CLASSICAL CAPACITY OF A BOSONIC MEMORY CHANNEL WITH GAUSS-MARKOV NOISE

Joachim Schäfer

Diplomarbeit

Erstgutachter: Prof. Dr. Nicolas J. Cerf
Zweitgutachter: Prof. Dr. Arno Rauschenbeutel
Eingereicht zum: 23. Januar 2009

Université Libre
de Bruxelles

Johannes-Gutenberg
Universität Mainz

Summary

Today's economy relies on a worldwide telecommunication network and the integrity of the sensitive data which is exchanged. Recent discoveries in quantum information science show that most classical encryption algorithms no longer protect the transmitted data between a sender and receiver if an undesirable eavesdropper is in possession of a quantum computer. Therefore, quantum communication is introduced to ensure the safety of the transmission by the laws of quantum mechanics.
In this work we investigate the capacity of a bosonic quantum channel with a Gauss-Markov correlated noise, which defines a *memory* for the channel. At first we construct a covariance matrix which corresponds to such a memory and determine its properties in the asymptotic limit. Then we apply the specified memory model to a classical Gaussian channel where we derive an explicit expression for the capacity in the full input power domain. In the main body, the quantum mechanical part of this work, we at first treat the monomodal anisotropic bosonic channel and determine the analytical solution for the optimal transmission rate when the input power is above a certain threshold. Below this threshold we apply an approximative solution and present the difference to the real capacity which is determined numerically. The findings of the one dimensional case are than applied for the infinite dimensional bosonic channel with the previously constructed memory model. Above a certain power threshold again the full solution is derived, whereas in the low power regime we test several ansatz for the given problem. In addition we conclude that a certain degree of entanglement at the input is needed to reach the capacity. Furthermore, we show that in the classical limit of the quantum mechanical solution (above the power threshold) the corresponding classical results are recovered.

Zusammenfassung

Die weltweite Wirtschaft beruht heutzutage auf der Sicherheit eines globalen Kommunikationsnetzwerkes und der Integrität der ausgetauschten Daten. Neuste Forschungsergebnisse der Quanteninformationstheorie zeigen, dass die meisten klassischen Verschlüsselungsverfahren, die übertragenen Daten, welche zwischen einem Sender und Empfänger ausgetauscht werden, nicht mehr schützen, sobald eine Person, die die Daten abhört, über einen Quantencomputer verfügt. Die Quantenkommunikation bietet hier Abhilfe, da sie die Sicherheit der Übertragung auf Grund der Postulate der Quantenmechanik garantiert.
In dieser Diplomarbeit untersuchen wir die Kapazität eines bosonischen Quantenkanals mit Gauß-Markow korreliertem Rauschen, welches dem Kanal ein sogenanntes *Gedächtnis* definiert. Zunächst konstruieren wir eine Kovarianzmatrix welche solch einem Prozess zugeordnet werden kann und bestimmmen seine Eigenschaften im asymptotischen Grenzfall. Danach wenden wir das spezifizierte Gedächtnismodell auf einen klassischen Gaußkanal an und bestimmen explizit seine Kapazität im vollen Eingangsleistungsbereich. Im Hauptteil dieser Arbeit behandeln wir zunächst den monomodalen, anisotropen bosonischen Kanal und ermitteln eine analytische Lösung der optimalen Transmissionsrate im Bereich oberhalb eines Eingangsleistungsschwellenwertes. Unterhalb dieses Schwellenwertes wenden wir eine approximative Lösung an und demonstrieren sein Nutzen gegenüber der echten Kapazität, welche numerisch bestimmt wird. Die Ergebnisse des eindimensionalen Falls werden daraufhin für den unendlichdimensionalen bosonischen Quantekanal mit gegebenem Gedächtnis erweitert. Oberhalb eines Eingangsleistungsschwellenwertes bestimmen wir ebenfalls die exakte Lösung für die Kapazität, wobei unterhalb diese Schwellenwertes verschiedene Ansätze getestet und numerisch verglichen werden. Des Weiteren stellen wir fest, dass ein gewisser Grad an Verschränkung zwischen den Eingangsmoden initialisiert werden muss, um die optimale Transmissionsrate zu erreichen. Im klassischen Grenzfall der Kapazität des quantenmechanischen Kanals finden wir letztendlich die vorher bestimmte Kapazität des klassischen Kanals wieder.

Résumé

L'économie d'aujourd'hui est basée sur un réseau de télécommunication et sur la sécurasisation des données sensibles qui sont échangées. Les découvertes récentes dans la science de l'information quantique montrent que la plupart des algorithmes de cryptage classiques ne protègent plus les données transmises entre deux interlocuteurs si un espion indésirable est en possession d'un ordinateur quantique. Donc, la communication quantique est introduite pour assurer la sécurité de la transmission à travers les lois de la mécanique quantique.
Dans ce travail nous étudions la capacité d'un canal quantique bosonique dans lequel le bruit est issu d'un processus de Gauss-Markov, qui modélise la *mémoire* du canal. Nous construisons premièrement une matrice de covariance qui correspond à cette mémoire et déterminons ses propriétés dans la limite asymptotique. Ensuite nous appliquons le modèle de mémoire spécifique à un canal Gaussien classique et dérivons une expression explicite pour la capacité dans le domaine de puissance d'entrée maximale. Dans la partie principale de ce travail, portant sur le canal Gaussien quantique, nous commençons par traiter le canal anisotropique bosonique monomode et déterminons la solution analytique pour le taux de transmission optimal quand la puissance d'entrée est au-dessus d'un seuil donné. En-dessous de ce seuil nous introduisons et testons une solution approchée, puis la comparons à la capacité exacte obtenue numériquement. Les résultats du cas unidimensionel sont ensuite appliqués au canal bosonique à dimension infinie avec le modèle de mémoire construit précédemment. Au dessus d'un certain seuil de puissance la solution au problème est établie, alors que dans le regime faible puissance nous introduisons et testons plusieurs approches de solution. Nous en concluons qu'un certain degré d'intrication en entrée est nécessaire pour atteindre la capacité. De plus nous montrons que les résultats du cas classique se déduisent de ceux du cas quantique (au dessus du seuil de puissance) dans la limite classique.

Acknowledgments

I would like to thank Nicolas Cerf for offering me the great opportunity to do my diploma thesis in this fascinating branch of quantum information science. Moreover, the visit at MIT and all the fruitful discussion we had convinced me to continue research in the field of continuous variables.

I am also grateful to my supervisor Evgueni Karpov who patiently answered all of my more or less clever questions and helped correcting this thesis, and to David Daems who introduced me to my topic. I thank all my other colleagues at QuIC for their support: Loïc Magnin for his constant help, Jérémie Roland for all his explanations at the beginning, Xavier Lacour for his help with the French translation of the summary and Louis-Philippe Lamoureux for correct reading parts of this work and improving my English during this year.

I really appreciated the cooking skills of my flatmate François Stevens during these last weeks when I came home late and hungry and the permanent will of my flatmate Michael Poss (a mathematician) to reply to all my semi physical, semi mathematical questions.

Last but not least I thank my brother Jan who helped me to design an efficient program structure in Matlab and finalizing the outlook of this work.

Gewidmet meinen Eltern Eugen und Marion
und meinen Brüdern Jan und Marc.

Contents

Chapter 1

Introduction: A Quantum of Security

In its young history quantum information science has originated a broad field of experimental demonstrations such as quantum teleportation and dense coding and technical applications such as true random number generation and safe communication.
One could consider the gedankenexperiment of Stephen Wiesner in 1970, namely his idea of *quantum money* as the hour of birth of this branch of science: if a bank encodes his notes in *non-orthogonal quantum states* any attempt to copy them will fail due to the *no cloning theorem*, a principle of quantum mechanics, which forbids exact copying of such states. However, as this "cash" was neither practical nor realizable, it took another 13 years for the concept to be published [31]. The first impact in quantum information theory appeared one year later[1] and is now refered to as the BB84 protocol [2]. Charles H. Bennett and Gilles Brassard introduced a simple, yet secure way of exchanging a secret key encoded in quantum states. Any attempt by an eavesdropper to tap the communication will due to the laws of quantum mechanics, disturb the signal and thus, be detected afterwards. If this key is then used only once to encrypt a message of the same length (*one-time pad*), then full privacy is assured. In classical communication however we are not protected by principles of quantum physics and always have to fear that the exchanged data has been eavesdropped. Therefore, most of encryption algorithms which are used today rely on the complexity of a *private key* which is used to encrypt the sent information and is only in the possession of the sender and receiver. The encryption method RSA (developed 1977 by Ronald L. Rivest, Adi Shamir and Leonard Adleman) for instance uses a product of two large prime numbers to generate the private key. Since the cost of decomposing an integer in its prime factors with common classical algorithms increases (sub-)exponentially[2] with the size of the number, this algorithm is regarded as *safe* for a certain time frame if the assumed number pair is *large enough*. In principle, this holds in reality, as long as there exists no *quantum* computer; a machine which can run *quantum* algorithms. Indeed, in 1994 Peter W. Shor introduced an algorithm which, when run on such a machine, decomposes integers in its prime factors in polynomial time [27]. This implies that if a quantum

[1] Although the protocol was actually already developed in 1983.

[2] The general number field sieve algorithm can factor an integer of size N in $\mathcal{O}(\exp((\log N)^{\frac{1}{3}}(\log\log N)^{\frac{2}{3}}))$ [32]

computer existed today, tomorrow a vast amount of classical encrypted information (using RSA for example) would be no longer protected. This threat alone justifies the inevitable convergence towards quantum communication and cryptography.

The first protocols for Quantum Key Distribution (QKD) and quantum cryptography, such as the BB84 protocol, are only based on *Discrete Variables* (DV). That means that the transmitted information is encoded in single particle states, such as the polarization of a single photon or the spin of an atom. Such 2-dimensional quantum states are called *qubits* (quantum bit) and live in a two dimensional Hilbert space. In contrast to the classical bit which can only be in the state "0" or "1", a qubit can remain in any *superimposed* state between two corresponding quantum *eigenstates* $|0\rangle$ and $|1\rangle$, such as $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$. A sender, Alice, has therefore the freedom to encode her information in two non-diagonal bases, where she switches randomly between the two during her transmission to a receiver, Bob. If an eavesdropper, Eve, now tries to tap the transmission and randomly applies (in half of the cases) the wrong basis when measuring the captured particle, its state irreversibly *collapses* by the laws of quantum mechanics to the basis she chose. If Alice and Bob compare after the transmission the bases they used (not the information they exchanged) via a public channel they can with certainty determine whether Eve was listening or not.

Since the beginning of this century several QKD protocols have been introduced which are not based on DV, but *Continuous Variables* (CV), i.e. the quadratures of the quantized electromagnetic field [20, 13, 30]. The main reason to introduce continuous variables in quantum communication is its practical feasibility. Preparing and manipulating quantum states with for instance optical laser modes is a task which is relatively easy in comparison to the generation and handling of single photon states. Several CV protocols can be implemented by using linear optics such as beam splitters and phase shifters and efficiently measured using homodyne detection instead of single photon detectors in the case of DV protocols. In addition, optical squeezers provide a simple way to generate systems with a finite (imperfect) degree of *entanglement*, whereas a perfect entangled state in DV comes to the cost of a lower success rate when generated. Furthermore, when using CV for quantum cryptography the best possible attacks (of an eavesdropper) have been already further specified than in DV cryptography, such as the proof of the optimality of Gaussian attacks on Gaussian channels. [9].
The mathematical treatment of CV states, although they live in an infinite dimensional Hilbert space, can reduce to simple finite linear algebra operations. One example are Gaussian states (which are emitted by optical lasers), where infinite dimensional quantum states are completely described by finite dimensional covariance matrices.

The investigation of the classical transmission rate[3] of a specific quantum channel where the information is encoded in such states is the subject of this work. We structured the thesis as follows: In **Chapter 2** we give at first a short introduction to the needed mathematical tools of probability theory and define the Markov process. Afterwards we construct the main noise model which is described by a Gauss-Markov covariance matrix and which is in the center of focus.

[3] I.e. the amount of classical information (bits) which is transmitted per use through the channel.

In **Chapter 3** the main issues of classical information theory are addressed and the derived noise model is applied to a classical Gaussian additive channel. We determine and discuss the full solution of the capacity for this channel.
In the main body of this work, namely **Chapter 4**, we determine at first an explicit solution for the optimal transmission rate (above a certain input power threshold) in the mono modal case and apply the results to the central issue of this thesis: the bosonic memory channel with Gauss-Markov noise.

Notation and previous knowledge: While an introduction to probability theory and several basics of linear algebra is provided, most basics of quantum mechanics will not be repeated in this work. Throughout this work we try to distinguish between scalars, vectors, etc. in the following way

x	$\hat{=}$	scalar
$\boldsymbol{x}, \boldsymbol{M}$	$\hat{=}$	vector, matrix
$\hat{p}$	$\hat{=}$	operator
$\|\psi\rangle, \langle\phi\|$	$\hat{=}$	Hilbert space vector

In addition we remark that $\log(x)$ always denotes the logarithm to the basis 2 and that $\ln(x)$ denotes the natural logarithm.

A publication based on the results of this work is in preparation.

Chapter 2

Construction of the Gauss-Markov memory

In the following section we introduce the necessary mathematical background and derive a covariance matrix for the Gauss-Markov process.

2.1 Probability Theory

This chapter provides the reader with fundamental definitions from probability theory.

2.1.1 Random Variables, density functions, conditional probabilities

Continuous variables

All random variables in the following definitions, if not stated otherwise, live in the whole real space. Most of the definitions are taken from [14].
Let $(\Omega, \mathcal{F}, P)$ be a *probability space*, with a *sample space* Ω (whose elements are colled the set of *outcomes*), a σ-algebra $\mathcal{F}$ being the collection of *measurable subsets* of Ω (whose elements are called *events*) and a *probability measure* P, a function assigning a real value between 0 and 1, called *probability*, to each event. A n-dimensional function $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ is called a (n-dimensional) random variable if

$$\{\omega : \boldsymbol{X}(\omega) \leq \boldsymbol{r}\} \in \mathcal{F}, \forall \boldsymbol{r} \in \mathbb{R}^n. \tag{2.1.1}$$

For a one dimensional random variable X we define the *distribution function*

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty. \tag{2.1.2}$$

Furthermore, we introduce a *density function* f_X:

$$F_X(x) = \int_{-\infty}^{x} dy \; f_X(y), \quad -\infty < x < \infty. \tag{2.1.3}$$

We state that

$$P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} dx\ f_X(x),$$

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} dx\ f_X(x) = 1,$$

$$P(X = x) = \int_{x}^{x} dx'\ f_X(x') = 0.$$

For a n-dimensional random variable $\boldsymbol{X}$ we the *joint distribution function* reads

$$F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n), \quad x_k \in \mathbb{R}. \tag{2.1.4}$$

Its *joint density function* is equivalently to the one dimensional case defined as

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = \frac{\partial^n F_{X_1,...,X_n}(x_1, ..., x_n)}{\partial x_1\, \partial x_2,\, ...\, \partial x_n}. \tag{2.1.5}$$

For simplicity we give the following definitions in the two-dimensional case. The same definitions hold for $X \to \boldsymbol{X}$.
The *marginal density function* is defined as

$$f_Y(y) = \int_{-\infty}^{\infty} dx\ f_{X,Y}(x, y). \tag{2.1.6}$$

In addition, the *conditional density function* for Y knowing the outcome of X, assuming that $f_X(x) > 0$:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{\infty} dy'\ f_{X,Y}(x, y')}. \tag{2.1.7}$$

The *conditional distribution function* for Y knowing the outcome of X. following (2.1.3), reads:

$$F_{Y|X=x}(y) = \int_{-\infty}^{y} dy'\ f_{Y|X=x}(y'). \tag{2.1.8}$$

The conditional probability for a n-dimensional random variable $\boldsymbol{X}$, with $X_1 \to X_2 \to ... \to X_n$, follows from the *multiplication rule*

$$\begin{aligned} P(X_1 \cap X_2 \cap X_3 \cap ... \cap X_n) = P(X_1) \cdot P(X_2 \,|\, X_1) \cdot P(X_3 \,|\, X_2 \cap X_1) \\ \cdot\ ...\ \cdot\ P(X_n \,|\, X_1 \cap ... \cap X_{n-1}). \end{aligned} \tag{2.1.9}$$

At last we define the *expectation value* for the random variable X

$$\mathrm{E}\,(X) = \int_{-\infty}^{\infty} dx\ x \cdot f_X(x), \tag{2.1.10}$$

and its *variance*

$$\mathrm{Var}(X) = \mathrm{E}\left(X - \mathrm{E}\left(X\right)\right).$$

In the following we will call any n-dimensional random variable $\boldsymbol{X}$ a *random vector* and use the short term notations

$$\mathrm{E}\left(\boldsymbol{X}\right) := \begin{pmatrix} \mathrm{E}\left(X_1\right) \\ \mathrm{E}\left(X_2\right) \\ \vdots \\ \mathrm{E}\left(X_n\right) \end{pmatrix} = \boldsymbol{\mu},\ \mathrm{Var}(\boldsymbol{X}) := \begin{pmatrix} \mathrm{Var}(X_1) \\ \mathrm{Var}(X_2) \\ \vdots \\ \mathrm{Var}(X_n) \end{pmatrix} = \sigma^2_{\boldsymbol{X}},\ \cdots$$

Discrete variables

For a random variable X with discrete alphabet $\mathcal{X}$ we define a probability mass function

$$p_X(x) = P(X = x), \quad x \in \mathcal{X},$$

for two random variables $\{X, Y\}$ the joint and conditional probability mass function

$$p_{X,Y}(x,y) = P(X = x, Y = y), \quad p_{Y|X=x}(y) = \frac{p_{X,Y}(x,y)}{p_X(x)}.$$

To keep definitions readable we will sometimes use a short term notation

$$p(x) \equiv p_X(x),\ f(x) \equiv f_X(x),\ f(x|y) \equiv f_{X|Y=y}(x),\ p(x,y) \equiv p_{X,Y}(x,y),\ \text{etc.}$$

2.1.2 Covariance

The *covariance* is a measure for correlations or anti correlations between two random variables X, Y and is defined as

$$\begin{aligned}\mathrm{Cov}(X,Y) &:= \mathrm{E}\left((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\right)\\ &= \mathrm{E}(X\cdot Y) - \mathrm{E}(X)\,\mathrm{E}(Y).\end{aligned}$$

A positive covariance results from positive outcomes for X and Y, a negative covariance from negative outcomes. If for two random variables X, Y the covariance $\mathrm{Cov}(X,Y)=0$ then X, Y are uncorrelated.

For a random vector $\boldsymbol{X}$ with $\mathrm{Var}(\boldsymbol{X}) \geq 0$ the *covariance matrix* is defined as

$$\boldsymbol{\Lambda} := \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \cdots & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \mathrm{Cov}(X_2,X_3) & \cdots & \mathrm{Cov}(X_2,X_n) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathrm{Cov}(X_{n-1},X_1) & \cdots & \mathrm{Cov}(X_{n-1},X_{n-2}) & \mathrm{Var}(X_{n-1}) & \mathrm{Cov}(X_{n-1},X_n) \\ \mathrm{Cov}(X_n,X_1) & \cdots & \cdots & \mathrm{Cov}(X_n,X_{n-1}) & \mathrm{Var}(X_n) \end{pmatrix}.$$

From the definition of the covariance it follows that $\boldsymbol{\Lambda}$ is always symmetric and $\boldsymbol{\Lambda}_{ij} \in \mathbb{R}$.
In addition we state that every covariance matrix is nonnegative-definite
(= positive-semidefinite) and cite here the proof from [14].

