. Starting from (4.1) and using the chain rule and the equation (3.6) one has

$$\begin{aligned} I(S;Y) &= \lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left\{\log_2\left(\frac{p(y_1^n|s_1^n)}{p(y_1^n)}\right)\right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left\{\log_2\left(\frac{\prod_{k=1}^{n} p(y_k|s_k)}{\prod_{k=1}^{n} p(y_k|y_1^{k-1})}\right)\right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left\{\log_2\left(\frac{p(y_k|s_k)}{p(y_k|y_1^{k-1})}\right)\right\}. \end{aligned}$$

Exploiting the Markovian property in (3.8) one can replace $p(y_k|s_k)$ with $p(y_k|s_k, y_1^{k-1})$ and then conclude

$$\begin{aligned} I(S;Y) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left[h(Y_k|Y_1^{k-1}) - h(Y_k|S_k, Y_1^{k-1}) \right] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(S_k; Y_k|Y_1^{k-1}). \end{aligned}$$

□

From an computational point of view, if X is a stationary ergodic process, by the Shannon-McMillan-Breiman (SMB) theorem, one can generate a long sequence $\{x_k\}$ and compute the entropy

$$h(X) = -\lim_{n\to\infty} \frac{1}{n} \log_2 \left(p(x_k|x_1^{k-1}) \right),$$

without the mean operator. Therefore, from (4.4) we have

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log_2 \left(\frac{p(y_k|s_k)}{p(y_k|y_1^{k-1})} \right). \tag{4.6}$$

The numerator in the logarithm of (4.6) is derivable from the measurement equation in (3.2), while the denominator $p(y_k|y_1^{k-1})$ needs to Bayesian tracking. It can be calculated from the Chapman-Kolmogorov equation in (3.28). The Equation (4.5) shows that the k-th update step of Bayesian tracking extracts all the information about the state that is present in the k-th measurement given all the measurements up to time $k-1$.
Starting from (4.2), the information rate in (4.3) between the two studied process with memory becomes

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E} \left\{ \log_2 \left(\frac{p(s_k|s_{k-1}, y_k^n)}{p(s_k|s_{k-1})} \right) \right\} \tag{4.7}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(S_k; Y_k^n|S_{k-1}). \tag{4.8}$$

Proof. Starting from (4.2) and using the chain rule and the equation (3.5)

one has

$$\begin{aligned} I(S;Y) &= \lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left\{ \log_2 \left(\frac{p(s_0^n|y_1^n)}{p(s_0^n)} \right) \right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \mathsf{E}\left\{ \log_2 \left(\frac{p(s_0|y_1^n) \prod_{k=1}^n p(s_k|s_0^{k-1}, y_1^n)}{p(s_0) \prod_{k=1}^n p(s_k|s_{k-1})} \right) \right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathsf{E}\left\{ \log_2 \left(\frac{p(s_k|s_0^{k-1}, y_1^n)}{p(s_k|s_{k-1})} \right) \right\}, \end{aligned}$$

where the term $\log_2 (p(s_0)) - \log_2 (p(s_0|y_1^n))$ has been deleted because, considering the limit operation, it does not affect the sum. Thanks to the state-space approach

$$p(s_k|s_0^{k-1}, y_1^n) = p(s_k|s_{k-1}, y_k^n)$$

and then one conclude

$$\begin{aligned} I(S;Y) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathsf{E}\left\{ \log_2 \left(\frac{p(s_k|s_{k-1}, y_k^n)}{p(s_k|s_{k-1}} \right) \right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n I(S_k; Y_k^n|S_{k-1}). \end{aligned}$$

□

Computationally, from (4.7) we have

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \log_2 \left(\frac{p(s_k|s_{k-1}, y_k^n)}{p(s_k|s_{k-1})} \right) \tag{4.9}$$

The denominator in the logarithm of (4.9) is derivable from the state transition equation in (3.1), while the numerator $p(s_k|s_{k-1}, y_k^n)$ needs to be manipulated. It can be seen as

$$p(s_k|s_{k-1}, y_k^n) = \int_{s_{k+1}^n \in \mathcal{S}_{k+1}^n} p(s_k^n|s_{k-1}, y_k^n), \mathrm{d}s_{k+1}^n \tag{4.10}$$

where $p(s_k^n|s_{k-1}, y_k^n)$ can be calculated recursively as in (3.75) but starting with a known state s_{k-1},

$$p(s_k^j|s_{k-1}, y_k^j) \propto p(y_j|s_j) p(s_j|s_{j-1}) p(s_k^{j-1}|s_{k-1}, y_k^{j-1}) \tag{4.11}$$

with $j \geq k$, starting from $p(s_k^{j-1}|s_{k-1}, y_k^{j-1}) = 1$ when $k = j$ and obtaining

$$p(s_k^n|s_{k-1}, y_k^n) \propto \prod_{j=k}^{n} p(y_j|s_j)p(s_j|s_{j-1}). \tag{4.12}$$

The equation (4.11) is similar to the one-step Bayesian recursion in (3.32) but without the integral operation. The operation in (4.10), that must be done $\forall k$ and it is the result of the forward recursion in (4.11) and then an integration to explore the past, is called *Bayesian smoothing*.

The information rate $I(S;Y)$ in the form of (4.6) is equivalent to that in the form of (4.9) and they are exact. If we have all the needed probabilities, we can perform a Monte-Carlo simulation of the two processes and then compute the information rate as sample mean of logarithms of probability ratios. The main obstacle is the exact knowledge of $p(y_k|y_1^{k-1})$ and $p(s_k^n|s_{k-1}, y_k^n)$ in (4.6) and (4.9), respectively, $\forall k$. In Subsection (4.1.2) one can see as overcoming this difficulty to compute the information rate between the state and the measurement for a general model defined in the framework of the state-based approach.

Example: Exact information rate by Kalman filter

When the measurement and the transition state Equations are linear and the noisy processes are Gaussian, the information rate $I(S;Y)$ can be exactly computed by the Kalman filter. Starting from the definition (4.5) one has

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left[h(Y_k|Y_1^{k-1}) - h(Y_k|S_k, Y_1^{k-1}) \right].$$

If Y_1^{k-1} is know then also the predictive mean $\overline{m}_k$ is know and, according to the equation (3.55), one can write

$$h(U_k) = h(Y_k|Y_1^{k-1}).$$

Directly from the measurement equation, one has

$$h(N_k) = h(Y_k|S_k, Y_1^{k-1}).$$

The two last equations lead to

$$I(S;Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left[h(U_k) - h(N_k) \right].$$

Taking into account the differential entropy of a m-variate Gaussian distribution with covariance matrix $\mathbf{\Gamma}_k$ that is

$$h(X_k) = \frac{1}{2}\log_2\left((2\pi e)^m \det\left(\mathbf{\Gamma}_k\right)\right), \tag{4.13}$$

one concludes

$$\begin{aligned} I(S;Y) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log_2\left(\frac{\det\left(\mathbf{H}_k\overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T+\mathbf{R}_k\right)}{\det\left(\mathbf{R}_k\right)}\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log_2\det\left(\left(\mathbf{R}_k^{-1}\left(\mathbf{H}_k\overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T+\mathbf{R}_k\right)\right)\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log_2\det(\mathbf{I}+\mathbf{R}_k^{-1}\mathbf{H}_k\overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T), \end{aligned} \tag{4.14}$$

where $\mathbf{H}_k\overline{\mathbf{\Sigma}}_k\mathbf{H}_k^T+\mathbf{R}_k$ is the covariance matrix of U_k and the properties of the determinant of the inverse matrix and of the matrix product have been used.
In conclusion, when the model is linear and with Gaussian noise, as in (3.49) and (3.50), the information rate between S and Y can be computed as sum of log-det with matrices tracked by the Kalman filter as argument.

4.1.2 Upper and lower bounds to the information rate

If is not possible to perform an exact Bayesian tracking of the state, one has an approximation of $p(y_k|y_1^{k-1})$ and $p(s_k^n|s_{k-1},y_k^n)$ needed in (4.6) and (4.9), respectively, $\forall k$. Consequently, the computation of the information rate between the state process and the measurement process cannot be exacted.

Let $p(x)$ the distribution of a continuous random variable X (the same can be done if the variable is discrete), and let $q(x)$ an other generic probability distribution with the same domain $\mathcal{X}$ of $p(x)$. The Gibbs' inequality say that

$$h(X) = -\int_{\mathcal{X}} p(x)\log_2 p(x) \leq -\int_{\mathcal{X}} p(x)\log_2 q(x) = \overline{h}(X) \tag{4.15}$$

with equality if and only if $p(x) = q(x)$, $\forall x \in \mathcal{X}$. Note that $\overline{h}(X)$ is defined as an upper bound on the actual entropy $h(X)$. In other words, the entropy of a distribution $p(x)$ is less than or equal to its cross entropy with any other distribution $q(x)$. The difference between the left and right

side part of the inequality in (4.15) is the already defined in Equation (2.17) Kullback-Leibler Divergence or relative entropy:

$$\text{KLD}(p(x)|q(x)) = \int_{\mathcal{X}} p(x) \log_2 \left(\frac{p(x)}{q(x)} \right) \mathrm{d}x \geq 0. \tag{4.16}$$

Every time that we do not know the actual probability distribution, we can realize a Monte-Carlo simulation with the actual processes and compute an upper bound on the wanted entropy using an approximated version of the probability distribution.

Let $q(y_k|y_1^{k-1})$ and $q(s_k^n|s_{k-1}, y_k^n)$ the approximations of the distribution $p(y_k|y_1^{k-1})$ and $p(s_k^n|s_{k-1}, y_k^n)$, respectively. Using (4.6) and (4.9), the Gibbs' inequality leads to the upper bounds

$$\begin{aligned} \overline{h}(Y) &= -\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log_2 \left(q(y_k|y_1^{k-1}) \right) \\ &\geq h(Y) \end{aligned} \tag{4.17}$$

and

$$\begin{aligned} \overline{h}(S|Y) &= -\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log_2 \left(q(s_k|s_{k-1}, y_k^n) \right) \\ &\geq h(S|Y), \end{aligned} \tag{4.18}$$

and then we can write upper and lower bounds to the information rate as

$$\overline{I}(S;Y) = \overline{h}(Y) - h(Y|S) \tag{4.19}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log_2 \left(\frac{p(y_k|s_k)}{q(y_k|y_1^{k-1})} \right) \tag{4.20}$$

and

$$\underline{I}(S;Y) = h(S) - \overline{h}(S|Y) \tag{4.21}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log_2 \left(\frac{q(s_k|s_{k-1}, y_k^n)}{p(s_k|s_{k-1})} \right), \tag{4.22}$$

respectively. If these two bounds are close to each other one has obtained the actual information rate $I(S;Y)$. The goal becomes to find a good approximation of $p(y_k|y_1^{k-1})$ and $p(s_k|s_{k-1}, y_k^n)$ in order to shrink the bounds' gap. The conditional probability $p(y_k|y_1^{k-1})$ can be calculated in an exact manner by Bayesian tracking with the Chapman-Kolmogorov equation in

(3.28). If we compute an approximated tracking, we track $q(s_k|y_1^{k-1})$ in place of the actual predictive distribution, leading to

$$q(y_k|y_1^{k-1}) \simeq \int_{s_k \in \mathcal{S}_k} p(y_k|s_k) q(s_k|y_1^{k-1}) \, \mathrm{d}s_k. \tag{4.23}$$

The conditional probability $p(s_k|s_{k-1}, y_k^n)$ can be exactly calculated in an exact manner by Bayesian smoothing with the equation (4.10). In (4.18) $n \to \infty$ is required to calculate the upper bound $\overline{h}(S|Y)$. Using a limited set of measurements y_k^{k+l} with $l > 0$ and then adopting in (4.18) an approximation of $p(s_k|s_{k-1}, y_k^{k+l})$ with limited l in place of $n \to \infty$, one is still performing an upper bound of $h(S|Y)$. In other words, with $q(s_k|s_{k-1}, y_k^{k+l})$ one has

$$\overline{h}(S_k|S_{k-1}, Y_k^{k+l}) \geq h(S_k|S_{k-1}, Y_k^{k+l}) \geq h(S_k|S_{k-1}, Y_k^n),$$

where the last inequality holds because removing conditions the entropy can only increase. As in (4.10), the approximation of $p(s_k|s_{k-1}, y_k^{k+l})$ can be seen as

$$q(s_k|s_{k-1}, y_k^{k+l}) = \int_{s_{k+1}^{k+l} \in \mathcal{S}_{k+1}^{k+l}} q(s_k^{k+l}|s_{k-1}, y_k^{k+l}) \, \mathrm{d}s_{k+1}^{k+l}, \tag{4.24}$$

where l is called *time-lag* of the smoother. In the state-base approach, the state sequence $\{S_k, S_{k+1}, S_{k+2}, \cdots, S_{k+l}\}$ is a reversible transformation of the process noise sequence $\{V_{k-1}, V_k, V_{k+2}, \cdots, V_{k+l-1}\}$ given the initial state S_{k-1}, therefore in some practical case it's better to evaluate the wanted approximation by

$$\begin{aligned} q(s_k|s_{k-1}, y_k^{k+l}) &= q(v_{k-1}|s_{k-1}, y_k^{k+l}) \\ &= \int_{s_{k+1}^{k+l} \in \mathcal{S}_{k+1}^{k+l}} q(v_{k-1}^{k+l-1}|s_{k-1}, y_k^{k+l}) \, \mathrm{d}v_k^{k+l-1}. \end{aligned} \tag{4.25}$$

To find a good approximation of $p(y_k|y_1^{k-1})$ and $p(s_k|s_{k-1}, y_k^n)$, the parametric/non-parametric methods described in the Subsections 3.1.3 and 3.1.5 respectively, can be used.

4.1.3 Upper and lower bounds to the information rate by particle filtering

The particle filtering, widely described in the Subsection 3.1.6, is suitable to provide an approximation of the needed probabilities that appears in (4.17) and (4.18). In the following line we can see as the particle filter can compute the upper bounds on $h(Y)$ and on $h(S|Y)$.

Evaluation of $\overline{h}(Y)$ As shown in [Dauwels and Loeliger, 2008], the normalization factor α_k defined in (3.78) that normalizes the weights of the particles in the update step can be a good approximation of $p(y_k|y_1^{k-1})$. Substituting the forward predictive distribution approximated by the particle filtering (3.81) in the approximated Chapman-Kolmogorov equation (4.23) one has:

$$q(y_k|y_1^{k-1}) \simeq \int_{s_k \in \mathcal{S}_k} p(y_k|s_k) \sum_{i=1}^{P} w_{k-1}^{(i)} \delta(s_k - s_k^{(i)}) \, \mathrm{d}s_k \tag{4.26}$$

$$= \sum_{i=1}^{P} p(y_k|s_k^{(i)}) w_{k-1}^{(i)}, \tag{4.27}$$

that is the factor (3.78). Therefore, it can be used in (4.17) to calculate the upper bound on $h(Y)$ and then the upper bound on the information rate $I(S;Y)$. The entire procedure for Monte-Carlo evaluation of $\overline{h}(Y)$ is equal to the classical particle filtering denoted with Algorithm 1 in the Subsection 3.1.6, followed by the final instruction $\overline{h}(Y) \leftarrow -n^{-1} \sum_{k=1}^{n} \log_2 \alpha_k$, that computes the sample mean of the logarithm of the normalization factors.

