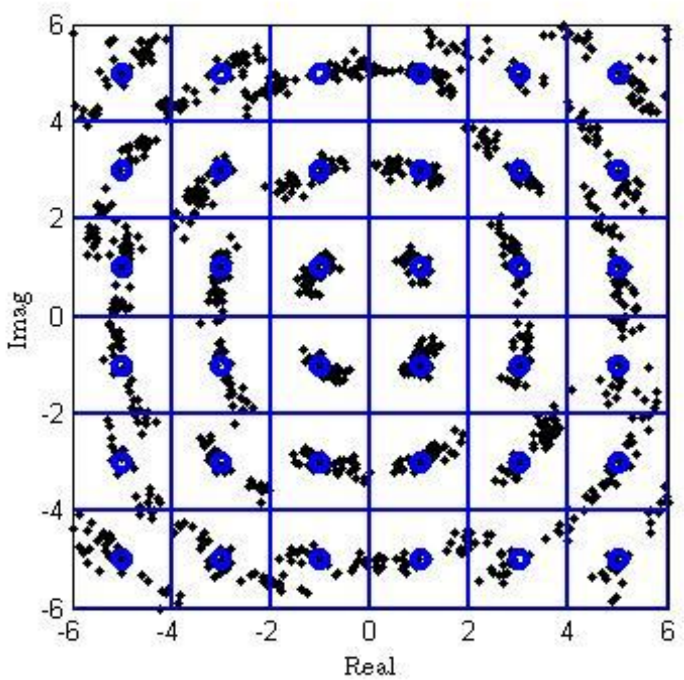
CHALMERS



Oscillator Phase Noise:
Capacity Bounds and Iterative Code-Aided Estimation

Master of Science Thesis in Communication Engineering

JOHAN SÖDER

Communication Systems and Information Theory
Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2010

Report No. EX025/2010

Thesis for the degree of Master of Science

Oscillator Phase Noise:

Capacity Bounds

and

Iterative Code-Aided Estimation

JOHAN SÖDER

CHALMERS

Communication Systems and Information Theory
Department of Signals and Systems
Chalmers University of Technology

Göteborg, Sweden 2010

Oscillator Phase Noise: Capacity Bounds and Code-Aided Iterative Estimation
JOHAN SÖDER

This thesis was prepared using LaTeX.

Communication Systems and Information Theory
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone + 46 (0)31-772 1000

Chalmers Reproservice
Göteborg, Sweden 2010

To Linda, my Baucis.

Abstract

Oscillator phase noise becomes a problem in digital communication systems operating at radio frequencies when the signal constellation is dense. Internal noise in the oscillators at the transmitter and receiver interacts with the transmitted data in a non-linear fashion, to cause a time-varying rotation of the signal space. In this thesis a receiver algorithm to counter the effects of the phase noise is derived and evaluated. The thesis also investigates theoretical performance bounds for communication with noisy oscillators.

In the first part of the thesis, the expectation-maximization framework from estimation theory is applied to the problem of phase noise estimation. A receiver structure is derived where the phase noise process is estimated jointly with data decoding. The algorithm works in cooperation with an iterative low-density parity-check code and uses soft information. Simulations are presented, showing large improvements in terms of bit error rates. The computational complexity of the derived algorithm is also reduced to implementable levels.

In the second part of the thesis, upper and lower bounds on the channel capacity, in terms of bits per channel use, are derived. The channel investigated consists of Wiener phase noise and additive white Gaussian noise (AWGN). The upper and lower bounds are shown to be tight, thus enclosing the true channel capacity. The channel capacity for a fixed level of phase noise follows the well known capacity for the AWGN channel, for low to medium signal to noise ratios (with respect to the additive noise). At some point, determined by the ratio of the additive noise variance and the innovation variance of the Wiener phase noise, the capacity for the phase noise constrained channel deviates from that of the channel impaired only by additive noise. In this region the increase in capacity gained by increasing the signal power is only 50% as compared with the channel without phase noise.

Contents

Acknowledgements

I would first and foremost like to thank my supervisor Associate Professor Thomas Eriksson for believing in me to have the stamina of pulling this off. I would also like to thank all the members of the Communication Systems and Information Theory group for being so open and friendly. Special thanks to Tilak for his help with introducing me to the C3SE cluster and for enjoying me with his crazy stories during the breaks.

My deepest gratitude to the people at the Microwave PDU, Ericsson AB, for letting me see how things work in "the real world". I thank Thomas Emanuelsson for inviting me and Sture Skalare for welcoming me into his department. Special thanks to Ph.D. Mats Rydström and Ph.D. Andreas Wolfgang, with Quamcom Technology AB, for many fruitful discussions as well as translations between the "company" and academia lingos. I would also like to thank Ph.D. Jan-Erik Porath, Ericsson, for supporting me while implementing the LDPC codes as well as taking such a big interest in my work. Thanks also to Helena Larsson for all the help with getting access cards, a desk with a computer etc., as well as joining me in the matlåde- lunches.

Last but not least a big thanks to friends and family for supporting me through all these years of studying.

Notation

Latin Symbols

Throughout this thesis, bold symbols denote vectors.

b	Uncoded bit
c	Coded bit
C	Channel capacity
E_b	Average bit energy
E_s	Average signal/symbol energy
f_c	Carrier frequency
f_{3dB}	3dB single-sided bandwidth of the oscillator PSD
$h(\cdot)$	Entropy
$I(\cdot;\cdot)$	Mutual information
$\mathbf{I}_n$	Identity matrix, $n \times n$
$\mathcal{L}(f)$	Single-sideband phase noise power
$n(t)$	Internal oscillator noise process
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and variance σ
$\mathcal{N}_{\mathcal{C}}(\mu, \sigma)$	Circular symmetric complex Normal distribution with mean μ and variance σ
P_{b_i}	Probability that bit i is a 1
P_c	Total power in carrier
r	Received (complex) symbol or received (real) amplitude (Paper B)
R	Transmitted amplitude
$R_x(\tau)$	Oscillator ACF
s	Transmitted (complex) symbol
$S_x(f)$	Oscillator PSD
T_s	Symbol interval
v_k	Rescaled discrete AWGN
w	Discrete AWGN

Greek Symbols

α	Soft symbol
Δ	Discrete innovation step
θ	Received phase (Paper B)
Θ	Transmitted phase
$\phi(t)$	Phase noise process, continuous time
ϕ_k	Phase noise process, discrete time
σ_Δ^2	Innovation variance, discrete phase noise process
σ_N^2	AWGN variance

Operators

$\arg\{\cdot\}$,$\angle$	Complex argument
$\lvert\cdot\rvert$	Absolute value
$\lvert\lvert\cdot\rvert\rvert^2$	Power
$\mathrm{Re}\{\cdot\}$	Real part
$\mathrm{Im}\{\cdot\}$	Imaginary part
$E[\cdot]$	Expected value
$\arg\max$	Maximizing argument

Superscripts

$\wedge$	Estimate
n	Iteration number
$-$	Average
$*$	Complex conjugate
H	Complex conjugate and transpose (Hermitian)
$\sim$	Vector of past three samples
Tx	Transmitter
Rx	Receiver

Subscripts

k,n	Discrete time index
$\parallel$	Projection onto transmitted vector
$\perp$	Orthogonal projection

Abbreviations

ACF	Autocorrelation Function
AWGN	Additive White Gaussian Noise
BER	Bit Error-Rate
LDPC	Low Density Parity-Check (Code)
LLR	Log-Likelihood Ratio
ML	Maximum Likelihood
PN	Phase Noise
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
SNR	Signal-to-Noise Ratio
SSB	Single Sideband

Part I

Introduction

Chapter 1

Background

The current evolution of the next generation wireless communication systems has as its main target a tenfold increase in the data rates offered to the users. To provide this, the pressure on the backbone networks is increasing accordingly. When pushing the systems to meet these requirements, limitations in the hardware reveal themselves. Links in the communication chain that has previously been approximated as behaving ideally can no longer be considered to do so. Fettweis [1] introduces the term "Dirty RF" and talks about a shift of paradigm when the analog (RF) and the digital design domains can no longer be kept separated. One of these "dirty" components is the local oscillator, transferring the baseband signal to the desired passband frequency registered for communication.

The hardware equipment is of course constantly being improved, but there will always be limitations. Also the improvements come at a cost, and one of the largest challenges of the communication systems of tomorrow is to deliver cost efficient services. A cheap solution to hardware obstacles is digital signal processing. Signal processing may be done in both the transmitter chain, e.g., predistortion, and in the receiving end, e.g, signal filtering. The focus for this thesis lies on the receiver structure.

This study derives and evaluates a digital signal processing algorithm for dealing with time varying phase noise from noisy oscillators. The algorithm works iteratively with soft information, in accordance with the "Turbo principle" employed by modern error correcting decoders. The algorithm is aided by working jointly with a low-density parity-check (LDPC) code to improve the estimates of the disturbing phase noise process.

To evaluate the performance of an algorithm, some benchmark tool is needed. This is provided by information theory and the channel capacity. Up until two decades ago channel capacity was merely theory; the achievable limits were far from what any actual systems could provide. In the beginning of the 1990s, however, the turbo codes were invented and the LDPC codes were rediscovered [2]. By the use of these error correcting codes, capacity-approaching results were obtained by real systems. Since then, information theory provides a benchmark tool for comparison. Knowing the possible achievable performance also provides designers and developers with knowledge of where to spend resources on further improving their systems. In this thesis capacity bounds are derived for a channel suffering from time varying oscillator phase noise.

1.1 Thesis Outline

Part I of the thesis provides some knowledge of oscillator phase noise and the impact it has on communication systems. Chapter 2 gives a short introduction to stochastic modeling of oscillator phase noise and Chapter 3 shows how the noise in the local oscillators impacts the digital communication system. In Chapter 4 the contributions of this thesis are summarized, and some suggestions for future work are given.

The main work of this thesis is provided in Part II, in the form of two included research papers.

Chapter 2

Phase Noise Modeling

2.1 Mathematical Oscillator Model

When communicating in a certain frequency band, the baseband pulses containing the information are carried by a high frequency radio wave. This wave has a frequency many times higher than the rate of the communication system, typically in the Gigahertz range. Denote the baseband signal $s(t)$ and the oscillator signal $x(t)$, the transmitted signal in baseband notation is then,

$$s_{\mathrm{Tx}}(t) = s(t)x(t). \tag{2.1}$$

Ideally the oscillator signal looks like,

$$x(t) = A\exp[j2\pi f_c t], \tag{2.2}$$

for the carrier frequency f_c and amplitude A. However, due to hardware impairments, a real oscillator is better modeled as

$$x(t) = (A + a(t))\exp\left[j(2\pi f_c t + \phi(t))\right], \tag{2.3}$$

where $a(t)$ is amplitude noise and $\phi(t)$ is phase noise.

A local oscillator is a self-resonating circuit. To mathematically analyze the noise terms this circuit may be modeled by a stochastic differential equation. This has been done very rigorously by Demir *et. al.* in [3] and [4]. They show that the amplitude noise decays over time, since the system is self stabilizing. The amplitude noise may thus be ignored and the following, normalized, oscillator model will be used from here on.

$$x(t) = \exp\left[j(2\pi f_c t + \phi(t))\right] = e^{j2\pi f_c}\, e^{j\phi(t)}. \tag{2.4}$$

This process is stationary, due to the wrapping of the phase. We denote the power spectral density (PSD) of the process $S_x(f)$, and the autocorrelation function (ACF) $R_x(\tau)$. It is immediately obvious that the oscillator PSD is determined by the spectrum of $\exp[j\phi(t)]$, translated to the carrier frequency, f_c.

When measuring the phase noise properties of an oscillator, the spectrum, $S_x(f)$ may be obtained through a spectrum analyzer [5]. This directly gives the single sideband phase noise power, $\mathcal{L}(f)$, a performance measure commonly used in practice. It is defined as the noise power in a 1Hz sideband at offset f from the carrier, divided by the total power in the carrier, P_c. The unit is dBc, decibel relative to carrier,

$$\mathcal{L}(f) = 10\log_{10}\frac{S_x(f_c+f)}{P_c}. \tag{2.5}$$

2.2 Phase Noise Model

Regarding disturbances in the phase of the oscillator, every time shifted oscillator is still a solution to the stochastic differential equation and there is nothing restoring the phase shift. Phase shifts thus accumulate over time and may be modeled as an integration of disturbances, [3],

$$\phi(t) = \int_0^t n(s)\,ds. \tag{2.6}$$

It is the characteristics of the the internal noise in the oscillator circuit, $n(t)$, that determines the behavior of the phase noise. It will be assumed to be a stationary Gaussian process with PSD $S_n(f)$.