Proof. for any vector $\boldsymbol{y} \in \mathbb{R}^n$,

$$\boldsymbol{y}^\mathsf{T}\,\boldsymbol{\Lambda}\,\boldsymbol{y} = \boldsymbol{y}^\mathsf{T}\,\mathrm{E}\left((\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\mathsf{T}\right)\boldsymbol{y} = \mathrm{Var}(\boldsymbol{y}^\mathsf{T}(\boldsymbol{X}-\boldsymbol{\mu})) \geq 0. \tag{2.1.11}$$

□

From the previous statement, if we assume $\boldsymbol{\Lambda}$ to be diagonalizable, follows

$$\boldsymbol{y}^\mathsf{T}\,\boldsymbol{\Lambda}\,\boldsymbol{y} = \boldsymbol{y}^\mathsf{T}\,U\,\widetilde{\boldsymbol{\Lambda}}\,U^\mathsf{T}\,\boldsymbol{y} = \widetilde{\boldsymbol{y}}^\mathsf{T}\,\widetilde{\boldsymbol{\Lambda}}\,\widetilde{\boldsymbol{y}} = \sum_i \boldsymbol{\Lambda}_i \widetilde{y_i}^2 \overset{eq.(2.1.11)}{\geq} 0, \tag{2.1.12}$$

with $\widetilde{\boldsymbol{\Lambda}} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$, where $\{\lambda_j\}$ denote the eigenvalues of $\boldsymbol{\Lambda}$ and where the columns of U are the normalized eigenvectors of $\boldsymbol{\Lambda}$.
As (2.1.12) holds, according to (2.1.11), for any $\widetilde{\boldsymbol{y}} \in \mathbb{R}^n$ we conclude that for the eigenvalues of any covariance matrix $\boldsymbol{\Lambda}$

$$\lambda_i \geq 0, \quad i = 1, 2, ..., n \tag{2.1.13}$$

holds.
For the following sections we require furthermore that any covariance matrix $\boldsymbol{\Lambda}$ is invertible. Since

$$\begin{aligned}\det \boldsymbol{\Lambda} &= \prod_i \lambda_i^{(\boldsymbol{\Lambda})} = \frac{1}{\det \boldsymbol{\Lambda}^{-1}} = \frac{1}{\prod_i \lambda_i^{(\boldsymbol{\Lambda}^{-1})}},\\ \lambda_i^{\boldsymbol{\Lambda}} &= \frac{1}{\lambda_i^{\boldsymbol{\Lambda}^{-1}}},\end{aligned} \tag{2.1.14}$$

we strengthen the condition in (2.1.13) to

$$\lambda_i^{(\mathbf{\Lambda})} > 0, \quad i = 1, 2, ..., n. \tag{2.1.15}$$

One might wonders if there are more conditions on a given matrix to call it a (classical) covariance matrix. The answer is no: Any real, symmetric, and positive matrix (all eigenvalues are positive) represents a possible physical correlation matrix. [3]
Since the inverse of a symmetric and positive definite matrix is again symmetric (and positive) the latter statement implies that the inverse of a covariance matrix is again a covariance matrix.

2.1.3 Gaussian distributions

A Gaussian distributed random vector

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K}_X)$$

is fully characterized by its *first* and *second moments*[1], that is the mean vector $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and the covariance matrix $\boldsymbol{K}_X$. Its density function reads:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sqrt{\det(\boldsymbol{K}_X)}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\, \boldsymbol{K}_X^{-1}\, (\boldsymbol{x}-\boldsymbol{\mu})\right\}, \quad \boldsymbol{x} \in \mathbb{R}^n. \tag{2.1.16}$$

Bivariate Gaussian Distribution

Let us assume two Gaussian distributed random variables X,Y, i.e.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}),$$

with the *mean vector*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

and the *covariance matrix*

$$\boldsymbol{\Lambda} = \begin{pmatrix} \sigma_x^2 & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \sigma_y^2 \end{pmatrix},$$

with σ_x^2, σ_y^2 being the variances of X, Y.
We calculate now the expectation value and the variance for Y, knowing the outcome of X. According to (2.1.10) and (2.1.7):

$$\mathrm{E}(Y|X=x) = \int_{-\infty}^{\infty} dy \;\cdot y \cdot f_{Y|X=x}(y) = \int_{-\infty}^{\infty} dy \;\cdot y \cdot \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} dy'\; f_{X,Y}(x,y')}.$$

[1] For all higher moments $\mathrm{E}(\boldsymbol{X}^k) = 0, k = 3, 4, ...\infty$.

Inserting the Gaussian distribution for X,Y as defined in (2.1.16) this leads after some steps to:

$$\mathrm{E}\,(Y|X=x) = \mu_y + \frac{\mathrm{Cov}(X,Y)}{\sigma_x^2}\,(x-\mu_x), \tag{2.1.17}$$

$$\mathrm{Var}(Y|X=x) = \sigma_y^2\,(1-\frac{\mathrm{Cov}(X,Y)^2}{\sigma_x^2\,\sigma_y^2}). \tag{2.1.18}$$

2.2 Stochastic Process, Markov property

The following definitions are taken from [8], but modified.

Given a probability space $(\Omega, \mathcal{F}, P)$ as introduced in section 1, we define a *discrete stochastic process*

$$\{X_t : t \in \mathbb{N}\} \tag{2.2.1}$$

as a function defined on $\mathbb{N} \times \Omega$ with values in $\mathbb{R}$, such that $\forall t \in \mathbb{N}$
$X_t : \Omega \to \mathbb{R}$ is a random variable, i.e.

$$\{\omega : X_t(\omega) \leq r\} \in \mathcal{F}, \forall r \in \mathbb{R}.$$

A stochastic process $\{X_t\}$ is a *Markov process* if

$$P(\underbrace{X_{t+1} \leq x_{t+1}}_{\text{future}} \mid \underbrace{X_t = x_t}_{\text{present}}, \underbrace{X_{t-1} = x_{t-1} ..., X_1 = x_1}_{\text{past}}) = P(\underbrace{X_{t+1} \leq x_{t+1}}_{\text{future}} \mid \underbrace{X_t = x_t}_{\text{present}}). \tag{2.2.2}$$

If we regard $t = 1, 2, ..., \infty$ as *discrete time steps* the latter can be interpreted as follows: The probability for the consecutive result of the process $\{X_t\}$ (that is after the next time step) depends only on the given present and not on the past. Using (2.1.2) and (2.1.8) this is equivalent to

$$\begin{gathered} F_{X_{t+1} \,|\, X_t = x_t, ..., X_1 = x_1}(x_{t+1}) = F_{X_{t+1} \,|\, X_t = x_t}(x_{t+1}) \\ \int_{-\infty}^{x_{t+1}} dz \; f_{X_{t+1} \,|\, X_t = x_t, X_{t-1} = x_{t-1}, ..., X_1 = x_1}(z) = \int_{-\infty}^{x_{t+1}} dz \; f_{X_{t+1} \,|\, X_t = x_t}(z). \end{gathered} \tag{2.2.3}$$

In the case of a discrete random variable $X_t \in \{s_1, s_2, ..., s_n\}$ with outcomes $\{s_i\}$ the stochastic process is called a *Markov chain*[2] if

$$P(X_{t+1} = s_k \,|\, X_t = s_j, X_{t-1} = s_i, ..., X_1 = s_h) = P(X_{t+1} = s_k \,|\, X_t = s_j). \tag{2.2.4}$$

The Markov chain is thus characterized by a *transition matrix*

$$p_{ij}(t) = P(X_{t+1} = s_j \,|\, X_t = s_i), \tag{2.2.5}$$

that is the probability that the random variable, actually (at time t) being in the state s_i with index i is next (at time $t+1$) found in the state s_j with index j. In addition we need to define the *initial distribution* $P(X_1 = s_i)$.
Furthermore, we conclude that the conditional probability (2.1.9) for the Markov chain $\{X_t\}$ reads

$$P(X_1 \cap X_2 \cap X_3 \cap ... \cap X_t) = P(X_t \,|\, X_{t-1}) \, P(X_{t-1} \,|\, X_{t-2}) \cdots P(X_2 \,|\, X_1) \, P(X_1). \tag{2.2.6}$$

[2] In the literature the term *Markov process* is also used in the discrete variable case, but we would like to stress the difference, when switching the domain.

2.2.1 Stochastic process with a trivariate Gaussian distribution

Let us assume the stochastic process $X \to Y \to Z$, with

$$\boldsymbol{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, V)$$

and the covariance matrix

$$V = \begin{pmatrix} 1 & -w & 0 \\ -w & 1 & -w \\ 0 & -w & 1 \end{pmatrix},$$

$$w \in \mathbb{R};\ 0 \leq w < \frac{1}{\sqrt{2}}.$$

Using (2.1.6), (2.1.9) and (2.2.3), the Markov property is fulfilled if

$$\begin{aligned} \int_{-\infty}^{z} dz'\ f_{Z|Y=y_0}(z') &= \int_{-\infty}^{z} dz'\ f_{Z|Y=y_0,\,X=x_0}(z') \\ \int_{-\infty}^{z} dz'\ \frac{f_{Y,Z}(y_0,z')}{f_Y(y_0)} &= \int_{-\infty}^{z} dz'\ \frac{f_{X,Y,Z}(x_0,y_0,z')}{f_X(x_0)\, f_{Y|X=x_0}(y_0)} \\ \int_{-\infty}^{z} dz'\ \frac{f_{Y,Z}(y_0,z')}{f_Y(y_0)} &= \int_{-\infty}^{z} dz'\ \frac{f_{X,Y,Z}(x_0,y_0,z')}{f_{X,Y}(x_0,y_0)}. \end{aligned} \tag{2.2.7}$$

Since this needs to hold for any integration limit z we conclude that

$$\begin{aligned} \frac{f_{Y,Z}(y,z)}{f_Y(y)} &\overset{!}{=} \frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} \\ \Leftrightarrow \frac{\int_{-\infty}^{\infty} dx'\ f_{X,Y,Z}(x',y,z)}{\int_{-\infty}^{\infty} dx'\ f_{X,Y}(x',y)} &= \frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} \end{aligned} \tag{2.2.8}$$

to fulfill the Markov property defined in (2.2.2). After some steps of calculation (using (2.1.16)) one observes that (2.2.8) is violated for the above defined covariance matrix V. Although one might expects, since $\mathrm{Cov}(X,Z) = 0$, by intuition that the Markov property should be fulfilled it is not the case since taking the inverse of V leads to cross terms between X and Z which do not cancel out.

2.2.2 Markov process with a multivariate Gaussian distribution

In the following we construct a Gauss-Markov process, that is a Markov process where all random variables are Gaussian distributed.
Let $\{X_t\} := X_1 \to X_2 \to ... \to X_n$ be a stochastic process as defined in (2.2.1) and the random vector $\boldsymbol{X}^\mathsf{T} := (X_1\ X_2\ ...\ X_i)^\mathsf{T}$ have a multivariate Gaussian distribution with zero mean, i.e.

$$\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}).$$

To fulfill a priori the Markov property we define

$$X_i := \phi\, X_{i-1} + Z_i, \quad 0 \leq \phi < 1$$

where Z_i are i.i.d. (independent and identically distributed) Gaussian random variables and ϕ is the strength of correlation between X_i and X_{i-1}. In addition we request that the variances of all X_i are equal, that is $\sigma_i^2 = \sigma_{i-1}^2 = \sigma_1^2 =: N$. This leads to:

$$\begin{aligned} \mathrm{E}(Z_i) &= \mathrm{E}(X_i) = 0, \\ \mathrm{Var}(Z_i) &= \mathrm{E}\left(Z_i^2\right) = (1-\phi^2).\, N \end{aligned}$$

Thus, the nearest neighbor covariance reads

$$\begin{aligned} \mathrm{Cov}(X_i, X_{i-1}) &= \mathrm{E}(X_i \cdot X_{i-1}) + \overbrace{\mathrm{E}(X_i) \cdot \mathrm{E}(X_{i-1})}^{=0} \\ &= \mathrm{E}((\phi\, X_{i-1} + Z_i) \cdot X_{i-1}) = \phi\, \mathrm{E}\left(X_{i-1}^2\right) + \mathrm{E}(Z_i)\ \mathrm{E}(X_{i-1}) \\ &= \phi\, N, \\ \mathrm{Cov}(X_i, X_{i-k}) &= \mathrm{Cov}(\phi\, X_{i-1} + Z_i, X_{i-k}) = \mathrm{Cov}(\phi\, X_{i-1}, X_{i-k}) \\ &= \cdots \\ &= \mathrm{Cov}(\phi^{i-k}\, X_{i-k}, X_{i-k}) = \phi^{i-k}\, N. \end{aligned} \tag{2.2.9}$$

Hence, the full covariance matrix reads

$$\boldsymbol{\Gamma} = N \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \cdots & \cdots & \phi^n \\ \phi & 1 & \phi & \phi^2 & \phi^3 & \cdots & \cdots & \phi^{n-1} \\ \phi^2 & \phi & 1 & \phi & \phi^2 & \cdots & \cdots & \phi^{n-2} \\ \phi^3 & \phi^2 & \phi & 1 & \phi & \cdots & \cdots & \phi^{n-3} \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \cdots & \cdots & \phi^{n-4} \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & & & & 1 & \phi \\ \phi^n & \phi^{n-1} & \cdots & \cdots & \cdots & \phi^2 & \phi & 1 \end{pmatrix} \tag{2.2.10}$$

The latter is a so called *Toeplitz matrix*, which simplifies its analysis. In the following section we give a general introduction about the properties of these matrices.

2.3 Toeplitz matrices

The following definitions can be found in [12].

2.3.1 Properties, inverse, asymptotic behavior

A matrix $\boldsymbol{M}$ is called *Toeplitz matrix* if $\boldsymbol{M}_{n\times n} = [t_{k,j}; k, j = 0, 1, ..., n-1]$, with $t_{k,j} = t_{k-j}$, has the form

$$\boldsymbol{M}_{n\times n} = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{n-1} \\ t_1 & t_0 & t_{-1} & & \vdots \\ t_2 & t_1 & t_0 & t_{-1} & \vdots \\ \vdots & & & \ddots & \\ t_{n-1} & & \cdots & & t_0 \end{pmatrix}.$$

The following theorems require several definitions for $\boldsymbol{M}$.
$\boldsymbol{M}$ is Hermitian if

$$\boldsymbol{M} = \boldsymbol{M}^\dagger = (\boldsymbol{M}^\mathsf{T})^*, \tag{2.3.1}$$

which is equivalent to $t_{-k} = t_k$.
$\boldsymbol{M}$ belongs to the *Wiener class* if $\{t_k\}$ is absolutely convergent, i.e.

$$\sum_{k=-\infty}^{\infty} |t_k| < \infty. \tag{2.3.2}$$

If (2.3.2) holds, than the Fourier series

$$f(x) = \sum_{k=-\infty}^{\infty} t_k\, e^{ikx}, \quad x \in [0, 2\pi]\,, \tag{2.3.3}$$

exists and $f(x)$ is Riemann integrable, with Fourier coefficients

$$t_k = \frac{1}{2\pi} \int_0^{2\pi} dx\; f(x) e^{-ikx}.$$

For the given sequence $\{t_k\}$ or the function $f(x)$ the matrix $\boldsymbol{M}$ is fully determined, thus $\boldsymbol{M} = \boldsymbol{M}(f)$.
The existence of the inverse of $\boldsymbol{M}$ requires that the essential infimum[3] of f is greater than zero, that is

$$\operatorname{ess\,inf} f > 0 \tag{2.3.4}$$

If the conditions (2.3.1), (2.3.2) and (2.3.4) are fulfilled, then the following statements hold:
If the spectrum $\lambda(\boldsymbol{M}) = \{\lambda_1, \lambda_2, ..., \lambda_n\}$ (i.e. the set of the eigenvalues of $\boldsymbol{M}_{n\times n}$) is strictly

[3]The largest essential lower bound of f; in (2.3.4) $f(x) \leq 0$ for only a negligible part of $x \in \mathbb{R}$.

positive, then its inverse is asymptotically Toeplitz, i.e. $\boldsymbol{M}(f)^{-1} \sim \boldsymbol{M}(1/f)$.

$$t_k^{(\boldsymbol{M}^{-1})} = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx \; \frac{e^{-ikx}}{f(x)}, \quad n \to \infty. \tag{2.3.5}$$

$$\lim_{n\to\infty} \sqrt[n]{\det \boldsymbol{M}} = \exp\left(\frac{1}{2\pi} \int_{0}^{2\pi} dx \; \ln\left(f(x)\right) \right). \tag{2.3.6}$$

2.3.2 Spectrum of the Markov covariance matrix

In the following we will apply the theorems of Toeplitz matrices to determine the spectrum of the covariance matrix defined in section 2.2.2.

Calculating the inverse

From the first view at (2.2.10) the spectrum is not immediate. To obtain the latter we demonstrate that the inverse, in the limit of large n, is a three diagonal Toeplitz matrix and hence can easily be characterized. The following is motivated by the statement (2.3.5). Let us assume the inverse of a covariance matrix $\boldsymbol{\Lambda}$ to be

$$\Lambda_{n\times n}^{-1} = \frac{1}{\sqrt{1-4\,w^2}} \begin{pmatrix} 1 & -w & 0 & 0 & 0 & \cdots & 0 \\ -w & 1 & -w & 0 & 0 & \cdots & 0 \\ 0 & -w & 1 & -w & 0 & \cdots & 0 \\ 0 & 0 & -w & 1 & -w & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & -w & 1 & -w \\ 0 & 0 & 0 & 0 & 0 & -w & 1 \end{pmatrix}, \tag{2.3.7}$$

$$w \in \mathbb{R};\; w \in [w_{\min}, w_{\max}]$$

The associated function as in (2.3.3) exists if (2.3.2) holds. Here

$$\sum_{k=-\infty}^{\infty} \left| t_k^{(\boldsymbol{\Lambda}^{-1})} \right| = \frac{1}{\sqrt{1-4\,w^2}} \left(1 + 2\,|w|\right) \overset{!}{<} \infty,$$

which confines the range $0 \leq |w| < 1/2$. We will show in the next section that this range forces the spectrum to be strictly positive.
Hence, we can define

$$f_{\boldsymbol{\Lambda}^{-1}}(x) = t_{-1}^{(\boldsymbol{\Lambda}^{-1})} e^{-ix} + t_0^{(\boldsymbol{\Lambda}^{-1})} + t_1^{(\boldsymbol{\Lambda}^{-1})} e^{ix} = \frac{1}{\sqrt{1-4\,w^2}} \left(1 - 2w\cos\left(x\right)\right).$$

Since the latter fulfills (2.3.4) we can, according to (2.3.5), acquire the diagonals of the inverse of $\boldsymbol{\Lambda}^{-1}$, that is the diagonals of $\boldsymbol{\Lambda}$ by computing

$$t_k^{((\boldsymbol{\Lambda}^{-1})^{-1})} = t_k^{(\boldsymbol{\Lambda})} = \frac{\sqrt{1-4\,w^2}}{2\pi} \int_{-\pi}^{\pi} dx \; \frac{e^{-ikx}}{1-2w\cos\left(x\right)} \tag{2.3.8}$$

Since $e^{-ikx} = \cos(kx) - i\sin(kx)$ and for any odd function $f(x)$

$$\int_{-a}^{a} dx\ f(x) = 0$$

holds, we rewrite (2.3.8)

$$t_k^{(\mathbf{\Lambda})} = \frac{\sqrt{1-4\,w^2}}{2\pi}\int_{-\pi}^{\pi} dx\ \frac{e^{-ikx}}{1-2w\cos(x)} = \frac{\sqrt{1-4\,w^2}}{\pi}\int_{0}^{\pi} dx\ \frac{\cos(kx)}{1-2\,w\ \cos x}. \tag{2.3.9}$$

The solution of 2.3.9 can be found in [11](3.613):

$$\begin{aligned} t_k^{(\mathbf{\Lambda})} &= \frac{\sqrt{1-4\,w^2}}{\pi}\int_{0}^{\pi} dx\ \frac{\cos(kx)}{1-2\,w\ \cos x} = \sqrt{1-4\,w^2}\ \frac{1}{\sqrt{1-4\,w^2}}\ \left(\frac{1-\sqrt{1-4\,w^2}}{2\,w}\right)^k \\ &= \left(\frac{1-\sqrt{1-4\,w^2}}{2\,w}\right)^k. \end{aligned} \tag{2.3.10}$$

One immediately recognizes these diagonals as the diagonals of $\mathbf{\Gamma}$ as defined in (2.2.10), except for the substitution

$$\phi := \frac{1-\sqrt{1-4\,w^2}}{2\,w}. \tag{2.3.11}$$

The domain of ϕ is determined by the domain of w, i.e.

$$\begin{aligned} w = \frac{\phi}{\phi^2+1}, \quad & 0 \le w < 1/2 \\ \Leftrightarrow\ & 0 \le (1-\phi)^2 < 1 \\ \Rightarrow\ & 0 \le \phi < 1. \end{aligned}$$

Hence, by replacing w in (2.3.7) by (2.3.11) we determine the inverse of $\mathbf{\Gamma}$ in the limit of infinite row/column dimensions, that is

$$\begin{aligned} \mathbf{\Gamma}^{-1} =& \frac{1}{N}\,\frac{1+\phi^2}{1-\phi^2}\begin{pmatrix} 1 & -\frac{\phi}{1+\phi^2} & 0 & 0 & \cdots & 0 \\ -\frac{\phi}{1+\phi^2} & 1 & -\frac{\phi}{1+\phi^2} & 0 & \cdots & 0 \\ 0 & -\frac{\phi}{1+\phi^2} & 1 & -\frac{\phi}{1+\phi^2} & \cdots & 0 \\ 0 & 0 & -\frac{\phi}{1+\phi^2} & 1 & \ddots & \\ \vdots & \vdots & & \ddots & \ddots & \end{pmatrix}, \\ & 0 \le |\phi| < 1. \end{aligned} \tag{2.3.12}$$

The latter is the expected result. If we regard again the condition for a stochastic process to be a Markov-process (2.2.3), i.e.

$$\int_{-\infty}^{x_{t+1}} dz\; f_{X_{t+1}\,|\,X_t=x_t,X_{t-1}=x_{t-1},\ldots,X_1=x_1}(z) = \int_{-\infty}^{x_{t+1}} dz\; f_{X_{t+1}\,|\,X_t=x_t}(z), \tag{2.3.13}$$

we actually see that this is for a multivariate Gaussian density function $f_{\boldsymbol{X}}(\boldsymbol{x})$ (we recall that the inverse of the covariance matrix appears in the exponent) only fulfilled if its inverse is three diagonal, since then only first order cross terms appear.