Evaluation of $\overline{h}(S|Z)$ For each time instant k, a particle filtering of l steps with all the particles initialized equal to s_{k-1} can be performed. In this way the particle sequence with the related weights gives an approximation of $p(s_k^{k+l}|s_{k-1}, y_k^{k+l})$ the argument needed in equation (4.24). Denoting with ℓ the lag index of the evolution of the particles. At time instant k and lag $\ell = 0$ the particles for $i = 1, \ldots, P$ are initialized as

$$s_{k,0}^{(i)} = f_{k-1}(s_{k-1}, v_{k-1,0}^{(i)})$$

with weight

$$w_{k,0}^{(i)} = \frac{p(y_k|s_{k,0}^{(i)})}{\sum_{j=1}^{P} p(y_k|s_{k,0}^{(j)})},$$

where the set $\{v_{k-1,0}^{(i)},\ i = 1, 2, \cdots, P\}$ is a set of independent samples of process noise, and s_{k-1} is the state visited at time $k-1$ by the realization (s_0^n, y_1^n). For each time lag $\ell = 1, \ldots, l$ the particles and their weights are updated as

$$s_{k,\ell}^{(i)} = f_{k-1+\ell}(s_{k,\ell-1}^{(i)}, v_{k-1,\ell}^{(i)}),$$

$$w_{k,\ell}^{(i)} = \frac{w_{k,\ell-1}^{(i)} p(y_{k+\ell}|s_{k,\ell}^{(i)})}{\sum_{j=1}^{P} w_{k,\ell-1}^{(j)} p(y_{k+\ell}|s_{k,\ell}^{(j)})},$$

where the sets $\{v_{k-1,\ell}^{(i)},\ i = 1, 2, \cdots, P\}$ for $\ell = 1, \ldots, l$ are sets of independent samples of the process noise. After l steps, the distribution tracked by the particle method can be approximated as

$$q(s_k^{k+l}|s_{k-1}, y_k^{k+l}) = \sum_{i=1}^{P} w_{k,l}^{(i)} \delta(s_k^{k+l} - s_{k,0:l}^{(i)})$$

where $s_{k,0:l}^{(i)} = \left(s_{k,0}^{(i)}, s_{k,1}^{(i)}, \ldots, s_{k,l}^{(i)}\right)$. Putting the last equation inside the integral in (4.24), one derives

$$q(s_k|s_{k-1}, y_k^{k+l}) = \sum_{i=1}^{P} w_{k,l}^{(i)} \delta(s_k - s_{k,0}^{(i)}), \tag{4.28}$$

that is the approximation of $p(s_k|s_{k-1}, y_k^{k+l})$ provided by the particle smoother: the Dirac functions are centered on the particles of the first step and have the last computed weight as multiplying factor.

Since the evaluation of $q(s_k|s_{k-1}, y_k^{k+l})$ in the point s_k visited by the realization requires that the inferred distribution is actually a probability density function, a smooth kernel should be used in place of the Dirac delta, leading to

$$q(s_k|s_{k-1}, y_k^{k+l}) = \sum_{i=1}^{P} w_{k,l}^{(i)} \kappa(s_{k,0}^{(i)}; s_k), \tag{4.29}$$

where the kernel $\kappa(\mu; x)$ is a probability density function over the space spanned by x with mean vector μ. The same Parzen technique has been adopted in the Example in Subsection 3.1.6. In practical case the particles can tracks the process noise. Starting from (4.25), the wanted approximation can be evaluated by

$$\begin{aligned} q(s_k|s_{k-1}, y_k^{k+l}) &= q(v_{k-1}|s_{k-1}, y_k^{k+l}) \\ &= \sum_{i=1}^{P} w_{k,l}^{(i)} \kappa(v_{k,0}^{(\rho(i))}; v_{k-1}), \end{aligned} \tag{4.30}$$

where $\rho(\cdot)$ is a function used in the resampling procedure. When the algorithm executes the particle resampling procedure, it is important to consider the right particles, and in the right order, of the set $\{v_{k,0}^{(i)},\ i = 1, 2, \cdots, P\}$ in such a way that, after l steps, the i-th particle $s_{k,l}^{(i)}$ was generated by $v_{k,0}^{(\rho(i))}$.

For example, if $P = 4$ and the particles $\{s_{k,l}^{(1)}, s_{k,l}^{(2)}, s_{k,l}^{(3)}, s_{k,l}^{(4)}\}$ are generated by $\{v_{k-1,0}^{(2)}, v_{k-1,0}^{(2)}, v_{k-1,0}^{(1)}, v_{k-1,0}^{(4)}\}$, respectively, then $\rho(1) = \rho(2) = 2$, $\rho(3) = 1$, and $\rho(4) = 4$ in (4.30).
The entire procedure for Monte-Carlo evaluation of $\overline{h}(S|Z)$ is reported in Algorithm 2.

Algorithm 2 Calculate $\overline{h}(S|Y)$

Generate $(s_0^{n+l}, y_1^{n+l}) \sim p(s_0^{n+l}, y_1^{n+l}) = \delta(s_0) \prod_{k=1}^{n+l} p(s_k|s_{k-1}) p(y_k|s_k)$
Compute v_0^{n+l-1} from s_0^{n+l}
for $k = 1, \ldots, n$ **do**
  Generate $v_{k-1,0}^{(i)} \sim p(v_{k-1})$ for $i = 1, \ldots, P$
  $\rho(i) = i$ for $i = 1, \ldots, P$
  $s_{k,0}^{(i)} \leftarrow f_{k-1}(s_{k-1}, v_{k-1,0}^{(i)})$ for $i = 1, \ldots, P$
  $w_{k,0}^{(i)} \leftarrow p(y_k|s_{k,0}^{(i)}) / \sum_{j=1}^{P} p(y_k|s_{k,0}^{(j)})$ for $i = 1, \ldots, P$
  for $\ell = 1, \ldots, l$ **do**
    if $\sum_{i=1}^{P} (w_{k,\ell-1}^{(i)})^2 > (0.3P)^{-1}$ **then**
      $\left(\{s_{k,\ell-1}^{(i)}\}, \{w_{k,\ell-1}^{(i)}\}, \{\rho(i)\}\right) \leftarrow$ resample $\left(\{s_{k,\ell-1}^{(i)}\}, \{w_{k,\ell-1}^{(i)}\}, \{\rho(i)\}\right)$
    end if
    Generate $v_{k-1,\ell}^{(i)} \sim p(v_{k+\ell-1})$ for $i = 1, \ldots, P$
    $s_{k,\ell}^{(i)} \leftarrow f_{k+\ell}(s_{k,\ell-1}^{(i)}, v_{k-1,\ell}^{(i)})$ for $i = 1, \ldots, P$
    $w_{k,\ell}^{(i)} \leftarrow w_{k,\ell-1}^{(i)} p(y_{k+\ell}|s_{k,\ell}^{(i)}) / \sum_{j=1}^{P} w_{k,\ell-1}^{(j)} p(y_{k+\ell}|s_{k,\ell}^{(j)})$ for $i = 1, \ldots, P$
  end for
  $q(v_{k-1}|y_k^{k+l}, s_{k-1}) \leftarrow \sum_{i=1}^{P} w_{k,l}^{(i)} \kappa(v_{k-1,0}^{(\rho(i))}; v_{k-1})$
end for
$\overline{h}(S|Y) \leftarrow -n^{-1} \sum_{k=1}^{n} \log_2 q(v_{k-1}|y_k^{k+l}, s_{k-1})$

From an operative point of view, the classical Algorithm 1 computes one particle filtering procedure of n steps, where n is the duration of the long Monte-Carlo sequence, while the Algorithm 2 performs n particle filtering, one for each time instant of the sequence and composed by l steps. In addition the Algorithm 1 does not have optimization parameters, while the Algorithm 2 has to optimize the kernel function. It is clear that the Algorithm 2 is computationally more expensive than the 1.

4.2 The information rate of the Discrete Model channel

In this Section it is reported as an example of Bayesian Tracking applied in order to compute the Information rate bounds. The Wiener's phase noise channel, i.e. the DM, bounds are computed as done in [Barletta et al., 2012a].

If one recalls Equations (1.1) and (2.10)

$$\begin{aligned} y_i &= a_i e^{j\varphi_i} + n_i \ , \text{ with} \\ \varphi_i &= \varphi_{i-1} + \sigma_{\text{PN}} \nu_i \\ \nu_i &\sim N(0,1) \ , \end{aligned}$$

where σ is considered to be the standard deviation of the DTFN in one symbol timespan. Without loss of generality, the i.i.d. sequence a_i has unitary power, i.e. $E[|a_i|^2] = 1$. Consider the discrete-state channel where the quantized phase noise is s in contrast to the continuous-state phase φ. The state s is the phase noise quantized into N bins as defined in Equations (3.68) and (3.69). The new auxiliary channel given by s is an approximated version of the real channel with a discretized state and it is easy to be simulated. Therefore a trellis-based recursions can be run in order to compute the entropies that are needed for the computation of the capacity. The considered capacity bounds are derived following the same steps of Subsection 4.1.2

$$\begin{aligned} \underline{I}(X;Y) = H(X) + \overline{h}(Y) - \overline{h}(X;Y) = H(X) - \overline{H}(X|Y) \leq \\ \leq H(X) - H(X|Y) = I(X;Y) \ . \end{aligned} \tag{4.31}$$

In order to derive the upper bound, one has to exploit the Bayes rule [Gelman et al., 2014] to write

$$\frac{P(Y|S,X)}{P(Y|X)} = \frac{P(S|X,Y)}{P(S|X)} \ . \tag{4.32}$$

If one considers Equation (4.32) and wants to write $h(Y|X)$, he can write that

$$h(Y|X) = h(Y|S,X) + P(S|X) - P(S|X,Y) \ ,$$

leading to the upper bound formula

$$\begin{aligned} \overline{I}(X;Y) = \overline{h}(Y) - h(Y|X,S) - H(S) + \overline{H}(S|X,Y) \geq \\ \geq h(Y) - h(Y|X,S) - H(S) + H(S|X,Y) = \\ = h(Y) - h(Y|X) = I(X;Y) \ . \end{aligned} \tag{4.33}$$

Note that all entropies $\overline{H}$ and differential entropies $\overline{h}$ are computed by auxiliary channel simulations. The entropies in the lower and upper bounds of Equations (4.31) and (4.33) respectively must bo computed as follows. The most trivial ones are $H(X)$ and $H(Y|X,S)$ and they do not need any

explanation, but the fact that a Quadrature Amplitude Modulation is considered.

$$H(X)_{\text{M-QAM}} = \log_2(M)$$
$$H(Y|X,S) = \log_2(\pi e \text{SNR}^{-1}) \ .$$

For $h(S)$, one can write that

$$h(S) = -\int_0^{2\pi} p(S_i|S_{i-1}) \log_2 p(S_i|S_{i-1}) dS_i \ ,$$

which can be numerically evaluated easily. Due to its folded nature in its tails $h(S)$ is upper bounded by the entropy of a Gaussian random variable with variance σ^2_{PN}, that is

$$h(S) \leq \frac{1}{2} \log_2(2\pi e \sigma^2_{\text{PN}}) \ . \tag{4.34}$$

When $\sigma^2_{\text{PN}} ll 0.1$, see Chapter 2, the effect of folding is negligible in the range $[0, 2\pi)$, therefore $h(S)$ is well approximated by the right side of the above inequality. In order to compute $\overline{h}(S)$, one should consider that if i is sufficiently large, the following holds

$$\overline{h}(S) = \lim_{i \to +\infty} \frac{1}{i} \log_2 \prod_{k=1}^{i} p(y_k|y_1^{k-1}) = \lim_{i \to +\infty} \frac{1}{i} \sum_{k=1}^{i} \log_2 p(y_k|y_1^{k-1}) \ . \tag{4.35}$$

The sequences x_1^i, y_1^i and s_0^i are simulated following the scheme reported in Figure 2.2. Note that in the channel goes the real phase φ, but in the trellis computation the quantized discrete-state s is considered. Note that the discrete-state Wiener phase noise s_k is memoryless given its previous value s_{k-1}, that leads to

$$P(s_0^i) = P(s_0) \prod_{k=1}^{i} P(s_k|s_{k-1}) \ .$$

The channel is assumed to be memoryless given the input and the state. Moreover the input process and the state process are independent and the input process are independent random variables, so one can write

$$p(y_1^i|x_1^i, s_0^i) = \prod_{k=1}^{i} p(y_k|x_k, s_k) \ , \tag{4.36}$$

and the channel probability law can be factorized into

$$p(y_1^i, s_0^i|x_1^i) = P(s_0^i)p(y_1^i|x_1^i, s_0^i) \ . \tag{4.37}$$

Manipulating the previous three equations (4.2), (4.36) and (4.37), one obtains

$$p(y_1^i, s_0^i|x_1^i) = P(s_0)\prod_{k=1}^{i} P(s_k|s_{k-1})p(y_k|x_k, s_k) \ . \tag{4.38}$$

This last equation permits to compute the entropies $\overline{h}(Y)$ and $\overline{h}(X,Y)$ with the probabilities $p(y_k|y_1^{k-1})$ and $p(x_k, y_k|x_1^{k-1}, y_1^{k-1})$ in a recursive way in trellis with S states, where S is the number of the states of the discrete-state channel. The computation of $\overline{H}(S|X,Y)$ from $P(s_k|s_{k-1}, x_k, y_k)$ derives easily from the trellis built to compute $p(x_k, y_k|x_1^{k-1}, y_1^{k-1})$. In what follows, $p(y_k|y_1^{k-1})$ is derived in such a form that can be simulated. Consider

$$\begin{aligned} P(s_k|y_1^k) &= \frac{p(y_k|s_k)P(s_k|y_k^{k-1})}{p(y_k|y_1^{k-1})} = \\ &= \frac{p(y_k|s_k)\sum_{s_{k-1}} P(s_k|s_{k-1})P(s_{k-1}|y_1^{k-1})}{p(y_k|y_1^{k-1})} \\ &= \frac{\sum_{x_k} p(y_k|x_k, s_k)P(x_k)\sum_{s_{k-1}} P(s_k|s_{k-1})P(s_{k-1}|y_1^{k-1})}{p(y_k|y_1^{k-1})} \ , \end{aligned} \tag{4.39}$$

where $p(y_k|x_k, s_k)$ is the pdf of AWGN noise, $P(s_k|s_{k-1})$ is the discrete-state s innovation law and $P(s_{k-1}|y_1^{k-1})$ the previous step metric. This is the recursion one needs to find $p(y_k|y_1^{k-1})$, which is the normalization factor such that $\sum_{s_k} P(s_k|y_1^k) = 1$. The same derivation is done for the other metric

$$P(s_k|x_1^l, y_1^k) = \frac{p(y_k|x_k, s_k)P(x_k)\sum_{s_{k-1}} P(s_k|s_{k-1})P(s_{k-1}|x_1^{k-1}, y_1^{k-1})}{p(x_k, y_k|x_1^{k-1}, y_1^{k-1})} \ , \tag{4.40}$$

where one can note the desired normalization factor $p(x_k, y_k|x_1^{k-1}, y_1^{k-1})$ needed to compute $\overline{h}(X,Y)$. It is important to note that for the trellis (4.39) we must not substitute the real x_i since the variable is saturated, while in the (4.40) one must use the real transmitted symbol. In the Upper Bound formula (4.33) it must be compute also

$$\overline{H}(S|X,Y) = -\lim_{i\to\infty} \frac{1}{i+1} \log_2 P(s_1^{i+1}|x_1^i, y_1^i) \ . \tag{4.41}$$