If $n(t)$ is a white process, $S_n(f) = C$, the phase noise process is the Wiener process (also known as a Brownian motion). It was shown in [3] that for this case the oscillator spectrum is a Lorentzian,

$$\mathcal{L}(f) = 10\log_{10}\frac{1/\pi f_{3\mathrm{dB}}}{1+\left(\frac{f}{f_{3\mathrm{dB}}}\right)^2}. \tag{2.7}$$

The spectrum is plotted in Figure 2.1. It is characterized by a single parameter, the 3db bandwidth, $f_{3\mathrm{dB}}$. The autocorrelation function is [4],

$$R_x(\tau) = e^{-2\pi f_{3\mathrm{dB}}\tau}. \tag{2.8}$$

2.3 Discrete Phase Noise

In a communication system the phase noise process will be sampled every T_s seconds, the transmission time interval. The discrete phase noise process is defined as,

$$\phi_k \triangleq \phi(kT_s). \tag{2.9}$$

In accordance with (2.6) the process may be written as a random walk,

$$\phi_k = \phi_{k-1} + \Delta_k. \tag{2.10}$$

Since the oscillator noise, $n(t)$, is assumed to be a Gaussian process, the innovation term, Δ_k, will be a discrete Gaussian random variable,

$$\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2). \tag{2.11}$$

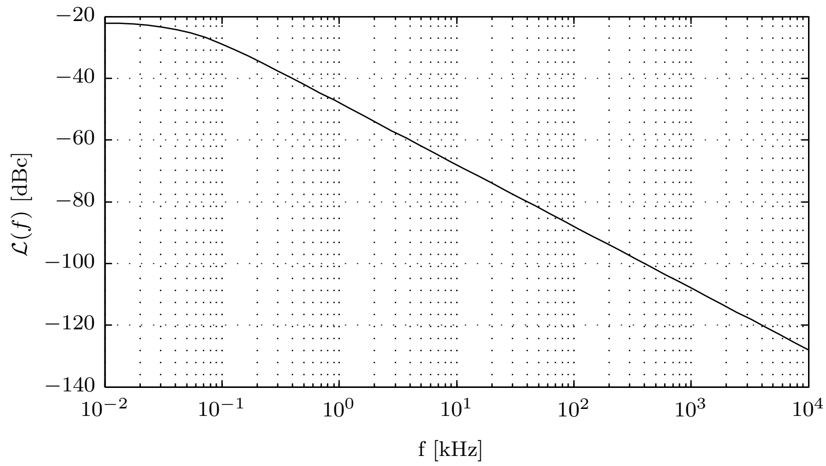


Figure 2.1: Lorentzian SSB Phase Noise power, $f_{3db} = 50\text{Hz}$

The variance can be related to the stochastic properties of the oscillator as,

$$\sigma_\Delta^2 = -2 \ln R_x(T_s). \tag{2.12}$$

If $n(t)$ is white the variance will be, from (2.8) and (2.12)

$$\sigma_\Delta^2 = 4\pi f_{3\text{dB}} T_s, \tag{2.13}$$

where $f_{3\text{dB}}$ is, as before, the single-sided 3dB bandwidth of the oscillator spectrum, and T_s is the symbol transmission interval. The discrete innovation process will of course also be white,

$$E[\Delta_k \Delta_l] = 0, k \neq l. \tag{2.14}$$

The phase noise model defined by (2.10), (2.11) and (2.14) will be referred to as the discrete Wiener phase noise model. It is restated in boxed equation 2.15 for clarity. A realization of Wiener phase noise is plotted in Figure 2.2.

Wiener Phase Noise model:

$$\begin{aligned} &\phi_k = \phi_{k-1} + \Delta_k \\ &\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2) \\ &E[\Delta_k \Delta_l] = 0, k \neq l \end{aligned} \tag{2.15}$$

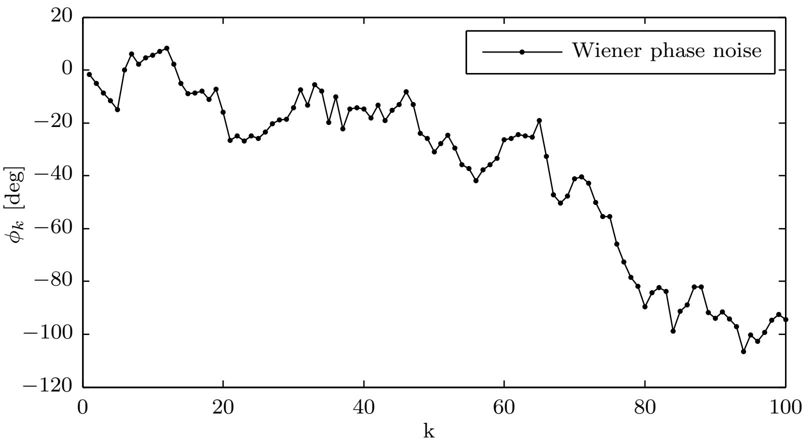


Figure 2.2: Wiener phase noise, $\sigma_\Delta = 6^\circ$

Chapter 3

Digital Communication with Phase Noise

In a digital communication system a pulse shape is transmitted every T_s seconds, the symbol transmission rate. The transmitted pulse shape reaching the receiver at time kT_s is represented in baseband notation by a complex symbol, s_k. The symbols are chosen from a discrete set of points in the complex plane, $s_k \in \mathcal{S} \in \mathbb{C}$. $\mathcal{S}$ is the signal space, e.g., an M-QAM constellation as displayed in Figure 3.1, see, e.g., [6] for the basics of digital communication.

A simple channel modeling thermal noise affecting the signal is the additive white Gaussian noise (AWGN) channel. The received symbol at time kT_s is,

$$r_k = s_k + \tilde{w_k}. \tag{3.1}$$

The additive noise, $\tilde{w_k}$, is complex, circular symmetric, white Gaussian noise with variance σ_N^2 per dimension,

$$\begin{aligned} \tilde{w_k} &\sim \mathcal{N}_{\mathcal{C}}(0, 2\sigma_N^2) \\ E[\tilde{w_k}\tilde{w_l}] &= 0, k \neq l. \end{aligned} \tag{3.2}$$

An example of received symbols from the AWGN channel are plotted in Figure 3.2.

To include the effects of oscillator phase noise the transmitted and received symbols are multiplied by a complex phasor, $e^{j\phi}$, see [7] for details. This will cause a rotation of the signal space, illustrated in Figure 3.3.

Let ϕ_k^{Tx} and ϕ_k^{Rx} denote the discrete phase noise sample at the transmitter and the receiver, respectively. The received signal at discrete time k will now be

$$r_k = (s_k e^{j\phi_k^{\mathrm{Tx}}} + \tilde{w_k})e^{j\phi_k^{\mathrm{Rx}}} \tag{3.3}$$

$$= s_k e^{j(\phi_k^{\mathrm{Tx}} + \phi_k^{\mathrm{Rx}})} + w_k \tag{3.4}$$

$$= s_k e^{j\phi_k} + w_k. \tag{3.5}$$

In (3.4) $w_k \triangleq \tilde{w_k} \exp[j\phi_k^{\mathrm{Rx}}]$. A rotation of the circular symmetric additive noise does not change the stochastic properties, so w_k has the same probabilistic definition as $\tilde{w_k}$ in (3.2). In (3.5) the total phase noise process is introduced,

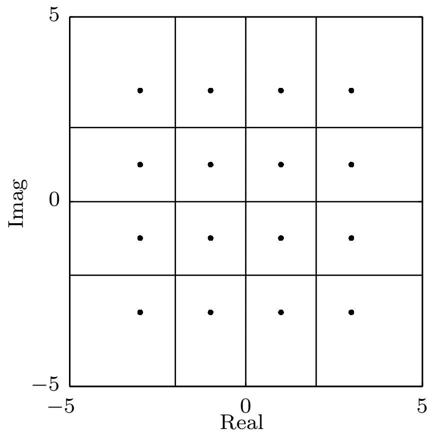


Figure 3.1: 16-QAM signal constellation

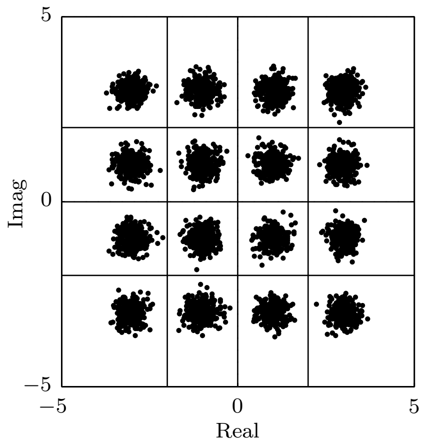


Figure 3.2: Received constellation affected by AWGN

$\phi_k \triangleq \phi_k^{\text{Tx}} + \phi_k^{\text{Rx}}$. The result of this channel containing AWGN and Wiener phase noise is displayed in Figure 3.4. It is also depicted as a block diagram in Figure 3.5.

As seen in (3.5) the transmitter and receiver phase noise samples add together and may be described by one process, the properties of which will now be examined more closely. By the random walk statement of the discrete phase noise process (2.10),

$$\begin{aligned}\phi_k &= \phi_k^{\text{Tx}} + \phi_k^{\text{Rx}} && (3.6)\\ &= \sum_{l=0}^{k} \Delta_l^{\text{Tx}} + \sum_{l=0}^{k} \Delta_l^{\text{Rx}} && (3.7)\\ &= \sum_{l=0}^{k} (\Delta_l^{\text{Tx}} + \Delta_l^{\text{Rx}}) && (3.8)\\ &\triangleq \sum_{l=0}^{k} \Delta_l. && (3.9)\end{aligned}$$

So the total phase noise process is also a discrete Wiener process. Assuming independence between the transmitter and receiver oscillators, the innovation variance for the total phase noise process will be the sum of the variances,

$$\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2(\text{Tx}) + \sigma_\Delta^2(\text{Rx})). \quad (3.10)$$

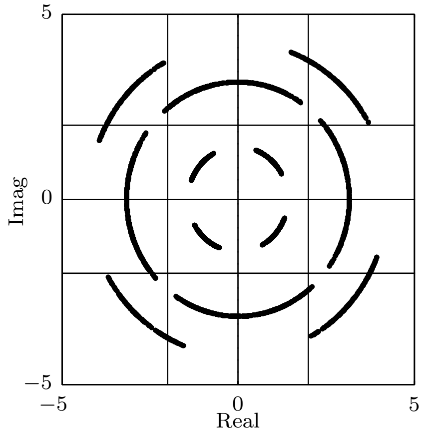


Figure 3.3: Constellation rotated by PN

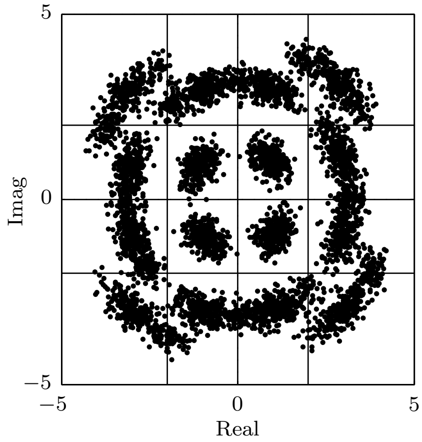


Figure 3.4: Constellation affected by AWGN and PN

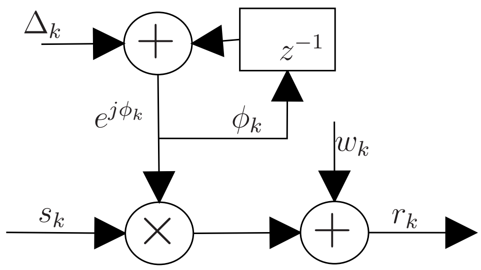


Figure 3.5: Block diagram of channel model

Chapter 4

Contributions and Future Work

This chapter summarizes the contributions made in the included papers. Some directions and suggestions for future work are also given.