Eigenvalues, trace, determinant

Now that we know that the inverse of $\mathbf{\Gamma}$ (for large n) does not have a complex structure it will be straight forward to find its eigenvalues. For simplicity we determine the spectrum and the determinant of (2.3.7) and afterwards substitute as in (2.3.11) to obtain the asymptotic spectrum.
The spectrum of the three-diagonal Toeplitz matrix

$$T = \begin{pmatrix} a_0 & a_1 & 0 & 0 & \cdots \\ a_{-1} & a_0 & a_1 & 0 & \cdots \\ 0 & a_{-1} & a_0 & a_1 & \cdots \\ 0 & 0 & a_{-1} & a_0 & \ddots \\ \vdots & \vdots & & \ddots & \ddots \end{pmatrix}, \tag{2.3.14}$$

can be found in [4]:

$$\lambda_i^{(T)} = a_0 + 2\sqrt{a_1 a_{-1}}\cos\left(\frac{\pi\, i}{n+1}\right), \tag{2.3.15}$$

where n is the dimension of T. Thus, we can immediately write out the spectrum of (2.3.7) by taking $a_0 = 1, a_1 = a_{-1} = -w$:

$$\lambda_i^{(\Lambda^{-1})} = \frac{1}{\sqrt{1-4\,w^2}}\left(1 - 2w\,\cos\left(\frac{\pi\, i}{n+1}\right)\right). \tag{2.3.16}$$

Regarding (2.3.16) we deduce that for any $w > 0, w \in \mathbb{R}$

$$\begin{aligned} \lambda_i^{(\Lambda^{-1})} &< \lambda_{i+1}^{(\Lambda^{-1})} \quad i = 1, 2, ..., n, \\ \lambda_1^{(\Lambda^{-1})}(n) &> \lambda_1^{(\Lambda^{-1})}(n+1) \quad n = 1, 2, ..., \infty, \end{aligned} \tag{2.3.17}$$

since (2.3.16) is strictly increasing with increasing i and

$$1 - 2w\cos(\pi/n) > 1 - 2w\cos(\pi/(n+1)), \quad n \in \mathbb{N}. \tag{2.3.18}$$

Thus if the greatest eigenvalue of $\mathbf{\Lambda}^{-1}$ is greater than zero for the limit $n \to \infty$ for fixed w, all eigenvalues of $\mathbf{\Lambda}^{-1}$ will be greater than zero for any $n \in \mathbb{N}$. We obtain that

$$\lim_{n\to\infty} \lambda_1^{(\Lambda^{-1})} = \frac{1-2w}{\sqrt{1-4\,w^2}}. \tag{2.3.19}$$

Thus,

$$\begin{aligned} &\lambda_i^{(\Lambda^{-1})} > 0, \quad i = 1, 2, ..., n, \\ &0 \leq w < \frac{1}{2}. \end{aligned} \tag{2.3.20}$$

By performing the substitution as in (2.3.11) we obtain the eigenvalue spectrum of $\mathbf{\Gamma}^{-1}$, that is

$$\lambda_j^{(\mathbf{\Gamma}^{-1})} = \frac{1}{N} \frac{1}{1-\phi^2} \left(1 + \phi^2 - 2\phi \, \cos\left(\frac{\pi \, j}{n+1}\right) \right), \tag{2.3.21}$$

and for $n \to \infty$ the spectrum of $\mathbf{\Gamma}$:

$$\lim_{n\to\infty} \lambda_j^{(\mathbf{\Gamma})} = \lambda^{(\mathbf{\Gamma})}(x) = N \, \frac{1-\phi^2}{1+\phi^2-2\phi\cos(x)}, \quad 0 \leq \phi < 1, \tag{2.3.22}$$

where we introduced now the spectral parameter $x \in [0, \pi]$. In addition, using (2.3.6) and [11](4.225) we find the nth square root of the determinant of $\mathbf{\Gamma}$ in the asymptotic limit, that is

$$\begin{aligned} \lim_{n\to\infty} \sqrt[n]{\det \mathbf{\Gamma}} &= \exp\left(\frac{1}{2\pi} \int_0^{2\pi} dx \; \ln \lambda^{(\mathbf{\Gamma})}(x) \right) \\ &= \exp\left(\frac{1}{\pi} \int_0^{\pi} dx \; \log N \, \frac{1-\phi^2}{1+\phi^2-2\phi\cos(x)} \right) \\ &= \exp\left(\log\left(N\left(1-\phi^2\right)\right) + \frac{1}{\pi} \underbrace{\int_0^{\pi} dx \; \log\left(1+\phi^2-2\,\phi\,\cos x\right)}_{=0} \right) \\ &= N(1-\phi^2). \end{aligned} \tag{2.3.23}$$

One notices that all eigenvalues $\lambda_j^{(\mathbf{\Gamma})}$ are strictly greater than 0 if $0 \leq \phi < 1$, as claimed in the previous section. Regarding (2.3.23), one observes that the determinant of $\mathbf{\Gamma}$ converges to 0 in the limit of large n if $N < \frac{1}{1-\phi^2}$, diverges if $N > \frac{1}{1-\phi^2}$ and is constant otherwise. As the trace of a matrix is the sum over all eigenvalues for finite dimension n it is the integral over the eigenvalue function in the asymptotic limit.

$$\lim_{n\to\infty} \mathrm{Tr}\left(\mathbf{\Gamma}_{n\times n}\right) = \frac{1}{\pi} \int_0^{\pi} dx \; \lambda^{(\mathbf{\Gamma})}(x) = N. \tag{2.3.24}$$

Thus, we know that the trace of $\mathbf{\Gamma}$ is independent of the correlation strength ϕ.

Chapter 3

Classical Part

3.1 Classical Information Theory

In what follows all random variables are considered discrete. The definitions in the continuous case are equivalent, where all sums over probability distributions are replaced by integrals over probability density functions. Most of the following definitions are taken from [6].

3.1.1 Entropy and Mutual Information

The *measure of information* for a discrete random variable X is related to its *Shannon entropy*, which is defined as

$$H(X) = H(p) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{3.1.1}$$

The latter has two equivalent meanings: it is the average amount of bits gained by measuring X or equivalently the amount of bits needed to obtain X, which is also called the *uncertainty* in X. Furthermore, Shannon showed that the entropy is the lower bound on the number of bits required to store the outcome of X.
A simple example for (3.1.1) is the fair coin toss: $X = h$ with probability $1/2$, $X = t$ with probability $1/2$. The information obtained by observing the coin is $H(X) = 1$ bit.
For a pair of random variables $\{X, Y\}$ the *joint entropy* is defined as

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y), \tag{3.1.2}$$

that is the measure of uncertainty of X and Y or more precisely the average amount of bits gained by measuring X and Y. In addition, we define the *conditional entropy*

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x)\, H(Y|X = x) \\ &= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= H(X,Y) - H(X), \end{aligned} \tag{3.1.3}$$

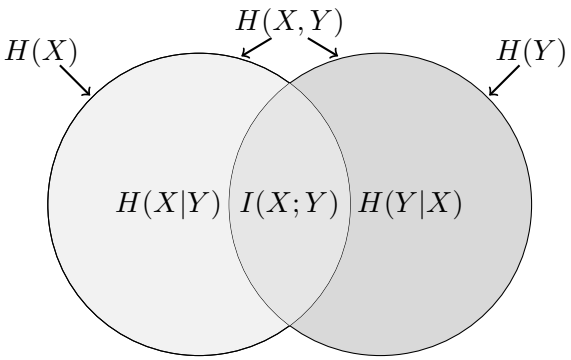

Figure 3.1: The relationship between the joint entropy, the conditional entropy and the mutual information

with $p(y|x) = p(x,y)/p(x)$.
The latter is the total uncertainty of X and Y reduced by the information gained by measuring X or equivalently the remaining bits needed to measure Y (on average) after obtaining X.
Knowing the previous two definitions we are now ready to define the amount of information X and Y have in common, that is the *mutual information* of X and Y

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \end{aligned} \tag{3.1.4}$$

In other words the mutual information is the reduced uncertainty of X due to the knowledge of Y.
The relations between the different entropies are outlined in figure 3.1.

In the case of continuous random variables the discrete entropies in (3.1.4) have to be replaced by the differential entropies

$$\begin{aligned} h(X) &= -\int_{\mathbb{R}} f_X(x)\, \log f_X(x)\, dx\ , \\ h(X|Y) &= -\int_{\mathbb{R}} f_{X,Y}(x,y)\, \log f_{X|Y=y}(x)\, dx\, dy\ , \end{aligned} \tag{3.1.5}$$

where $f_X(x)$ and $f_{X|Y=y}(x)$ are defined as in (2.1.3), (2.1.7) respectively.

3.1.2 Channel, Memory and Capacity

In the following we introduce the general definition of a channel system as in [6]. However, in this work we will not treat different encoding and decoding schemes for the channel systems we investigate.
Consider two parties, commonly referred to as Alice (the sender) and Bob (the receiver), who would like to share information. For this purpose Alice encodes her message, sends

Figure 3.2: A classical channel

it via a (noisy) *channel* where at its end Bob receives the encoded information, tries to decode it, and thus, reconstruct the original message. In this work we only consider *discrete channels* (classical or non-classical), in other words all messages are sent in discrete steps. In the classical case the exact definition reads:
A *discrete memoryless channel* is a system composed of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$ and a probability transition matrix $\mathcal{P}(y|x) \geq 0$ specifying the probability that Bob successfully receives y when x was sent, with $\sum_y \mathcal{P}(y|x) = 1, \forall x \in \mathcal{X}$.
A channel is *memoryless* if the output for the k-th use of the channel only depends on the k-th input; otherwise we are using the term *channel with memory.*
Suppose now Alice uses the previous defined channel n times, or equivalently transmits sequences $\boldsymbol{x} = (x_1, x_2, ..., x_n)^\intercal$ of length n. The channel then extends to

$$(\mathcal{X}^n, \mathcal{P}(\boldsymbol{y}|\boldsymbol{x}), \mathcal{Y}^n),$$

where $\boldsymbol{x} \in \mathcal{X}^n, \boldsymbol{y} \in \mathcal{Y}^n$.
We define a (M, n)-*code* for the channel $(\mathcal{X}^n, \mathcal{P}(\boldsymbol{y}|\boldsymbol{x}), \mathcal{Y}^n)$ as a system containing:

- An index set $\mathcal{I} = \{1, 2, ..., M\}$, where each index represents a message W.
- An encoding function $E^n : \mathcal{I} \to \mathcal{X}^n$, assigning to each message W an input string $\boldsymbol{X}$ with length n.
- A decoding function $D^n : \boldsymbol{Y} \to \mathcal{I}$ returning an estimated message $W' = D^n(\boldsymbol{Y})$ from the received output $\boldsymbol{Y}$,
  with a probability of error $\epsilon_i = P(W' \neq W) = P(D^n(\boldsymbol{Y}) \neq W | \boldsymbol{X} = E^n(W))$.

The channel system is sketched in figure 3.2.

The immediate question which emerges from the previous definition is how many bits can Alice and Bob share per transmission? Therefore we first need to define the *rate* of an (M, n) code

$$R = \frac{\log M}{n} \text{ bits per transmission}^1. \tag{3.1.6}$$

A rate R is *achievable* if there exists a sequence of $(2^{nR}, n)$ codes such that the maximal probability of error $\max_{i \in \mathcal{I}} \epsilon_i \to 0$, as $n \to \infty$.
The *channel capacity* for a discrete memoryless channel is the supremum of all achievable rates. In terms of the mutual information between Alice and Bob, the capacity is defined as

$$C = \max_{p(x)} I(X; Y), \tag{3.1.7}$$

[1]The unit bits will mostly not suppressed throughout this work

where $I(X;Y)$ is maximized over all possible input distributions $p(x)$ and C is quantified in bits.
The channel coding theorem: All rates below the capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\max_i \epsilon_i \to 0$. Conversely, any sequence of $(2^{nR}, n)$ codes with $\max_i \epsilon_i \to 0$ must have $R \leq C$

In the following we will sketch the idea of the channel coding theorem. (The exact proof can be found in [6]).
First, we state the *Asymptotic Equipartition Property* (AEP), which follows simply from the weak law of large numbers.
If the set of random variables $X_1, X_2, ..., X_n$ is i.i.d. (= independent and identically distributed) then according to the probability mass function $p(x)$,

$$-\frac{1}{n}\log p(x_1, x_2, ..., x_n) = -\frac{1}{n}\sum_i \log p(X_i = x_i) \to H(X), \tag{3.1.8}$$

where $X \sim p(x)$ and $p(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$. Using the previous statement we define a *typical* set $\mathcal{A}^n \subseteq \mathcal{X}^n$ to be composed by all outcomes $\boldsymbol{x} \in \mathcal{X}^n$ with

$$2^{-n(H(X)+\epsilon)} \leq p(\boldsymbol{x}) \leq 2^{-n(H(X)-\epsilon)}, \epsilon > 0. \tag{3.1.9}$$

According to (3.1.8) the probability for an event to be typical tends to 1 as $n \to \infty$. Since

$$1 = \sum_{\boldsymbol{x}\in\mathcal{X}^n} p(\boldsymbol{x}) \approx \sum_{\boldsymbol{x}\in\mathcal{X}^n} 2^{-nH(X)},$$

we conclude that there are in total approximately $2^{nH(X)}$ outcomes of the set $X_1, X_2, ..., X_n$.

If we suppose now that Alice's input sequence is i.i.d., then we can similarly estimate that there are about $2^{nH(Y|X)}$ possible $\boldsymbol{Y}$ sequences (conditional on the input $\boldsymbol{X}$) Bob receives. The number of outcome sequences regarded independently from the input is approximately $2^{nH(Y)}$. By dividing the number of possible outcomes by the number of possible outcomes conditioning on the input we calculate *the number of distinguishable sequences* of length n which is less than or equal to

$$\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)}. \tag{3.1.10}$$

Thus, the maximum number of distinguishable sequences is determined by maximizing the exponent of (3.1.10) over different input probability mass functions $p(x)$, hence finding the capacity.

3.1.3 Classical Gaussian channel

We will now introduce the general formulation for a *classical Gaussian channel*. The model for the latter is well known and has applications in satellite and deep space communications.

Most of the notations in this sections are taken from [6].
Assume Alice with input X_i communicates via a (discrete, additive) Gaussian channel with noise Z_i (at time i) with Bob who receives the output Y_i. Then the latter simply reads

$$Y_i = X_i + Z_i, \tag{3.1.11}$$

where $X_i \in \mathcal{X}, Y_i \in \mathcal{Y}, Z_i$ are random variables, X_i is independent from Z_i and

$$\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_Z),$$

with the noise covariance matrix $\boldsymbol{K}_Z$. As already defined in section 3.1.2, the channel defined in (3.1.11) has a memory if $\boldsymbol{K}_Z$ has off-diagonal elements. If we want to represent a physical model we have to introduce a constraint on the input. Otherwise the transmission rate per use, and hence the capacity of the channel would be infinite. Therefore we impose a power constraint P for any codeword $(x_1, x_2, ..., x_n)$ transmitted over the channel, such that

$$\sum_{i=1}^{n} x_i^2 \leq n\, P. \tag{3.1.12}$$

Using definition (3.1.7) we state the capacity of the Gaussian channel

$$C = \max_{p(x):\frac{1}{n}\sum\limits_i \mathrm{Var}(X_i)\leq P} I(\boldsymbol{X};\boldsymbol{Y}). \tag{3.1.13}$$

We rewrite

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) =& h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}) = h(\boldsymbol{Y}) - h(\boldsymbol{X}+\boldsymbol{Z}|\boldsymbol{X}) \\ =& h(\boldsymbol{Y}) - h(\boldsymbol{Z}), \end{aligned} \tag{3.1.14}$$

because $\boldsymbol{X}$ and $\boldsymbol{Z}$ are statistically independent. $h(\boldsymbol{Y}), h(\boldsymbol{Z})$ denote the joint entropies $h(Y_1, Y_2, ..., Y_n)$, $h(Z_1, Z_2, ..., Z_n)$.
Before treating the capacity of the Gaussian channel with memory we briefly discuss, for a better understanding, the single variate case and the case of parallel, uncorrelated Gaussian channels.

The single variate case

Let us assume the noise Z_i is drawn i.i.d. from a Gaussian distribution with variance N. Hence, the covariance matrix $\boldsymbol{K}_Z$ for the noise is replaced by the single parameter N and the expressions (3.1.11), (3.1.14) reduce to the single variate case

$$Y = X + Z.$$

The differential entropy of the noise $Z \sim \mathcal{N}(0, N)$ can be calculated in a few steps using (3.1.5) and (2.1.16):

$$h(Z) = \frac{1}{2}\log(2\pi e N) \quad \text{bits},$$

where e is Euler's number. Assuming we saturate the input power constraint P we then find the variance of Y to be

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(X+Z) = \mathrm{Var}(X) + \mathrm{Var}(Z) \\ &= P + N \end{aligned} \tag{3.1.15}$$

By using theorem 9.6.5[6] which states that for a random vector $\boldsymbol{X} \in \mathbb{R}^n$ with zero mean and $\boldsymbol{M}_{ij} = \mathrm{Cov}(X_i, X_j)$ its entropy is bounded by the entropy of a Gaussian distribution with covariance matrix $\boldsymbol{M}$, i.e.

$$h(\boldsymbol{X}) \leq \frac{1}{2} \log\left((2\pi e)^n \det \boldsymbol{M}\right),$$

we can conclude that

$$\begin{aligned} I(X;Y) = h(Y) - h(Z) &\leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N \\ &= \frac{1}{2}\log\left(1 + \frac{P}{N}\right). \end{aligned}$$

Hence, the maximal transmission rate per use of the channel is obtained by using a Gaussian distributed input with variance P which reads

$$C = \max_{p(x):\mathrm{Var}(X)\leq P} I(X;Y) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right). \tag{3.1.16}$$

We notice that the capacity is only depending on the signal to noise ratio (SNR) and not on input and noise separately. In the unconstrained case (arbitrary high input power), as mentioned before, C can be infinite, as well as when the noise variance is vanishing, which allows one to transmit any real number (the input X is Gaussian distributed) with no error.