If one writes

$$P(s_1^{i+1}|x_1^i, y_1^i) = P(s_{i+1}) \prod_{k=1}^{i} P(s_k|s_{k+1}, x_1^i, y_1^i) = \\ = P(s_{i+1}) \prod_{k=1}^{i} P(s_k|s_{k+1}, x_1^k, y_1^k) \ , \tag{4.42}$$

the $\overline{H}(S|X,Y)$ can be computed with the same trellis of Equation (4.40) with some manipulations by Bayes rule

$$P(s_k|s_{k+1}, x_1^k, y_1^k) = \frac{P(s_{k+1}|s_k, x_1^k, y_1^k)P(s_k|x_1^k, y_1^k)}{P(s_{k+1}|x_1^k, y_1^k)} = \\ = \frac{P(s_{k+1}|s_k)P(s_k|x_1^k, y_1^k)}{P(s_{k+1}|x_1^k, y_1^k)} \ , \tag{4.43}$$

where $P(s_k|x_1^k, y_1^k)$ is the metric of the trellis (4.40), $P(s_{k+1}|s_k)$ the state change probability and $P(s_{k+1}|x_1^k, y_1^k)$ a normalization factor. Note that the real quantized state s_k must be put in the formula in order to have that $P(s_k|s_{k+1}, x_1^k, y_1^k)$ is a number, and not a function of s_k. In the simulations the parameters to be chosen are k and the number of the phase quantizer bins N. Obviously one should expect that the Upper Bound (UB) and the Lower Bound (LB) are closer the higher is the value N. In Figure 4.1 the information rate bounds of a QPSK modulation with the Discrete Model Channel is reported versus SNR with three different values of

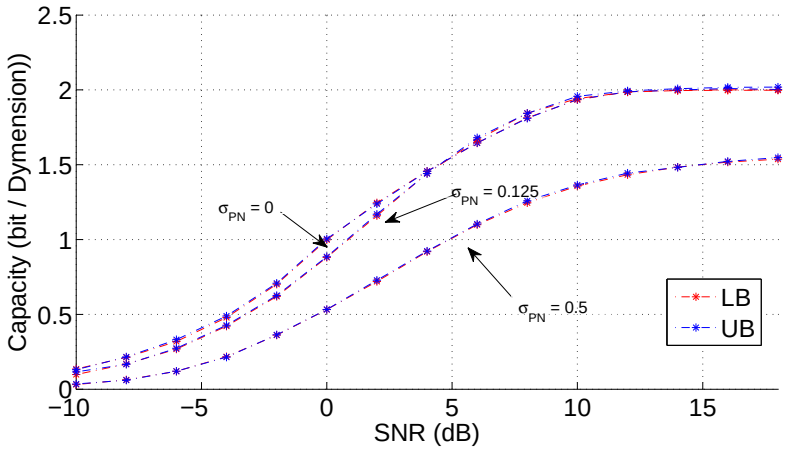


Figure 4.1: *UB and LB of the information rate transferred through a DM channel with QPSK transmission.*

$\sigma_{\text{PN}} = \{0, 0.125, 0.5\}$. One should note the good agreements of the two bounds, therefore he can conclude that the real information rate is well estimated. Note the very little loss of information rate from no phase noise and $\sigma_{\text{PN}} = 0.125$, that is equal to the threshold of validity of the DM. This is a very interesting result of this thesis work, i.e. at most there is a loss of less than 1 dB in the SNR from the information rate point of view between strong phase noise and no phase noise at all with a QPSK modulation. Then one should design demodulators that can extract, e.g. with Bayesian tracking, all the available information from the received data. QPSK modulation is considered since with such phase noise one should not use higher-order modulations. The almost absence of performance degradation can be interpreted also from a theoretical point of view. Indeed, if one considers that if the trellis can follow the phase, and the continuous-phase is quantized in a sufficient way, the DM has only the AWGN impairment
However, in the case of phase noise that strongly impairs the transmission, like $\sigma_{\text{PN}} = 0.5 > \overline{\sigma}_{\text{PN}}$ that is a lot higher than the validity threshold of the model, one has a consistent loss of performance. Particularly, the curve has a floor that is lower than 2 bits even with high SNR. This suggests that such phase noise can not be perfectly tracked by any system, even the strongest Bayesian tracking one, and its effect cannot be totally erased.

4.3 Iterative demodulation and decoding without Pilot Symbols

In this Section the Discrete Model (DM) of first order Wiener phase noise channel is considered. However, while in Section 4.2 the information bounds are derived for such channel, in this Section Bayesian Tracking is exploited to design a receiving demodulator and decoder that can bootstrap a transmission without the aid of Pilot Symbols [Pecorino et al., 2015]. The proposed system can work without any losses compared to competitor schemes, but can achieve the information rate, like the one already exploited in the previous Chapter [Barletta et al., 2013] and [Colavolpe et al., 2005]. Moreover, it is proposed a way to reduce the strong computational requirements of the complete trellis-based algorithm, by exploring only some states of the trellis at each step. Finally 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.

4.3.1 Transmission Setup and State-based Approach

Consider the Discrete Model of Equation (1.1) and (2.10) recalled below

$$\begin{aligned} \overline{y}_i &= a_i e^{j\varphi_i} + n_i \ , \\ \varphi_i &= \varphi_{i-1} + \sigma_{\text{PN}} \nu_i \\ \sigma_{\text{PN}} &= \sigma^2 T \\ \nu_i &\sim N(0,1) \ . \end{aligned}$$

Here, a_i is made by points drawn from a two-dimensional M-ary constellation $\mathcal{A}$ carved out from the grid of integers $\mathcal{Z}^2$. One can write the *state transition probability* of the hidden phase as

$$p(\varphi_i|\varphi_{i-1}) = \sum_{k=-\infty}^{+\infty} g(\varphi_{i-1} + 2k\pi, \sigma_{\text{PN}^2}; \varphi_i) \ , \tag{4.44}$$

and the *measurement* or *channel probability*

$$\begin{aligned} p(y_i|\varphi_i) &= \sum_{a_i \in \mathcal{A}} p(a_i, y_i|\varphi_i) = \sum_{a_i \in \mathcal{A}} p(a_i) p(y_i|a_i, \varphi_i) = \\ &= \sum_{a_i \in \mathcal{A}} p(a_i) g_c(a_i e^{j\varphi_i}, \text{SNR}^{-1}; y_i) \ , \end{aligned} \tag{4.45}$$

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

The attention now is focused onto the Information Rate. Consider a transmission scheme where the source is the vector A_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 of Equation (4.5) as

$$I(A;Y) \equiv \lim_{N \to +\infty} \frac{1}{N} I(A_1^N; Y_1^N) = \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} I(A_i; Y_1^N | A_1^{i-1}) \ . \tag{4.46}$$

Note that the information extracted about A_k takes advantage of the knowledge of past source symbols A_1^{k-1}. A scheme that allows to make use of past source symbols is the joint equalization and decoding scheme proposed in [EyuboĞlu, 1988] 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 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 (4.46) 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. However this approach can lead to unsatisfactory performance when the error probability on preliminary decisions is high. Whatever use of Y_1^N is made, renouncing to the observations A_1^{N-1} one renounces to some information rate, as it is apparent from the following where inequality holds because conditioning does not increase entropy

$$\begin{aligned} I(A_i; Y_1^N | A_1^{i-1}) &= H(A_i|A_1^{i-1}) - H(A_i|Y_1^N, A_1^{i-1}) = \\ &= H(A_i) - H(A_i|Y_1^N, A_1^{i-1}) \leq \qquad (4.47) \\ &\leq H(A_i) - H(A_i|Y_1^N) = \\ &= I(A_i; Y_1^N) \; . \qquad (4.48) \end{aligned}$$

4.3.2 Demodulation by Bayesian Inference on the state

The $\mathcal{Z}^2/4\mathcal{Z}^2$ transmitted constellation is well explained in the published article [Pecorino et al., 2015], along with the hybrid iterative demodulation technique, that exploits a trellis-based demodulation of the signal at the first step to initialize the known Colavolte, Barbieri and Caire (CBC) [Colavolpe et al., 2005] algorithm, which is run at the following iteration steps as reported in Figure 4.2. Note that the Soft-In Soft-Out (SISO) decoding is

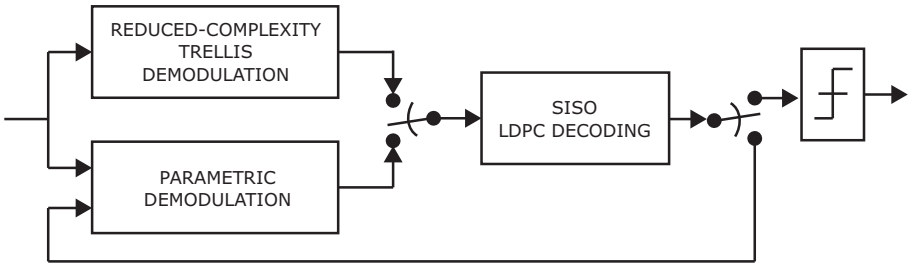


Figure 4.2: *Proposed Iterative Demodulation and Decoding Algorithm Block Scheme.*

based on Low Density Parity Check (LDPC) [MacKay, 2003]. This Subsection focuses the attention on trellis-based Bayesian Inference of the channel state to demodulate the received signal. The proposed trellis can work without pilot symbols and this is the most important gain compared to other

works in the literature, achieving higher data rates compared to pilot-aided transmissions.

Ideal demodulation, which exploits the conditions A_1^{i-1} appearing in the conditional entropy (4.47), can be performed as

$$\begin{aligned} p(a_i|y_1^N, a_1^{i-1}) &= \int_0^{2\pi} p(\varphi_i, a_i|y_1^N, a_1^{i-1}) d\varphi_i = \\ &= \int_0^{2\pi} p(\varphi_i|y_1^N, a_1^{i-1}) p(a_i|y_i, \varphi_i) d\varphi_i \propto \\ (a) &\propto \int_0^{2\pi} p(y_1^{i-1}, a_1^{i-1}|\varphi_i) p(\varphi_i|y_i^N) p(a_i|y_i, \varphi_i) d\varphi_i \propto \quad (4.49) \\ &\propto \int_0^{2\pi} \frac{p(\varphi_i)|y_1^{i-1}, a_1^{i-1}}{p(\varphi_i)} p(\varphi_i|y_i^N) p(a_i|y_i, \varphi_i) d\varphi_i \propto \\ &\propto \int_0^{2\pi} p(\varphi_i)|y_1^{i-1}, a_1^{i-1} p(\varphi_i|y_i^N) p(a_i|y_i, \varphi_i) d\varphi_i \ , \quad (4.50) \end{aligned}$$

where $\propto$ indicates that a factor independent of a_i and φ_i has been brought outside the integral and not considered, the product $p(y_1^{i-1}, a_1^{i-1}|\varphi_i)p(\varphi_i|y_i^N)$ in step (a) is obtained from $p(\varphi_i|y_1^N, a_1^{i-1})$ by derivation similar to those of [Bahl et al., 1974], and the last step holds because $p(\varphi_i)$ is uniform. The forward and backward probability distributions of the phase appearing inside the integral (4.50) are iteratively computed by Bayesian tracking as

$$p(\varphi_{i+1}|y_1^i, a_1^i) \propto \int_0^{2\pi} p(\varphi_i|y_1^{i-1}, a_1^{i-1}) p(y_i, a_i|\varphi_i) p(\varphi_{i+1}|\varphi_i) d\varphi_i \quad (4.51)$$

$$p(\varphi_i|y_i^N) \propto p(y_i|\varphi_i) \int_0^{2\pi} p(\varphi_{i+1}|y_{i+1}^N) p(\varphi_i|\varphi_{i+1}) d\varphi_{i+1} \quad (4.52)$$

where the *blind* channel probability $p(y_i|\varphi_i)$ is computed from Equation (4.45) considering a uniform distribution on $\mathcal{A}$, while the *data-aided* channel probability $p(y_i, a_i|\varphi_i) = p(y_i|a_i, \varphi_i)p(a_i)$ is computed by taking as $p(a_i)$ an indicator function that is non-zero only in the input symbol visited by the actual realization.

However, as it is written is Equation (4.48) in Subsection 4.3.1 for the Information Rate, when input data are not available one is forced to resort to non-data-aided demodulation, that is similar to (4.50), but all the a_1^{i-1} are dropped

$$p(a_i|y_1^N) \propto \int_0^{2\pi} p(\varphi_i)|y_1^{i-1} p(\varphi_i|y_i^N) p(a_i|y_i, \varphi_i) d\varphi_i \ . \quad (4.53)$$

In iterative demodulation and decoding, the first is based on (4.53), but, after the first decoding, one can use the probability distribution p(a_i) coming from the decoder as extrinsic information in (4.50) - but not to compute the intrinsic term $p(a_i|y_i, \varphi_i)$ inside the integral (4.53)- to get the transition metric of forward and backward recursions.

In practice, the continuous-state channel model is intractable. Nevertheless, 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 phase evolution in the discrete-state space, where trellis' states are the centroids of the bins, like it is done in Subsection 3.1.5. However, trellis-based demodulation can be complex, and its complexity increases by a factor equal to the number of iterations. Therefore here trellis-based demodulation at the first demodulation step followed by the classical CBC algorithm, that is far less computational expensive than the trellis-based technique, and it is briefly recalled in what follows.

The probability distribution of the hidden phase given the observation is modeled as a Tikhonov distribution. In CBC, the incoming phase distribution is obtained by tracking the parameter of the Tikhonov distribution. The two-step forward recursion for the tracked parameter is

$$\overline{m}_i^f = m_{i-1}^f + 2\frac{y_{i-1}\alpha_{i-1}^*}{\text{SNR}^{-1} + \beta_{i-1}^2 - |\alpha_{i-1}|^2} , \tag{4.54}$$

$$m_i^f = \frac{\overline{m}_i^f}{1 + \sigma_{\text{PN}}^2|\overline{m}_i^f|} , \tag{4.55}$$

$$\alpha_i = \sum_{a_i \in \mathcal{A}} a_i p(a_i) ,$$

$$\beta_i = \sum_{a_i \in \mathcal{A}} |a_i|^2 p(a_i)$$

The parametric approximation to the wanted probability distribution is computed as

$$p(a_i|y_1^N) \propto e^{-|a_i|^2\text{SNR}} I_0\left(\left|m_i^f + m_i^b + 2\text{SNR}y_i a_i^*\right|\right) , \tag{4.56}$$

where m_i^b is obtained from a backward recursion similar to the forward one and $I_0(\cdot)$ is the 0-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

$$I_0(z) \approx \frac{e^z}{\sqrt{2\pi z}} , \quad z \gg 0 , \tag{4.57}$$

leading to

$$p(a_i|y_1^N) \propto \frac{e^{-|a_i|^2 \mathrm{SNR}\left|m_i^f + m_i^b + 2\mathrm{SNR} y_i a_i^*\right| - \lambda}}{\sqrt{2\pi \left|m_i^f + m_i^b + 2\mathrm{SNR} y_i a_i^*\right|}} , \tag{4.58}$$

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_i and β_i are computed by assuming uniform distribution for $p(a_i)$. 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 Equation (4.54) one realizes that the algorithm cannot bootstrap.