4.1 Paper A

4.1.1 Summary

In Paper A the problem of estimating the phase noise process at the receiver is considered. By doing so the constellation may be de-rotated, in order to alleviate the performance decreasing effects of the phase noise. Since the phase noise interacts non-linearly with the transmitted data the unknown data sequence is considered as a nuisance parameter. The expectation-maximization framework is employed to find the ML-estimate of the phase noise process over a block of data consisting of one codeword. The error correcting code used is an LDPC code.

The derivation shows that the soft information iterated in the LDPC decoder may be used to compute estimates of the phase noise process. These two operations, decoding and tracking, should then work iteratively to further improve both the data and the phase noise estimates.

The algorithm was simulated and shown to give good results. Also some propositions lowering the computational complexity to reasonable levels were investigated.

4.1.2 Future Work

Complexity

The computational complexity of the proposed algorithm should be evaluated more thoroughly. First of all the optimal operating settings, in terms of low complexity, while maintaining performance, should be found. Then the increased complexity of the proposed scheme as compared to a conventional (non-iterative) phase noise tracker could be derived analytically.

BER Floor?

To see the performance at lower bit-error rates the algorithm should be simulated at a higher signal-to-noise ratio. This will show if the BER curve follows the curve for a system without phase noise, which looks like a waterfall, or if it will floor out. In order to do this the algorithm must be implemented in C-code or something similar to obtain reasonable simulating times.

Colored phase noise innovation

The model used was the Wiener phase noise model. The problem could be extended to tracking time varying phase noise with colored (correlated) innovation steps. This would not change the overall structure of the algorithm. The only thing needed is to extend the Kalman filter with more states. These states could be the covariance matrix for the noise, or the phase noise could be reformulated as an AR-process and the states would then be the taps.

4.2 Paper B

4.2.1 Summary

Paper B derives upper and lower bounds on the capacity for the AWGN and Wiener phase noise channel. The approach consists of stating the input-output relations of the phase and amplitude of the signal, and then making a high SNR approximation. To resolve the infinite memory of the Wiener phase noise process, differential decoding of the phase is employed. To find the optimal distributions an optimization theorem of functionals from the calculus of variations is used.

The results are expressions from which the capacity bounds may be numerically found for given levels of phase noise and signal-to-noise ratios. The channel capacity for the limiting cases when the phase noise dominates the additive noise, and vice versa, are also derived to ensure consistency.

The numerical results show that the upper and lower bounds are close to each other, thus enclosing the channel capacity. For low enough levels of phase noise the capacity is the same as for the AWGN channel. When the phase noise becomes a limiting factor, however, the achievable rate increase gained by increasing the signal-to-noise ratio rapidly decreases to 50%.

4.2.2 Future Work

Distributions

The derived output distributions should be investigated to see if it is possible to construct corresponding input distributions. For high SNR the optimal input distributions may be approximated by the derived optimal output distributions. The mutual information can then be investigated through simulations.

Discrete constellations

In the paper a system with continuous, complex, inputs from an arbitrary distribution of our choice is assumed. An interesting question is what effect a discrete constraint on the input distribution would have on the capacity. Practical systems always work with discrete signal constellations. Given the number of points in the constellation the optimal placement of these with respect to maximizing the mutual information could be derived. This would be very valuable knowledge since it could be easily implemented in an existing system.

Another interesting question is what the mutual information for some fixed, commonly used, constellations is. For example an M-QAM constellation should be possible to investigate analytically.

Bibliography

[1] G. Fettweis, M. Löhning, D. Petrovic, M. Windisch, P. Zillmann, and W. Rave, "Dirty rf: A new paradigm," *International Journal of Wireless Information Networks*, vol. 14, no. 2, pp. 133–148, 2007.

[2] M. Fossorier and S. Olcer, "Guest editorial - capacity approaching codes, iterative decoding algorithms, and their applications," *IEEE, Communications Magazine*, vol. 41, pp. 100 – 101, aug. 2003.

[3] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, pp. 655–674, May 2000.

[4] A. Demir, "Computing timing jitter from phase noise spectra for oscillators and phase-locked loops with white and 1/f noise," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, pp. 1869 –1884, sept. 2006.

[5] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*. Wiley-IEEE Press, 2005.

[6] U. Madhow, *Fundamentals of Digital Communication*. Cambridge University Press, 2008.

[7] E. Alpman, "Estimation of oscillator phase noise for mpsk-based communication systems over awgn channels," Master's thesis, Chalmers University of Technology, 2004.

Part II

Included papers

Paper A

Iterative Code-aided Estimation of Time Varying Phase Noise

Iterative Code-aided Estimation of Time Varying Phase Noise

Johan Söder and Thomas Eriksson
Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden
soderj@student.chalmers.se , thomase@chalmers.se

***Abstract*—The problem of estimating time varying Wiener phase noise by using the iterative structure of the low-density parity-check decoder is investigated. An algorithm performing joint iterative phase noise estimation and data decoding is derived from the expectation-maximization framework. The algorithm is simulated for a high order QAM system and shown to give huge improvements in terms of decreased bit error rates. Furthermore suggestions to decrease the computational complexity of the algorithm are discussed, tested and proved to work.**

I. Introduction

Carrier phase noise due to noisy oscillators is a serious source of errors in digital communication systems. The impact of phase noise increases with the number of points in the signal constellation, as well as with the increasing baud rates required in modern systems. The phase noise interacts with the transmitted symbols in a non-linear manner and rotates the signal space. To alleviate these effects the phase noise process should be estimated in order to de-rotate the signal space.

By using the iterative structures of receivers employing turbo or low-density parity-check (LDPC) codes, huge improvements may be gained in these estimates. The main idea is to perform the noise estimation jointly with the iterative data decoding, and let the improved knowledge of the transmitted data and the reduced estimation errors of the noise help each other to further improvement. This has previously been done to provide improved estimates of synchronization parameters such as the symbol delay and constant frequency and phase offsets of the carrier. The method has been labeled ”Turbo synchronization” and was initially done *ad hoc*, see e.g. [1]. The approach has since been formalized with the use of the expectation-maximization (EM) framework in [2] and by the use of factor graphs in [3].

This paper is concerned with the tracking of time-varying carrier phase noise in a system using an LDPC code, [4]. Previous work on tracking a constant carrier phase offset using the turbo approach includes [1], [2] and [3]. Time-varying phase noise in a turbo coded system has been investigated in e.g. [5].

The case investigated here consists of a high order QAM-constellation used in single carrier transmission. This is of great practical interest since most work in the literature focuses on very small constellation sizes. The phase noise model used is the commonly accepted Wiener phase noise process, equivalent to a Lorentzian oscillator spectrum, behaving as $1/f^2$ for high frequency offsets [6]. The transmitter and channel models are presented in Section II.

Fig. 1. Block diagram of system model

The approach of this paper is to start off with the EM algorithm, described in Section III, and apply it to the problem of tracking time varying phase noise, extending the work done in [2] on a constant phase offset. This is done in Section IV. In doing so it turns out that information given by the standard LDPC decoder, ignoring the phase noise, can be used to track the time varying phase noise. The tracking is performed by a Kalman Smoother based on a linearized measurement equation previously suggested for a turbo coded system in [5]. In Section V, the joint iterative phase noise tracking and LDPC decoding algorithm is presented. To show the performance of the algorithm simulations has been performed, and the results are presented in Section VI.

II. System model

The transmitter and channel are drawn in Figure 1. At the transmitter the bits, $\mathbf{b}$, are encoded with an LDPC block code to produce a codeword, $\mathbf{c}$. The encoded bits are modulated onto complex symbols, $\boldsymbol{s} = [s_1, \ldots, s_L]$, from an M-QAM constellation, $\mathcal{S}$. These symbols are rotated by oscillator phase noise and then transmitted over an additive noise channel. The output, r_k, from the Wiener phase noise and AWGN channel, with input s_k at discrete time k, is,

$$r_k = s_k e^{j\phi_k} + w_k. \quad (1)$$

The additive noise, w_k, is complex white circular symmetric Gaussian noise with power σ_N^2 per dimension,

$$w_k \sim \mathcal{N}_{\mathcal{C}}(0, 2\sigma_N^2), \quad (2)$$
$$E[w_k w_l] = 0, k \neq l. \quad (3)$$

The phase noise, ϕ_k, is defined as

$$\phi_k = \phi_{k-1} + \Delta_k, \quad (4)$$
$$\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2), \quad (5)$$
$$E[\Delta_k \Delta_l] = 0, k \neq l. \quad (6)$$

This discrete time Wiener model corresponds to a sampled version of a continuous time Brownian motion process, i.e., integrated white noise. Samples are taken every T_s second, the transmission symbol interval. The continuous time process of the corresponding oscillator has a Lorentzian spectrum [6]. This spectrum is fully characterized by a single parameter, such as the 3-dB single sided bandwidth, f_{3dB}. The innovation variance for the discrete phase noise process is $\sigma_\Delta^2 = 4\pi f_{3dB} T_s$.

III. Expectation-Maximization

The Expectation-Maximization (EM) framework is an iterative approach to solving an estimation problem where the observation depends not only on the parameter to be estimated but also on some, unobserved, nuisance parameter. Let the observation be $y = y(x, \theta)$ where θ is the parameter to be estimated and x is the nuisance parameter. Then y is called the incomplete observation, and the set $\{y, x\}$ the complete observation. We assume here that x and θ are independent. The Maximum Likelihood (ML) estimate of θ from the marginal likelihood of the complete observation is

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log f_{y|\theta}(y|\theta). \tag{7}$$

The marginal likelihood function $f_{y|\theta}(y|\theta)$ is the total likelihood function averaged over the unobserved nuisance parameter,

$$f_{y|\theta}(y|\theta) = \int f_{y,x|\theta}(y, x|\theta) dx = \int f_x(x) f_{y|x,\theta}(y|x, \theta) dx. \tag{8}$$

The EM-algorithm consists of iterating between an expectation (E) and a maximization (M) step defined by

$$\mathrm{E} : \Lambda(\theta|\hat{\theta}^n) = E_{x|y,\hat{\theta}^n}[\log f_{y|x,\theta}(y|x, \theta)] \tag{9}$$

$$\mathrm{M} : \hat{\theta}^{n+1} = \arg\max_{\theta} \Lambda(\theta|\hat{\theta}^n). \tag{10}$$

The algorithm has been shown to converge to the ML-solution [7].

Inspecting the steps, we see that the E-step computes the conditional a posteriori expectation of the log likelihood function (LLF), given the current estimate of the parameter, $\hat{\theta}^n$. This is a function of θ, the conditioning argument in the LLF, and $\hat{\theta}^n$ which is the conditioning argument in the expectation. Note that $\hat{\theta}^n$ is kept fixed through the E-step. It is instead updated in the M-step which maximizes over θ to find the next estimate, $\hat{\theta}^{n+1}$

IV. Phase Noise Estimation

For the Wiener phase noise and AWGN channel, the parameter to be estimated is the phase noise sequence, $\boldsymbol{\phi}$, and the nuisance parameter is the vector of L transmitted symbols, $\boldsymbol{s}$. The incomplete observation is the received vector, $\boldsymbol{r}$, and the complete observation set is $\{\boldsymbol{r}, \boldsymbol{s}\}$.