Parallel uncorrelated channels

In the following we discuss the problem of finding the capacity of a system of parallel, uncorrelated channels. Later on we will see that even in the presence of correlations the approach will be similar. Suppose we have n Gaussian channels in parallel, that is

$$\begin{aligned} Y_1 &= X_1 + Z_1 \\ Y_2 &= X_2 + Z_2 \\ &\vdots \\ Y_n &= X_n + Z_n, \end{aligned} \tag{3.1.17}$$

where $Z_i \sim \mathcal{N}(0, N_i)$ and $Z_1, Z_2, ..., Z_n$ are statistically independent. Furthermore we introduce a global power constraint for the system

$$\sum_{i=1}^{n} \mathrm{Var}(X_i) \leq P_{\mathrm{gl}}. \tag{3.1.18}$$

For the mutual information the upper bound can be determined in some steps:

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= h(\boldsymbol{Y}) - h(\boldsymbol{Z}) \\ &= h(\boldsymbol{Y}) - \sum_i h(Z_i) \\ &\leq \sum_i h(Y_i) - h(Z_i) \\ &\leq \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right), \end{aligned} \tag{3.1.19}$$

where the third line is due to the fact, that the joint entropy is always smaller or equal to the sum of the separate entropies, (see figure 3.1).
Equality in (3.1.19) is achieved by picking a Gaussian distributed input

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K}_X),$$

with $\boldsymbol{K}_X = \mathrm{diag}(P_1, P_2, ..., P_n)$.
Hence, finding the capacity (3.1.13) is equivalent to finding the best power distribution $(P_1^*, P_2^*, ..., P_n^*)$ fulfilling (3.1.18) and

$$P_i^* \geq 0, \quad \forall i. \tag{3.1.20}$$

The latter, a set of inequality constraints, combined with (3.1.18) are also known as the Kuhn-Tucker conditions and the method for solving this problem is a generalization of the Lagrange multipliers method. Defining the functional

$$h(P_1, P_2, ..., P_n) = \sum_{i=1}^{n} \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) + \mu\left(\sum_{i=1}^{n} P_i - P_{\mathrm{gl}}\right)$$

differentiating it with respect to P_i while setting the result to 0, we receive

$$\frac{\partial}{\partial P_i} h(\{P_i\}) = \frac{1}{2}\frac{1}{P_i + N_i} + \mu = 0,$$

Thus, respecting (3.1.20) we conclude that

$$P_i^* = (\mu - N_i)^+, \tag{3.1.21}$$

such that

$$\sum_i (\mu - N_i)^+ = P_{\mathrm{gl}}, \tag{3.1.22}$$

with

$$(x)^+ = \begin{cases} x, & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.1.23}$$

Thus, for a given noise distribution one has to look for a "water-filling" level μ, respecting (3.1.22). In other words, to optimize the overall transmission rate, one has to induce very

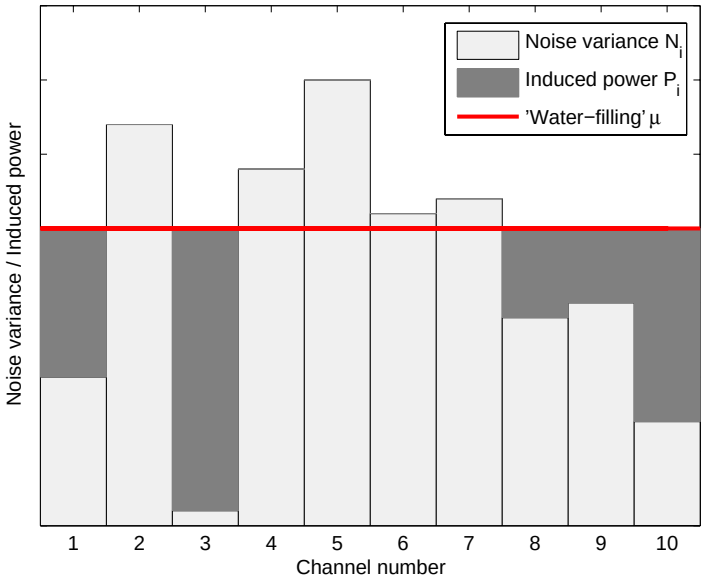

Figure 3.3: Induced power P_i vs. channel number i, such that equation (3.1.22) is fulfilled.

little or no power at all for very noisy channels and a great amount of power for low-noise channels, (see figure 3.3). With (3.1.19) and (3.1.21) we determine the capacity for the system of parallel uncorrelated channels, that is

$$C = \frac{1}{2n} \sum_i \log \left(1 + \frac{(\mu - N_i)^+}{N_i} \right). \tag{3.1.24}$$

Channel with memory

In the previous chapter we treated n channels in parallel with uncorrelated noise. By adding off-diagonal elements to the noise covariance matrix we introduce a dependence between the channels. This model can then be regarded in two ways: Either as a system of n parallel uncorrelated channels which is used one time or equivalently n consecutive uses of a correlated channel, i.e. a channel with memory.
We rewrite the energy constraint introduced in (3.1.12) as

$$\mathrm{Tr}\left(\boldsymbol{K}_X\right) \leq n\,P, \tag{3.1.25}$$

where $\boldsymbol{K}_X$ is the covariance matrix of the input and n is the number of uses of the channel or equivalently the row and column dimension of $\boldsymbol{K}_X$.
As already shown in the previous section, to maximize the mutual information, we have to chose a Gaussian distributed input. The noise and the input are still assumed to be statistically independent, and thus, (3.1.15) extends to

$$\boldsymbol{K}_Y = \boldsymbol{K}_X + \boldsymbol{K}_Z, \tag{3.1.26}$$

and

$$h(\boldsymbol{Y}) = \frac{1}{2}\log\left((2\pi e)^n \det\left(\boldsymbol{K}_X + \boldsymbol{K}_Z\right)\right).$$

Thus, since $h(\boldsymbol{Z})$ is given by the channel system, the mutual information (3.1.14) is optimized by maximizing $h(\boldsymbol{Y})$ and hence, the problem is reduced to maximizing $\det\left(\boldsymbol{K}_X + \boldsymbol{K}_Z\right)$. We may rewrite the latter

$$\begin{aligned}\det\left(\boldsymbol{K}_X + \boldsymbol{K}_Z\right) &= \det\left(\boldsymbol{K}_X + \boldsymbol{U}_Z \Lambda_Z \boldsymbol{U}_Z{}^{\mathsf{T}}\right)\\ &= \det\left(\boldsymbol{U}_Z\right)\det\left(\boldsymbol{U}_Z{}^{\mathsf{T}}\boldsymbol{K}_X\boldsymbol{U}_Z + \boldsymbol{\Lambda}_Z\right)\det\left(\boldsymbol{U}_Z{}^{\mathsf{T}}\right)\\ &= \det\left(\widetilde{\boldsymbol{K}}_X + \boldsymbol{\Lambda}_Z\right),\end{aligned}$$

where $\boldsymbol{\Lambda}_Z = \mathrm{diag}(\lambda_1^{(Z)}, \lambda_2^{(Z)}, ..., \lambda_n^{(Z)})$ is diagonalized by the unitary operation $\boldsymbol{U}_Z$, and $\widetilde{\boldsymbol{K}}_X = \boldsymbol{U}_Z{}^{\mathsf{T}}\boldsymbol{K}_X\boldsymbol{U}_Z$. Since

$$\mathrm{Tr}\left(\boldsymbol{U}_Z{}^{\mathsf{T}}\boldsymbol{K}_X\boldsymbol{U}_Z\right) = \mathrm{Tr}\left(\boldsymbol{U}_Z{}^{\mathsf{T}}\boldsymbol{U}_Z\boldsymbol{K}_X\right) = \mathrm{Tr}\left(\boldsymbol{K}_X\right),$$

the energy constraint on the input trace remains unchanged. Using Hadamard's inequality, which states that for the determinant of any positive definite matrix $\boldsymbol{M}$

$$\det \boldsymbol{M} \leq \prod_i \boldsymbol{M}_{ii},$$

holds, with equality if $\boldsymbol{M}$ is diagonal, we deduce that

$$\det\left(\widetilde{\boldsymbol{K}}_X + \boldsymbol{\Lambda}_Z\right) \leq \prod_i \left(\widetilde{\boldsymbol{K}}_{X,ii} + \lambda_i^{(Z)}\right), \tag{3.1.27}$$

where $\lambda_i^{(Z)}$ are the eigenvalues of $\boldsymbol{K}_Z$.
Hence, in order to maximize the left hand side of (3.1.27), we have to chose $\boldsymbol{K}_X$ such that it diagonalizes in the same basis as $\boldsymbol{K}_Z$, because

$$\begin{aligned} \widetilde{\boldsymbol{K}}_X &= \boldsymbol{U}_Z{}^{\mathsf{T}} \boldsymbol{K}_X \boldsymbol{U}_Z \stackrel{!}{=} \boldsymbol{\Lambda}_X \\ \Leftrightarrow \boldsymbol{K}_X &= \boldsymbol{U}_Z \boldsymbol{\Lambda}_X \boldsymbol{U}_Z{}^{\mathsf{T}}. \end{aligned}$$

where $\boldsymbol{\Lambda}_X = \mathrm{diag}(\lambda_1^{(Z)}, \lambda_2^{(Z)}, ..., \lambda_n^{(Z)})$. Now, similarly to the case of uncorrelated, parallel channels, but in the rotated basis, we have to find the water-filling level μ, with

$$\widetilde{\boldsymbol{K}}_{X,ii} = (\mu - \lambda_i^{(Z)})^+,$$

such that

$$\mathrm{Tr}\left(\widetilde{\boldsymbol{K}}_{X,ii}\right) = n\,P. \tag{3.1.28}$$

Thus, in order to express the optimal input covariance matrix in the previous basis (as in (3.1.26)) one has to find the unitary operation which diagonalizes the noise $\boldsymbol{K}_Z$. The capacity is then calculated as in (3.1.24).

3.2 Classical channel with Gauss-Markov memory

We now consider the case of a stationary process, that is a process whose joint probability distribution and hence, mean and variance do not change when shifted in time. A covariance matrix for such a process was already introduced in (2.2.10). We state from [6] that in the limit of $n \to \infty$ the density of eigenvalues on the real line tends to the power spectrum of the stochastic process. Since the eigenvalue spectrum in this limit is no longer discrete, but continuous, the constraint (3.1.28) has to be rewritten in terms of an integral, that is

$$\frac{1}{\kappa}\int dx\ (\mu - \lambda^{(Z)}(x))^+ = P, \tag{3.2.1}$$

where $\lambda^{(Z)}(x)$ denotes the continuous eigenvalue spectrum of the noise and the integral is taken over the whole domain of the spectrum, normalized by κ. The previous water-filling is therefore transferred to the spectral domain and the capacity reads

$$C = \frac{1}{\kappa}\int dx\ \log\left(1 + \frac{(\mu - \lambda^{(Z)}(x))^+}{\lambda^{(Z)}(x)}\right). \tag{3.2.2}$$

In some sense we have either infinite uses or a set of infinite correlated channels, labeled by x, which will be used one time. We will now calculate (3.2.2) for the noise model introduced in (2.2.10).
First, we varify that the eigenvalue spectrum of the channel indeed tends to the power spectrum. Therefore we rewrite the results we already found in section 2.3.
The eigenvalue function of (2.2.10), found in (2.3.22), read

$$\lambda^{\mathrm{env}}(x) \equiv \lambda^{(\boldsymbol{\Gamma})}(x) = N\,\frac{1-\phi^2}{1+\phi^2-2\phi\cos x}. \quad 0 \le \phi < 1, x \in [0,\pi]$$

$\lambda^{\text{env}}(x)$ is continous and it has a continuous derivative in $[0,\pi]$. Thus, we can expand the latter in a Fourier series (see [11][0.325]) of the form

$$\frac{a_0}{2}+\sum_{k=1}^{\infty} a_k\,\cos(kx),$$

with the coefficients

$$a_k=\frac{2}{\pi}\int_0^{\pi} dt\;\lambda^{\text{env}}(t)\cos(kt).$$

These coefficients were already determined in (2.3.10) (with the substitution introduced in (2.3.11)). Thus, we recover in the limit case, the power spectral density function of (2.2.10)

$$1+\sum_{k=1}^{\infty} 2\,\phi^k\cos(kx)\equiv\frac{1}{N}\,\lambda^{\text{env}}(x),$$

which confirms the previous statement.
Since the function $\lambda^{\text{env}}(x)$ is decreasing for $0\leq\phi<1$, we can simplify the left hand side of (3.2.1)

$$\frac{1}{\pi}\int_0^{\pi}(dx\;\mu-\lambda^{\text{env}}(x))^{+}=\frac{1}{\pi}\int_{\alpha}^{\pi} dx\;\mu-\lambda^{\text{env}}(x). \tag{3.2.3}$$

To determine the position α and thus μ we need to solve the equation

$$(\pi-\alpha)\,\mu-\int_{\alpha}^{\pi} dx\;\lambda^{\text{env}}(x)=\pi\,P, \tag{3.2.4}$$

see figure 3.6.
In order to simplify the latter we calculate the primitive of $\lambda^{\text{env}}(x)$, that is:

$$\int dx\;\lambda^{\text{env}}(x)=2N\arctan\left(\frac{1+\phi}{1-\phi}\tan\left(\frac{x}{2}\right)\right)\equiv L^{(\mathbf{\Gamma})}(x). \tag{3.2.5}$$

Since

$$L^{(\mathbf{\Gamma})}(\pi)=\pi N,$$

we can rewrite (3.2.4) as:

$$(\pi-\alpha)\,\mu+L^{(\mathbf{\Gamma})}(\alpha)=\pi\,(P+N). \tag{3.2.6}$$

Once α and μ are determined the capacity (3.2.2) reads

$$C^{(\text{cl})}=\frac{1}{2\pi}\int_{\alpha}^{\pi} dx\;\log\left(\frac{\mu}{\lambda^{\text{env}}(x)}\right) \tag{3.2.7}$$

Regarding the latter in the uncorrelated case $\phi = 0$, i.e. when the noise spectrum is constant, that is

$$\lambda^{\text{env}}(\phi = 0) = N,$$

we find $\alpha = 0$ and the water-filling $\mu = P + N$. The capacity then reads

$$\begin{aligned} C^{(\text{cl})} &= \frac{1}{2\pi} \int_0^\pi \log\left(1 + \frac{P + N - N}{N}\right) \\ &= \frac{1}{2} \log\left(1 + \frac{P}{N}\right), \end{aligned} \tag{3.2.8}$$

which is just the single variate expression (3.1.16).
For each correlation strength $\phi \neq 0$ we have to distinguish between two cases: the case when the input power is sufficient to use the whole spectrum, and the case when it is insufficient.

3.2.1 Sufficient power to use the whole spectrum

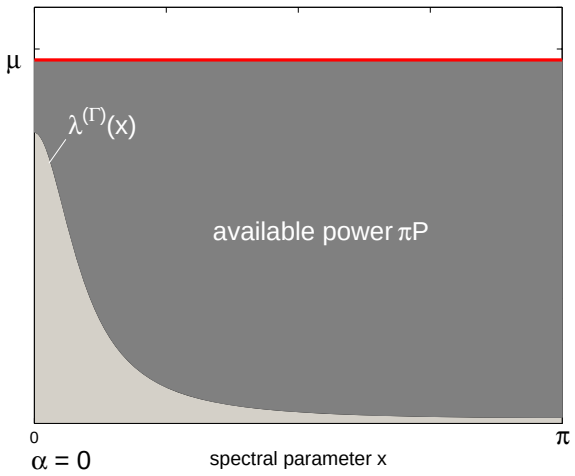

Figure 3.4: The input power πP suffices to "fill" the spectrum completely. $\alpha = 0$, hence $\mu = P + N$. Here $\phi \neq 0$.

For a given input power per channel P we know we will be able to have to include all channels in the communication at the output if

$$\begin{aligned} \pi P &\geq \pi \lambda^{\text{env}}(0) - \int_0^\pi dx\ \lambda^{\text{env}}(x) \\ P &\geq N\left(\frac{1+\phi}{1-\phi} - 1\right), \end{aligned} \tag{3.2.9}$$

where the integral was already solved in (2.3.24), holds. Therefore we define

$$P_{\mathrm{s}}(\phi) = \left(\frac{1+\phi}{1-\phi} - 1\right), \tag{3.2.10}$$

as the input power which is needed to use the full spectrum. The first line of inequality (3.2.9) is sketched in figure 3.4. We observe that the power which is needed to use the whole channel spectrum tends to infinity when the correlation parameter $\phi \to 1$. This is obvious, since the highest noise eigenvalue (or the channel with the strongest noise), tends to infinity as well in this limit, i.e.

$$\lim_{\phi\to 1} \lambda^{\mathrm{env}}(0) = N \lim_{\phi\to 1} \frac{1+\phi}{1-\phi} \to \infty. \tag{3.2.11}$$

If we assume that inequality (3.2.9) is indeed fulfilled we set $\alpha := 0$ and

$$\mu = P + \frac{1}{\pi}\int_0^\pi \lambda^{\mathrm{env}}(x) = P + N, \tag{3.2.12}$$

which is just the total average power of the system. The capacity (3.2.7) reads

$$\begin{aligned} C_{\mathrm{s}}^{(\mathrm{cl})} &= \frac{1}{2}\log\left(P+N\right) - \underbrace{\frac{1}{2\pi}\int_0^\pi dx\ \log\lambda^{\mathrm{env}}(x)}_{=\frac{1}{2}\log\left(N(1-\phi^2)\right),(2.3.23)}, \quad P \geq P_{\mathrm{s}}, \\ &= \frac{1}{2}\log\left(\frac{1}{1-\phi^2}\left(1+\frac{P}{N}\right)\right), \quad P \geq P_{\mathrm{s}}, \end{aligned} \tag{3.2.13}$$

where the integral is simply the nth square root of the determinant of the noise covariance matrix in the asymptotic limit, which was already found in (2.3.23). We conclude that in the limit of full correlations $\phi \to 1$ the channel capacity $C_{\mathrm{s}}^{(\mathrm{cl})}$ tends to infinity, since the logarithm is a monotonically increasing function. This is actually not really surprising since a full correlated channel would correspond to a channel with the same noise over all uses. This noise could therefore be dropped and would lead to an infinite transmission rate.

We plotted in figure 3.5 the capacity vs. the correlation strength ϕ for different noise strengths N and for a power $P = P_{\mathrm{s}}(\phi) + 1$, i.e.

$$C_{\mathrm{s}}^{(\mathrm{cl})}(P = P_{\mathrm{s}} + 1) = \frac{1}{2}\log\left(\frac{1}{(1-\phi)^2} + \frac{1}{(1-\phi^2)N}\right). \tag{3.2.14}$$

We observe that with increasing correlation strength the capacities get closer and closer, since the noise difference (one and two, respectively, orders of magnitude) becomes less and less important due to the dominance of the two diverging terms.

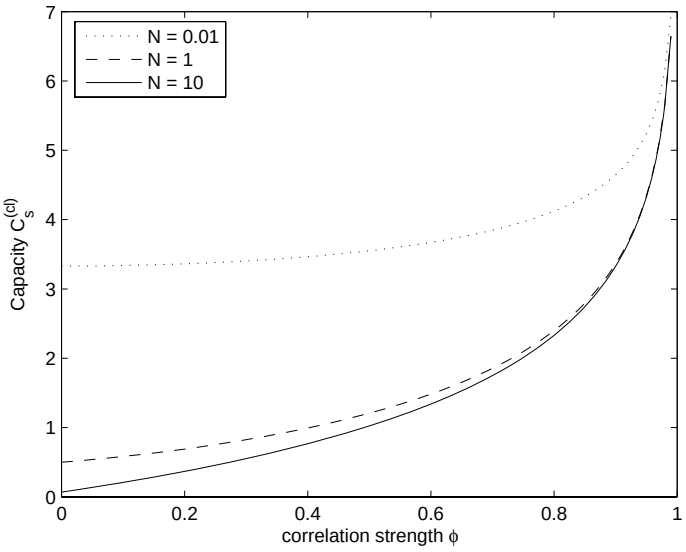

Figure 3.5: Capacity $C_{\rm s}^{(\rm cl)}$ vs. correlation strength ϕ. For all curves we took $P = P_{\rm s}(\phi)+1$. The dotted, dashed and solid line correspond to $N = 0.01, 1, 10$.