4.3.3 Trellis-Based Algorithm for Simulations

Like it is done in Subsection 3.1.5, the continuous-phase φ is discretized into $|\mathcal{S}|$ bins. In this thesis work, uniform and time-invariant quantization is considered. The state transition probability results to be

$$p(s_i|s_{i-1}) = \int_{\mathcal{R}(s_{i-1}^i)} \frac{p(\varphi_i|\varphi_{i-1})}{\Delta} d\varphi_i d\varphi_{i-1} , \tag{4.59}$$

where $\Delta = \pi/|\mathcal{S}|$ is half the bin width and $\mathcal{R}(s_{i-1}^i)$ the two-dimensional quantization region whose centroids is s_{i-1}^i. Since $p(s_i|s_{i-1})$ depends only on the difference $(s_i - s_{i-1})$, it takes its values in a set of $|\mathcal{S}|$ numbers that can be precomputed by Equation (4.59). The probability $p(y_i|s_i, a_i)$ is similar to (4.45) with centroids of the bin s_i in place of the continuous phase φ_i.

As it is said in the previous Subsection, trellis-based demodulation is computationally expensive. If one wants to reduce the trellis complexity he can reduce the state transition frequency by merging n steps of the random walk. This can work for very small step sizes, i.e. $\sigma_{\mathrm{PN}} \cdot n < \overline{\sigma}_{\mathrm{PN}}^2$, see Chapter 2. Another way to reduce the computation load is found out by assuming that at the time instant i, only a small range of the phase domain has non-negligible probability. Based on this, one can reduce the state bin set from $\mathcal{S}$ to a subset $\mathcal{S}' \subset \mathcal{S}$, where $|\mathcal{S}'|$ is up to the user

$$\mathcal{S}'_i = \{-(|\mathcal{S}'| - 2)\Delta, -(|\mathcal{S}'| - 4)\Delta, \cdots, (|\mathcal{S}'| - 2)\Delta\} + \mu_i , \tag{4.60}$$

where intuitively μ_i is the predicted centroid of interest for the phase state at the time i. The last way to reduce complexity with complex constellation is to reduce the investigated points of $\mathcal{A}$ of Equation (4.45). Actually, at intermediate-to-high SNR only N_p constellation points give non-negligible contribution to the sum. To further simply signal processing, each constellation point is associated with a look-up table that contains the coordinates of nearest constellation's points. After this, only the hard decision on the de-rotated signal and the points in the table associated with the hard-decision are considered in (4.45), thus avoiding the search of the nearest constellation points to the received signal.

The complete derivation and exposition of the techniques above can be found in the published paper [Pecorino et al., 2015], whose results are briefly resumed in the next Subsection.

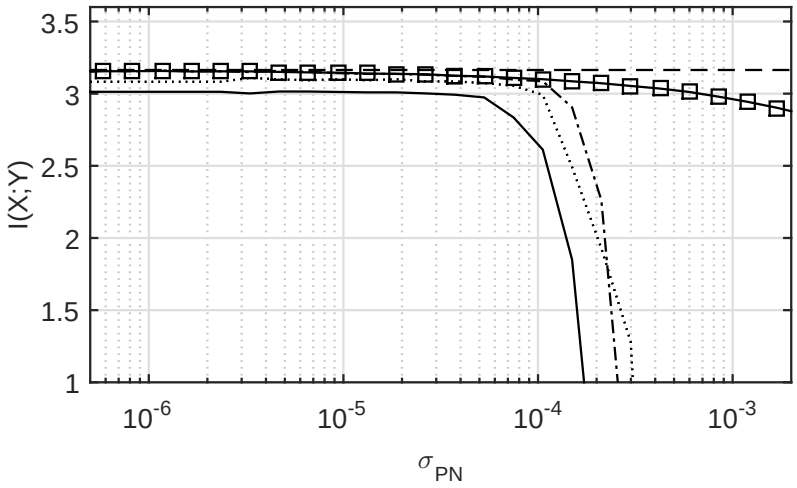
4.3.4 Simulation Results

In this Subsection are shown the results of [Pecorino et al., 2015], where the proposed algorithm is compared also with [Kamiya and Sasaki, 2013] results, where Pilot Symbols were requested in order to demodulate the signal. In the following table are resumed the parameters used to run the simulation The used LDPC code is well explained in [Beermann et al., 2011]. Figures 4.3 and 4.4 report the achievable information rate versus σ_{PN} with 16-QAM and 64-QAM with two different values of SNR each. In the Figures are reported the information rates of

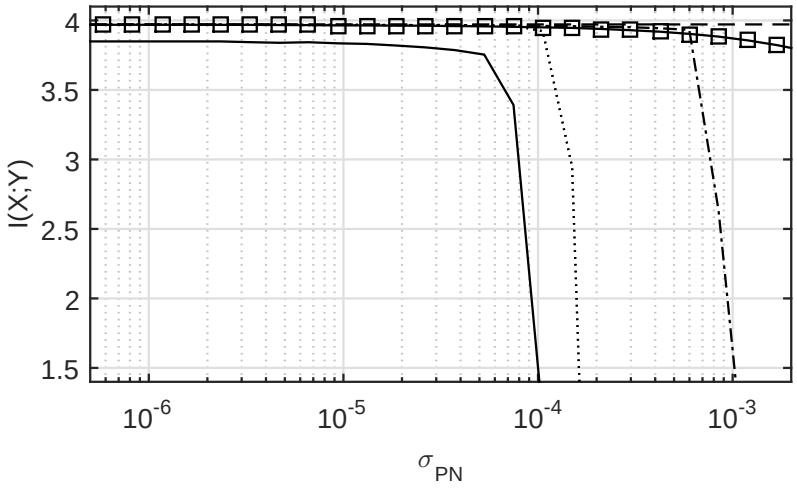
Table 4.1: *Simulation Parameters of Trellis-based demodulation*

	16-QAM	64-QAM
Half bin Width, Δ	$\pi/16$	$\pi/64$
Number of bins, $\mathcal{S}'$	8	8
DNumber of nearest QAM points, N_p	16	9
Number of merged phase steps, n	73	73
LDPC code, (n, k)	4088,3066	4088,3066
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

- pure AWGN channel without phase noise (Dashed line),
- Full version of the trellis with data-aided recursions (Solid line with squares),

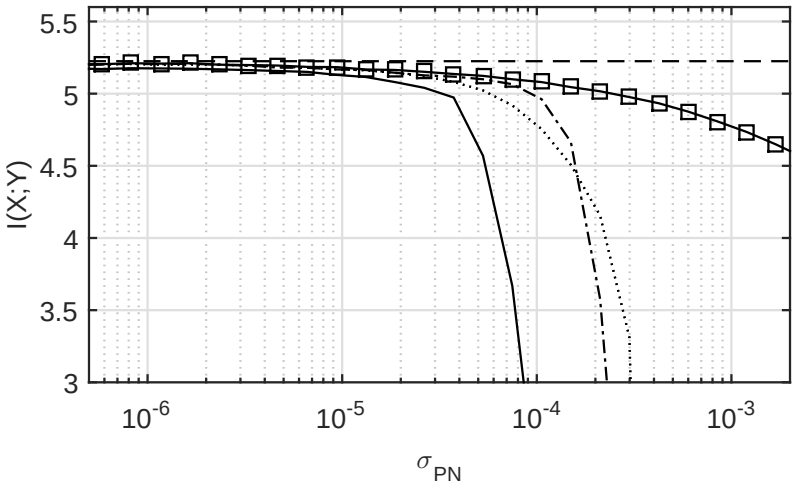


(a) SNR = 10 dB

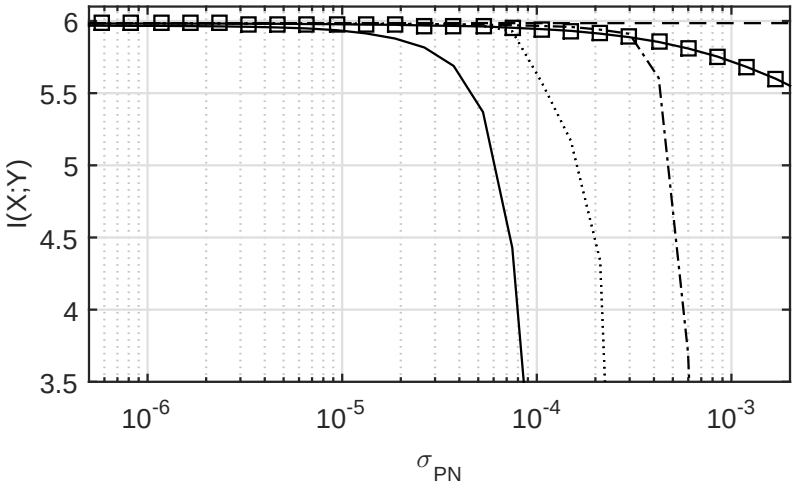


(b) SNR = 16 dB

Figure 4.3: *Achievable information rate versus σ_{PN} for the phase noise channel with 16-QAM and two values of SNR. Dashed line: pure AWGN. 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.*



(a) SNR = 17 dB



(b) SNR = 23 dB

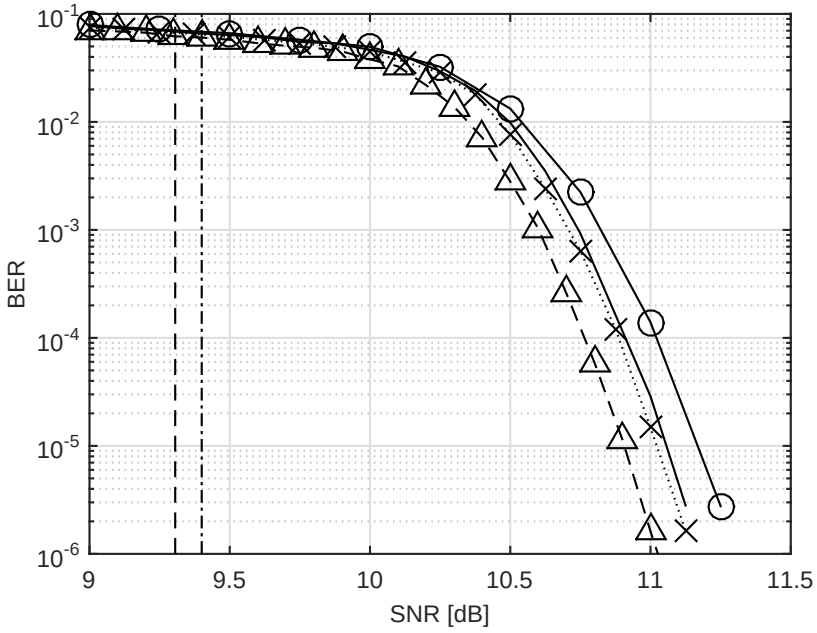
Figure 4.4: *Achievable information rate versus σ_{PN} for the phase noise channel with 64-QAM and two values of SNR. Dashed line: pure AWGN. Solid line with squares: full trellis, forward data-aided recursion. Dash-dotted line: full trellis with forward-backward non-data-aided recursions. Solid line: reduced complexity trellis with forward-backward non-data-aided recursions.*

- Full version of the trellis without data-aided recursions (Dash-dotted line),
- Reduced trellis without data-aided recursion (Solid line).

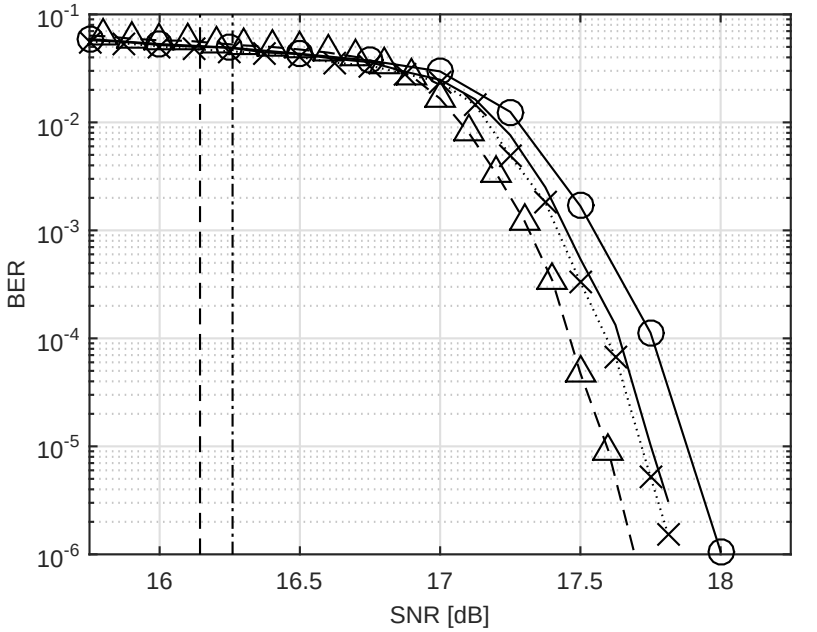
Note that all systems can achieve the information rate computed by Bayesian tracking with data-aided recursions. Then, the non-data-aided loses track of the phase at high phase noise values. This is intuitively true since when the system begins to be stressed by phase noise and decisions $\hat{a}_i \neq a_i$, the performance of data-aided begins to differ from the non-data-aided one. The reduced complexity trellis that exploits the minimal state with parameters of Table 4.1 has more weaknesses than the full-trellis but it can still extract the full information under some phase noise values. In Figures 4.5 the BER of the full system, i.e. trellis-based initialization and CBC demodulation and decoding, are reported:

- Performance limit of AWGN channel (Dashed Line),
- Performance limit of AWGN and phase noise channel (Dash-dotted line),
- Pure AWGN (Dashed line with Triangles),
- Hybrid iterative demodulation and decoding without pilot symbols (Solid line),
- Full trellis with data-aided forward recursion and non-data aided backward (Dotted line with Crosses),
- Iterative demodulation and decoding of [Kamiya and Sasaki, 2013], $1/25$ Pilot rate (Solid line with circles).

Apparently, the proposed systems can work with $1 - 1.5$ dB losses from the Information Rate optimum SNR. It is apparent from Fig. 5 that, although the proposed method does not make use of pilot symbols and it is initialized with the loose trellis first step, it can still achieve better performances than the Pilot-aided one of [Kamiya and Sasaki, 2013]. Also, despite the dramatic complexity reduction, the performance with reduced trellis is virtually optimal compared to the full-trellis one. Actually, we have found that system complexity can be reduced without any appreciable performance degradation until cycle slips starts appearing.



(a) 16-QAM



(b) 64-QAM

Figure 4.5: *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 dataaided forward recursion and non-data-aided backward recursion. Solid line with circles: iterative demodulation and decoding of Kamiya and Sasaki [Kamiya and Sasaki, 2013] with pilot rate 1/25.*

CHAPTER 5

Chromatic Dispersion impact onto phase noise in DDO-OFDM

In this Chapter the issue of Wiener phase noise is dealt by considering its effects in another field of interest, Direct Detection (DD) Optical OFDM . In the first Section a referenced model for the phase noise effects in DDO OFDM, well known in the literature, is derived [Peng, 2010]. However, in this thesis the model is criticized since it imposes a strong assumption in its derivation. After exposing the weak points, a MC simulator have been setup with two DDO OFDM Scenarios to compare the performances between them and the literature model. At last, measurements in [Schmidt et al., 2008] are compared with both [Peng, 2010] and the MC performance in a "Scenario 0".

5.1 Single Sideband DDO-OFDM - A mathematical model

With reference to the work of [Peng, 2010], the author proposes a the mathematical derivation of the Phase Noise (PN) Effects in Single Side-Band (SSB) DDO-OFDM Transmission that improves the previous work in two aspects:

- The derivation is done in a continuous-time domain. More precise results are obtained with integrals compared to simple discrete-time sum.
- The results are reported in a cleaner way if compared to [Peng, 2010], and the dependencies on the system's parameters are more intelligible.