A. E-step analysis

Given the transmitted data and the phase noise process, the received signal has a Gaussian distribution due to the additive noise, see the system model in Section II. The conditional likelihood function is thus,

$$f(\boldsymbol{r}|\boldsymbol{s}, \boldsymbol{\phi}) = \frac{\exp\left[-\frac{1}{4\sigma_N^2}(\boldsymbol{r} - \boldsymbol{s}e^{j\boldsymbol{\phi}})(\boldsymbol{r} - \boldsymbol{s}e^{j\boldsymbol{\phi}})^H\right]}{(4\pi\sigma_N^2)^{L/2}}. \tag{11}$$

Taking the logarithm and dropping terms independent of $\boldsymbol{\phi}$ we get the conditional log-likelihood function (LLF),

$$\log f(\boldsymbol{r}|\boldsymbol{s}, \boldsymbol{\phi}) \propto -2\mathrm{Re} \sum_{k=1}^{L} s_k^* r_k e^{-j\phi_k}. \tag{12}$$

Now taking the conditional expectation over $\boldsymbol{s}$ given $\boldsymbol{r}$ and $\hat{\boldsymbol{\phi}}^n$, the current phase noise estimate, gives:

$$\Lambda(\boldsymbol{\phi}|\hat{\boldsymbol{\phi}}^n) = E_{\boldsymbol{s}|\boldsymbol{r},\hat{\boldsymbol{\phi}}^n}\left[-2\mathrm{Re} \sum_{k=1}^{L} s_k^* r_k e^{-j\phi_k}\right] \tag{13}$$

$$= -2\mathrm{Re} \sum_{k=1}^{L} E_{\boldsymbol{s}|\boldsymbol{r},\hat{\boldsymbol{\phi}}^n}\left[s_k^* r_k e^{-j\phi_k}\right] \tag{14}$$

$$= -2\mathrm{Re} \sum_{k=1}^{L} E_{\boldsymbol{s}|\boldsymbol{r},\hat{\boldsymbol{\phi}}^n}\left[s_k\right]^* r_k e^{-j\phi_k} \tag{15}$$

Next we define the a posteriori average vector of the transmitted symbols, conditioned on $\hat{\boldsymbol{\phi}}^n$,

$$\alpha_k \triangleq E_{\boldsymbol{s}|\boldsymbol{r},\hat{\boldsymbol{\phi}}^n}\left[s_k\right] = \sum_{a_l \in \mathcal{S}} a_l \Pr(s_k = a_l|\boldsymbol{r}, \hat{\boldsymbol{\phi}}^n) \tag{16}$$

The sum is over the points in the signal constellation, $\mathcal{S}$. $\Pr(s_k = a_k|\boldsymbol{r}, \hat{\boldsymbol{\phi}}^n)$ are the marginal a posteriori probabilities (APPs) of the transmitted symbols.

We can now write (15) as

$$\Lambda(\boldsymbol{\phi}|\hat{\boldsymbol{\phi}}^n) = -2\mathrm{Re} \sum_{k=1}^{L} \alpha_k^* r_k e^{-j\phi_k} \tag{17}$$

The E-step may thus be computed by first finding the marginal APPs of the transmitted symbol vector, given the current estimate of the phase noise process. These should then be used to construct a vector of weighted, or ”soft”, symbols, $\boldsymbol{\alpha}$.

To find the marginal APPs a summation over all possible sequences of data symbols has to be performed. For an uncoded system with independent and equiprobable symbols this is possible to compute. For a coded system, however, it will be a very large set to sum over, and it is also questionable if all codewords have the same probability.

B. Soft decoding as the E-step

An iterative decoder working with soft information is marginalizing the APPs of the individual bits. For a memoryless Gaussian channel the decoder converges to the true values of the APPs. Since the channel at hand has memory, convergence is not guaranteed. But the soft output from the decoder may be used as approximations of the APPs for the transmitted bits. If a simple mapping from bits to symbols is available these bit APPs may then be used to construct the soft symbol vector, $\boldsymbol{\alpha}$.

As an example, consider the Gray mapping from bits to symbols in 16 QAM,

$$a = b_1(2 - b_2) - jb_3(2 - b_4), \tag{18}$$

where $b_i \in \{-1, 1\}$. By assuming the bits to be independent, which is true for the data bits and reasonable for the parity bits if an interleaver is used between the encoder and the modulator, α_k may be broken down into the probabilities of the individual bits (the subscript k on the bits is dropped for notational clarity) [1],

$$\begin{aligned} \alpha_k =& (P_{b_1} + P_{b_3})\,(1 + j + 2\,(P_{b_2} + jP_{b_4})) \\ & (P_{b_3} - P_{b_1})\,(-1 + j - 2\,(P_{b_2} - jP_{b_4}))\,, \end{aligned} \tag{19}$$

where $P_{b_i} \triangleq \Pr(b_i = 1|\boldsymbol{r}, \hat{\boldsymbol{\phi}}^n)$, i.e., the a posteriori probability of the bits conditioned on the current phase noise estimate.

The LDPC decoder works with the log likelihood ratios (LLRs) of the APPs, defined as

$$\mathrm{L}_i \triangleq \log \frac{\Pr(b_i = 1|\boldsymbol{r}, \hat{\boldsymbol{\phi}}^n)}{\Pr(b_i = -1|\boldsymbol{r}, \hat{\boldsymbol{\phi}}^n)} = \log \frac{P_{b_i}}{1 - P_{b_i}}. \tag{20}$$

By inverting (20) the APP may be expressed as a function of the LLR,

$$P_{b_i} = \frac{e^{L_i}}{1 + e^{L_i}} \tag{21}$$

Let $\beta_i \triangleq \tanh(\mathrm{L}_i/2)$, by inserting (21) into (19) the following expression of the soft symbols is obtained:

$$\alpha_k = \beta_1(2 - \beta_2) - j\beta_3(2 - \beta_4). \tag{22}$$

We call this a soft modulation of the LLRs. It is easily verified that as the APPs converge to one or zero $\beta_i \to 1$ or -1 and, consequently $\alpha_k \to a$, a symbol in the constellation.

C. M-step

The M-step should maximize the a posteriori likelihood function $f(\boldsymbol{r}|\boldsymbol{s} = \boldsymbol{\alpha}, \boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$. A general solution to this problem is given by the Kalman Smoother, which suits this problem perfectly since it is already given in state-space form;

$$\text{State}: \qquad \phi_k = \phi_{k-1} + \Delta_k \tag{23}$$

$$\text{Observation}: \qquad r_k = \alpha_k e^{j\phi_k} + w_k. \tag{24}$$

The only problem is that the observation equation is nonlinear and we must use a linearized version, giving suboptimal solutions. It will be shown, however, that this still gives good results.

To be able to make a small angle approximation of the phase noise process, we begin by removing the mean of the phase rotation over one codeword, as proposed in [5]. The ML estimator for the mean is [5],

$$\hat{\bar{\boldsymbol{\phi}}} = \arg\left\{\sum r_k \alpha_k^*\right\}. \tag{25}$$

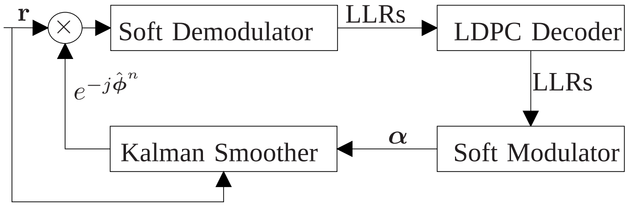


Fig. 2. Block diagram of receiver algorithm

Now the parameter to be estimated is $\boldsymbol{\theta} \triangleq \boldsymbol{\phi} - \hat{\bar{\boldsymbol{\phi}}}$. First consider the real and imaginary parts of the received signal,

$$\begin{aligned} \mathrm{Re}\{r_k\} &= \mathrm{Re}\{\alpha_k\}\cos(\theta_k) - \mathrm{Im}\{\alpha_k\}\sin(\theta_k) + \mathrm{Re}\{w_k\} \\ \mathrm{Im}\{r_k\} &= \mathrm{Re}\{\alpha_k\}\sin(\theta_k) + \mathrm{Im}\{\alpha_k\}\cos(\theta_k) + \mathrm{Im}\{w_k\}. \end{aligned} \tag{26}$$

Assuming small values of the phase noise deviation from the mean, the following approximations can be made:

$$\sin(\theta_k) \approx \theta_k \tag{27}$$

$$\cos(\theta_k) \approx 1. \tag{28}$$

Applying the approximations to (26) gives linear functions of θ_k:

$$\begin{aligned} \mathrm{Re}\{r_k\} &\approx \mathrm{Re}\{\alpha_k\} - \mathrm{Im}\{\alpha_k\}\theta_k + \mathrm{Re}\{w_k\} \\ \mathrm{Im}\{r_k\} &\approx \mathrm{Re}\{\alpha_k\}\theta_k + \mathrm{Im}\{\alpha_k\} + \mathrm{Im}\{w_k\}. \end{aligned} \tag{29}$$

The following measurement equation, removing the $\cos(\theta_k)$ term, was proposed by [5],

$$\begin{aligned} y_k =\;& \mathrm{Im}\{r_k\alpha_k^*\} && (30) \\ =\;& \mathrm{Im}\{r_k\}\mathrm{Re}\{\alpha_k\} - \mathrm{Re}\{r_k\}\mathrm{Im}\{\alpha_k\} && (31) \\ \approx\;& (\mathrm{Re}\{\alpha_k\}\theta_k + \mathrm{Im}\{\alpha_k\} + \mathrm{Im}\{w_k\})\mathrm{Re}\{\alpha_k\} \\ & - (\mathrm{Re}\{\alpha_k\} - \mathrm{Im}\{\alpha_k\}\theta_k - \mathrm{Re}\{w_k\})\mathrm{Im}\{\alpha_k\} && (32) \\ =\;& (\mathrm{Re}\{\alpha_k\}^2 + \mathrm{Im}\{\alpha_k\}^2)\theta_k \\ & + \mathrm{Im}\{w_k\}\mathrm{Re}\{\alpha_k\} - \mathrm{Re}\{w_k\}\mathrm{Im}\{\alpha(n)\} && (33) \\ \triangleq\;& ||\alpha_k||^2\theta_k + v_k, && (34) \end{aligned}$$

where $v_k \sim \mathcal{N}(0, E_s\sigma_N^2/2)$ and E_s is the average symbol energy. The measurement equation is now linear and a regular Kalman Smoother may be used.

V. The Joint Iterative PN Tracking and LDPC Decoding Algorithm

The joint iterative phase noise tracking and LDPC decoding algorithm is described in Algorithm 1 and depicted as a block diagram in Figure 2.

When implementing the proposed algorithm some design parameters needs to be decided upon. Firstly the EM-algorithm does not converge to the global maximum if the initial guess is not close enough to the optimum. The procedure thus needs to be bootstrapped with a good enough guess. To achieve this, pilots are inserted into the data stream every p_r symbol. The pilots are known data symbols providing a more stable phase noise estimate since there is no uncertainness of the nuisance parameter.

Algorithm 1: The Joint Iterative PN Tracking and LDPC Decoding Algorithm

```
input      : r
output     : b̂
parameters : Nloops, Nits, σ_N^2, σ_Δ^2
s^(0) ← r
for n ← 1 to Nloops do
    LLRs ← SoftDemodulator(s^(n−1), σ_N^2)
    for n ← 1 to Nits do
        LLRs ← LDPCDecoder(LLRs)
    end
    α ← SoftModulator(LLRs)
    φ̂^(n) ← KalmanSmoother(α, r, σ_Δ^2, σ_N^2)
    s^(n) ← r^(0) exp[−jφ̂^n]
end
b̂ ← HardDemodulator(s^(n), σ_N^2)
```

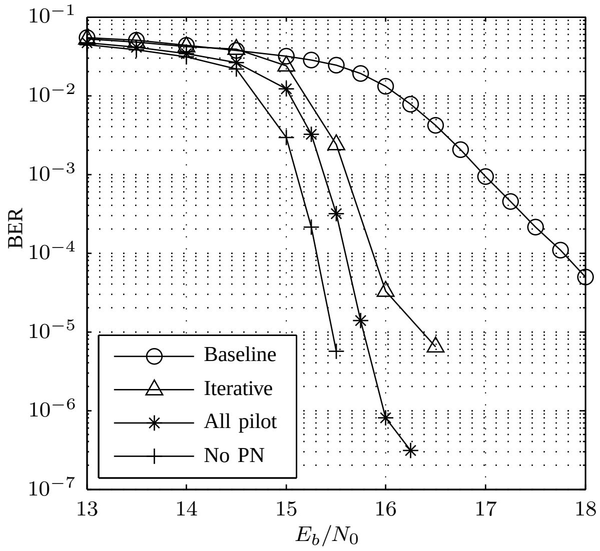


Fig. 3. BER curves

Another question related to convergence is how many iterations the LDPC decoder should perform before passing the LLRs on to the Kalman Smoother. To be a good approximation of the marginal APPS of the transmitted bits, the decoder should be run to convergence. This is, however, not feasible in a real system since the processing load and the delay time will be very high. It was investigated here as a first step to see the performance of the algorithm. The next step is of course to decrease the number of decoder iterations and see how many are needed for convergence. A trick to speed up convergence was proposed in [2] for a turbo coded system. The idea is to avoid restarting the decoder between iterations of the total algorithm. For the LDPC decoder this would mean that the extrinsic information in the node checksums are kept between phase noise estimate updates. Only the input LLRs of the individual bits are replaced.