3.2.2 Insufficient power to use the whole spectrum

If the power is insufficient to use the whole channel spectrum (see figure 3.6)

$$P < P_{\rm s} \tag{3.2.15}$$

is fulfilled. Hence, one has to solve (3.2.6), where

$$\mu = \lambda^{\rm env}(\alpha). \tag{3.2.16}$$

The capacity then reads

$$C_{\rm is}^{(\rm cl)} = \frac{1}{2\pi}(\pi - \alpha)\log\lambda^{\rm env}(\alpha) - \frac{1}{2\pi}\int_{\alpha}^{\pi} dx\ \log\lambda^{\rm env}(x). \tag{3.2.17}$$

As in the previous section we will have a look at the limit of $C_{\rm is}^{(\rm cl)}$ if the channel becomes fully correlated. We have noticed already in (3.2.11) that the noisiest channel is diverging in the limit of $\phi \to 1$. Since

$$\frac{1}{\pi N}\int_{0}^{\pi} dx\ \lambda^{\rm env}(x) = \frac{1}{N}\lim_{n\to\infty}\frac{1}{n}{\rm Tr}\left(\widetilde{\boldsymbol{K}}_X\right) = 1,$$

the eigenvalue spectrum for $\phi \to 1$ tends everywhere else to zero in order to keep the normalization satisfied. Thus, α will tend to 0 and the water-filling μ will tend to the

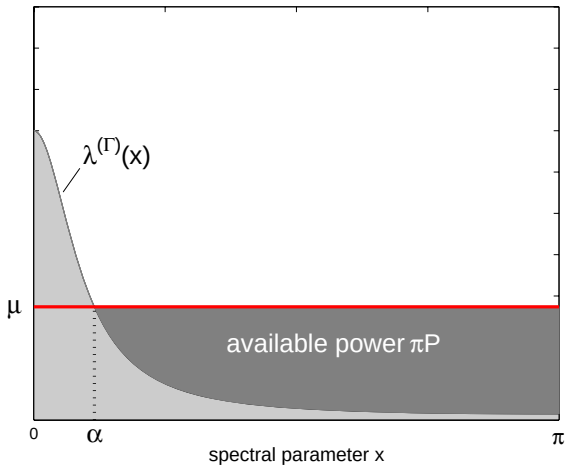

Figure 3.6: The input power πP is insufficient to "fill" the whole spectrum. The water-filling level μ is found at the position α by solving equation (3.2.6). Here $\phi \neq 0$.

power constraint P. This is due to the following reasoning: in $L^{(\mathbf{\Gamma})}(x)$ (see (3.2.5)) the function $\frac{1+\phi}{1-\phi}$ tends stronger to infinity (with $\phi \to 1$) than $\tan\frac{x}{2}$ to zero (with $x \to 0$), see figure 3.7, 3.8. Therefore,

$$\lim_{\phi\to 1} L^{(\mathbf{\Gamma})}(x) = \pi N$$

and using (3.2.6) we conclude that

$$\lim_{\phi\to 1} \mu = P. \tag{3.2.18}$$

The behavior of the capacity (3.2.17) in this limit can be computed in several steps:

$$\begin{aligned}
\lim_{\phi\to 1} C_{\text{is}}^{(\text{cl})} &= \lim_{\phi\to 1} \frac{1}{2\pi}(\pi-\alpha)\log\mu - \lim_{\phi\to 1}\frac{1}{2\pi}\int_{\alpha}^{\pi} dx\ \log\lambda^{\text{env}}(x) \\
&= -\lim_{\epsilon\to 0}\frac{1}{2\pi}\int_{0}^{\pi} dx\ \log\frac{1-(1-\epsilon)^2}{1+(1-\epsilon)^2-2(1-\epsilon)\cos(x)} + c_\mu \\
&= -\lim_{\epsilon\to 0}\frac{1}{2\pi}\int_{0}^{\pi} dx\ \log\frac{\epsilon}{(1-\epsilon)(1-\cos(x))} + c_\mu \\
&= \frac{1}{2}\lim_{\epsilon\to 0}\log\frac{\epsilon}{1-\epsilon}\underbrace{-\frac{1}{2\pi}\int_{0}^{\pi} dx\ \log\left(1-\cos(x)\right)}_{=1/2} + c_\mu \\
&\to \infty.
\end{aligned} \tag{3.2.19}$$

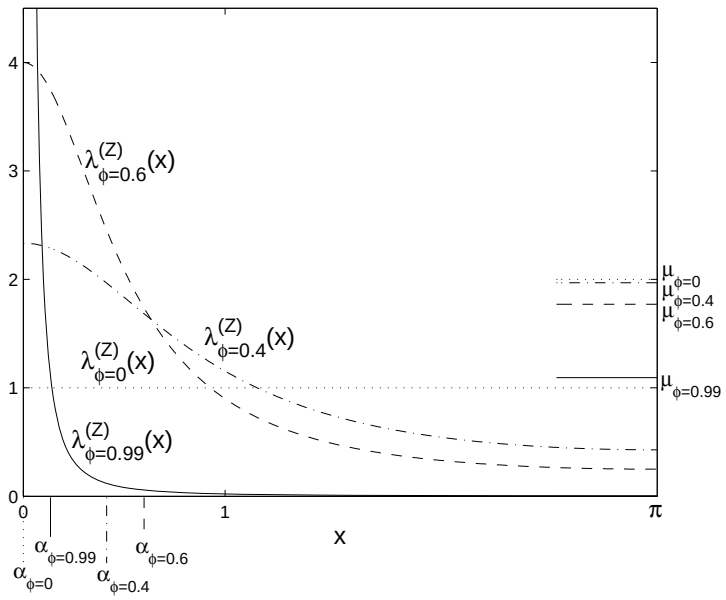

Figure 3.7: $\lambda^{\text{env}}(x)$ vs. x for $\phi = \{0, 0.4, 0.6, 0.99\}$, with $P = 1$, $N = 1$ for all plots (for all ϕ we have $P \leq N((1+\phi)/(1-\phi))$). With increasing ϕ $\lambda^{\text{env}}(x)$ is shifted more and more towards zero. In addition, the different water-filling levels μ at positions α are displayed. With increasing ϕ the water-filling sinks towards P. One observes that α increases and decreases again with increasing ϕ (for more details, see figure 3.9).

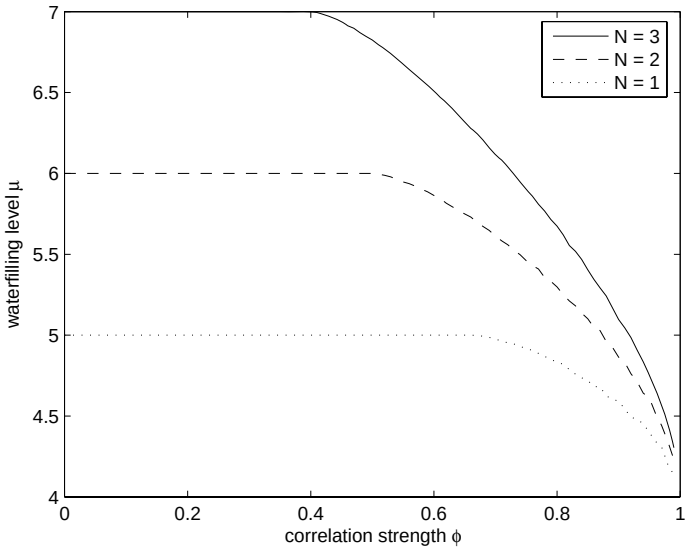

Figure 3.8: The water-filling level μ vs. correlation strength ϕ. The solid, dashed and dotted curves correspond to $N = 3, 2, 1$. For all graphs we fixed $P = 4$.

where all terms of order ϵ^2 are neglected and $c_\mu = \frac{1}{2} \log P$. We also observe that in the case where not all channels can be used, the overall output rate does diverge in the case of full correlations. In figure 3.8 and 3.9 the water filling level μ and position α are plotted vs. the correlation ϕ for different noises $N = 1, 2, 3$ and with constant input power $P = 4$. α is zero in the region where P is sufficient to fill the whole spectrum, according to (3.2.6). For $\phi \to 1$ α tends for all noise values to zero. We observe that α attains a maximum for some correlation strength ϕ. This somehow reflects the behavior of the noise spectrum which can be completely filled at the beginning, maximally filled at some correlation strength and then almost completely filled again, when all but the highest eigenvalue tend to zero with $\phi \to 1$. The water filling level is constant in the region where the complete spectrum can be used, i.e. (3.2.12) holds. In the region where P becomes insufficient to fill the whole spectrum μ decreases towards the limit (3.2.18).

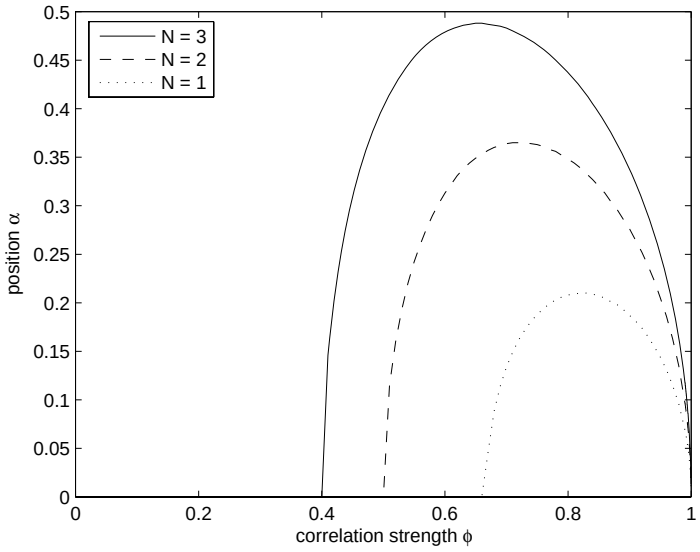

Figure 3.9: Position α vs. correlation strength ϕ. The solid, dashed and dotted curves correspond to $N = 3, 2, 1$. For all graphs we fixed $P = 4$.

The capacities and positions α are plotted vs. the input power P in figure 3.10 for different noise strengths N and a constant correlation strength ϕ. We started the plot at $P = 0.01$. α decreases to zero and at the point where $\alpha = 0$ we recover $P = P_{\rm s}$. The capacities are plotted in the region where $\alpha > 0$ with $C^{(\rm cl)} = C_{\rm is}^{(\rm cl)}$ and where $\alpha = 0$ with $C^{(\rm cl)} = C_{\rm s}^{(\rm cl)}$.

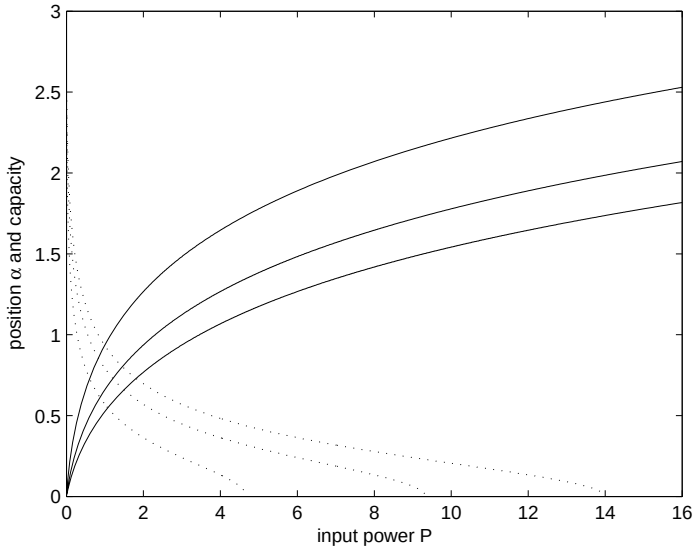

Figure 3.10: Capacities and positions α vs. input power P for different fixed noises. The dotted curves show α and correspond from left to right to $N = 1, 2, 3$. The solid curves show $C_{\rm is}^{\rm (cl)}, C_{\rm s}^{\rm (cl)}$, respectively, and correspond from top to bottom to $N = 1, 2, 3$. For all curves we took $\phi = 0.7$.

Chapter 4

Quantum Part

This chapter concerns the central topic of this work: the bosonic channel with Gauss-Markov correlated noise. After treating the monomodal anisotropic bosonic channel explicitly we apply most of its results to the multimode channel with the previously introduced memory model.

4.1 Quantum Optics

The following definitions are taken from [29] and [3].

4.1.1 Quantization of the electromagnetic field

The quantization of the electromagnetic field requires the definition of the source free[1] Maxwell equations, that is

$$\nabla \cdot \boldsymbol{B} = 0 \tag{4.1.1}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial}{\partial t}\boldsymbol{B} \tag{4.1.2}$$

$$\nabla \cdot \boldsymbol{D} = 0 \tag{4.1.3}$$

$$\nabla \times \boldsymbol{H} = \frac{\partial}{\partial t}\boldsymbol{D}, \tag{4.1.4}$$

with the electric field $\boldsymbol{E}$, the magnetic field $\boldsymbol{B}$, the electric displacement field $\boldsymbol{D}$ and the magnetizing field $\boldsymbol{H}$. By introducing a non-physical vector potential $\boldsymbol{A}$

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \tag{4.1.5}$$

$$\boldsymbol{E} = -\frac{\partial}{\partial t}\boldsymbol{A}, \tag{4.1.6}$$

and the Coulomb gauge $\nabla \cdot \boldsymbol{A} = 0$, we retrieve by inserting (4.1.5) in (4.1.4) the well known wave equation

$$\nabla^2 \boldsymbol{A}(\boldsymbol{r}, t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\boldsymbol{A}(\boldsymbol{r}, t), \tag{4.1.7}$$

[1] There charge density and the current density are fixed to zero, i.e. $\rho \equiv 0$, $\boldsymbol{j} \equiv 0$.

where c is the speed of light. By decomposing $\boldsymbol{A}(\boldsymbol{r},t)$ in its orthonormal modes and inserting $\boldsymbol{A}(\boldsymbol{r},t)$ in (4.1.6) we receive the solution of (4.1.7) in terms of the electric field, that is

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r},t) &= \sum_{\lambda=1}^{2}\sum_{k} E_k\, \boldsymbol{e}_k^{(\lambda)} \left[\alpha_{k,\lambda} e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega_k t)} + \alpha_{k,\lambda}^* e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega_k t)}\right], \\ E_k &= \sqrt{\frac{\hbar\omega_k}{2\epsilon_0}},\end{aligned} \tag{4.1.8}$$

with the propagation vector $\boldsymbol{k}$, the polarization vector $\boldsymbol{e}_k^{(\lambda)}$ (where λ stands for the two possible transversal polarizations, $\lambda = 1, 2$), the angular frequency ω_k and the complex amplitudes $\alpha_{k,\lambda}, \alpha_{k,\lambda}^*$ of the mode k, the reduced Planck constant $\hbar$ and the vacuum permittivity ϵ_0.
The electromagnetic field is now quantized by replacing the complex amplitudes by annihilation and creation operators:

$$\alpha_{k,\lambda}, \alpha_{k,\lambda}^* \to \hat{a}_{k,\lambda}, \hat{a}_{k,\lambda}^\dagger.$$

Since photons are bosons the latter fullfill the boson commutation relations

$$\begin{aligned}[\hat{a}_{k,\lambda}, \hat{a}_{k',\lambda'}^\dagger] &= \delta_{kk'}\delta_{\lambda\lambda'} \\ [\hat{a}_{k,\lambda}, \hat{a}_{k',\lambda'}] &= [\hat{a}_{k,\lambda}^\dagger, \hat{a}_{k',\lambda'}^\dagger] = 0.\end{aligned} \tag{4.1.9}$$

Furthermore, the Hamiltonian of the quantized electromagnetic field reads[2]

$$\hat{H} = \sum_k \hbar\omega_k \left(\hat{N}_k + \frac{1}{2}\right), \tag{4.1.10}$$

with the number operator

$$\hat{N}_k = \hat{a}_k^\dagger \hat{a}_k, \tag{4.1.11}$$

and the energy of the vacuum $\frac{1}{2}\hbar\omega_k$. Hence, we have decomposed the electromagnetic field into harmonic oscillators, each with energy $\hbar\omega_k(\hat{N}_k + 1/2)$.

For simpler readability we will from now on change the physical units to natural units, i.e. setting

$$\hbar = c = ... \equiv 1.$$

4.1.2 The Quadrature operators

The Hamiltonian (4.1.10) for the mode k in terms of its "position" and "momentum" operators[3], $\hat{X}_k, \hat{P}_k$, reads

$$\hat{H}_k = \frac{1}{2}\left(\hat{P}_k + \omega_k^2 \hat{X}_k^2\right), \tag{4.1.12}$$

[2]For simplicity we drop from now on the sum over the two possible polarizations $\lambda = 1, 2$

[3]These operators do not correspond to the position and momentum operators (=observables) of a particle as in "ordinary" quantum mechanics. The naming stems from the fact that they obey the same commutation relations.

with

$$\begin{aligned}\hat{X}_k &= \sqrt{\frac{1}{2\omega_k}}\left(\hat{a}_k + \hat{a}_k^\dagger\right), \\ \hat{P}_k &= -i\sqrt{\frac{\omega_k}{2}}\left(\hat{a}_k - \hat{a}_k^\dagger\right),\end{aligned} \tag{4.1.13}$$

which fulfill the well-known commutation relation

$$\left[\hat{X}_k, \hat{P}_{k'}\right] = i\delta_{kk'}. \tag{4.1.14}$$

From now on we will always use the dimensionless notation

$$\begin{aligned}\hat{x}_k &\equiv \sqrt{\frac{\omega_k}{2}}\hat{X}_k = \Re(\hat{a}_k) = \hat{a}_k + \hat{a}_k^\dagger \\ \hat{p}_k &\equiv \sqrt{\frac{1}{2\omega_k}}\hat{P}_k = \Im(\hat{a}_k) = \frac{1}{2i}(\hat{a}_k - \hat{a}_k^\dagger),\end{aligned} \tag{4.1.15}$$

with the commutation relation

$$[\hat{x}_k, \hat{p}_{k'}] = \frac{i}{2}\delta_{kk'}.$$

These operators represent the *quadratures* of the mode k, which corresponds in the classical regime to the real and imaginary parts of the oscillator's complex amplitude.
As a result from the commutation relation (4.1.2) follows the famous Heisenberg uncertainty princple:

$$\Delta\hat{x}_k \Delta\hat{p}_k \geq \frac{1}{2}, \tag{4.1.16}$$

where $\Delta\hat{A}$ denotes the uncertainty of the predicted measurement outcome of the observable $\hat{A}$:

$$\Delta\hat{A} = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2. \tag{4.1.17}$$

4.1.3 Fock/Number states

The *Fock state* or *number state* is a quantum state with a well-defined number of particles (here photons). Thus, by Heisenberg's uncertainty principle, it has a completely random phase. It is the eigenstate of the number operator defined in (4.1.11), i.e.

$$\hat{a}_k^\dagger \hat{a}_k |n_k\rangle = n_k |n_k\rangle,$$

where $|n_k\rangle$ denotes a fock state with photon number $n_k = 0, 1, ..., \infty$ associated to the mode k. The ground state (or vacuum state) is defined by

$$\hat{a}_k |0\rangle = 0.$$

The actions of the annihilation and creation (or "ladder") operators read

$$\begin{aligned}\hat{a}_k |n_k\rangle &= \sqrt{n_k}\,|n_k - 1\rangle \\ \hat{a}_k^\dagger |n_k\rangle &= \sqrt{n_k + 1}\,|n_k + 1\rangle.\end{aligned} \tag{4.1.18}$$

Furthermore, the number states provide an orthogonal,

$$\langle n_k | m_k \rangle = \delta_{mn}$$

and complete

$$\sum_{n=0}^{\infty} |n_k\rangle \langle n_k| = 1$$

basis which spans the so called *Fock space*.

4.1.4 Coherent states

In contrary to the Fock states *coherent states* have a precise phase and thus, a completely random number of photons. The coherent states belong to the class of *minimal uncertainty states*. These are states which saturate Heisenberg's uncertainty principle, i.e.

$$\Delta\hat{x}\Delta\hat{p} = \frac{1}{2}. \tag{4.1.19}$$

In addition, the quadratures of a coherent state have equal uncertainties, that is

$$\Delta\hat{x} = \Delta\hat{p}.$$

A coherent state $|\zeta\rangle$ is generated by "displacing the vacuum" in the phase space by the complex number ζ, that is

$$\hat{D}(\zeta)\,|0\rangle = |\zeta\rangle\,.$$

with the unitary *displacement operator*

$$\hat{D}(\zeta) = e^{\zeta\hat{a}^\dagger - \zeta^*\hat{a}}, \tag{4.1.20}$$

where $\hat{a}^\dagger, \hat{a}$ are the creation and annihilation operator. Furthermore, the displacment operator fulfills the following properties:

$$\begin{aligned} \hat{D}^\dagger(\zeta) &= \hat{D}^{-1}(\zeta) = \hat{D}(-\zeta), \\ \hat{D}^\dagger(\zeta)\,\hat{a}\,\hat{D}(\zeta) &= \hat{a} + \zeta, \\ \hat{D}^\dagger(\zeta)\,\hat{a}^\dagger\,\hat{D}(\zeta) &= \hat{a} + \zeta^*. \end{aligned} \tag{4.1.21}$$

In addition we state that coherent states are eigenstates of the annihilation operator, i.e.

$$\hat{a}\,|\zeta\rangle = \zeta\,|\zeta\rangle\,.$$

At last, we remark that coherent states are non-orthogonal, that is

$$|\,\langle\zeta_1|\zeta_2\rangle\,|^2 = e^{-|\zeta_1-\zeta_2|^2}.$$

However, coherent states satisfy the completeness relation

$$\frac{1}{\pi}\int \mathrm{d}^2\zeta\ \ |\zeta\rangle\langle\zeta| = 1,$$

where $\mathrm{d}^2\zeta\ = d\Re(\zeta)\,d\Im(\zeta)$.