Consider the continuous-time SSB OFDM signal

$$s(t) = e^{j2\pi f_A t}\left(A + \sum_{k=1}^{N_d} d_k e^{j2\pi[(k+N_d)\Delta f]t}\right) , \tag{5.1}$$

where A and f_A are the carrier complex amplitude and frequency. Note that f_A can be set to zero since now without loss of generality. d_k are the N_d complex modulated symbols on the k-th subcarrier, Δf the subcarrier spacing. The laser phase noise is a continuous-time Wiener process of Equation (2.2)

$$\varphi(t) = \sigma \int_0^t \nu(\tau) d\tau ,$$

where it has been set $\varphi(0) = 0$. Note that $\varphi(t) \sim N(0, \sigma t)$ and the laser phase noise linewidth can be derived as easily as $\gamma_{\text{PN}} = \sigma^2/(2\pi)$. Hence, the transmitted signal is affected by phase noise, becoming

$$s_T(t) = \left(A + \sum_{k=1}^{N_d} d_k e^{j2\pi[(k+N_d)\Delta f]t}\right) e^{j\varphi(t)} . \tag{5.2}$$

The chromatic dispersion is modeled as a different time-delay of the process with respect of the central frequency of the signal, i.e.

$$s_R(t) \approx A e^{j\Phi(t)} + \sum_{k=1}^{N_d} d_k \cdot e^{j2\pi[(k+N_d)\Delta f](t-T_k)+j\Phi(t-T_k)} , \tag{5.3}$$

where $T_k = [DL\lambda^2(k+N_d)\Delta f/c]$ and the approximation holds if the phase noise γ_{PN} is sufficiently small compared to Δf, see Chapter 2. This is the focal point of the derivation, and it will be deepened later. Before resuming the derivation, in Figure 5.1 the transmission and CD effect are trivially shown. From Equation (5.3), one can define the phase noise that affects the k-th subcarrier as

$$\begin{aligned} &\rho_k(t) = \varphi(t - T_k) - \varphi(t) , \\ &\rho_k(t) \sim N(0, \beta_k = \sigma^2 T_k) . \end{aligned} \tag{5.4}$$

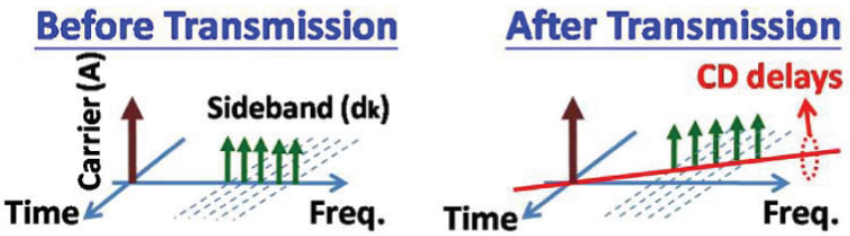


Figure 5.1: *OFDM Transmission and different delays due to Chromatic Dispersion.*

When the signal is received at the photodiode, it computes the intensity, leading to the converted electrical signal

$$|s_R(t)|^2 \approx 2\,\mathrm{Re}\left\{A^* \sum_{k=1}^{N_d} d_k e^{j2\pi f_k t + j\rho_k(t) + j\theta_k}\right\} + |A|^2 + \sum_{k=1}^{N_d} |d_k|^2 + \text{SSBI} \tag{5.5}$$

where $f_k = [(k + N_d)\Delta f]$ and the sideband-sideband beat interference and $\theta_k = -2\pi f_k T_k$ the CD phase rotation on k-th sub [Peng et al., 2009a]. Given (5.5), assuming rectangular filter shaping, the k-th FFT demodulated signal is

$$R(k) = D_k \psi_k(0) + \sum_{m=1, m\neq k}^{N_d} D_m \psi_m(k-m) \text{ , with } D_k = A^* d_k e^{j\theta_k} \text{ , } \tag{5.6}$$

where the disturb comes from the non-zero samples of the sampled process $e^{j\rho_k(n)}$ FFT

$$\psi_k(p) = \frac{1}{N} \sum_{n=0}^{N-1} e^{j[\frac{2\pi pn}{N} + \rho_k(n)]} \text{ .} \tag{5.7}$$

In the thesis it is also considered the effect of phase noise in a continuous-time manner. Therefore the (5.7) can be written as

$$\psi'_k(p) = \frac{1}{T_O} \int_{T_O} \exp\left(j\frac{2\pi pt}{T_O}\right) e^{j\rho_k(t)} dt \text{ .} \tag{5.8}$$

where T_O is the OFDM symbol period. The Power Spectral Density (PSD) of $\rho_k(t)$ has been derived in [Qi et al., 2006]

$$P_k(f) = e^{-\beta_k}\left[\delta(f) - \frac{\sin(\beta_k)}{\pi f}\right] + \frac{1 - e^{-\beta_k}\left[\cos(\beta_k) - \frac{f}{\gamma_{\text{PN}}}\sin(\beta_k)\right]}{\pi\gamma_{\text{PN}}\left(1 + \left(\frac{f}{\gamma_{\text{PN}}}\right)^2\right)} \text{ ,} \tag{5.9}$$

Note that is different from $\varphi(t)$ PSD, that is the phase noise process in Coherent Optical OFDM with the Lorentzian spectrum of Equation (2.5) [Magarini et al., 2011].
Consider Equation (5.6), one can split the disturb on the desired signal D_k into three term, as it reported in Fig. 5.2 [Peng, 2010] and below

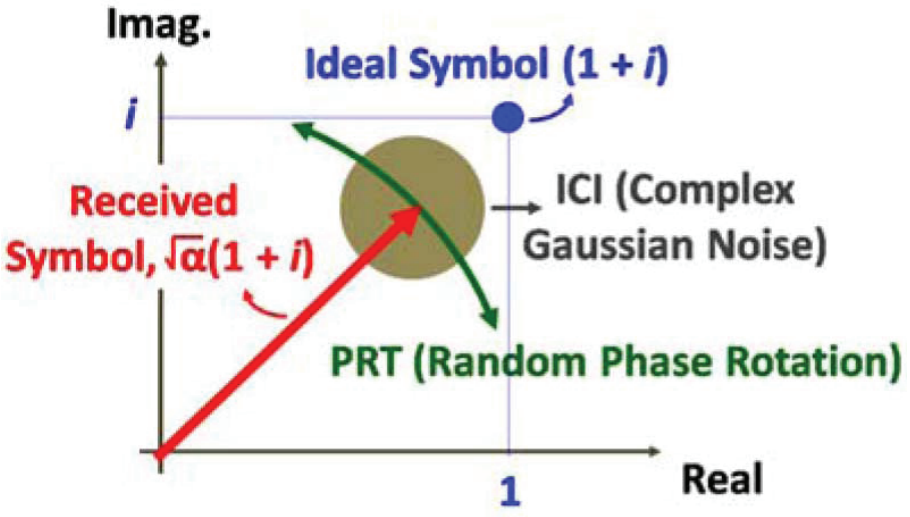


Figure 5.2: *Effects of the PN on the received subcarrier signal: (1-red) power degradation α, (2-green) phase rotation term (PRT), and (3-brown) inter-carrier interference (ICI).*

1. Power Degradation (PD),
2. Phase Rotation Term (PRT),
3. Inter-Carrier Interference (ICI).

In the next three Subsections, those three effects are derived in a continuous-time manner, as opposed to what is done in previous works.

5.1.1 Power Degradation (PD)

Under small phase noise (SPN) assumption [Mandelli et al., 2014], one can easily compute the ratio α_k between the k-th subcarrier received symbol power with PN and the one without PN. This power loss is due to the power shifting of the signal to PRT and ICI components. Define

$$\alpha_k = |E\left[\psi_k(0)\right]|^2 = \left| E\left[\frac{1}{T_O} \int_T^{T+T_O} e^{j\rho_k(t)} dt \right] \right|^2 , \tag{5.10}$$

where T is the time at the OFDM symbol start. If SPN $\rho_k(t) \to 0$, one can approximate the expectation in (5.10) as follows, by manipulating $e^{j\rho_k(t)}$

$$\begin{aligned} & E\left[\frac{1}{T_O}\int_T^{T+T_O} e^{j\rho_k(t)}dt\right] \approx \\ \approx & \frac{1}{T_O}E\left[\int_T^{T+T_O} 1dt + j\int_T^{T+T_O}\rho_k(t)dt - \frac{1}{2}\int_T^{T+T_O}\rho_k^2(t)dt\right] = \\ = & 1 - \frac{1}{2T_O}\int_T^{T+T_O} E\left[\rho_k^2(t)\right]dt = 1 - \pi\gamma T_k \ . \end{aligned} \tag{5.11}$$

The (5.10) becomes

$$\alpha_k \approx |1 - \pi\gamma T_k|^2 \approx 1 - 2\pi\gamma T_k = 1 - \beta_k \ . \tag{5.12}$$

The equation above underlines that if SPN holds, $\alpha_k + \beta_k \approx 1$, meaning that the power loss of the k-th subcarrier will turn to the PN power of k-th subcarrier.

5.1.2 Phase Rotation Term (PRT)

PRT is the 0-order interference $\psi_k(0)$. Deriving from 5.8 and with SPN, one can write

$$\psi_k(0) = \frac{1}{T_O}\int_T^{T+T_O} e^{j\rho_k(t)}dt \approx 1 + j\frac{1}{T_O}\int_T^{T+T_O}\rho_k(t)dt \tag{5.13}$$

In (5.13) it is clear that PRT only affects signal imaginary part and therefore results in a small phase rotation (the second order term is taken into account in PD, from the third are omitted). Assume without loss of generality that $T = 0$. Since $\rho_k(t)$ is a zero-mean Gaussian variable, PRT will be also a

zero-mean Gaussian variable with variance

$$\begin{aligned}
\sigma^2_{k,\text{PRT}} &\approx \frac{1}{T_O^2} E\left[\int_0^{T_O} \rho_k(t)dt\right]^2 = \\
&= \frac{1}{T_O^2} E\left[\int_0^{T_O} dt \int_{t-T_k}^{t} n(\nu)d\nu\right]^2 = \\
&= \frac{1}{T_O^2} E\left[T_k \int_0^{T_O-T_k} n(\nu)d\nu + \int_0^{T_k} \nu\, n(\nu - T_k)d\nu + \int_0^{T_k} \nu\, n(T_O - \nu)d\nu\right]^2 = \\
&= \frac{1}{T_O^2}\left\{T_k^2 \int_0^{T_O-T_k} E\left[n^2(\nu)\right] d\nu + \int_0^{T_O} \nu^2 E\left[n^2(\nu - T_k)\right] d\nu + \right. \\
&\left. + \int_0^{T_O} \nu^2 E\left[n^2(T_O - \nu)\right] d\nu\right\} = \\
&= \frac{2\pi\gamma}{T_O^2}\left[T_k^2(T_O - T_k) + \frac{2}{3}T_k^3\right] = \\
&= \frac{2\pi\gamma}{T_O^2}\left[T_O T_k^2 - \frac{1}{3}T_k^3\right] . \qquad (5.14)
\end{aligned}$$

In the following list are reported the main differences between this work and [Peng, 2010] together with the arguable points.

- This is a continuous-time derivation. It is more precise and there are not problems of dimensionality of spare terms in (5.14) like the ones in the (11) of [Peng, 2010].

- The number of subcarriers is hidden since now, and the derivation does not take into account that so it does not depend neither on N nor N_d. However, note that T_O is inversely proportional to the number N of FFT points, which is strictly related to N_d, given the total band B, i.e. $T_O = B/N$.

- In order to compare this work to [Peng, 2010] one should define the sampling period T_s, the frequency spacing Δf, the number FFT points N, the number of subcarriers N_d and the total considered OFDM band B. Therefore, the parameters are defined below, along with M_k that is

the number of delay FFT samples due to the CD:

$$\Delta f = T_O^{-1}\,, \tag{5.15}$$

$$T_s = \frac{T_O}{N}\,, \tag{5.16}$$

$$B \geq 2N_d\Delta f = \frac{2N_d}{T_O}\,, \tag{5.17}$$

$$M_k = \frac{T_k}{T_s}\,. \tag{5.18}$$

Note that the number of FFT points N must be such that $N \geq N_d$. In what follows the considered band will be considered at its minimum value $B = 2N_d\Delta f$. Exploiting the last four Equation, one can arrange the (5.14) into

$$\sigma^2_{k,\text{PRT}} \approx \frac{2\pi\gamma T_s}{N^2}\left[N\,M_k^2 - \frac{1}{3}M_k^3\right]\,, \tag{5.19}$$

that is very similar to previous derivation, without that questionable + 1 that came from the discrete-time sampling. Then, if $N \gg M_k$, one can write

$$\begin{aligned}\sigma^2_{k,\text{PRT}} &\approx \frac{2\pi\gamma T_s}{N}M_k^2 = 2\pi\gamma\frac{T_k^2}{NT_s} = 2\pi\gamma T_k^2\Delta f = \\ &= 2\pi\gamma\left(\frac{DL\lambda}{c}\right)^2(k+N_d)^2(\Delta f)^3 = \\ &= 2\pi\gamma\left(\frac{DL\lambda}{c}\right)^2\left(B\cdot\frac{k+N_d}{2N_d}\right)^2\Delta f\end{aligned} \tag{5.20}$$

From (5.20), one can split the effects on the PRT into three terms:

1. Laser ans Fiber parameters: $2\pi\gamma(DL\lambda/c)^2$,
2. Frequency offset with respect to the carrier, that is the total band multiplied by the relative position of the k-th subcarrier. This is the only term that depends on k: $(B(k+N_d)/2N_d)^2$,
3. The subcarrier frequency offset Δf.

5.1.3 Inter-Carrier Interference (ICI)

Not yet derived with integrals, the noise behavior of ICI has long been considered as a zero-mean complex Gaussian random variable [Yi et al., 2008, Tomba, 1998]. The ICI power is also a function of the subcarrier

index and its variance, when normalized to the subcarrier symbol power, can be exactly obtained without using the small PN assumption, leading to the upper-bounded to the signal power minus both the PD and PRT power

$$\sigma^2_{k,\text{ICI}} \leq 1 - (\alpha_k + \sigma^2_{k,\text{PRT}}) \; . \tag{5.21}$$

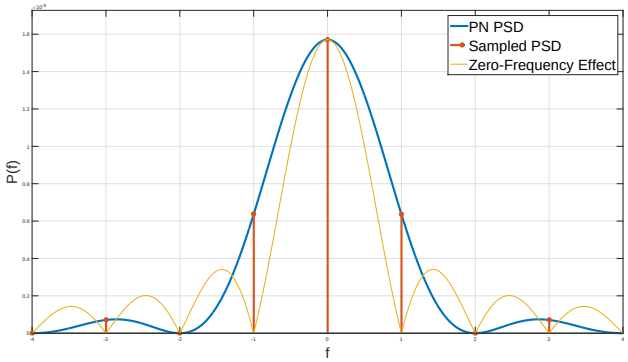
This results is really intuitive, since the power loss $(1 - \alpha_k)$ is almost given by the two orthogonal effects PRT and ICI variances, given the independence of the transmitted subcarriers.

Equations (5.20) is worthy to be investigated. As it is said in the end of Subsection 5.1.2, $\sigma^2_{k,\text{PRT}}$ linearly depends on Δf. With constant band B, it seems that increasing N will reduces the rotation due to the Phase Noise. This is only partly true. The PRT power in (5.20) can be seen also as the integration of a constant PSD (constant because $M_k \ll N$, or $T_k \ll T_O$) in the frequency span Δf.