VI. Simulations and Results

The derived algorithm was simulated for a 256-QAM system, with a rate 7/8 LDPC code from the NASA Goddard technical standard [8], having a block length of 7154 coded bits. Pilots were inserted every 20th symbol. The phase noise process was a Wiener process as described in Section II, with innovation variance $\sigma_\Delta^2 = 10^{-4}$. At the receiver joint iterative phase noise tracking and LDPC decoding was performed on each code block.

In Figure 3 the results of the simulations are shown. The LDPC decoder makes 20 iterations before outputting information to the Kalman smoother. Then the decoder is restarted with the input rotated according to the new estimate of the phase noise process. For comparison, a baseline case is plotted. This refers to performing the phase noise estimation separated from the decoding. In detail the signal is first demodulated in a soft manner to retrieve bit LLRs, these are then used to construct soft symbols which are fed to a Kalman Smoother, which provides a phase estimate. This phase estimate is used to rotate the original received signal before passing it to the LDPC decoder. The LDPC decoder in the baseline case makes a large number of iterations to give a fair comparison since the iterative algorithm multiplies the total number of decoding iterations performed.

To evaluate the possible performance of the algorithm a simulation was performed where the Kalman Smoother had access to the transmitted symbols. This ”all pilot” scenario should show what performance in terms of bit error-rate the algorithm gives when it reaches convergence. For comparison the BER curve resulting from removing all phase noise in the system is also shown.

To improve the speed of the algorithm, the number of iterations required for convergence was investigated. Figure 4 shows the improvement in BER for each new iteration of the algorithm, for a fixed signal-to-noise ratio. It also displays the effect of decreasing the number of decoding iterations performed inside the algorithm. As seen it is possible to go down to 5 decoder iterations per algorithm iteration with small losses in performance.

Figure 5 shows the same experiment, but here the LDPC decoder is not restarted between phase noise updates. With this setting it is possible to go all the way down to one decode iteration per total algorithm iteration. Since the decode iterations are the heaviest operations in the algorithm, the total computational complexity is approximately given by the total number of decode iterations. The number of operations performed after 30 algorithm iterations with one decode iteration is thus approximately the same as the number of operations after 10 algorithm iterations with 3 decode iterations.

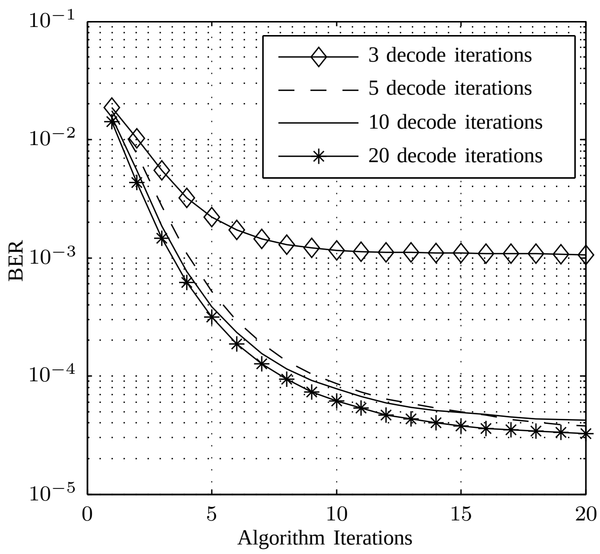


Fig. 4. BER performance for each algorithm iteration at $E_b/N_0 = 16$dB

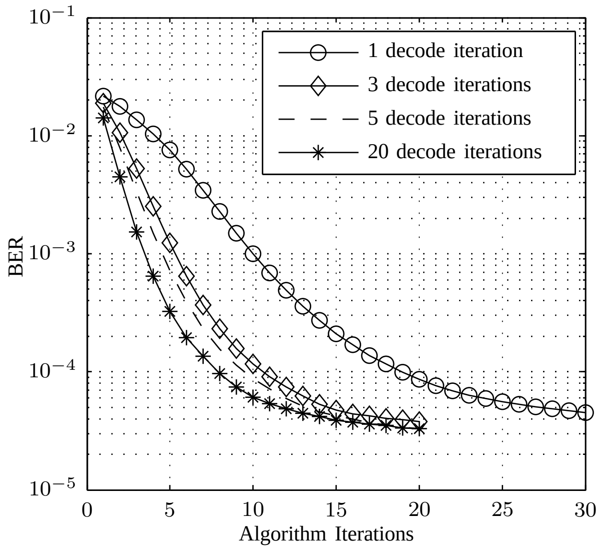


Fig. 5. BER performance for each algorithm iteration at $E_b/N_0 = 16$dB, keeping internal decoder information

VII. Conclusion

The expectation maximization framework has been applied to the problem of estimating time varying Wiener phase noise in a digital communication system. The derivation motivates the proposed joint iterative phase noise tracking and LDPC decoding algorithm. The algorithm uses the soft information iterated in the decoder to estimate the phase noise process through a Kalman Smoother. The phase noise estimates are used to de-rotate the received signal returning better input to the decoder. This process is then iterated to improve both the phase noise estimates and the quality of the decoded data.

The performance of the proposed algorithm was evaluated for a 256-QAM system with a rate 7/8 LDPC code and shown to give performance gains, in terms of significantly lower bit error rates. To obtain reasonable computational complexity it was shown that the algorithm converges faster if the LDPC decoder is never reset. The phase can be estimated, and the received signal de-rotated, every decode iteration.

References

[1] V. Lottici and M. Luise, “Embedding carrier phase recovery into iterative decoding of turbo-coded linear modulations,” *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 661–669, April 2004.

[2] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, “A theoretical framework for soft-information-based synchronization in iterative (turbo) receivers,” *EURASIP J. Wirel. Commun. Netw.*

[3] C. Herzet, N. Noels, V. Lottici, H. Wymeersch, M. Luise, M. Moeneclaey, and L. Vandendorpe, “Code-aided turbo synchronization,” *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1255–1271, june 2007.

[4] R. Gallager, “Low-density parity-check codes,” *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21 –28, january 1962.

[5] T. Shehata and M. El-Tanany, “Joint iterative detection and phase noise estimation algorithms using kalman filtering,” in *CWIT 2009. 11th Canadian Workshop on Information Theory*, May 2009, pp. 165–168.

[6] A. Demir, A. Mehrotra, and J. Roychowdhury, “Phase noise in oscillators: a unifying theory and numerical methods for characterization,” *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655–674, May 2000.

[7] T. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, nov 1996.

[8] “Goddard technical standard: Low density parity check code for rate 7/8,” NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, Tech. Rep. GSFC - STD - 9100, 2006.

Paper B

On the Capacity of the Phase Noise and AWGN Channel

On the Capacity of the Phase Noise and AWGN Channel

Johan Söder and Thomas Eriksson
Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden
soderj@student.chalmers.se , thomase@chalmers.se

***Abstract*—The channel capacity of a channel consisting of additive white gaussian noise (AWGN) and time varying Wiener phase noise, equivalent to a random walk, is investigated. This channel is of interest since it models the effects noisy oscillators has on a communication system operating at radio frequencies.**

Upper and lower bounds on the capacity are derived through functional optimization of the output distributions. To deal with the infinite memory of the Wiener phase noise process differential decoding of the phase is employed. The capacity bounds are evaluated numerically and it turns out they are extremely tight, thus enclosing the true capacity well. The results are also shown to be consistent with capacities derived for asymptotical levels of phase noise. The procedure followed may be repeated for given values of the Wiener phase noise innovation variance and signal-to-noise ratio(SNR).

The capacity for the AWGN and Wiener phase noise channel follows the well known AWGN capacity curve for low values of SNR. At some point, determined by the ratio of the Wiener phase noise innovation variance and the SNR, the curve deviates. After this point the increase in achievable data rate gained by increasing the SNR is only half of that for the pure AWGN channel.

I. Introduction

INFORMATION THEORY has been used since it was introduced by Shannon in 1948 to calculate the capacity of different communication channels in terms of maximum data rate achievable for a fixed time interval and spectral allocation. However, it took until the 1990s for real communication systems to approach the capacity limit, with the arrival of the turbo codes [1] and the revival of the LDPC codes [2],[3]. Since then, the importance of capacity has increased as it provides a benchmark comparison tool for communication systems.

Phase noise (PN) is a reality in any RF communication system, where local oscillators are used to translate the baseband signal to passband and back down again at the transmitter and receiver, respectively. It has become a more important factor with the increased popularity of the orthogonal frequency division multiplexing (OFDM) as well as with the increased constellation sizes in single carrier systems. For OFDM systems the phase noise tends to destroy the orthogonality of the subcarriers [4].

In the research literature, little work is found on the capacity of a channel with phase noise. In [5], the capacity is derived for a system with residual white phase noise after a phase locked loop. A bound for a phase noise process with memory is derived by Lapidoth in [6]. It is assumed, however, that the phase noise is a stationary process. That assumption is lifted here by considering time varying phase noise generated by a Wiener process, which is a commonly accepted model [7].

This paper derives upper and a lower bounds on the capacity of the additive white Gaussian noise (AWGN) and Wiener PN channel. The approach is initialized by replacing the classical I and Q-channel with an equivalent amplitude-phase channel in Section II. This shows more explicitly the effects of the phase noise and also allows for some high SNR approximations simplifying the calculations. Channel capacity is defined as a functional optimization problem in Section III and a theorem from the calculus of variations is stated, to be used later.

Using the amplitude-phase channel, the mutual information between the input and the output is examined in Section IV. To strengthen the methods used and familiarize the reader with the approach used, two extreme cases are then investigated. The capacity is derived for the case of negligible phase noise in Section V. The result is the well known AWGN capacity formula. In Section VI the case of a phase noise dominated channel is investigated.

In Section VII lower and upper bounds on the mutual information for arbitrary levels of phase noise are derived. The infinite memory of the Wiener phase noise process is resolved by the use of differential decoding. There is however a dependence on the amplitude distribution in the bounds on the mutual information. To make the bounds tight they are optimized over all amplitude distributions satisfying an energy constraint in Section VIII. Finally, in Section IX, numerical results are presented showing that the bounds are very close to each other, thus enclosing the true capacity. The bounds also match the extreme cases derived.

II. System Model

A. The channel

The output, y_k, from the Wiener phase noise and AWGN channel with input x_k at discrete time k is

$$y_k = x_k e^{j\phi_k} + w_k. \tag{1}$$

The additive noise, w_k, is complex white circular symmetric Gaussian noise with power σ_N^2 per dimension,

$$[w_1, w_2, \cdots, w_n] \sim \mathcal{N}_{\mathcal{C}}(0, 2\sigma_N^2 \mathbf{I}_n). \tag{2}$$

The phase noise, ϕ_k, is defined as

$$\phi_k = \phi_{k-1} + \Delta_k, \tag{3}$$
$$[\Delta_1, \Delta_2, \cdots, \Delta_n] \sim \mathcal{N}(0, \sigma_\Delta^2 I_n). \tag{4}$$

This discrete time Wiener model corresponds to a sampled version of a continuous time Brownian motion process. Samples are taken every T_s second, the transmission symbol interval. The continuous time process of the corresponding oscillator has a Lorentzian spectrum [7]. This spectrum is fully characterized by a single parameter; the 3dB single sided bandwidth, f_{3dB}. The innovation variance for the discrete phase noise process is $\sigma_\Delta^2 = 4\pi f_{3dB} T_s$.