4.1.5 Squeezed states

The *squeezed state* also belongs to the class of minimal uncertainty states. Unlike the coherent state with symmetric uncertainties, its quantum fluctuations in one variable are reduced at the expense of the fluctuations in the other, such that (4.1.19) is always fulfilled. The squeezed states are generated by using the unitary *squeezing operator*

$$S(\epsilon) = \exp\left[\frac{1}{2}(\epsilon^* \hat{a}^2 - \epsilon\, \hat{a}^{\dagger 2})\right], \tag{4.1.22}$$

where $\epsilon = re^{2i\phi}$, with the *squeezing parameter* r and the *squeezing angle* ϕ. One may create an arbitrary squeezed state by squeezing the vacuum and afterwards displacing it, that is

$$|\zeta, \epsilon\rangle = D(\zeta)S(\epsilon)\,|0\rangle\,.$$

4.1.6 Gaussian states

Gaussian states may be represented in the Fock basis or the coherent basis (i.e. the state space). However, it is often easier to switch to the phase space where we treat the Gaussian state in terms of its quadratures. Similarly to the classical case, where a Gaussian distributed random variable is fully characterized by its first and second moments (see section 2.1.3), in the quantum case a *Gaussian state* is fully characterized by the first and second moments of its quadrature operators. The following expressions are taken from [17] and [3].
At first, let us group the quadratures of an n-mode state together in one $2n$-dimensional column vector

$$\hat{\boldsymbol{R}} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n, \hat{p}_1, \hat{p}_2, ..., \hat{p}_n)^{\mathsf{T}}, \tag{4.1.23}$$

where the quadratures satisfy the commutation relations, as defined in (4.1.14), and thus

$$[\hat{R}_k, \hat{R}_{k+n}] = \frac{i}{2} J_{k,k+n}, \quad k = 1, ..., n$$

with the *symplectic (or commutation) matrix*

$$\boldsymbol{J}_{2n\times 2n} = \begin{pmatrix} \mathbb{0}_{n\times n} & \mathbb{1}_{n\times n} \\ -\mathbb{1}_{n\times n} & \mathbb{0}_{n\times n} \end{pmatrix}. \tag{4.1.24}$$

Next, we define the mean

$$\boldsymbol{m} = \langle \hat{\boldsymbol{R}} \rangle = \mathrm{Tr}\left(\rho\, \hat{\boldsymbol{R}}\right)$$

with the density operator ρ and furthermore, the covariance matrix

$$\boldsymbol{\gamma} = \mathrm{Tr}\left(\left(\hat{\boldsymbol{R}} - \boldsymbol{m}\right)\, \rho \left(\hat{\boldsymbol{R}} - \boldsymbol{m}\right)^{\mathsf{T}}\right) \tag{4.1.25}$$

The density operator ρ is called Gaussian, if

$$\mathrm{Tr}\left(\rho\, \exp\left(i\hat{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{z}\right)\right) = \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\boldsymbol{\gamma}\boldsymbol{z} + i\boldsymbol{m}^{\mathsf{T}}\boldsymbol{z}\right), \tag{4.1.26}$$

holds, where the right hand side is called the *quantum characteristic function*, and $\boldsymbol{z}^{\mathsf{T}} = [x_1, x_2, ..., x_n; y_1, y_2, ..., y_n]^{\mathsf{T}}$ is a 2n-dimensional real column vector. In addition to the conditions on a covariance matrix in the classical case, as already mentioned in section 2.1.2, a covariance matrix $\boldsymbol{\gamma}$ generating a Gaussian state ρ must obey Heisenberg's uncertainty principle, expressed by

$$\boldsymbol{\gamma} - \frac{i}{2}\boldsymbol{J} \geq 0. \tag{4.1.27}$$

In addition we state, that an n-mode Gaussian state ρ is pure, if

$$\det \boldsymbol{\gamma}_{2n\times 2n} = \frac{1}{16^n}$$

In the following we present the covariance matrices for one mode states which will be used often throughout this work.

Vacuum The vacuum has zero mean and is centered at the origin of the phase space. The variances are the so called *vacuum fluctuations*, i.e. the saturated Heisenberg inequality. The covariance matrix therefore reads

$$\boldsymbol{\gamma}_{\text{vac}} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.1.28}$$

Coherent state The coherent state is simply the displaced vacuum. The second moments are unchanged and read as in (4.1.28).

Squeezed state The squeezed and displaced vacuum with squeezing factor r leads to the covariance matrix

$$\boldsymbol{\gamma}_{\text{squ}} = \frac{1}{2}\begin{pmatrix} e^{2r} & 0 \\ 0 & e^{-2r} \end{pmatrix}. \tag{4.1.29}$$

If $r < 0$ the state is squeezed in the $\hat{x}$-quadrature and anti-squeezed in the $\hat{p}$-quadrature, if $r > 0$ the state is anti-squeezed in $\hat{x}$ and squeezed in $\hat{p}$.

Thermal state The thermal state, i.e. the state of the harmonic oscillator in thermal equilibrium with average number of photons $\overline{n}$ with zero mean is given by the covariance matrix

$$\boldsymbol{\gamma}_{\text{th}} = \frac{1}{2}\begin{pmatrix} 2\overline{n}+1 & 0 \\ 0 & 2\overline{n}+1 \end{pmatrix}. \tag{4.1.30}$$

For a thermal state with zero average photons or equivalently zero temperature we recover the ground state of the oscillator which is again the state of the vacuum (4.1.28). Schematic representations of the mentioned one mode states are shown in figure 4.1.

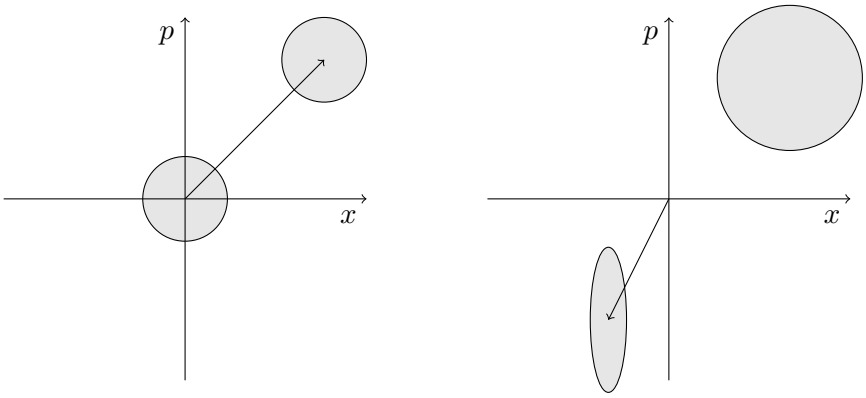

Figure 4.1: Left image: Phase space (schematic) representation of the uncertainty of the vacuum state (centered around zero) and a coherent state, i.e. the displaced vacuum. Right image: Phase space (schematic) representation of a displaced squeezed state (ellipse) and a thermal state (big circle).

Gaussian unitary operations When applying a unitary operation U on a Gaussian density operation ρ by

$$\rho \to U\,\rho\,U^\dagger, \tag{4.1.31}$$

the corresponding covariance matrix $\boldsymbol{\gamma}$ undergoes a *symplectic transformation* $S \in Sp(2N, \mathbb{R})$ [1], that is

$$\boldsymbol{\gamma} \to S\,\boldsymbol{\gamma}\,S^\mathsf{T}. \tag{4.1.32}$$

Both, the unitary and the symplectic transformation preserve the canonical commutation relations. For the symplectic transformation this condition reads

$$S\,\boldsymbol{J}\,S^\mathsf{T} = \boldsymbol{J}. \tag{4.1.33}$$

Within the set of symplectic transformation there are the so called *passive* transformations which preserve the trace of $\boldsymbol{J}$ and therefore do not change the number of photons of the state. An example for such a transformation is the action of a 50:50 beam splitter on two incoming modes, given by

$$S_{BS} = \begin{pmatrix} \sqrt{T}\mathbb{1} & \sqrt{1-T}\mathbb{1} \\ \sqrt{1-T}\mathbb{1} & \sqrt{T}, \end{pmatrix} \tag{4.1.34}$$

where T is the transmittivity.

4.2 Quantum information theory

4.2.1 A few words about entanglement

The maybe most interesting phenomenon of quantum mechanics is *entanglement* between quantum states, which leads to a "spooky action at a distance" (Einstein). This property of nature, although already discussed by Erwin Schrödinger in 1935 [25], is still an everyday topic in metaphysics or philosophy as it is lacking any intuitive explanation. In discrete variables we define an entangled state as follows. Suppose we have two photons A and B which both can be in one of the two eigenstates $|0\rangle$ or $|1\rangle$ (for example the horizontal and vertical polarization). We denote the state of photon A by $|\psi\rangle_A$ and the state of photon B by $|\psi\rangle_B$ where both live in the two dimensional Hilbert spaces $\mathcal{H}_\mathcal{A}$, $\mathcal{H}_\mathcal{B}$, respectively. Let us assume that both photons are always polarized orthogonally. If we do not know the initial state of A and B simultaneously and superimpose both states then the composed system can be for instance in the following state:

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|0\rangle_A |1\rangle_B + |1\rangle_A |0\rangle_B \right), \tag{4.2.1}$$

where the total Hilbert space $\mathcal{H}_{\mathcal{AB}} \ni |\Psi\rangle_{AB}$ is not separable, i.e. $\mathcal{H}_{\mathcal{AB}} \neq \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. This means that if one detects now that particle A is horizontally polarized then particle B will always be detected vertically polarized and vice versa. This "spooky action" is instantaneous and has been confirmed experimentally over distances of over 100km [28]. Since no information is exchanged this is not a violation of the speed of light. However, as long as one does not perform a measurement on either of the particles the total system remains entangled. In contrary to the perfect entangled DV state (4.2.1) the entangling of for instance laser modes in CV will be imperfect.[4] Since we will be dealing with entangled Gaussian states later in this work we introduce now the *two-mode squeezed vacuum state*, which is a two-mode entangled system. Assume we prepare two one mode squeezed states as introduced in (4.1.29), where one mode is squeezed in the $\hat{x}$ quadrature, the other one in the $\hat{p}$ quadrature. The covariance matrix of the total system then reads

$$\boldsymbol{\gamma} = \frac{1}{2} \begin{pmatrix} e^{2r} & 0 & 0 & 0 \\ 0 & e^{-2r} & 0 & 0 \\ 0 & 0 & e^{-2r} & 0 \\ 0 & 0 & 0 & e^{2r} \end{pmatrix}, \tag{4.2.2}$$

with squeezing parameter r. By sending both modes through a 50:50 beam splitter (see figure 4.2) we will receive the entangled output system[5]:

[4] However, we remark that the perfection of entanglement in DV comes together with a lower success rate for the generation of such states, while in CV finite degree entangled states can be easily prepared.

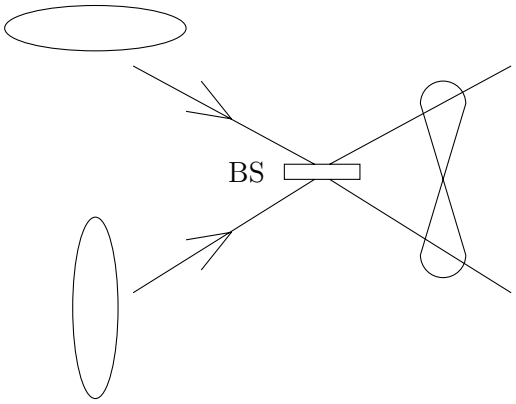

Figure 4.2: Two Gaussian modes, one squeezed in x, one squeezed in p enter a 50:50 beam splitter. The two outgoing modes are in a two mode entangled state.

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}\begin{pmatrix} \cosh 2r & \sinh 2r & 0 & 0 \\ \sinh 2r & \cosh 2r & 0 & 0 \\ 0 & 0 & \cosh 2r & -\sinh 2r \\ 0 & 0 & -\sinh 2r & \cosh 2r \end{pmatrix}, \tag{4.2.3}$$

We observe that the position quadratures are correlated, while the momentum quadratures are anti-correlated, i.e.

$$\Delta\hat{x}_1 - \hat{x}_2 = \Delta\hat{p}_1 + \hat{p}_2 = e^{-2r}, \tag{4.2.4}$$

In the limit of infinitely squeezed two modes (before the beam splitter), i.e. $r \to \infty$ the correlations and anti-correlations become perfect and one retrieves the EPR pair, introduced in the famous gedankenexperiment by Einstein, Podolski and Rosen in 1935 [7] which lead them to the belief "that the description of reality as given by a wave function is not complete".

4.2.2 Von Neumann entropy

Parts of the following content are taken from [23].
Similarly to the Shannon entropy which defines the uncertainty of a random variable the *von Neumann entropy* measures the uncertainty of a quantum state. At first, we define the *density operator* ρ as a composition of an ensemble $\{|x\rangle\} \in \mathcal{H}$ according to a probability distribution $p(x)$, that is

$$\rho = \sum_x p(x)\,|x\rangle\,\langle x|. \tag{4.2.5}$$

ρ is Hermitian ($\rho^\dagger = \rho$), positive-semidefinite and has trace one. In addition, we remark that a quantum state defined by ρ is in a *pure* state if $\text{Tr}\left(\rho^2\right) = 1$ and in a *mixed* state

[5] This is done mathematically by applying the beam splitter operation (4.1.34) on the covariance matrix (4.2.2)

otherwise.
For a given quantum state ρ we define the von Neumann entropy as

$$S(\rho) = -\mathrm{Tr}\left(\rho \log \rho\right). \tag{4.2.6}$$

Given the eigenvalue spectrum $\{\lambda_i\}$ of ρ the latter can be expressed as

$$\begin{aligned} S(\rho) &= -\sum_i \lambda_i \log \lambda_i \\ &= H(\lambda), \end{aligned}$$

where H denotes the Shannon entropy as defined in (3.1.1). As two examples of (4.2.6) we remark the two states which minimize and maximize it, respectively.
The pure state which can be expressed in terms of $|\psi\rangle$, that is $\rho_\psi = |\psi\rangle\langle\psi|$, has no entropy. That means, that the system is with certainty in one eigenstate state, namely $|\psi\rangle\langle\psi|$, and hence all eigenvalues of ρ_ψ except one are zero and thus, $S(\rho_\psi) = 0$.
In contrary, the completely mixed state (in a d-dimensional space) $\overline{\rho} = \sum_x 1/d\,|x\rangle\langle x|$ has a constant eigenvalue spectrum and the maximum entropy $\log d$. Hence, it is equally probable to find the system in any of its basis states $\{|x\rangle\}$.
For a state $\rho_{AB} \in \mathcal{H}_{\mathcal{AB}}$, composed by two systems A and B we state the joint entropy

$$S(\rho_{AB}) = -\mathrm{Tr}\left(\rho_{AB} \log \rho_{AB}\right)$$

and equivalent to the classical definitions, the conditional entropy and mutual information

$$S(\rho_A|\rho_B) = S(\rho_{AB}) - S(\rho_B), \tag{4.2.7}$$

$$I(\rho_A;\rho_B) = S(\rho_B) - S(\rho_B|\rho_A). \tag{4.2.8}$$

In addition we state the subadditivity property:

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B), \tag{4.2.9}$$

where ρ_A, ρ_B, are defined via the *partial trace:*

$$\mathrm{Tr}_A(\rho_{AB}) = \rho_B. \tag{4.2.10}$$

Equality in (4.2.9) holds if and only if $\rho_{AB} = \rho_A \otimes \rho_B$, i.e. if the system is seperable. Unlike the Shannon entropy the von Neumann entropy has some counterintuitive properties. In the classical case, given two random variables X and Y, the uncertainty of X can never be higher than the total uncertainty of X and Y, that is

$$H(X) \leq H(X,Y).$$

In the quantum case, however, this inequality can be violated. Consider the entangled state (analog to (4.2.1))

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),$$

where $|00\rangle$ ($|11\rangle$) denotes that both qubits are in the state $|0\rangle$ ($|1\rangle$). While, $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is in a pure state ($S(\rho_{AB}) = 0$) the system A may be in a maximally mixed state $\rho_A = 1/2(|0\rangle\langle 0| + |1\rangle\langle 1|)$ ($S(\rho_A) = 1$). Thus, in this example

$$S(\rho_B|\rho_A) < 0.$$

4.2.3 Quantum channels

Quantum channels have a wide application in the concept of quantum information theory. At the same time they can be used to describe the transmission of quantum states over distances (using for instance optical fibers) and the process of saving quantum information on a storage medium (using for instance trapped ions). The following definitions are taken from [22] and [19].
As for classical channels, we assume again a sender Alice and a receiver Bob. Instead of random variables we are now dealing with observables and density operators. The observables of a quantum system are given by bounded linear operators on the Hilbert space $\mathcal{H}$, denoted by $\mathcal{B}^*(\mathcal{H})$. The density operators ρ, representing the physical states, are living in the dual space $\mathcal{B}(\mathcal{H})$, $\rho \in \mathcal{B}(\mathcal{H})$.
We denote Alice's input register by $\mathcal{A} = \mathcal{B}(\mathcal{H}_\mathcal{A})$, Bob's output register by $\mathcal{B} = \mathcal{B}(\mathcal{H}_\mathcal{B})$ and the set of environment states by $\mathcal{E} = \mathcal{B}(\mathcal{H}_\mathcal{E})$.
We define a *quantum channel* T in the Schrödinger picture[6] as a completely positive trace preserving map, which transforms Alice's input states (living in $\mathcal{A}$) into Bob's output states (living in $\mathcal{B}$) as

$$T : \mathcal{A} \to \mathcal{B}. \tag{4.2.11}$$

Given the input state $\rho_A \in \mathcal{A}$, the environment state $\rho_E \in \mathcal{E}$ the act of T in a general form reads

$$T[\rho_A] = \mathrm{Tr}_E\Big(U_{AE}\,(\rho_A \otimes \rho_E)\,U_{AE}^\dagger\Big), \tag{4.2.12}$$

where U_{AE} is an unitary operator describing the common evolution of the inital system and the environment. In other words, the global system, including a generally noisy environment, evolves according to U_{AE} and is then reduced to the output by "tracing out" the environment.
(4.2.11) can be equivalently represented in the *Kraus form* as

$$T[\rho_A] = \sum_j V_j \rho_A V_j^\dagger, \tag{4.2.13}$$

with the *Kraus operators* V_j fulfilling $\sum_j V_j^\dagger V_j = 1$.
There is a class of channels for which $V_j \to \sqrt{p_j}\,V_j$, that means that the operator V_j is applied on ρ_A with probability p_j, which will come in handy for further channel definitions. Hence,

$$T'[\rho_A] = \sum_j p_j\, V_j \rho_A V_j^\dagger. \tag{4.2.14}$$

Before we move on to further channel definitions we impose a rather intuitive physical constraint: throughout this work we assume all channels to be *causal channels*, that are channels for which outputs up to some (discrete) time step t do not depend on inputs at times $t' > t$.[7]

[6] While in the Schrödinger picture states evolve in time and observables are kept constant, in the Heisenberg picture the opposite holds.
[7] For a formal definition one should take a look at [19].

Memoryless channels

Let us now assume Alice choses a message of length n from a classical alphabet and wants to transmit it to Bob using a quantum channel. In general this message is then encoded in a quantum state ρ_A sent through the channel and afterwards decoded[8].
We define the *memoryless quantum channel*, that is a channel which acts on each input state independently of the previous input, in terms of the general representation (4.2.12) as

$$T^{(n)}[\rho_A] = \mathrm{Tr}_E\Big(U_{A_nE_n} \cdot U_{A_{n-1}E_{n-1}} \cdots U_{A_1E_1} \left(\rho_A \otimes \rho_E\right) U^\dagger_{A_1E_1} \cdot U^\dagger_{A_2E_2} \cdots U^\dagger_{A_nE_n}\Big), \tag{4.2.15}$$

where ρ_A represents a possibly entangled input state for the n channel uses, the unitary operations $U_{A_kE_K}$ are all identical, and the environment is a product state $\rho_E = |0\rangle_{E_1}\langle 0|_{E_1} \otimes \cdots \otimes |0\rangle_{E_n}\langle 0|_{E_n}$. Hence, for memoryless channels the statement

$$T^{(n)}[\rho_A] = T^{\otimes n}[\rho_A]$$

holds, where $T[\rho_A]$ reads as in (4.2.14).
In terms of the Kraus form (4.2.14) with probability distribution p_{i_k} per use k, the memoryless channel can be expressed as

$$T^{(n)}[\rho_A] = \sum_{i_1}\sum_{i_2}\cdots\sum_{i_n} p_{i_n}p_{i_{n-1}}\cdots p_{i_1}(V_{i_n}\otimes ... \otimes V_{i_1})\,\rho_A\,(V^\dagger_{i_1}\otimes ... \otimes V^\dagger_{i_n}). \tag{4.2.16}$$

Channels with memory

A *quantum channel with memory*, as already mentioned in the classical case, has a noise which is correlated between subsequent uses of the channel. Depending on the noise model the representations may differ. In the case where the correlated noise state $\rho_{E,\mathrm{corr}}$ can be generated from the previous product state ρ_E by a unitary operation Ω_E as

$$\rho_{E,\mathrm{corr}} = \Omega_E\,\rho_E\,\Omega^\dagger_E,$$

we can write the action of n channel uses similar to (4.2.15)

$$T^{(n)}[\rho_A] = \mathrm{Tr}_E\Big(U_{A_nE_n}\cdots U_{A_1E_1}\left(\rho_A\otimes\rho_{E,\mathrm{corr}}\right)U^\dagger_{A_1E_1}\cdots U^\dagger_{A_nE_n}\Big), \tag{4.2.17}$$

but, due to the correlations in $\rho_{E,\mathrm{corr}}$

$$T^{(n)}[\rho_A] \neq T^{\otimes n}[\rho_A].$$

In the Kraus form the joint probability distribution $p_{i_1,i_2,...,i_n}$ no longer factorizes and hence

$$T^{(n)}[\rho_A] = \sum_{i_1}\cdots\sum_{i_n} p_{i_1,i_2,...,i_n}(V_{i_n}\otimes ... \otimes V_{i_1})\,\rho_A\,(V^\dagger_{i_1}\otimes ... \otimes V^\dagger_{i_n}). \tag{4.2.18}$$

[8] As in the classical case we do not treat encoding and decoding methods in this work but just the action of the channel.