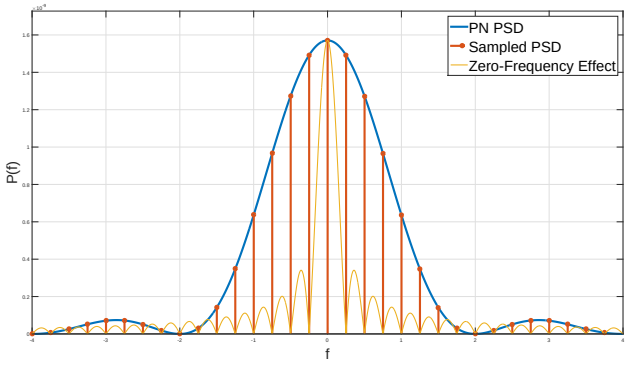
The more N increases, the less Δf will be. Try to consider the whole system in the frequency domain, as it is shown in (5.6) and (5.8). With the same system parameter, the PSD of the Phase Noise (5.9) is the same. If one increases N, he is sampling the PSD of the k-th sub closer and closer in the frequency domain. From the model expressed by (5.6), the PSD of the k-th subcarrier baseband demodulated is given by taking each sample of $P_k(f)$ and convolving this discrete-frequency signal with a sinc-shape signal with first zero at $f' = \Delta f$. From Figures 5.3 it is clear now how the same Phase Noise power spreads differently with increasing N: remind that the power is the area of the considered curves. If N is small, a lot of power goes to the PRT, reducing the ICI power since the closer subcarrier are far away from the signal. On the contrary, if N is big, PRT power is less, since ICI draws more available power to itself.
Let $N \to \infty$: this is the case where all phase noise of k-th sub spreads onto the other subcarriers, where all the phase noise that affects the k-th sub comes from the other subcarriers. In this case $\sigma^2_{k,\text{PRT}} \to 0$.

5.2 Simulation, measurements and performance analysis

In the beginning of this Section the SA results are computed, from Equations (5.6) and (5.7). Therefore the SA performance are compared with a Monte Carlo (MC) simulator that goes deeper in the system. In this Section two scenarios - Scenario 1 and 2 - of short reach usable lasers are presented. The first is given by a DFB and second by a VCSEL of 1 MHz and 10 MHz linewidths respectively. At last the SA approach and the MC simulator are



(a) Small N

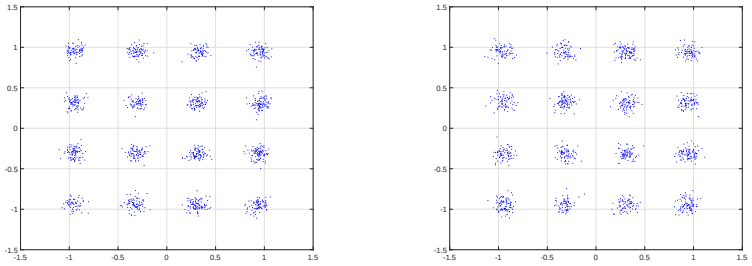


(b) Much higher N

Figure 5.3: *Effect on PRT's statistic due to the sampling of the Phase Noise PSD.*

Table 5.1: *Semi-Analytical (SA) DDO-OFDM transmission Parameters in Figure 5.4*

Scenario	0- [Schmidt et al., 2008]	1 (DFB)	2 (VCSEL)
γ_{PN}	200 kHz	1 MHz	10 MHz
Fiber length, L	$\approx$ 400 km	$\approx$ 800 km	100 km
Cyclic prefix, (CP)	20%	20%	20%
M-QAM	16 km	16 km	16 km
Chromatic Dispersion, D	16 ps/(nm $\cdot$ km)	16 ps/(nm $\cdot$ km)	16 ps/(nm $\cdot$ km)
Laser wavelength, λ	1550 nm	1550 nm	1550 nm
Transmitted Rate, R_b	20 Gbit/s	20 Gbit/s	20 Gbit/s
Number of data Subcarriers, N_d	128	128	128



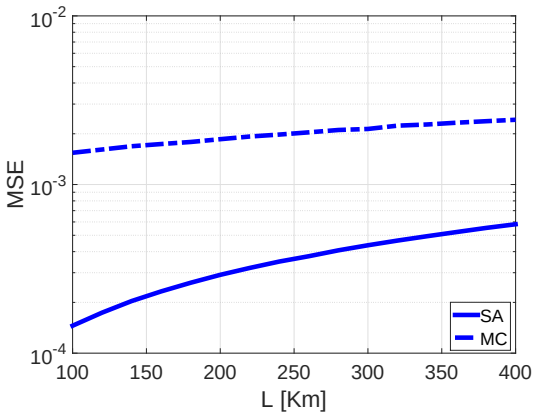
(a) Scenario 1 (b) Scenario 2

Figure 5.4: *Received 16-QAM constellation transmitted over a L km fiber in the SA model.*

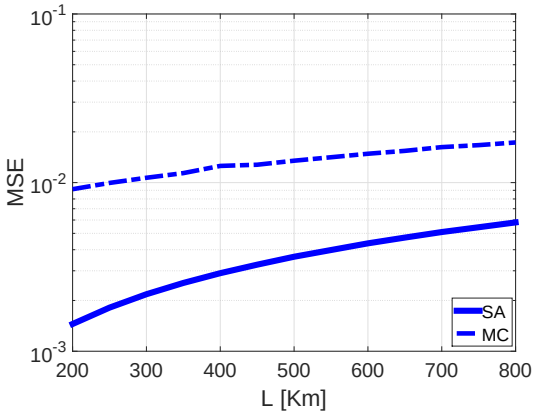
compared with the 200 KHz linewidth measurements of [Schmidt et al., 2008] that will be called Scenario 0.

5.2.1 Semi-Analytical Model

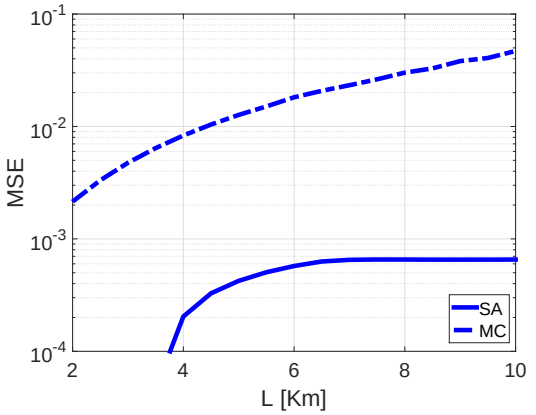
The considered scenarios are taken from the DDO-OFDM measurements of [Schmidt et al., 2008] , where only γ_{PN} is changed. Hence the parameters are reported in Table 5.1: Consequently the OFDM subcarriers spacing is computed as $\Delta f = (1 + CP)\, R/[N_d \log_2(M)] \approx 46$ MHz. SA simulation are run by taking random generations of the process $\Phi(t)$ and d_k of the well-known Wiener Phase noise given by Equation (2.2) and then computing the (5.6) and the (5.7) for all $k = 1, ..., N_d$. In Figures 5.4 are reported the received constellation of all subcarriers if one considers the semi-analytic model (5.6). Note the good performances of the modulation, with a negligible Bit Error Rate (BER). In Figure 5.5 the Mean Square Error (MSE) of the transmission is plotted versus the Fiber Length (L) defined



(a) Scenario 0



(b) Scenario 1



(c) Scenario 2

Figure 5.5: *MSE of the received signal with the SA and MC simulator.*

as

$$\text{MSE} = \mathsf{E}\left\{|R - D|^2\right\} \ , \tag{5.22}$$

where R and D are the generic received $R(k)$ and transmitted D_k symbols respectively and no AWGN is considered. One could conclude from this results that DDO-OFDM transmission is possible in a Short Reach and even Metro optical access with low quality lasers, like DFB for 1 MHz and VCSEL for 10 MHz respectively. However, as it is written in previous Sections, the SA Model has weaknesses if γ is not negligible compared to Δf. Hence the authors analyze in the following Subsections the scenario in a more precise fashion - i.e. in exploiting a MC approach versus the SA presented in this Subsection.

5.2.2 Monte Carlo Framework

The Monte Carlo simulation is run following the block scheme of Figure 5.6.

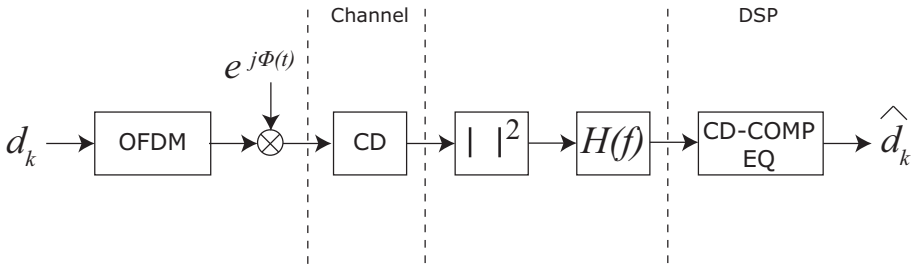


Figure 5.6: *DDO-OFDM Monte Carlo Simulator Block Scheme*

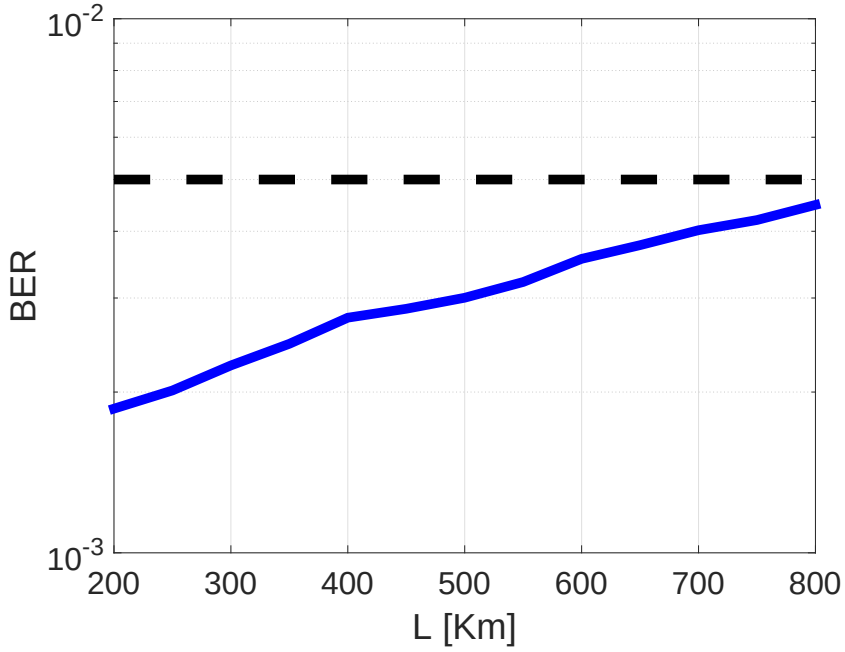
In the following list the features of the simulator are deepened:

- the OFDM block shapes the signal as it is written in (5.1), with the insertion of Cyclic Prefix (CP). Note that the signal is upsampled by a factor of 16 to model the continuous-time,
- Chromatic Dispersion is applied as a multiplication in the frequency domain by $\exp(-j2\pi DL\lambda^2 f^2/c)$
- Optical Filter must be a constant in the band of interest, in order to avoid distortion that can not be equalized after the photodiode,
- The receive electrical filter is free as long as it is shaped around the band of the signal,

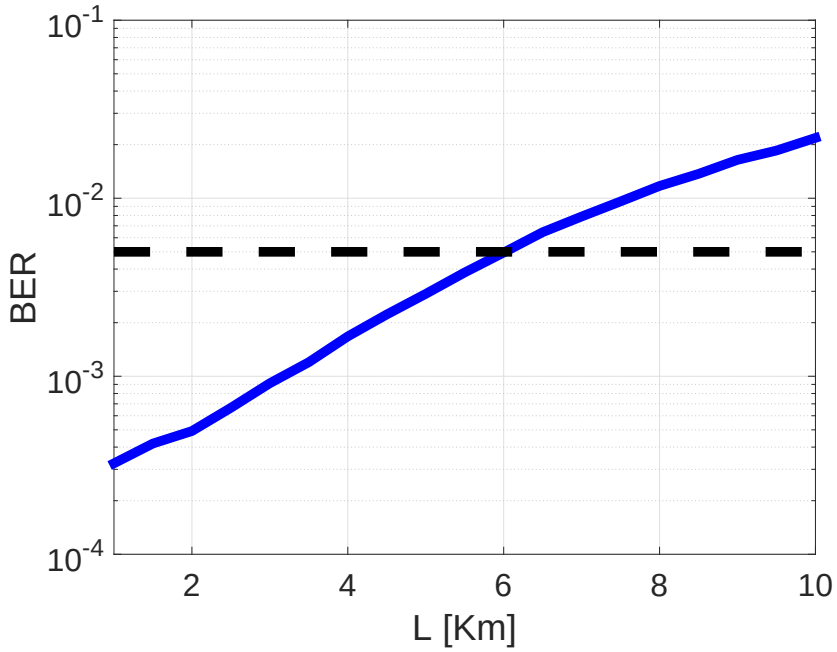
- The signal is then sampled, and, without Amplified Spontaneous Emissions (ASE), the Digital Signal Processing (DSP) at the receiver operates a Minimum Mean Square Error Equalization that performs CD Compensation and removes the Electrical Filter memory in the frequency domain.

In what follows there are reported two different set of results coming from the MC simulator. The first one regards the transmission setup presented previously in the semi-analytic Subsection.
As one can see from Figure 5.5 the MC MSE is completely different from the SA ones with much lower distances than the ones on the Figure 5.4. The MSE is much higher, about one or two order of magnitude in all scenarios. Indeed CD distorts the signal in the channel, since it is not an impulse train in the frequency domain due to convolution with the phase noise Fourier transform. This proves what the authors claim, i.e. with large linewidth lasers the SA models in the literature are not valid anymore. However MC simulator can show the performance of such channels. In the literature there are several works that deal with long-haul SSB DDO-OFDM [Peng et al., 2009b, Schuster et al., 2008]. However in these works CD is corrected by dispersion compensating fibers, that are necessary when working with long transmissions. In short reach networks dispersion fibers are not necessary . This arise the CD distortion that is point out in this work. Particularly the authors claim that CD is the limiting factor of SSB DDO-OFDM with a big number of subcarriers N_d. This is because the (5.6) is not valid anymore with a laser linewidth comparable with the subcarrier spacing Δf.

In Figure 5.7 is also reported the Bit Error Rate of the MC simulator with additive noise modeled as equivalent Additive White Gaussian Noise (AWGN) at the input of the demodulator. A SNR equal to 20 dB is chosen according to the Scenario 0 of [Schmidt et al., 2008]. Particularly, if one considers a $5 \cdot 10^{-3}$ threshold on the BER, the transmission fulfills the requirement until $L = 800$ km in Scenario 1 and $L = 6$ km in the 2^{nd} one. This work seems to be in contrast with the previous [Peng, 2010]. However, if one tries to reproduce the measurements of [Schmidt et al., 2008] with the MC simulator they fit. In particular at 400 km the simulated $\text{BER}_S \approx 7{\cdot}10^{-4}$ is a little bit better than the measured one $\text{BER}_M \approx 2{\cdot}10^{-3}$ of that paper. The small difference is trivially due to amplifier noise and experimental impairments. If one has still some issues in validating this work's results compared to [Peng, 2010], he should consider that the MSE of that framework, even in Scenario 0, is always more than one order of magnitude smaller. Hence the performance of such systems seems to be too optimistic to the measurements present in the literature.