B. Amplitude and phase input-output relations

The input, x_k, to the channel (1) is a complex number. We denote it's amplitude, $|x_k|$, with R_k and it's phase, $\angle x_k$, with Θ_k so that $x_k = R_k e^{j\Theta_k}$. In the same way r_k and θ_k are the output amplitude and phase, respectively. The additive noise, w_k, is divided in two orthogonal parts, one parallel with the transmitted vector, $w_{k,\parallel}$ and one orthogonal to it, $w_{k,\perp}$. The input-output relationships between the transmitted and received phase and amplitude are:

$$r_k \triangleq |y_k| = \sqrt{(R_k + w_{k,\parallel})^2 + w_{k,\perp}^2} \tag{5}$$
$$\theta_k \triangleq \angle y_k = \Theta_k + \arctan \frac{w_{k,\perp}}{R_k + w_{k,\parallel}} + \phi_k. \tag{6}$$

These two real-valued channels, equivalent to the complex channel (1), has the advantage of being additive in terms of the phase noise. The impact of the amplitude on the total noise added to the transmitted phase is also made more visible.

C. Approximations

For high signal-to-noise ratio (SNR) the amplitude dominates the noise terms, $R_k >> w_{k,\perp/\parallel}^2$. In this case the orthogonal noise, $w_{k,\perp}$, will not change the amplitude much. This may be seen by expanding (5),

$$r_k = \sqrt{(R_k + w_{k,\parallel})^2 + w_{k,\perp}^2} \tag{7}$$
$$= \left(R_k + w_{k,\parallel}\right)\sqrt{1 + \frac{w_{k,\perp}^2}{(R_k + w_{k,\parallel})^2}} \tag{8}$$
$$\approx R_k + w_{k,\parallel}. \tag{9}$$

Using the above approximation and the fact that $\arctan(x) \approx x$ for small x, which is the case in (6) for high SNR, the following approximative relation for the received phase may be used,

$$\theta_k = \Theta_k + \arctan \frac{w_{k,\perp}}{R_k + w_{k,\parallel}} + \phi_k \tag{10}$$
$$\approx \Theta_k + \arctan \frac{w_{k,\perp}}{r_k} + \phi_k \tag{11}$$
$$\approx \Theta_k + \frac{w_{k,\perp}}{r_k} + \phi_k. \tag{12}$$

To summarize, the following approximative input-output relationships for the phase and amplitude have been derived in (7) - (12):

$$r_k \approx R_k + w_{k,\parallel} \tag{13}$$
$$\theta_k \approx \Theta_k + \frac{w_{k,\perp}}{r_k} + \phi_k. \tag{14}$$

The approximations are valid for high SNR. These equations provide additive noise relationships for the phase noise, ϕ_k, as well as the white noise, w_k. The received phase does, however, have a non-linear dependence on the received amplitude.

III. Theoretical foundations

A. Definition of Capacity

The capacity of a channel is defined as the maximum of the mutual information between the input and the output over all input distributions satisfying an energy constraint [8]. Since the problem at hand involves a channel with memory, we consider capacity per channel use. This is defined as the maximum mutual information between a long input vector, $\mathbf{x}$, and the corresponding output vector, $\mathbf{y}$, divided by the length of the vectors.

$$C = \max_{f_{\mathbf{x}}(\mathbf{x})} \frac{1}{n} I(\mathbf{x}; \mathbf{y})$$
$$\text{subj.to } \frac{1}{n} E\left[||\mathbf{x}||^2\right] = E_s. \tag{15}$$

B. Calculus of Variations

The problem stated in (15) is a functional optimization problem. A functional is a mapping from a function to a real number, e.g. an integral. A theorem from the calculus of variations will be used in this paper, for proof see, e.g., [9].

Theorem 1: (Functional optimization)
If the functional $I(u)$ is defined by:

$$I(u) = \int_\Omega K(\mathbf{x}, u(\mathbf{x}))\, d\mathbf{x}, \tag{16}$$

where $u(\mathbf{x})$ is a real valued function of a real vector argument $\mathbf{x}$,

$$u : \Omega \subset \mathbb{R}^N \to \mathbb{R}, \tag{17}$$

and K and L_i, $i = 1, 2, ..., n$, are real valued functions with continuous first partial derivatives,

$$K, L_i : \Omega \times \mathbb{R} \to \mathbb{R},\ i = 1, 2, ..., n. \tag{18}$$

Then a necessary condition for u to be a stationary point of $I(u)$ under the n constraints

$$\int_\Omega L_i(\mathbf{x}, u(\mathbf{x}))\, d\mathbf{x} = 0,\ i = 1, 2, ..., n\,, \tag{19}$$

is satisfying the simplified Euler-Lagrange equation:

$$\frac{\partial K}{\partial u} + \sum_{i=1}^{n} \lambda_i \frac{\partial L_i}{\partial u} = 0. \tag{20}$$

The Lagrange multipliers λ_i should be chosen to fulfill the constraint.

IV. Mutual information

The functional to be optimized is the mutual information per channel use between an input vector $\mathbf{x}$ and an output vector $\mathbf{y}$, as stated in (15). The input and the output are related through the Wiener phase noise and AWGN channel stated in (1). By using the approximative amplitude and phase input-output relations, (13) and (14), this may be rewritten,

$$\frac{1}{n}I(\mathbf{x};\mathbf{y}) = \frac{1}{n}I(\mathbf{r},\boldsymbol{\theta};\mathbf{R},\boldsymbol{\Theta}) \tag{21}$$

$$=\frac{1}{n}\left(h(\mathbf{r},\boldsymbol{\theta}) - h(\mathbf{r},\boldsymbol{\theta}|\mathbf{R},\boldsymbol{\Theta})\right) \tag{22}$$

$$=\frac{1}{n}\left(h(\mathbf{r}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\mathbf{r}|\mathbf{R},\boldsymbol{\Theta}) - h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})\right) \tag{23}$$

$$=\frac{1}{n}\left(h(\mathbf{r}) - h(\mathbf{r}|\mathbf{R}) + h(\boldsymbol{\theta}|\mathbf{r}) - h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})\right) \tag{24}$$

$$\leq\frac{1}{n}\left(h(\mathbf{r}) - h(\mathbf{r}|\mathbf{R}) + h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})\right) \tag{25}$$

$$=\frac{1}{n}\left(I(\mathbf{r};\mathbf{R}) + I(\boldsymbol{\theta};\mathbf{r},\mathbf{R},\boldsymbol{\Theta})\right). \tag{26}$$

In (23) the chain rule for entropy has been used. Under the high SNR approximations the received amplitudes, $\mathbf{r}$, are conditionally independent of the transmitted phases, $\boldsymbol{\Theta}$, given the transmitted amplitudes, $\mathbf{R}$. This makes (24) true. The inequality, (25), comes from the fact that conditioning reduces entropy. Equality is obtained if $\boldsymbol{\theta}$ is independent of $\mathbf{r}$ which is true if the transmitted phases, $\boldsymbol{\Theta}$, are uniformly distributed, $\Theta_k \sim U(0, 2\pi)$.

Looking at (26) we see two information channels. The first one, $I(\mathbf{r};\mathbf{R})$, will be referred to as the amplitude channel and the second one, $I(\boldsymbol{\theta};\mathbf{r},\mathbf{R},\boldsymbol{\Theta})$ as the phase channel. The phase channel information may be thought of as the information gained by decoding the phase after the the amplitude has already been decoded.

V. AWGN dominated

As a special case we first consider the case when the additive noise dominates over the phase noise. Assume that $\sigma_N^2 >> \sigma_\Delta^2$, then the approximative phase channel in (14) may be further approximated as:

$$\theta_k \approx \Theta_k + \frac{w_{k,\perp}}{r_k} \tag{27}$$

$$= \Theta_k + N_k. \tag{28}$$

For notational convenience the process $N_k \triangleq w_{k,\perp}/r_k$ has been introduced. It is the orthogonal part of the additive noise scaled with the received amplitude.

A. Mutual Information

First consider the amplitude information channel in (26),

$$\frac{1}{n}I(\mathbf{r};\mathbf{R}) = \frac{1}{n}(h(\mathbf{r}) - h(\mathbf{r}|\mathbf{R})) \tag{29}$$

$$= \frac{1}{n}h(\mathbf{r}) - \frac{1}{n}h(\mathbf{w}_\parallel) \tag{30}$$

$$= \frac{1}{n}h(\mathbf{r}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \tag{31}$$

$$\leq h(r) - \frac{1}{2}\log_2 2\pi e\sigma_N^2. \tag{32}$$

Equality is achieved in the last step if the received amplitude symbols are independent.

Consider now the phase information channel in (26),

$$\frac{1}{n}I(\boldsymbol{\theta};\mathbf{r},\mathbf{R},\boldsymbol{\Theta}) = \frac{1}{n}(h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})) \tag{33}$$

The entropy of the received phases may be bounded as,

$$\frac{1}{n}h(\boldsymbol{\theta}) \leq \frac{1}{n}\sum_{k=0}^{n} h(\theta_k) \leq \log_2 2\pi, \tag{34}$$

where equality is achieved for independent, uniformly distributed phases.

The conditional entropy may be written as an expectation over the distribution of the received amplitudes, if the approximative channel ignoring the phase noise, (27), is used,

$$\frac{1}{n}h(\boldsymbol{\theta}|\boldsymbol{\Theta},\mathbf{r},\mathbf{R}) \approx \frac{1}{n}h(\boldsymbol{\Theta}+\mathbf{N}|\boldsymbol{\Theta},\mathbf{r},\mathbf{R}) \tag{35}$$

$$= \frac{1}{n}h(\mathbf{N}|\boldsymbol{\Theta},\mathbf{r},\mathbf{R}) \tag{36}$$

$$= \frac{1}{n}h(\frac{\mathbf{w}_\perp}{\mathbf{r}}|\mathbf{r}) \tag{37}$$

$$= \frac{1}{n}E_{\mathbf{r}}\left[h(\frac{\mathbf{w}_\perp}{\mathbf{r}})\right] \tag{38}$$

$$= E_r\left[\frac{1}{2}\log_2 2\pi e\frac{\sigma_N^2}{r^2}\right]. \tag{39}$$

In (39) the independence of the amplitudes was used.

To summarize, the mutual information is bounded by the following expression, depending on the distribution of the received amplitudes,

$$\frac{1}{n}I(\boldsymbol{\theta},\mathbf{r};\mathbf{R},\boldsymbol{\Theta}) \leq h(r) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \tag{40}$$

$$+\log_2 2\pi - E_r\left[\frac{1}{2}\log_2 2\pi e\frac{\sigma_N^2}{r^2}\right] \tag{41}$$

$$= h(r) - E_r\left[\log_2\frac{1}{r}\right] - \log_2 e\sigma_N^2. \tag{42}$$

Equality in (40) is achieved when the received amplitudes are independent and the received phases are independent and uniformly distributed.l

B. Optimization

Since the capacity is defined as the maximum of the mutual information an optimization needs to be done over all possible amplitude distributions. Applying *Theorem 1* to the mutual information in (42) under the constraints of a normalized pdf and limited power, the following distribution is obtained,

$$\Rightarrow f(r) = Kre^{-\lambda r^2}. \tag{43}$$

So r is Rayleigh distributed, as expected since this is actually nothing but the regular AWGN channel. The capacity is,

$$C = \log_2\left(1 + \frac{E_s}{2\sigma_N^2}\right), \tag{44}$$

which is the well known AWGN channel capacity.