Markovian correlated quantum channel

The *Markovian correlated quantum noise channel* of length n is defined (in the Kraus Form) as

$$\begin{aligned} T^{(n)}[\rho_A] = &\sum_{i_1} \cdots \sum_{i_n} p_{i_n|i_{n-1}} p_{i_{n-1}|i_{n-2}} \cdots p_{i_2|i_1} p_{i_1} \\ &\times (V_{i_n} \otimes ... \otimes V_{i_1})\, \rho_A\, (V_{i_1}^\dagger \otimes ... \otimes V_{i_n}^\dagger), \end{aligned} \tag{4.2.19}$$

where V_{ik} are the Kraus operators for the k-th use of the channel with probability distribution p_{ik} and transition matrix $p_{i_k\,|\,i_{k-1}}$ as in (2.2.5). Regarding the way the probabilities factor we recognize the relation between the joint and the conditional probabilities as in (2.2.6). In other words, the noise has just a nearest neighbor memory.[9]

Binary symmetric channel

In order to understand the Kraus representation of the Markovian correlated channel (4.2.19), let us have short look on a simple example, i.e. a simplification the model presented in [21].
Assume the input to be a qubit in a pure state ρ. Now we define a set of Kraus operators

$$V_i = \sqrt{p_i}\sigma_i,$$

where $i = 0, x, y, z$, $\sum_i p_i = 1$ and σ_i are the Pauli operators

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{4.2.20}$$

and with $\sigma_0 = \mathbb{1}$. A possible action of the channel (i.e. the interaction with the environment) could be a rotation by the angle π around the axis $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ with probability $\{p_x, p_y, p_z\}$ and no rotation with probability p_0.
By setting $\{p_0, p_x, p_y, p_z\} = \{1-p, 0, 0, p\}$ and inserting the Pauli operators into (4.2.14) we define the "binary symmetric channel", where the input qubit is flipped with probability p and transmitted unchanged with probability $1-p$, that is

$$T[\rho] = p\,\sigma_z\,\rho\,\sigma_z^\dagger + (1-p)\,\rho.$$

Now, let us define the transition matrix

$$p_{i_k\,|\,i_{k-1}} = (1-\chi)\,p_{i_k} + \chi\,\delta_{i_k,i_{k-1}}, \quad i = 0, z; k = 1, 2, ..., n,$$

where $\chi \in [0,1]$ and $p_{0_k} = 1-p, p_{z_k} = p$ and assume the Markovian model (4.2.19). Hence, with probability χ the qubit at time step k will be effected by the same noise as the qubit at time step $k-1$ and with probability $1-\chi$ the noise at time step k is uncorrelated with the previous noise. Thus, if the qubit $k-1$ has been flipped, the qubit k will be flipped again with probability χ, if the qubit $k-1$ was transmitted unchanged, the qubit k will be transmitted unchanged with probability χ.

[9] In the literature this channel is often called "memoryless", because the noise of the future just depends on the present and has no connection to the past (i.e. a one step connection). Here, however, we use the term *channel with memory*, as in [21].

4.2.4 Classical capacity

In the following we will focus on the transmission of classical information between Alice and Bob via quantum states. Therefore we treat only the *classical capacity* of quantum channels, which will be defined in terms of the amount of classical bits which can be shared by the sender and receiver (analog to the classical case) and leave out the definition of the *quantum capacity*, which is defined in terms of shared quantum information. For further information one should take a look at [18].

The Holevo bound

If Alice and Bob intend to share classical information via quantum states, then there is a fundamental upper bound on the mutual information of Alice and Bob, called the *Holevo bound* [23]:
Alice prepares a state ρ_X, where $X = 0, 1, ...n$ with probabilities $p_0, ..., p_n$. Bob performs a positive operator-valued measurement, that means he applies an operator $\hat{E}_y$ of the set $\{\hat{E}_0, ..., \hat{E}_m\}$, with

$$\hat{E}_y = \hat{E}_y^\dagger \geq 0, \quad \sum_y \hat{E}_y = \hat{I},$$

on Alice's state, with measurement outcome Y with the conditional probability

$$P(Y|X) = \mathrm{Tr}\left(\hat{E}_Y \rho_X\right).$$

The *Holevo bound* states, that for any such measurement of Bob

$$I(X;Y) \leq S(\overline{\rho}) - \sum_x p_x\, S(\rho_x), \tag{4.2.21}$$

with the mixed state $\overline{\rho} = \sum_x p_x\, \rho_x$, the mutual information $I(X;Y)$ (see (3.1.4) and the von Neumann entropy $S(\rho)$ (see (4.2.6)). The right hand side of (4.2.21) is called the *Holevo χ-quantity*

$$\chi(p_X, \rho_X) = S(\overline{\rho}) - \sum_x p_x\, S(\rho_x). \tag{4.2.22}$$

One-shot capacity

Given a quantum channel T we introduce the *one-shot classical capacity* or *product state capacity* of T [26]

$$\begin{aligned} C_1(T) &= \max_{\rho_i, p_i}\ \chi(p_i, T[\rho_i]) \\ &= \max_{\rho_i, p_i}\ S\left(\sum_i p_i\, T[\rho_i]\right) - \sum_i p_i\, S(T[\rho_i]), \end{aligned} \tag{4.2.23}$$

where χ denotes the Holevo quantity (4.2.22) and the maximum is taken over all ensembles of $\{p_i, \rho_i\}$ of possible pure input states[10] ρ_i which fulfill the energy constraint[11]

$$\sum_i p_i \operatorname{Tr}\left(\rho_i \, \hat{a}^\dagger \hat{a}\right) \leq \overline{n}. \tag{4.2.24}$$

The term "one-shot" means that only one invocation of T is needed to calculate (4.2.23). If Alice uses n copies of T, then $C_1(T)$ is the upper bound of all rates at which Alice can transmit classical information to Bob using only product states $(\rho_1 \otimes \rho_2 ... \otimes \rho_n)$ as input. Hence, in order to call (4.2.23) the capacity of $T^{\otimes n}$, Alice has to use only one copy of T at a time, and therefore any entanglement between the input states $\rho_1, \rho_2, ...$ is not allowed.

Capacity of entangled channels

We mentioned in the previous section that, in order for expression (4.2.23) to hold as the capacity, Alice cannot entangle their subsequent uses. However, we can consider n copies of T and prepare them in one entangled state (i.e. entangle the channels with each other). This entangled system can be regarded as a new single channel $T^{(n)} \neq T^{\otimes n}$. For this channel we state the definition [26]

$$C_n(T^{(n)}) = \frac{1}{n} \max_{\rho_i, p_i} \chi(p_i, T^{(n)}[\rho_i]), \tag{4.2.25}$$

where the maximum is now taken over all (pure) input ensembles, even entangled states, fulfilling (4.2.24). Thus, (4.2.25) denotes the upper bound of the amount of bits which Alice can send per letter to Bob, if Alice is allowed to use the elementary channels in entangled blocks of length n.
For the memoryless lossy bosonic channel and a more general class of memoryless Gaussian channels (including the additive channel) it was proven that entanglement is not useful [10, 15]. That means that the capacity of such multimode channels is achieved by using a sequence of (not entangled) coherent states as input[12].
For memory channels the first evidence that entanglement enhances the transmission rate (in comparison to separable product states) was found in [21], that is the general depolarization qubit channel with memory as treated in section 4.2.3. Furthermore, entanglement was shown to be useful for the two mode Gaussian additive channel with correlated noise [5] and recently for a non-Markovian correlated noise model in [24].
For the limit $n \to \infty$ we define the *asymptotic capacity*

$$C(T^{(n)}) = \lim_{n \to \infty} C_n(T^{(n)}). \tag{4.2.26}$$

[10] In [26] it was shown, that a pure input is better than any mixed input, since any mixed state can again be expressed in terms of pure states.

[11] If $\{\hat{a}^\dagger, \hat{a}\}$ are the creation and annihilation operators of laser modes than (4.2.24) stands for the maximum mean number of photons at the input of the channel.

[12] Throughout this work we do not question the conjecture that Gaussian states achieve the capacity of Gaussian channels.

4.2.5 Bosonic Gaussian channel

In the following we will present a quantum analog to the classical Gaussian channel presented in section 3.1.3 - 3.1.3.
More specifically we will treat the *bosonic channel*, that is a continuous-variable quantum channel acting on a bosonic field, as for instance the electromagnetic field. This electromagnetic field is assumed to contain n Gaussian modes (for instance light waves fulfilling the boundary conditions of an optic resonator) which are described by a $2n$ by $2n$ covariance matrix, since each mode has a $\hat{x}$ and a $\hat{p}$ quadrature. The quantum input will still undergo a classical noise, which is physically a displacement of the modes in the phase space. Therefore the noise is now also represented by a $2n$ by $2n$ covariance matrix. We will encode than our classical information in terms of displacements of the input modes, as well. The classical information encoded per mode is given by a complex number whose real part is the x-displacement and whose imaginary part stands for the p-displacement (see figure 4.3 for a schematic represenation of such an encoding). Thus, in addition to the covariance matrix of the quantum input we will need to introduce a covariance matrix for the classical displacement of the modes. In a presence of a memory of the channel it was already shown in [5] and [24] that entanglement between the modes can increase the classical transmission rate. However, it is not yet fully understood how much energy to devote to the preparation of the quantum input (entanglement does not come for free!) and how much energy to leave for the classical modulation of the signal such that the overall transmission rate is optimal. The investigation of this problem will be the central question of the following chapters.
If needed, the reader should have a look at appendix (A), where we treated the entropy of Gaussian states, before moving on to the following sections.

General model

We introduce now the general mathematical description of the bosonic channel for an arbitrary number of modes whose act will throughout this work always be described by a displacement of the input.
The classical information we would like to transmit through the channel is given by the vector

$$\boldsymbol{\alpha} = (\Re\{\alpha_1\}, ..., \Re\{\alpha_n\}; \Im\{\alpha_1\}, ..., \Im\{\alpha_n\})^{\mathsf{T}}, \tag{4.2.27}$$

and encoded in the quantum input state denoted by $\rho^{\text{in}}_{\boldsymbol{\alpha}}$. This means, that each letter α_j is encoded in a displacement according to the density function

$$f(\alpha_1, ..., \alpha_n) = \frac{1}{\pi^n \sqrt{\det(\boldsymbol{\gamma}_{\text{mod}})}} \exp[-\frac{1}{2}\boldsymbol{\alpha}^\dagger \boldsymbol{\gamma}_{\text{mod}}^{-1} \boldsymbol{\alpha}], \tag{4.2.28}$$

where $\boldsymbol{\gamma}_{\text{mod}}$ denotes the covariance matrix of the modulation. The classical Gaussian noise of the channel is given by the density function

$$f_{\text{env}}(\beta_1, ..., \beta_n) = \frac{1}{\pi^n \sqrt{\det(\boldsymbol{\gamma}_{\text{env}})}} \exp[-\frac{1}{2}\boldsymbol{\beta}^\dagger \boldsymbol{\gamma}_{\text{env}}^{-1} \boldsymbol{\beta}]. \tag{4.2.29}$$

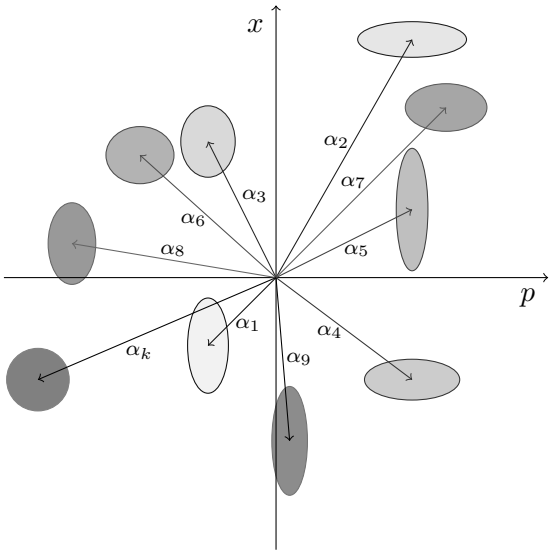

Figure 4.3: Schematic representation for the encoding of the classical information $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n)^{\mathsf{T}}$ in a sequence of displacements in the phase space.

with covariance matrix $\boldsymbol{\gamma}_{\text{env}}$ and

$$\boldsymbol{\beta} = (\Re\{\beta_1\}, ..., \Re\{\beta_n\}; \Im\{\beta_1\}, ..., \Im\{\beta_n\})^{\mathsf{T}}. \tag{4.2.30}$$

Then, the act of the bosinic channel T on the quantum input reads

$$\begin{aligned} T^{(n)}[\rho^{\text{in}}_{\boldsymbol{\alpha}}] = \int d^2\beta_1 \int d^2\beta_2 ... \int d^2\beta_n \, f_{\text{env}}(\beta_1, \beta_2, ..., \beta_n) \\ \times D(\beta_n) \otimes ... \otimes D(\beta_1) \, \rho^{\text{in}}_{\boldsymbol{\alpha}} \, D^\dagger(\beta_1) \otimes ... \otimes D^\dagger(\beta_n), \end{aligned} \tag{4.2.31}$$

where $d^2\beta_j = d\Re\{\beta_j\} \, d\Im\{\beta_j\}$, $D(\beta_j) = e^{\beta_j a_j^\dagger - \beta_j^* a_j}$ denotes the displacement operator (4.1.20). Compared to the Kraus representation (4.2.14), now an integral replaces the sum because we are working with probability density functions (see 2.1.3) instead of probability distributions. The channel (4.2.31) is *memoryless* when the noise of all uses is uncorrelated and hence, $\boldsymbol{\gamma}_{\text{env}}$ is diagonal (analog to the classical case in section 3.1.3). Then the n-variate density function factorizes to a product of n bivariate density functions (two quadratures for each channel). In general, we assume $\boldsymbol{\gamma}_{\text{env}}$ to be non-diagonal and thus $T^{(n)}$ to be a *channel with memory* (examples will be discussed in the following sections). The input $\rho^{\text{in}}_{\boldsymbol{\alpha}}$ is a multimode (possibly entangled) Gaussian state with characteristic function (4.1.26)

$$\text{Tr}\left(\rho \, \exp\left(i\hat{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{z}\right)\right) = \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\boldsymbol{\gamma}_{\text{in}}\boldsymbol{z}\right), \tag{4.2.32}$$

with input covariance matrix $\boldsymbol{\gamma}_{\text{in}}$, $\boldsymbol{R} = (\hat{x}_1, ..., \hat{x}_n; \hat{p}_1, ..., \hat{p}_n)^{\text{T}}$, where $\hat{x}_j, \hat{p}_j$ denote the quadratures of the input modes. As we have two components for the input of the channel,

the quantum input and the classical modulation of the input, we have to denote two quantum outputs, namely the quantum and overall (classically modulated) output state:

$$\begin{aligned} \rho_{\boldsymbol{\alpha}}^{\text{out}} &= T[\rho_{\boldsymbol{\alpha}}^{\text{in}}], \\ \overline{\rho} &= \int d^2\alpha\, f(\alpha)\, T[\rho_{\boldsymbol{\alpha}}^{\text{in}}], \end{aligned} \tag{4.2.33}$$

where $\overline{\rho}$ contains the actual information and $\rho_{\boldsymbol{\alpha}}^{\text{out}}$ is just the unmodulated input state which suffered the noise. It was shown in [17] that a Gaussian input which is modulated with a Gaussian density function is again a Gaussian state. Hence, we can write the channel operation (4.2.31) completly in terms of the covariance matrices, i.e.

$$\begin{aligned} \boldsymbol{\gamma}_{\text{out}} &= \boldsymbol{\gamma}_{\text{in}} + \boldsymbol{\gamma}_{\text{env}}, \\ \overline{\boldsymbol{\gamma}} &= \boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_{\text{mod}}, \end{aligned} \tag{4.2.34}$$

where $\boldsymbol{\gamma}_{\text{out}}, \overline{\boldsymbol{\gamma}}$ denote the covariance matrices of $\rho_{\boldsymbol{\alpha}}^{\text{out}}, \overline{\rho}$, respectively. In the following sections we will denote the eigenvalue spectra of $\boldsymbol{\gamma}_{\text{in}}, \boldsymbol{\gamma}_{\text{env}}, \boldsymbol{\gamma}_{\text{out}}$ and $\overline{\boldsymbol{\gamma}}$ by

$$\begin{aligned} \{\lambda_j^{\text{in}}\} = \{\lambda_{x,j}^{\text{in}}, \lambda_{p,j}^{\text{in}}\},\ \{\lambda_j^{\text{env}}\} = \{\lambda_{x,j}^{\text{env}}, \lambda_{p,j}^{\text{env}}\}, \\ \{\lambda_j^{\text{out}}\} = \{\lambda_{x,j}^{\text{out}}, \lambda_{p,j}^{\text{out}}\},\ \{\overline{\lambda}_j\} = \{\overline{\lambda}_{x,j}, \overline{\lambda}_{p,j}\}, \end{aligned} \tag{4.2.35}$$

where $j = 1, 2, ..., n$. Since the channel is fully characterized by covariance matrices we get an expression for the saturated energy constraint (4.2.24) which is in general easy to treat, that is

$$\frac{1}{2n}\left(\text{Tr}\left(\boldsymbol{\gamma}_{\text{in}}\right) + \text{Tr}\left(\boldsymbol{\gamma}_{\text{mod}}\right) - 1\right) = \overline{n}. \tag{4.2.36}$$

where n is the dimension of the system (or equivalently the number of uses), $\overline{n}$ is the maximum mean photon number of the channel and "$\frac{1}{2}$" the vacuum energy which we do not regard as invested input power. In the following it is necessary to distinguish between the fraction of power used to generate the quantum input and the fraction devoted to classical modulation. Therefore, we define the fraction of power used for entanglement (for finite n) η_n by

$$\eta_n \equiv \frac{1}{2n\overline{n}}\text{Tr}\left(\boldsymbol{\gamma}_{\text{in}}\right) - \frac{1}{2}, \tag{4.2.37}$$

where $0 \leq \eta \leq 1$, and where we subtract the vacuum energy $1/2$. Hence, for the transmission of classical information we have in average $(1-\eta_n)\overline{n}$ photons left. Thus, the saturated energy constraint on the modulation reads

$$\frac{1}{2n}\text{Tr}(\boldsymbol{\gamma}_{\text{mod}}) = (1-\eta_n)\overline{n}. \tag{4.2.38}$$

The Holevo quantity of the bosonic channel with n uses of the channel reads

$$\begin{aligned} \chi_n &= S(\overline{\rho}) - \int d^2\alpha\, f(\alpha)\, S(\rho_{\boldsymbol{\alpha}}^{\text{out}}) \\ &= \sum_{j=1}^{n} g\left(|\overline{\nu}_j| - \frac{1}{2}\right) - g\left(|\nu_j^{\text{out}}| - \frac{1}{2}\right), \end{aligned} \tag{4.2.39}$$

where $\nu_j^{\text{out}}, \overline{\nu}_j$ denote the symplectic eigenvalues of the output and mixed state (4.2.34), and where we used the entropy relation (A.0.12) for a Gaussian state ρ, namely

$$S(\rho) = \sum_j g\left(|\nu_j| - \frac{1}{2}\right). \tag{4.2.40}$$

Hence, the capacity of the bosonic channel for n-uses of the channel reads

$$C_n(T) = \frac{1}{n} \max_{\boldsymbol{\gamma}_{\text{in}}, \boldsymbol{\gamma}_{\text{mod}}} \chi_n. \tag{4.2.41}$$

Mono-modal channel with thermal noise

We will shortly review the steps to determine the optimal transmission rate of the mono-modal bosonic channel where the input undergoes a thermal noise, i.e. a noise identical in the $\hat{x}$ and $\hat{p}$ quadrature. Therefore, the noise reads as in (4.2.29) with covariance matrix

$$\boldsymbol{\gamma}_{\text{env}} = \begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}, \quad N \in \mathbb{R}. \tag{4.2.42}$$

This covariance matrix is, as already mentioned before, a classical covariance matrix and is therefore not restricted by the Heisenberg inequality. This means that the variance N can indeed be zero[13].
The mono-modal bosonic thermal channel than reads

$$T[\rho] = \int d^2\beta \, f_{\text{env}}(\beta) \, D(\beta) \, \rho_\alpha^{\text{in}} \, D^\dagger(\beta). \tag{4.2.43}$$

The optimal input for the single-mode displacement channel reads [16]

$$\rho_\alpha^{\text{in}} = |\alpha\rangle \langle\alpha| , \tag{4.2.44}$$

with coherent states $|\alpha\rangle$, which is pure and has the covariance matrix

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.2.45}$$

Since we can with no loss of generality assume the means to be zero the input is completely characterized by (4.2.45). One varifies, using (A.0.12), that $S(\rho_\alpha^{\text{in}}) = 0$. The energy of the input is given by the normalized trace of its covariance matrix, i.e.

$$\frac{1}{2}\text{Tr}\left(\boldsymbol{\gamma}_{\text{in}}\right) = \frac{1}{2}, \tag{4.2.46}$$

which is simply the energy of the vacuum and is by definition not taken into account as invested power. Thus, we can use all photons for modulation, i.e.

$$\boldsymbol{\gamma}_{\text{mod}} = \begin{pmatrix} \overline{n} & 0 \\ 0 & \overline{n} \end{pmatrix}. \tag{4.2.47}$$

[13] However, this would correspond to the unphysical case of infinite capacity.