(a) Scenario 1



(b) Scenario 2

Figure 5.7: *MC Bit Error Rate versus L. SNR = 20 dB*

Concluding the analysis, Coarse linewidth lasers are considered in this Chapter due to the increasing demand of economic and small optical systems in short reach Networks without CD Compensation fibers. The author has considered lasers like VCSELs of 10 MHz and DFB 1 MHz line bandwidth, comparing them with the 200 kHz of [Schmidt et al., 2008] in the same scenario. With phase noise introduced by the laser, Direct Detection Optical OFDM (DDO-OFDM) becomes one solution with strong advantages like system complexity and resistance to carrier phase noise. However, like it is pointed out in previous works, like [Peng, 2010], Chromatic Dispersion negates the full cancellation of the phase noise at the receiver. Here the mathematical model for the performance of DDO OFDM affected by phase noise is analyzed, along with its weak points when the laser linewidth is not small compared to the subcarrier spacing Δf. A Monte Carlo simulator is proposed to analyze the BER loss due to CD in DDO-OFDM with coarse lasers. This confirms that CD strongly affects the performances of the transmission in Metro Access scenario, limiting to some kilometers ($L < 6$ km) the transmission of a 20 Gbit/s stream with a 16-QAM with a 10 MHz line laser. Note the exorbitant loss of performance when the laser linewidth of VCSEL begin to being comparable with the subcarrier spacing Δf. From these poor results for such strong linewidth laser, like VCSEL of Scenario 2, one can say that it is not worth anymore to implement a SSB transmission. Accordingly, a DSB transmission with no substantial losses compared to the SSB one can be performed. However, for a 1 MHz DFB laser the achievable length is $L < 800$ km. Still the results disagree with [Peng, 2010] because of the MSE being far bigger with the MC simulator. However, one should note that the MSE performance in Figure 5.5 shows that the derivation and the results in the literature coming from Equation (5.6) are not valid even if the laser linewidth $\gamma_{\text{PN}} \ll \Delta f$. This happens since the phase noise spectrum, which is convolved in the frequency domain with the OFDM signal Fourier Transform, widens the frequency components of the OFDM subcarriers. The signal is not frequency-impulsive anymore, hence Chromatic Dispersion does not only impose a delay, but also distortion that cannot be neglected.

CHAPTER 6

Conclusion

Several Wiener phase noise channel issues have been analyzed in this thesis work. From the investigation of validity the commonly assumed discrete-time model, Bayesian tracking is presented to derive the information rate of that channel and to design a complete iterative demodulation and decoding algorithm for it. Then the analysis is expanded to SSB DDO-OFDM transmission and CD impact on Wiener phase noise in such scenario. In the following pages the thesis' key points are resumed.

This thesis begins from a discussion about Wiener phase noise channels. Typically the performance of those channels is evaluated assuming a Discrete-time Model (DM) reported in Equation (1.1). At some point, we asked ourself if anyone has ever put a threshold of validity at that model especially when the transmission is impaired by strong phase noise. Since no works in the literature have dealt with the problem, in Chapter 2 the mismatch between two models is investigated. The Discrete-Time Frequency Noise (DTFN) processes of both the DM commonly assumed in the literature and the continuous-time Complete Model (CM) have been compared with two statistical tests about their whiteness and gaussianity. One can conclude that their statistical behavior agrees when the standard deviation of the DTFN increment σ_{PN} is below a threshold of DM validity $\overline{\sigma}_{\text{PN}} = 0.1$

rad. This result gives the author relief in claiming that DM can be almost always used to approximate CM, since normal values of σ_{PN} are way lower than the threshold of validity of the DM. Then the MSE due to the mismatch and the different BERs of CM and DM have been measured, proving that not only DM is a good approximation if the condition of validity is respected, being the mismatch error 20 dB below the power of the signals in the worst case, i.e. $\sigma_{\text{PN}} \to \overline{\sigma}_{\text{PN}}$ in Figure 2.6. Moreover DM assumption is conservative, being DM's BER is a little bit higher than CM's, see Figure 2.9. At last phase noise complex exponential PSDs are estimated. From the stronger high frequency components of DM one can conclude that the DM phase noise process is a bit more difficult to track with a carrier recovery system for its higher PSD components with high normalized frequencies.

After the validation of DM has been done, the next step was that of defining the its performance's limits. Being the DM a discrete-time continuous-state system, the tracking of the state is cast on the framework of state-based problems in the Bayesian tracking theory. In Chapter 3, Bayesian tracking applied to state-based approach is presented with some examples and a scenario where this theory shows its limits. In the discrete-time Wiener phase noise channel, phase noise is the hidden state that must be estimated. Bayesian tracking is introduced in Chapter 3, together with the example of Non linear optical channel propagation, where application of this theory does not provide meaningful results. Indeed, if one wants to model all the physical effects in the fiber with a multi-dimensional Markovian state, he has strong limits in term of computational capacity. Moreover, even if only two fiber sections can be simulated limiting the transmission to very short tracks, the approximation of CD and the channel impulse response given by the limited state totally compromises the whole analysis.

However, the single state DM does not give any of the problems above and it can be used together with Bayesian Tracking techniques. Those allow not only to track the hidden phase noise, but also to derive the information rate bounds of that channel. Accordingly, in Chapter 4, the Information rate of such channel is computed in Figure 4.1. One can conclude that with low to medium phase noise σ_{PN} the information rate at high SNR still achieve the AWGN upper bound after a different transient between DM and AWGN. However, if the phase noise is really huge, there is a lower floor at high SNR with respect to the AWGN. Bayesian tracking has allowed also to design a demodulation and decoding algorithm to achieve the predicted information rates. In contrast to previous published algorithms, the proposed one does not require Pilot-Aided symbols to bootstrap. Furthermore, the heavy computational requirements of the algorithm are reduced by smart

techniques that allow practical implementation.

Moving our analysis towards another applicative field of interest where phase noise impairs the transmission, in Chapter 5 the focus is set onto OFDM optical transmission. Direct-Detection systems are considered to elude the Wiener phase noise and carrier recovery. However, when dispersive compensating fibers are not considered, CD has an impact on the phase noise that cannot be neglected. The author takes the literature analysis of this phenomenon [Peng, 2010] to a more complete mathematical derivation. Enlightening previous works weaknesses, the Semi-Analytical approach of the literature is compared to Monte Carlo simulations in order to have more realistic performance. A huge difference is found out, particularly with large linewidth lasers, e.g. VCSEL where the transmission is limited to few kilometers. With DFB lasers having smaller linewidths it is possible to achieve hundreds of kilometers transmissions, even if the results are worse than that obtained by Semi-Analytical analysis in the literature. The Monte Carlo simulator is at last compared with the experimental measurements of [Schmidt et al., 2008], finding a good match between the two.

As future works one can continue the analysis of Chapter 5 with experimental measurements of its own. Short reach and metro access networks with DFB transmission can be studied without dispersion compensating fibers. Moreover Data center application where VCSEL are strongly implemented can benefit from the analysis of CD and phase noise interactions. Since it has been concluded that VCSEL linewdiths impair so badly the performance, a possible investigation of Dual Side-Band OFDM solutions can be done.

Bibliography

[Agrawal, 2007] Agrawal, G. P. (2007). *Nonlinear fiber optics*. Academic press.

[Alves et al., 2014] Alves, T. M., Alberto, A., Cartaxo, A. V., et al. (2014). Direct-detection multi-band ofdm metro networks employing virtual carriers and low receiver bandwidth. In *Optical Fiber Communications Conference and Exhibition (OFC), 2014*, pages 1–3. IEEE.

[Amann et al., 2012] Amann, M.-C., Wong, E., and Mueller, M. (2012). Energy-efficient high-speed short-cavity vcsels. In *Optical Fiber Communication Conference*, pages OTh4F–1. Optical Society of America.

[Amblard et al., 2003] Amblard, P.-O., Brossier, J.-M., and Moisan, E. (2003). Phase tracking: what do we gain from optimality? particle filtering versus phase-locked loops. *Signal Processing*, 83(1):151–167.

[Arizono and Ohta, 1989] Arizono, I. and Ohta, H. (1989). A test for normality based on kullback-leibler information. *The American Statistician*, 43(1):20–22.

[Armstrong and Lowery, 2006] Armstrong, J. and Lowery, A. (2006). Power efficient optical ofdm. *Electronics Letters*, 42(6):370–372.

[Arulampalam et al., 2002] Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188.

[Ascheid and Meyr, 1982] Ascheid, G. and Meyr, H. (1982). Cycle slips in phase-locked loops: A tutorial survey. *Communications, IEEE Transactions on*, 30(10):2228–2241.

[Bahl et al., 1974] Bahl, L., Cocke, J., Jelinek, F., and Raviv, J. (1974). Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *Information Theory, IEEE Transactions on*, 20(2):284–287.

[Barletta et al., 2013] Barletta, L., Bergamelli, F., Magarini, M., Carapellese, N., and Spalvieri, A. (2013). Pilot-aided trellis-based demodulation. *Photonics Technology Letters, IEEE*, 25(13):1234–1237.

[Barletta and Kramer, 2014a] Barletta, L. and Kramer, G. (2014a). On continuous-time white phase noise channels. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 2426–2429. IEEE.

[Barletta and Kramer, 2014b] Barletta, L. and Kramer, G. (2014b). Signal-to-noise ratio penalties for continuous-time phase noise channels. In *Int. Conf. on Cognitive Rad. Oriented Wireless Net.(CROWN)*.

[Barletta and Kramer, 2015] Barletta, L. and Kramer, G. (2015). Lower bound on the capacity of continuous-time wiener phase noise channels. *arXiv preprint arXiv:1503.03223*.

[Barletta et al., 2011] Barletta, L., Magarini, M., and Spalvieri, A. (2011). Estimate of information rates of discrete-time first-order markov phase noise channels. *Photonics Technology Letters, IEEE*, 23(21):1582–1584.

[Barletta et al., 2012a] Barletta, L., Magarini, M., and Spalvieri, A. (2012a). The information rate transferred through the discrete-time wiener's phase noise channel. *Journal of Lightwave Technology*, 30(10):1480–1486.

[Barletta et al., 2012b] Barletta, L., Magarini, M., and Spalvieri, A. (2012b). A new lower bound below the information rate of wiener phase noise channel based on kalman carrier recovery. *Optics express*, 20(23):25471–25477.

[Barletta et al., 2012c] Barletta, L., Magarini, M., and Spalvieri, A. (2012c). Staged demodulation and decoding. *Optics express*, 20(21):23728–23734.

[Beermann et al., 2011] Beermann, M., Breddermann, T., and Vary, P. (2011). Rate-compatible ldpc codes using optimized dummy bit insertion. In *Wireless Communication Systems (ISWCS), 2011 8th International Symposium on*, pages 447–451. IEEE.

[Bellantoni and Dodge, 1967] Bellantoni, J. and Dodge, K. (1967). A square root formulation of the kalman-schmidt filter. *AIAA journal*, 5(7):1309–1314.

[Beppu et al., 2015] Beppu, S., Kasai, K., Yoshida, M., and Nakazawa, M. (2015). 2048 qam (66 gbit/s) single-carrier coherent optical transmission over 150 km with a potential se of 15.3 bit/s/hz. *Optics express*, 23(4):4960–4969.

[Bisplinghoff et al., 2011] Bisplinghoff, A., Cabirol, C., Langenbach, S., Sauer-Greff, W., and Schmauss, B. (2011). Soft decision metrics for differentially encoded qpsk. In *European Conference and Exposition on Optical Communications*, pages Tu–6. Optical Society of America.

[Blackman, 2004] Blackman, S. S. (2004). Multiple hypothesis tracking for multiple target tracking. *Aerospace and Electronic Systems Magazine, IEEE*, 19(1):5–18.

[Candy, 2007] Candy, J. V. (2007). Bootstrap particle filtering. *Signal Processing Magazine, IEEE*, 24(4):73–85.

[Caron et al., 2007] Caron, F., Davy, M., Duflos, E., and Vanheeghe, P. (2007). Particle filtering for multisensor data fusion with switching observation models: Application to land vehicle positioning. *Signal Processing, IEEE Transactions on*, 55(6):2703–2719.

[Carpenter et al., 1999] Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):2–7.

[Christiansen, 1994] Christiansen, G. S. (1994). Modeling of prml timing loop as a kalman filter. In *Global Telecommunications Conference, 1994. GLOBECOM'94. Communications: The Global Bridge., IEEE*, volume 2, pages 1157–1161. IEEE.

[Colavolpe et al., 2005] Colavolpe, G., Barbieri, A., and Caire, G. (2005). Algorithms for iterative decoding in the presence of strong phase noise. *Selected Areas in Communications, IEEE Journal on*, 23(9):1748–1757.

[Creal, 2012] Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric Reviews*, 31(3):245–296.

[Dauwels and Loeliger, 2008] Dauwels, J. and Loeliger, H.-A. (2008). Computation of information rates by particle methods. *IEEE Transactions on Information Theory*, 54(1):406–409.

[Del Moral, 1996] Del Moral, P. (1996). Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4):555–581.

[Demir et al., 2000] Demir, A., Mehrotra, A., and Roychowdhury, J. (2000). Phase noise in oscillators: a unifying theory and numerical methods for characterization. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 47(5):655–674.

[Djurić et al., 2003] Djurić, P. M., Kotecha, J. H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F., and Miguez, J. (2003). Particle filtering. *Signal Processing Magazine, IEEE*, 20(5):19–38.

[Driessen, 1994] Driessen, P. F. (1994). Dpll bit synchronizer with rapid acquisition using adaptive kalman filtering techniques. *IEEE Transactions on Communications*, 42(9):2673–2675.

[Duffie, 2010] Duffie, D. (2010). *Dynamic asset pricing theory*. Princeton University Press.

[Dunn and Clark, 1986] Dunn, O. J. and Clark, V. A. (1986). *Applied statistics: analysis of variance and regression*. John Wiley & Sons, Inc.

[Essiambre et al., 2010] Essiambre, R., Kramer, G., Winzer, P., Foschini, G., and Goebel, B. (2010). Capacity limits of optical fiber networks. *Lightwave Technology, Journal of*, 28(4):662–701.

[EyuboĞlu, 1988] EyuboĞlu, M. V. (1988). Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving. *Communications, IEEE Transactions on*, 36(4):401–409.

[Fatadin et al., 2009] Fatadin, I., Ives, D., and Savory, S. J. (2009). Blind equalization and carrier phase recovery in a 16-qam optical coherent system. *Journal of lightwave technology*, 27(15):3042–3049.

[Fearnhead, 2005] Fearnhead, P. (2005). Using random quasi-monte-carlo within particle filters, with application to financial time series. *Journal of Computational and Graphical Statistics*, 14(4):751–769.

[Foschini and Vannucci, 1988] Foschini, G. J. and Vannucci, G. (1988). Characterizing filtered light waves corrupted by phase noise. *Information Theory, IEEE Transactions on*, 34(6):1437–1448.

[Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis.

[Ghozlan and Kramer, 2013a] Ghozlan, H. and Kramer, G. (2013a). Multi-sample receivers increase information rates for wiener phase noise channels. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pages 1897–1902. IEEE.

[Ghozlan and Kramer, 2013b] Ghozlan, H. and Kramer, G. (2013b). On wiener phase noise channels at high signal-to-noise ratio. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2279–2283. IEEE.

[Goebel et al., 2011] Goebel, B., Essiambre, R.-J., Kramer, G., Winzer, P. J., and Hanik, N. (2011). Calculation of mutual information for partially coherent gaussian channels with applications to fiber optics. *Information Theory, IEEE Transactions on*, 57(9):5720–5736.