VI. Phase Noise dominated

Consider now the case of a phase noise dominated system. Under the assumption that $\sigma_\Delta^2 >> \sigma_N^2$ the following phase channel may be used:

$$\theta_k = \Theta_k + \phi_k. \tag{45}$$

A. Mutual Information

When the phase noise dominates over the additive noise there will be no dependence between the phase channel and the amplitudes,

$$\begin{aligned}\frac{1}{n}I(\boldsymbol{\theta};\mathbf{r},\mathbf{R},\boldsymbol{\Theta}) &= \frac{1}{n}(h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\boldsymbol{\Theta},\mathbf{r},\mathbf{R})) &(46)\\ &= \frac{1}{n}(h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\boldsymbol{\Theta})) &(47)\\ &\le \log_2 2\pi - \frac{1}{n}h(\boldsymbol{\Theta}+\boldsymbol{\phi}|\boldsymbol{\Theta}) &(48)\\ &= \log_2 2\pi - \frac{1}{n}h(\boldsymbol{\phi}) &(49)\\ &= \log_2 2\pi - \frac{1}{n}h(\boldsymbol{\Delta}) &(50)\\ &= \log_2 2\pi - \frac{1}{2}\log_2 2\pi e\sigma_\Delta^2. &(51)\end{aligned}$$

For the amplitude channel, the distribution of the amplitudes may now be chosen without care of the phase channel,

$$\begin{aligned}\frac{1}{n}I(\mathbf{r};\mathbf{R}) = \frac{1}{n}h(\mathbf{r}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 &\quad(52)\\ \le h(r) - \frac{1}{2}\log_2 2\pi e\sigma_N^2, &\quad(53)\end{aligned}$$

where equality is reached for independent amplitude symbols. The total mutual information is thus,

$$\begin{aligned}\frac{1}{n}I(\mathbf{r},\boldsymbol{\theta};\mathbf{R},\boldsymbol{\Theta}) \le & 2\pi - \frac{1}{2}\log_2 2\pi e\sigma_\Delta^2\\ & + h(r) - \frac{1}{2}\log_2 2\pi e\sigma_N^2. \quad(54)\end{aligned}$$

B. Optimization

The only function involved in the maximization is the amplitude entropy, $h(r)$. Applying *Theorem 1* to the problem of maximizing $h(r)$ under power and normalization constraints the result is

$$\Rightarrow f(r) = Ke^{-\lambda r^2}\ ,\ r > 0, \tag{55}$$

which is a half-normal (or folded normal) distribution, the entropy of which is 1 bit less than the regular, bilateral, normal distribution,

$$h(r) = \frac{1}{2}\log_2 2\pi e(E_s + 2\sigma_N^2) - 1. \tag{56}$$

Combining equations (54) and (56) the capacity may be written in closed form,

$$C = \frac{1}{2}\log_2\left(1 + \frac{E_s}{2\sigma_N^2}\right) - \frac{1}{2} - \frac{1}{2}\log_2\frac{e\sigma_\Delta^2}{2\pi}. \tag{57}$$

The result is displayed in Figure 1.

VII. General case

Now the general case will be investigated, where no assumptions are made regarding the relative impacts of the phase noise and the additive noise.

A. The Phase Channel

The phase channel in (26) contains two entropies:

$$\frac{1}{n}I(\boldsymbol{\theta};\mathbf{r},\mathbf{R},\boldsymbol{\Theta}) = \frac{1}{n}\left(h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})\right) \tag{58}$$

The entropy of the received phase, $h(\boldsymbol{\theta})$, may be bounded as,

$$\frac{1}{n}h(\boldsymbol{\theta}) \le \frac{1}{n}\sum_{k=0}^{n} h(\theta_k) \le \log_2 2\pi. \tag{59}$$

Equality is achieved for independent, uniformly distributed phases.

For the conditional entropy of the received phase, $h(\boldsymbol{\theta}|\mathbf{r},\mathbf{R},\boldsymbol{\Theta})$, the approximative phase channel, (14), will now be used. It is stated here again with $N_k = w_{k,\perp}/r_k$, as before,

$$\theta_k \approx \Theta_k + N_k + \phi_k \tag{60}$$

To resolve the infinite memory of the phase noise process the conditional entropy of the received phase may now be rewritten as:

$$\begin{aligned}\frac{1}{n}h(\boldsymbol{\theta}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}) \approx & \frac{1}{n}h(\boldsymbol{\Theta}+\mathbf{N}+\boldsymbol{\phi}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}) &(61)\\ = & \frac{1}{n}h(\mathbf{N}+\boldsymbol{\phi}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}) &(62)\\ = & \frac{1}{n}h(\{N_k - N_{k-1} + \Delta_k\}_{k=1}^{n}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}) &(63)\\ \triangleq & \frac{1}{n}h\left(\{a_k\}_{k=1}^{n}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}\right). &(64)\end{aligned}$$

In (63) the noise vector has been rearranged into the difference between consecutive noises, as in differential decoding. The new process $a_k \triangleq N_k - N_{k-1} + \Delta_k$, with $N_0 = 0$, is a stationary process and in the limit of $n \to \infty$ the average joint entropy may be replaced with an equivalent expression for the entropy rate, see p.74 in [8],

$$\frac{1}{n}h\left(\{a_k\}_{k=1}^{n}\right) \to h\left(a_n|\{a_k\}_{k=1}^{n-1}\right). \tag{65}$$

By conditioning on only one previous sample of a_k the conditional entropy of the received phase may be bounded. Continuing (64) using (65) we get:

$$\begin{aligned}\frac{1}{n}h(\boldsymbol{\theta}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}) &\approx \frac{1}{n}h\left(\{a_k\}_{k=1}^{n}|\mathbf{r},\boldsymbol{\Theta},\mathbf{R}\right) &(66)\\ &\to h\left(a_n|\{a_k\}_{k=1}^{n-1},\mathbf{r},\boldsymbol{\Theta},\mathbf{R}\right) &(67)\\ &\le h\left(a_n|a_{n-1},\mathbf{r},\boldsymbol{\Theta},\mathbf{R}\right) &(68)\\ &= h\left(a_n|a_{n-1},r_n,r_{n-1},r_{n-2}\right) &(69)\\ &= h(a_n,a_{n-1}|\tilde{\mathbf{r}}) - h(a_{n-1}|\tilde{\mathbf{r}}). &(70)\end{aligned}$$

The equality (69) is true because there is no dependence on the transmitted phases or amplitudes, as long as the received amplitudes are known. Furthermore there is only dependence on the past three amplitudes. To get to (70) the chain rule of entropy has been used and the vector $\tilde{\mathbf{r}} \triangleq \{r_n, r_{n-1}, r_{n-2}\}$ denotes the past three received amplitudes. The first term in (70) contains two samples of zero-mean correlated Gaussian noise with covariance matrix,

$$\Sigma = \begin{pmatrix} \sigma_\Delta^2 + \frac{\sigma_N^2}{r_n^2} + \frac{\sigma_N^2}{r_{n-1}^2} & \frac{\sigma_N^2}{r_{n-1}^2} \\ \frac{\sigma_N^2}{r_{n-1}^2} & \sigma_\Delta^2 + \frac{\sigma_N^2}{r_{n-1}^2} + \frac{\sigma_N^2}{r_{n-2}^2} \end{pmatrix}. \quad (71)$$

Expressing the entropies in (70) as expectations over $\tilde{\mathbf{r}}$ we get:

$$h(a_n, a_{n-1}|\tilde{\mathbf{r}}) - h(a_{n-1}|\tilde{\mathbf{r}}) \quad (72)$$

$$= E_{\tilde{\mathbf{r}}}\left[\frac{1}{2}\log(2\pi e)^2|\Sigma| - \frac{1}{2}\log 2\pi e(\sigma_\Delta^2 + \frac{\sigma_N^2}{r_{n-1}^2} + \frac{\sigma_N^2}{r_{n-2}^2})\right] \quad (73)$$

$$\triangleq E_{\tilde{\mathbf{r}}}\left[\frac{1}{2}\log 2\pi e\, g(\tilde{\mathbf{r}})\right]. \quad (74)$$

The function $g(\tilde{\mathbf{r}})$ is defined as:

$$g(\tilde{\mathbf{r}}) \triangleq \frac{|\Sigma|}{\sigma_\Delta^2 + \frac{\sigma_N^2}{r_{n-1}^2} + \frac{\sigma_N^2}{r_{n-2}^2}} \quad (75)$$

$$= \frac{(\sigma_\Delta^2 + \frac{\sigma_N^2}{r_n^2} + \frac{\sigma_N^2}{r_{n-1}^2})(\sigma_\Delta^2 + \frac{\sigma_N^2}{r_{n-1}^2} + \frac{\sigma_N^2}{r_{n-2}^2}) - \frac{\sigma_N^2}{r_{n-1}^4}}{\sigma_\Delta^2 + \frac{\sigma_N^2}{r_{n-1}^2} + \frac{\sigma_N^2}{r_{n-2}^2}}. \quad (76)$$

Combining (58) with (59), (70) and (74) a lower bound for the mutual information of the approximative phase channel is:

$$\frac{1}{n}I(\boldsymbol{\theta}; \boldsymbol{\Theta}, \mathbf{r}, \mathbf{R}) > \log_2 2\pi - E_{\tilde{\mathbf{r}}}\left[\frac{1}{2}\log 2\pi e\, g(\tilde{\mathbf{r}})\right] \quad (77)$$

To obtain an upper bound for the phase channel information, genie knowledge about a previous noise term, N_{n-1}, is added to (64), again using (65),

$$\frac{1}{n}h(\boldsymbol{\theta}|\mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}) \approx \frac{1}{n}h\left(\{a_k\}_{k=1}^n|\mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}\right) \quad (78)$$

$$\to h\left(a_n|\{a_k\}_{k=1}^{n-1}, \mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}\right) \quad (79)$$

$$> h\left(a_n|\{a_k\}_{k=1}^{n-1}, N_{n-1}, \mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}\right) \quad (80)$$

$$= h\left(N_n + \Delta_n|N_{n-1}, \mathbf{r}, \boldsymbol{\Theta}, \mathbf{R}\right) \quad (81)$$

$$= h\left(N_n + \Delta_n|r_n\right) \quad (82)$$

$$= E_r\left[\frac{1}{2}\log_2 2\pi e\left(\frac{\sigma_N^2}{r^2} + \sigma_\Delta^2\right)\right]. \quad (83)$$

The genie knowledge added in (80) decreases the entropy. In (81) a_n is conditionally independent of a_k for $k < n$ when N_{n-1} is known, and in (82) all other terms conditionally independent of $N_n + \Delta_n$ are removed. The final expression, (83), is thus an expectation over the last received amplitude of the entropy of the sum of two independent Gaussian random variables. Combining (83) with (59) an upper bound on the phase information channel, (58), is obtained:

$$\frac{1}{n}I(\boldsymbol{\theta}; \boldsymbol{\Theta}, \mathbf{r}, \mathbf{R}) < \log_2 2\pi - E_r\left[\frac{1}{2}\log_2 2\pi e\left(\frac{\sigma_N^2}{r^2} + \sigma_\Delta^2\right)\right]. \quad (84)$$

B. The Amplitude Channel

The amplitude channel in (26) may be expanded in the following way,

$$\frac{1}{n}I(\mathbf{r}; \mathbf{R}) = \frac{1}{n}\left(h(\mathbf{r}) - h(\mathbf{r}|\mathbf{R})\right) \quad (85)$$

$$\approx \frac{1}{n}\left(h(\mathbf{r}) - h(\mathbf{R} + \mathbf{w}_\parallel|\mathbf{R})\right) \quad (86)$$

$$= \frac{1}{n}\left(h(\mathbf{r}) - h(\mathbf{w}_\parallel)\right) \quad (87)$$

$$= \frac{1}{n}h(\mathbf{r}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2. \quad (88)$$

In (86) the approximate amplitude channel from (13) has been used. (87) comes from using the fact that the noise term is independent of the transmitted amplitudes and (88) from using that it is Gaussian.

For the lower bound on mutual information in the phase channel, (77), there might be something to gain from having dependent amplitudes over three consecutive symbols. To make the lower bound on the total mutual information tight, this should be exploited. Under this constraint the amplitude channel information may be bounded, continuing (88),

$$\frac{1}{n}h(\mathbf{r}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (89)$$

$$= \frac{1}{n}h(\tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \ldots, \tilde{\mathbf{r}}^{(n/3)}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (90)$$

$$= \frac{1}{n}\left(h(\tilde{\mathbf{r}}^{(1)}) + h(\tilde{\mathbf{r}}^{(2)}|\tilde{\mathbf{r}}^{(1)}) + \ldots\right) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (91)$$

$$\leq \frac{1}{3}h(\tilde{\mathbf{r}}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (92)$$

In (90) the vector of amplitudes has simply been divided into sets of three, $\tilde{\mathbf{r}}^{(1)} \triangleq \{r_1, r_2, r_3\}, \tilde{\mathbf{r}}^{(2)} \triangleq \{r_4, r_5, r_6\}$ etc. Next, in (91), the chain rule of entropy is applied. Equality is achieved in (92) if consecutive $\tilde{\mathbf{r}}^{(k)}$-vectors are independent. Combining (77) and (92) to get a lower bound, in accordance with the mutual information of the total channel, (26), yields:

$$\frac{1}{n}I(\mathbf{x}; \mathbf{y}) > \frac{1}{3}h(\tilde{\mathbf{r}}) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 + \log_2 2\pi - E_{\tilde{\mathbf{r}}}\left[\frac{1}{2}\log 2\pi e\, g(\tilde{\mathbf{r}})\right]. \quad (93)$$

When the upper bound for mutual information in the phase channel, (84), was derived, there were no limitations on the correlation between the amplitude symbols. We can therefore assume independent amplitudes, continuing from (88) this yields,

$$\frac{1}{n}\left(h(\mathbf{r})\right) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (94)$$

$$\leq h(r_n) - \frac{1}{2}\log_2 2\pi e\sigma_N^2 \quad (95)$$

Now, combining (84) with (95), according to the total mutual information, (26), gives an upper bound,

$$\frac{1}{n}I(\mathbf{x}; \mathbf{y}) < \log_2 2\pi - E_r\left[\frac{1}{2}\log_2 2\pi e\left(\frac{\sigma_N^2}{r^2} + \sigma_\Delta^2\right)\right] + h(r) - \frac{1}{2}\log_2 2\pi e\sigma_N^2. \quad (96)$$

The bound depends on the distribution of the received amplitudes, the phases are assumed to be uniformly distributed.