[Gray, 1996] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

[Hager and Belhumeur, 1998] Hager, G. D. and Belhumeur, P. N. (1998). Efficient region tracking with parametric models of geometry and illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(10):1025–1039.

[Haykin, 2004] Haykin, S. (2004). *Kalman filtering and neural networks*, volume 47. John Wiley & Sons.

[Hlinka et al., 2013] Hlinka, O., Hlawatsch, F., and Djuric, P. M. (2013). Distributed particle filtering in agent networks: A survey, classification, and comparison. *Signal Processing Magazine, IEEE*, 30(1):61–81.

[Hofmann and Bimberg, 2012] Hofmann, W. and Bimberg, D. (2012). Vcsel-based light sources-scalability challenges for vcsel-based multi 100-gb/s systems. *Photonics Journal, IEEE*, 4(5):1831–1843.

[Hofmann et al., 2012] Hofmann, W. H., Moser, P., and Bimberg, D. (2012). Energy-efficient vcsels for interconnects. *Photonics Journal, IEEE*, 4(2):652–656.

[Isard and Blake, 1998] Isard, M. and Blake, A. (1998). Condensation-conditional density propagation for visual tracking. *International journal of computer vision*, 29(1):5–28.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45.

[Kamiya and Sasaki, 2013] Kamiya, N. and Sasaki, E. (2013). Pilot-symbol assisted and code-aided phase error estimation for high-order qam transmission. *Communications, IEEE Transactions on*, 61(10):4369–4380.

[Kim et al., 1994] Kim, Y.-R., Sul, S.-K., and Park, M.-H. (1994). Speed sensorless vector control of induction motor using extended kalman filter. *Industry Applications, IEEE Transactions on*, 30(5):1225–1233.

[Koizumi et al., 2012] Koizumi, Y., Toyoda, K., Yoshida, M., and Nakazawa, M. (2012). 1024 qam (60 gbit/s) single-carrier coherent optical transmission over 150 km. *Optics express*, 20(11):12508–12514.

[Kwon and Lee, 2010] Kwon, J. and Lee, K. M. (2010). Visual tracking decomposition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1269–1276. IEEE.

[Le et al., 2014] Le, S. T., Kanesan, T., Giacoumidis, E., Doran, N. J., and Ellis, A. D. (2014). Quasi-pilot aided phase noise estimation for coherent optical ofdm systems. *Photonics Technology Letters, IEEE*, 26(5):504–507.

[Leoni et al., 2012] Leoni, P., Sleiffer, V., Calabrò, S., Kuschnerov, M., Jansen, S. L., Spinnler, B., and Lankl, B. (2012). On the performance of a soft decision fec scheme operating in highly non-linear regime. In *Signal Processing in Photonic Communications*, pages SpTu3A–6. Optical Society of America.

[Li and Jilkov, 2003] Li, X. R. and Jilkov, V. P. (2003). Survey of maneuvering target tracking. part i. dynamic models. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1333–1364.

[Liu and Bar-Ness, 2006] Liu, P. and Bar-Ness, Y. (2006). Closed-form expressions for ber performance in ofdm systems with phase noise. In *Communications, 2006. ICC'06. IEEE International Conference on*, volume 12, pages 5366–5370. IEEE.

[Lopes and Tsay, 2011] Lopes, H. F. and Tsay, R. S. (2011). Particle filters and bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168–209.

[Ma et al., 2009] Ma, Y., Yang, Q., Tang, Y., Chen, S., and Shieh, W. (2009). 1-tb/s per channel coherent optical ofdm transmission with subwavelength bandwidth access. In *National Fiber Optic Engineers Conference*, page PDPC1. Optical Society of America.

[MacKay, 2003] MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

[Magarini et al., 2012a] Magarini, M., Barletta, L., Spalvieri, A., Leven, A., Pepe, M., and Gavioli, G. (2012a). Impact of nonideal phase reference on soft decoding of differentially encoded modulation. *Photonics Technology Letters, IEEE*, 24(23):2179–2182.

[Magarini et al., 2012b] Magarini, M., Barletta, L., Spalvieri, A., Vacondio, F., Pfau, T., Pepe, M., Bertolini, M., and Gavioli, G. (2012b). Pilot-symbols-aided carrier-phase recovery for 100-g pm-qpsk digital coherent receivers. *Photonics Technology Letters, IEEE*, 24(9):739–741.

[Magarini et al., 2011] Magarini, M., Spalvieri, A., Vacondio, F., Bertolini, M., Pepe, M., and Gavioli, G. (2011). Empirical modeling and simulation of phase noise in long-haul coherent optical transmission systems. *Opt. Express*, 19(23):22455–22461.

[Mandelli et al., 2014] Mandelli, S., Magarini, M., and Spalvieri, A. (2014). Modeling the filtered and sampled continuous-time signal affected by wiener phase noise. In *Networks and Optical Communications-(NOC), 2014 19th European Conference on*, pages 173–178. IEEE.

[Mandelli et al., 2015] Mandelli, S., Magarini, M., Spalvieri, A., and Pecorino, S. (2015). On discrete-time modeling of the filtered and symbol-rate sampled continuous-time signal affected by wiener phase noise. *Optical Switching and Networking*.

[Marins et al., 2001] Marins, J. L., Yun, X., Bachmann, E. R., McGhee, R. B., and Zyda, M. J. (2001). An extended kalman filter for quaternion-based orientation estimation using marg sensors. In *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, volume 4, pages 2003–2011. IEEE.

[Mengali, 1997] Mengali, U. (1997). *Synchronization Techniques for Digital Receivers*. Applications of Communications Theory. Springer US.

[Meyr and Ascheid, 1990] Meyr, H. and Ascheid, G. (1990). *Synchronization in Digital Communications: Phase-, frequency-locked loops and amplitude control*, volume 1. Wiley.

[Minsky, 1967] Minsky, M. L. (1967). *Computation: finite and infinite machines*. Prentice-Hall, Inc.

[Morsy-Osman et al., 2011] Morsy-Osman, M., Zhuge, Q., Chen, L. R., and Plant, D. V. (2011). Feedforward carrier recovery via pilot-aided transmission for single-carrier systems with arbitrary m-qam constellations. *Optics express*, 19(24):24331–24343.

[Murray and Burmaster, 1995] Murray, D. M. and Burmaster, D. E. (1995). Residential air exchange rates in the united states: empirical and estimated parametric distributions by season and climatic region. *Risk Analysis*, 15(4):459–465.

[Nguyen and Shwedyk, 2009] Nguyen, H. H. and Shwedyk, E. (2009). *A First Course in Digital Communications*. Cambridge University Press.

[Nummiaro et al., 2003] Nummiaro, K., Koller-Meier, E., and Van Gool, L. (2003). An adaptive color-based particle filter. *Image and vision computing*, 21(1):99–110.

[Okuma et al., 2004] Okuma, K., Taleghani, A., De Freitas, N., Little, J. J., and Lowe, D. G. (2004). A boosted particle filter: Multitarget detection and tracking. In *Computer Vision-ECCV 2004*, pages 28–39. Springer.

[Patapoutian, 1999] Patapoutian, A. (1999). On phase-locked loops and kalman filters. *Communications, IEEE Transactions on*, 47(5):670–672.

[Patapoutian, 2002] Patapoutian, A. (2002). Application of kalman filters with a loop delay in synchronization. *Communications, IEEE Transactions on*, 50(5):703–706.

[Pattan, 2000] Pattan, B. (2000). *Robust Modulation Methods and Smart Antennas in Wireless Communications*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

[Pecorino et al., 2015] Pecorino, S., Mandelli, S., Barletta, L., Magarini, M., and Spalvieri, A. (2015). Bootstrapping iterative demodulation and decoding without pilot symbols. *Journal of Lightwave Technology*, 33(17):3613–3622.

[Peng, 2010] Peng, W.-R. (2010). Analysis of laser phase noise effect in direct-detection optical ofdm transmission. *Journal of Lightwave Technology*, 28(17):2526–2536.

[Peng et al., 2009a] Peng, W.-R., Wu, X., Arbab, V. R., Feng, K.-M., Shamee, B., Christen, L. C., Yang, J.-Y., Willner, A. E., and Chi, S. (2009a). Theoretical and experimental investigations of direct-detected rf-tone-assisted optical ofdm systems. *Journal of Lightwave Technology*, 27(10):1332–1339.

[Peng et al., 2009b] Peng, W.-R., Zhang, B., Feng, K.-M., Wu, X., Willner, A. E., and Chi, S. (2009b). Spectrally efficient direct-detected ofdm transmission incorporating a tunable frequency gap and an iterative detection techniques. *Lightwave Technology, Journal of*, 27(24):5723–5735.

[Perez et al., 2004] Perez, P., Vermaak, J., and Blake, A. (2004). Data fusion for visual tracking with particles. *Proceedings of the IEEE*, 92(3):495–513.

[Pfau et al., 2009] Pfau, T., Hoffmann, S., and Noé, R. (2009). Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m-qam constellations. *Journal of Lightwave Technology*, 27(8):989–999.

[Pham et al., 1998] Pham, D. T., Verron, J., and Roubaud, M. C. (1998). A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine systems*, 16(3):323–340.

[Pitt and Shephard, 1999] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599.

[Pouillot et al., 2010] Pouillot, R., Lubran, M. B., Cates, S. C., and Dennis, S. (2010). Estimating parametric distributions of storage time and temperature of ready-to-eat foods for us households. *Journal of Food Protection®*, 73(2):312–321.

[Punskaya et al., 2001] Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. J. (2001). Particle filtering for demodulation in fading channels with non-gaussian additive noise. *IEEE Transactions on Communications*, 49(4):579–582.

[Qi et al., 2006] Qi, G., Yao, J., Seregelyi, J., Paquet, S., Bélisle, C., Zhang, X., Wu, K., and Kashyap, R. (2006). Phase-noise analysis of optically generated millimeter-wave signals with external optical modulation techniques. *Journal of lightwave technology*, 24(12):4861–4875.

[Rabiner and Gold, 1975] Rabiner, L. R. and Gold, B. (1975). Theory and application of digital signal processing. *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1975. 777 p.*, 1.

[Rabiner et al., 1983] Rabiner, L. R., Levinson, S. E., and Sondhi, M. M. (1983). On the application of vector quantization and hidden markov models to speaker-independent, isolated word recognition. *Bell System Technical Journal, The*, 62(4):1075–1105.

[Rigatos, 2009] Rigatos, G. G. (2009). Particle filtering for state estimation in nonlinear industrial systems. *Instrumentation and measurement, IEEE transactions on*, 58(11):3885–3900.

[Ross et al., 2008] Ross, D. A., Lim, J., Lin, R.-S., and Yang, M.-H. (2008). Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3):125–141.

[Särkkä et al., 2007] Särkkä, S., Vehtari, A., and Lampinen, J. (2007). Rao-blackwellized particle filter for multiple target tracking. *Information Fusion*, 8(1):2–15.

[Sasiadek et al., 2000] Sasiadek, J., Wang, Q., and Zeremba, M. (2000). Fuzzy adaptive kalman filtering for ins/gps data fusion. In *Intelligent Control, 2000. Proceedings of the 2000 IEEE International Symposium on*, pages 181–186. IEEE.

[Schmidt et al., 2009] Schmidt, B., Zan, Z., Du, L., and Lowery, A. (2009). 100 gbit/s transmission using single-band direct-detection optical ofdm. In *Optical Fiber Communication - incudes post deadline papers, 2009. OFC 2009. Conference on*, pages 1–3.

[Schmidt et al., 2008] Schmidt, B. C., Lowery, A., and Armstrong, J. (2008). Experimental demonstrations of electronic dispersion compensation for long-haul transmission using direct-detection optical ofdm. *J. Lightwave Technol.*, 26(1):196–203.

[Schuster et al., 2008] Schuster, M., Randel, S., Bunge, C. A., Lee, S. C. J., Breyer, F., Spinnler, B., and Petermann, K. (2008). Spectrally efficient compatible single-sideband modulation for ofdm transmission with direct detection. *IEEE Photonics Technology Letters*, 20(9):670.

[Shieh et al., 2008] Shieh, W., Bao, H., and Tang, Y. (2008). Coherent optical ofdm: theory and design. *Optics Express*, 16(2):841–859.

[Silverman, 1986] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

[Simon, 2006] Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.

[Sinkin et al., 2003] Sinkin, O. V., Holzlöhner, R., Zweck, J., and Menyuk, C. R. (2003). Optimization of the split-step fourier method in modeling optical-fiber communications systems. *Lightwave Technology, Journal of*, 21(1):61–68.

[Smith et al., 2013] Smith, A., Doucet, A., de Freitas, N., and Gordon, N. (2013). *Sequential Monte Carlo methods in practice*. Springer Science & Business Media.

[Spalvieri, 2006] Spalvieri, A. (2006). Optimal loop filter of the discrete-time pll in the presence of phase noise. In *Computers and Communications, 2006. ISCC'06. Proceedings. 11th IEEE Symposium on*, pages 1013–1018. IEEE.

[Spalvieri and Barletta, 2011] Spalvieri, A. and Barletta, L. (2011). Pilot-aided carrier recovery in the presence of phase noise. *Communications, IEEE Transactions on*, 59(7):1966–1974.

[Spalvieri and Magarini, 2008] Spalvieri, A. and Magarini, M. (2008). Wiener's analysis of the discrete-time phase-locked loop with loop delay. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 55(6):596–600.

[Tamir et al., 2009] Tamir, D. E., Shaked, N. T., Wilson, P. J., and Dolev, S. (2009). High-speed and low-power electro-optical dsp coprocessor. *JOSA A*, 26(8):A11–A20.

[Tomba, 1998] Tomba, L. (1998). On the effect of wiener phase noise in ofdm systems. *Communications, IEEE Transactions on*, 46(5):580–583.

[Van Der Merwe et al., 2000] Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E. (2000). The unscented particle filter. In *NIPS*, pages 584–590.

[Walpole et al., 1993] Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (1993). *Probability and statistics for engineers and scientists*, volume 5. Macmillan New York.

[Wang and Papageorgiou, 2005] Wang, Y. and Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended kalman filter: a general approach. *Transportation Research Part B: Methodological*, 39(2):141–167.

[Wu and Bar-Ness, 2004] Wu, S. and Bar-Ness, Y. (2004). Ofdm systems in the presence of phase noise: consequences and solutions. *Communications, IEEE Transactions on*, 52(11):1988–1996.

[Yi et al., 2008] Yi, X., Shieh, W., and Ma, Y. (2008). Phase noise effects on high spectral efficiency coherent optical ofdm transmission. *Journal of Lightwave Technology*, 26(10):1309–1316.

[Zan et al., 2008] Zan, Z., Premaratne, M., and Lowery, A. (2008). Laser rin and linewidth requirements for direct detection optical ofdm. In *Lasers and Electro-Optics, 2008 and 2008 Conference on Quantum Electronics and Laser Science. CLEO/QELS 2008. Conference on*, pages 1–2.

[Zhang et al., 2012] Zhang, F., Li, Y., Wu, J., Li, W., Hong, X., and Lin, J. (2012). Improved pilot-aided optical carrier phase recovery for coherent-qam. *Photonics Technology Letters, IEEE*, 24(18):1577–1580.

[Zhou et al., 2004] Zhou, S. K., Chellappa, R., and Moghaddam, B. (2004). Visual tracking and recognition using appearance-adaptive models in particle filters. *Image Processing, IEEE Transactions on*, 13(11):1491–1506.