VIII. Optimization of General case

To make the bounds on the mutual information into bounds on the capacity they have to be maximized over all input distributions fulfilling an energy constraint. It has already been established in Section IV that the mutual information is maximized for uniformly distributed, independent, phase symbols. When it comes to the amplitude, the bounds are dependent on the distribution of the received amplitude. The maximization will therefore be performed over probability distributions for the output amplitude.

A. Lower bound

Identifying the function corresponding to K in *Theorem 1* from the lower bound on the mutual information in (93) we find that:

$$K_L(\tilde{\mathbf{r}}, f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}})) = -\frac{1}{6} f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}}) \log_2 f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}})^2 g(\tilde{\mathbf{r}})^3 - \frac{1}{2}\log_2(e^2\sigma_N^2) \tag{97}$$

To obtain a normalized pdf and meet the power constraint the following constraint functions are identified:

$$\begin{aligned} L_{L,1}(r, f_{\tilde{\mathbf{r}}}(r)) &= f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}}) - 1 \\ L_{L,2}(r, f_{\tilde{\mathbf{r}}}(r)) &= ||\tilde{\mathbf{r}}||^2 f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}}) - 3(E_s + 2\sigma_N^2) \end{aligned} \tag{98}$$

Solving the Euler-Lagrange equation (20) yields:

$$f_{\tilde{\mathbf{r}}}(\tilde{\mathbf{r}}) = \kappa_L g(\tilde{\mathbf{r}})^{-\frac{3}{2}} e^{-\lambda_L ||\tilde{\mathbf{r}}||^2}. \tag{99}$$

κ_L and λ_L should be chosen to satisfy the normalization and energy constraints. They are, unfortunately, hard to derive analytically but may be found through numerical integration.

The capacity bound may now be expressed as a function of $K_L, \lambda_L, \sigma_\Delta^2, E_s$ and σ_N^2 (note that the only actual variables are of course σ_Δ^2 and $E_s/2\sigma_N^2$) by inserting (99) into (93):

$$C_L = -\frac{1}{3}\log_2 \kappa_L + \frac{\lambda_L}{\ln 2}(E_s + 2\sigma_N^2) - \frac{1}{2}\log_2 e^2\sigma_N^2 \tag{100}$$

B. Upper bound

Applying *Theorem 1* to the upper bound on mutual information in (96) with

$$\begin{aligned} K_U(r, f_r(r)) &= f_r(r)\log_2 f(r)\sqrt{\tfrac{\sigma_N^2}{r^2} + \sigma_\Delta^2} - \tfrac{1}{2}\log_2 \sigma_N^2 e^2 \\ L_{U,1}(r, f_r(r)) &= f_r(r) - 1 \\ L_{U,2}(r, f_r(r)) &= |r|^2 f_r(r) - (E_s + 2\sigma_N^2) \end{aligned} \tag{101}$$

$$\Rightarrow f(r) = \frac{\kappa_U}{\sqrt{\frac{\sigma_N^2}{r^2} + \sigma_\Delta^2}} e^{-\lambda_U r^2}. \tag{102}$$

κ_U and λ_U should be set to meet the constraints. This pdf is a mixture of a half-normal distribution and a Rayleigh distribution. The half-normal distribution was derived in section VI as the optimal distribution when the phase noise dominates over the additive noise, i.e., for high SNR when the phase noise level is fixed. In Section V the Rayleigh distribution was shown to be optimal for the case with insignificant levels of phase noise, i.e., for low SNR. The weights in (102) between these two distributions are the phase noise innovation variance, σ_Δ^2, and the white noise variance, σ_N^2.

The upper capacity bound is, from (102) and (93):

$$C_U = -\log_2 \kappa_U + \frac{\lambda_U}{\ln 2}(E_s + 2\sigma_N^2) - \frac{1}{2}\log_2 \sigma_N^2 e^2 \tag{103}$$

IX. Results

In Figure 1 the upper and lower bounds are plotted for various values of SNR for a fixed value of σ_Δ^2. The bounds are extremely close to each other and the true capacity is thus tightly enclosed. For comparison the capacity of the AWGN channel and the capacity of the phase noise dominated channel is also plotted.

For low SNR the capacity bounds approach the AWGN capacity, because the additive noise dominates over the phase noise. For high SNR the capacity bounds for the phase noise channel show a loss in slope by 50% as compared to the AWGN capacity. This is because the phase noise is dominating. Since the phase noise level is fixed in this setting, the capacity on the phase channel is a constant. The only increase in capacity with SNR is thus for the amplitude channel, meaning that half of the available dimensions are lost.

Figure 2 shows the upper bound for various levels of phase noise innovation variance, σ_Δ^2, and a fixed level of additive noise. As expected it approaches the capacity for the AWGN channel without phase noise when $\sigma_\Delta^2 << \sigma_N^2$.

In Figure 3 the upper bound on capacity is plotted against SNR for different levels of phase noise.

The optimal probability distributions of the received amplitude are displayed in Figure 4 for different values of SNR and a fixed σ_Δ^2. As discussed in Section VIII-B the distribution approaches the Rayleigh distribution for low SNR and a half-normal density function for high SNR.

X. Conclusion

In this paper a method to calculate a bound on the capacity of a channel disturbed by addtive white Gaussian noise of a given power and Wiener phase noise with given innovation variance has been presented. Numerical results show that for a fixed level of phase noise the capacity as function of SNR follows the well known capacity curve for the AWGN channel up to some point where the phase noise starts do dominate. This point is approximately given by $\sigma_\Delta^2 > 2\sigma_N^2/E_s$. In the phase noise dominating region the increase in capacity when increasing the SNR is only 50% of that in the case with no phase noise.

References

[1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261 –1271, oct 1996.

[2] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21 –28, january 1962.

[3] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 399 –431, mar 1999.

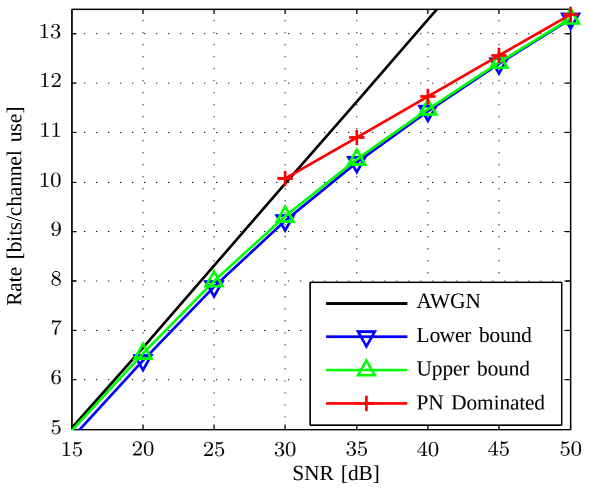


Fig. 1. Capacity bounds and AWGN Capacity, $\sigma_{\triangle}^2 = 10^{-3}$

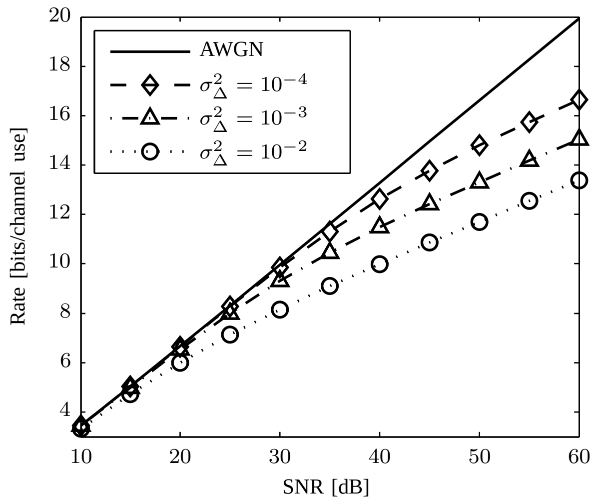


Fig. 3. Upper bound on Capacity for different $\sigma_{\triangle}^2$ and AWGN Capacity.

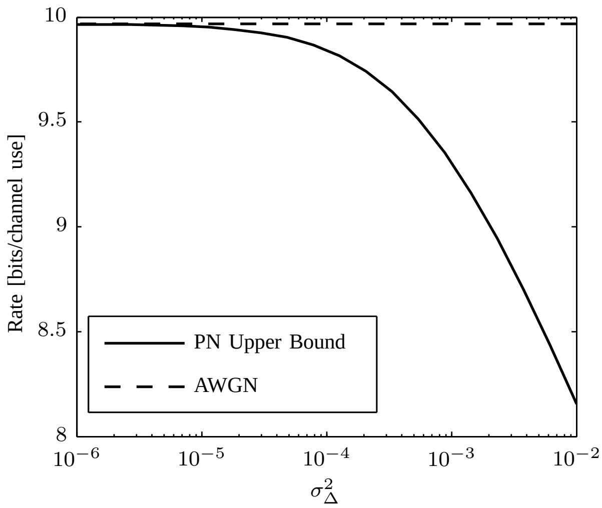


Fig. 2. Upper bound on Capacity and AWGN Capacity, SNR = 30dB.

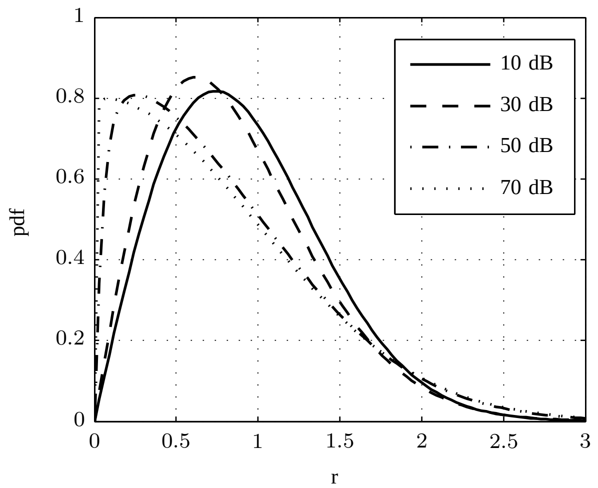


Fig. 4. Amplitude distributions from upper bound derivation, different SNRs, $\sigma_{\triangle}^2 = 10^{-3}$

[4] M. El-Tanany, Y. Wu, and L. Hazy, "Analytical modeling and simulation of phase noise interference in ofdm-based digital television terrestrial broadcasting systems," *Broadcasting, IEEE Transactions on*, vol. 47, no. 1, pp. 20–31, Mar 2001.

[5] T. Minowa, H. Ochiai, and H. Imai, "Phase noise effect on turbo-coded modulation over the coherent m-ary channel," in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 3, 2000, pp. 2232–2236 vol.3.

[6] A. Lapidoth, "On phase noise channels at high snr," in *Information Theory Workshop, 2002. Proceedings of the 2002 IEEE*, Oct. 2002, pp. 1–4.

[7] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655–674, May 2000.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, August 1991.

[9] F. W. Byron and R. W. Fuller, *Mathematics of Classical and Quantum Physics*. Dover, 1992.