The covariance matrices of the output and mixed state then read

$$\boldsymbol{\gamma}_{\text{out}} = \frac{1}{2}\begin{pmatrix} 1+2N & 0 \\ 0 & 1+2N \end{pmatrix}, \tag{4.2.48}$$

$$\overline{\boldsymbol{\gamma}} = \frac{1}{2}\begin{pmatrix} 1+2(N+\overline{n}) & 0 \\ 0 & 1+2(N+\overline{n}) \end{pmatrix}. \tag{4.2.49}$$

Both states, the quantum output with covariance matrix $\boldsymbol{\gamma}_{\text{out}}$ and the modulated output $\overline{\boldsymbol{\gamma}}$ are thermal states as defined in (4.1.30). Using (4.2.41) the one-shot capacity becomes

$$C_1 = g(\overline{n}+N) - g(N). \tag{4.2.50}$$

Classical limit: Let us have a look at the classical limit of (4.2.50), that is when $(N, \overline{n}) \to \infty$ with fixed power to noise ratio $\overline{n}/N = \text{const}$. Firstly we notice that the $g(x)$-function (A.0.4), i.e. the Gaussian entropy in the quantum case tends to the classical entropy expression if its argument tends to infinity:

$$\lim_{x\to\infty}\left[\log x - g(x)\right] = \lim_{x\to\infty}\left[\log x - x\log(x+1) - x\log x + \log(x+1)\right] = 0. \tag{4.2.51}$$

Using this limit behavior for (4.2.50), we confirm that

$$\lim_{\overline{n},N\to\infty,\,\overline{n}/N=c} C_1 = \log(\overline{n}+N) + \log(N) = \log\left(1+\frac{\overline{n}}{N}\right), \tag{4.2.52}$$

where c is a constant. We conclude that in the limit of a high input signal and a high noise (with constant signal to noise ratio) the quantum effects (i.e. vacuum fluctuations) vanish and we retrieve just twice the capacity of the single variate classical Gaussian channel (3.1.16), since we have in this limit just two classical Channels, one for the $\hat{x}$ and one for the $\hat{p}$ quadrature.

Mono-modal channel with phase dependent noise

Let us assume again just a mono-modal bosonic channel, but this time with a different noise in $\hat{x}$ and $\hat{p}$. The density function of the environment reads again as in (4.2.29) but has the covariance matrix

$$\boldsymbol{\gamma}_{\text{env}} = \begin{pmatrix} N_x & 0 \\ 0 & N_p \end{pmatrix}, \quad N_x, N_p \in \mathbb{R}, \tag{4.2.53}$$

where in general $N_x \neq N_p$. The capacity of a channel with such a noise which still appears to be simple, is not yet completely determined. In [33] the lossy mono-modal channel with such a noise was studied. It was assumed that the best input diagonalizes in the same basis as the noise by the following reasoning: the anisotropic channel (with asymmetry in the noise in $\hat{x}$ and $\hat{p}$) can be transformed into a thermal channel by active symplectic transformations[14]:

$$\boldsymbol{\gamma}_{\text{env}} = \begin{pmatrix} N_x & 0 \\ 0 & N_p \end{pmatrix} = S\boldsymbol{\gamma}'_{\text{env}}S^{\mathsf{T}} = S\begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}S^{\mathsf{T}}, \tag{4.2.54}$$

[14] We remark that active symplectic transformations change the trace and thus the energy constraint of the channel.

with $N = (N_x + N_p)/2$. For the thermal channel we found in the previous section the best input, namely coherent states with covariance matrix (4.2.45). Therefore the optimal input for the anisotropic channel will be a coherent state modified by symplectic transformations which balanced the $\hat{x}$ and $\hat{p}$ variances in the noise. This is a squeezed state and lives in the same basis as the noise, where the degree of squeezing will be determined in the following section. By this reasoning, we assume the input and modulation covariance matrices:

$$\boldsymbol{\gamma}_{\text{in}} = \begin{pmatrix} \gamma_x^{\text{in}} & 0 \\ 0 & \gamma_p^{\text{in}} \end{pmatrix}, \quad \boldsymbol{\gamma}_{\text{mod}} = \begin{pmatrix} \gamma_x^{\text{mod}} & \gamma_{xp}^{\text{mod}} \\ \gamma_{xp}^{\text{mod}} & \gamma_p^{\text{mod}} \end{pmatrix}, \tag{4.2.55}$$

with the energy constraint as in (4.2.36) and the requirement to have a pure input (as already mentioned in (4.2.4), that is

$$\det \boldsymbol{\gamma}_{\text{in}} = \frac{1}{4}. \tag{4.2.56}$$

If we recall in addition the total power constraint (4.2.36), then we can write out the total Lagrangian for the optimization problem, i.e.

$$\begin{aligned} \mathcal{L} = & g\left(\sqrt{(\gamma_x^{\text{in}} + N_x + \gamma_x^{\text{mod}})\left(\gamma_p^{\text{in}} + N_p + \gamma_p^{\text{mod}}\right) - (\gamma_{xp}^{\text{mod}})^2} - \frac{1}{2}\right) \\ & - g\left(\sqrt{(\gamma_x^{\text{in}} + N_x)\left(\gamma_p^{\text{in}} + N_p\right)} - \frac{1}{2}\right) \\ & - \mu\left(\gamma_x^{\text{in}} + \gamma_x^{\text{mod}} + \gamma_p^{\text{in}} + \gamma_p^{\text{mod}}\right) - \tau\left(\gamma_x^{\text{in}}\gamma_p^{\text{in}}\right), \end{aligned} \tag{4.2.57}$$

where we added the multiplier μ to respect the total energy constraint and τ to guarantee that the input is pure. Since we want to maximize (4.2.57) and since $g(x)$ is monotonically increasing we immediately see that we should set all off-diagonal elements in the modulation covariance matrix to zero, i.e. $\gamma_{xp}^{\text{mod}} = 0$. From this we already get that the optimal total modulation covariance matrix $\boldsymbol{\gamma}_{\text{mod}}$ should live as well in the basis as $\boldsymbol{\gamma}_{\text{env}}$. By applying

$$\nabla \mathcal{L} = 0, \tag{4.2.58}$$

with

$$\nabla^{\mathsf{T}} = \left(\frac{\partial}{\partial \gamma_x^{\text{mod}}}, \frac{\partial}{\partial \gamma_p^{\text{mod}}}, \frac{\partial}{\partial \gamma_x^{\text{in}}}, \frac{\partial}{\partial \gamma_p^{\text{in}}}\right), \tag{4.2.59}$$

we receive 4 equations:

$$\frac{\partial}{\partial \gamma_x^{\text{mod}}}\mathcal{L} = \left[\log\left(\frac{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} + \frac{1}{2}}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} - \frac{1}{2}}\right) + \frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} + \frac{1}{2}} - \frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} - \frac{1}{2}}\right] \frac{1}{2}\sqrt{\frac{\overline{\gamma}_p}{\overline{\gamma}_x}} - \mu = 0, \tag{4.2.60}$$

$$\frac{\partial}{\partial \gamma_p^{\text{mod}}}\mathcal{L} = \left[\log\left(\frac{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} + \frac{1}{2}}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} - \frac{1}{2}}\right) + \frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} + \frac{1}{2}} - \frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p} - \frac{1}{2}}\right] \frac{1}{2}\sqrt{\frac{\overline{\gamma}_x}{\overline{\gamma}_p}} - \mu = 0, \tag{4.2.61}$$

$$\begin{aligned}\frac{\partial}{\partial\gamma_x^{\text{in}}}\mathcal{L} = &\left[\log\left(\frac{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}+\frac{1}{2}}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}-\frac{1}{2}}\right)+\frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}+\frac{1}{2}}-\frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}-\frac{1}{2}}\right]\frac{1}{2}\sqrt{\frac{\overline{\gamma}_p}{\overline{\gamma}_x}}\\ &-\left[\log\left(\frac{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}+\frac{1}{2}}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}-\frac{1}{2}}\right)+\frac{1}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}+\frac{1}{2}}-\frac{1}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}-\frac{1}{2}}\right]\\ &\times\frac{1}{2}\sqrt{\frac{\gamma_p^{\text{out}}}{\gamma_x^{\text{out}}}}-\tau\,\gamma_p^{\text{in}}=0,\end{aligned}\tag{4.2.62}$$

$$\begin{aligned}\frac{\partial}{\partial\gamma_p^{\text{in}}}\mathcal{L} = &\left[\log\left(\frac{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}+\frac{1}{2}}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}-\frac{1}{2}}\right)+\frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}+\frac{1}{2}}-\frac{1}{\sqrt{\overline{\gamma}_x\,\overline{\gamma}_p}-\frac{1}{2}}\right]\frac{1}{2}\sqrt{\frac{\overline{\gamma}_x}{\overline{\gamma}_p}}\\ &-\left[\log\left(\frac{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}+\frac{1}{2}}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}-\frac{1}{2}}\right)+\frac{1}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}+\frac{1}{2}}-\frac{1}{\sqrt{\gamma_x^{\text{out}}\,\gamma_p^{\text{out}}}-\frac{1}{2}}\right]\\ &\times\frac{1}{2}\sqrt{\frac{\gamma_x^{\text{out}}}{\gamma_p^{\text{out}}}}-\tau\,\gamma_x^{\text{in}}=0,\end{aligned}\tag{4.2.63}$$

with

$$\begin{aligned}\overline{\gamma}_{x,p} &= \gamma_{x,p}^{\text{in}}+N_{x,p}+\gamma_{x,p}^{\text{mod}}\\ \gamma_{x,p}^{\text{out}} &= \gamma_{x,p}^{\text{in}}+N_{x,p}.\end{aligned}\tag{4.2.64}$$

The solution of this system of equations reads as follows. At first, we find that

$$\gamma_x^{\text{in}}+\gamma_x^{\text{env}}+\gamma_x^{\text{mod}}=\gamma_p^{\text{in}}+\gamma_p^{\text{env}}+\gamma_p^{\text{mod}}.\tag{4.2.65}$$

From the latter we conclude that the optimal total output state $\overline{\boldsymbol{\gamma}}$ is a thermal state. The covariance matrix for this state reads

$$\overline{\gamma}_x=\overline{\gamma}_p=\overline{n}+\frac{N_x+N_p}{2}+\frac{1}{2},\tag{4.2.66}$$

which is just the total power of the system analog to (4.2.49). Secondly, we obtain the optimal degree of squeezing for the input state:

$$\frac{\gamma_x^{\text{in}}}{\gamma_p^{\text{in}}}=\frac{\gamma_x^{\text{env}}}{\gamma_p^{\text{env}}}.\tag{4.2.67}$$

Thus, the anisotropy of the noise has to be precisely matched by the ratio of squeezing at the input. We assume for now that we have ineed input power to satisfy (4.2.66) and (4.2.67).

Since all matrices are diagonal we state that the variables $\gamma_x^{\text{in}},\gamma_p^{\text{in}},\gamma_x^{\text{mod}},\gamma_p^{\text{mod}}$ are the eigenvalues $\lambda_x^{\text{in}},\lambda_p^{\text{in}},\lambda_x^{\text{mod}},\lambda_p^{\text{mod}}$ of the input and modulation. From the condition that the input has to be pure (4.2.56) and with (4.2.67) we conclude that

$$\lambda_{x,p}^{\text{in}}=\frac{1}{2}\sqrt{\frac{N_{x,p}}{N_{p,x}}}.\tag{4.2.68}$$

Concerning the modulation eigenvalues, we see from (4.2.65) that they follow a *quantum water filling* with respect to both quadratures, whereas in contrary to the classical solution now the quantum input and the environment, i.e. the quantum output together play the role of an "extended noise". By rewriting (4.2.65) and (4.2.66) we get the optimal modulation for x and p, that is

$$\begin{aligned} \lambda_x^{\text{mod}} &= \mu - (\lambda_x^{\text{in}} + N_x), \\ \lambda_p^{\text{mod}} &= \mu - (\lambda_p^{\text{in}} + N_p), \end{aligned} \tag{4.2.69}$$

where the quantum waterfilling level μ is simply the total power of the system, i.e.

$$\mu = \overline{n} + \frac{N_x + N_p}{2} + \frac{1}{2}. \tag{4.2.70}$$

The solutions for the best input (4.2.68) and modulation (4.2.69), respectively, only hold if we have sufficient input power to squeeze the input in the noise like asymmetry and at the same to time to have enough power left to adjust the modulation, such that the overall output state is thermal, i.e. (4.2.65) is fulfilled. If this is indeed the case, we know the solution for the capacity, that is

$$C_1 = g\left(\overline{n} + \frac{N_x + N_p}{2}\right) - g(\sqrt{N_x\, N_p}), \quad \overline{n} \geq \overline{n}_{\text{thr}}, \tag{4.2.71}$$

Interestingly, the capacity of the phase dependent channel (4.2.71) is greater then the capacity of the thermal channel for same total noise $(N_x + N_p)/2 = N$, and grows with increasing asymmetry. This will be investigated further shortly. The threshold value $\overline{n}_{\text{thr}}$ can explicitly be determined as follows. If we increase $\overline{n}$ (starting at $\overline{n} = 0$) then $\overline{n} = \overline{n}_{\text{thr}}$ is exactly where the overall output starts to be a thermal state. With no loss of generality we assume now that

$$N_x > N_p. \tag{4.2.72}$$

We know that the best $\hat{p}$ quadrature will be squeezed and hence the best $\hat{x}$ quadrature anti squeezed (in order to match the noise ratio). Thus, below the threshold the asymmetry between the quantum output in x and p reads

$$\lambda_x^{\text{in}} + N_x > \lambda_p^{\text{in}} + N_p. \tag{4.2.73}$$

Now the modulation will try to compensate that asymmetry (in order to achieve a thermal state). Thus, up to the threshold value we will only induce a modulation for p and

$$\lambda_x^{\text{mod}} = 0, \quad \forall \overline{n} \leq \overline{n}_{\text{thr}}. \tag{4.2.74}$$

With this additional information and with (4.2.66) we determine

$$\overline{n}_{\text{thr}} = \frac{1}{2}\left(\sqrt{\frac{N_x}{N_p}} + N_x - N_p - 1\right). \tag{4.2.75}$$

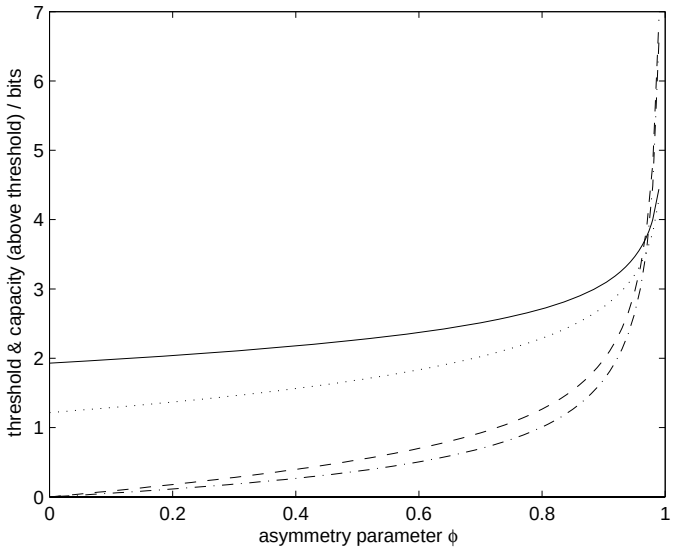

Figure 4.4: Capacity above threshold and threshold $\overline{n}_{\rm thr}$ vs. asymmetry parameter ϕ. The dashed and dashed-dotted curve show $\overline{n}_{\rm thr}$ for $N = 1/3, 1/100$. The dotted and solid curve the capacity (4.2.71) where $\overline{n} = \overline{n}_{\rm thr}(N = 1/3) + 1, \overline{n}_{\rm thr}(N = 1/100) + 1$.

Having the threshold value characterized, we can define for a simple comparison

$$\begin{aligned} N_x &= (1+\phi)N, \\ N_p &= (1-\phi)N, \end{aligned} \tag{4.2.76}$$

with the asymmetry parameter $\phi \in [0, 1]$ and $(N_x + N_p)/2 = N$. In figure 4.4 we plot (4.2.71) and (4.2.75) for different total noises N versus the asymmetry parameter ϕ, where the input power is always set equal to the threshold value $\overline{n} = \overline{n}_{\rm thr}$. The capacity with thermal noise is given at $\phi = 0$ and is always lower than the capacity with phase dependent noise where for this plot $\overline{n} > \overline{n}_{\rm thr}$. The threshold diverges for $\phi \to 1$ as $N_p \to 0$. In this limit we would need infinite energy to squeeze the input such that the uncertainty in one quadrature vanishes completely. Obviously, for a physical channel this limit cannot be achieved. The capacity (above threshold) increases with increasing ϕ as well, but is finite at $\phi = 1$, since the $g(x)$-function is fixed to zero for $x = 0$.

Capacity below the threshold As we treated in the previous part only the capacity above the threshold we shall now move on to an investigation of the more tricky domain where $\overline{n} < \overline{n}_{\rm thr}$. In fact, we will not be able to provide the reader with an analytical solution. However, we can observe and characterize the behavior of the optimal input squeezing and modulation by numerical optimization.
For this optimization we introduce a one mode squeezed state, characterized by the covari-

ance matrix

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}\begin{pmatrix} e^{2r} & 0 \\ 0 & e^{-2r} \end{pmatrix}, \tag{4.2.77}$$

with squeezing parameter $r \in [-r_{\max}, r_{\max}]$ whic is constraint by (4.2.36), i.e.

$$\frac{1}{2}\text{Tr}\left(\boldsymbol{\gamma}_{\text{in}}\right) - \frac{1}{2} = \sinh^2 r \leq \overline{n}. \tag{4.2.78}$$

In the last equation we did not include the power used for modulation, that means that if $r = r_{\max}$ then we used all power for preparing the quantum state and have no power left to transmit the actual information and thus the transmission rate would be zero. In order to have a simple characterization of the energy distribution between input and modulation we define the fraction of the power used for squeezing as in (4.2.37) by

$$\eta \equiv \frac{1}{\overline{n}}\sinh^2 r, \tag{4.2.79}$$

with $\eta \in [0,1]$. Thus, we know we have in average $(1-\eta)\overline{n}$ photons left to introduce a classical modulation. If we still assume that the noise in the $\hat{x}$-quadrature is stronger than noise in the $\hat{p}$-quadrature, then the modulation is actually easy to characterize. As we already mentioned below the threshold $\lambda_x^{\text{mod}} = 0$ and hence

$$\boldsymbol{\gamma}_{\text{mod}} = \begin{pmatrix} 0 & 0 \\ 0 & 2(1-\eta)\overline{n} \end{pmatrix} \quad \overline{n} < \overline{n}_{\text{thr}}, \tag{4.2.80}$$

where the factor two stems from (4.2.36). The capacity now reads

$$C = \max_{\eta\,\in\,[0,1]} g\left(|\overline{\nu}| - \frac{1}{2}\right) - g\left(|\nu^{\text{out}}| - \frac{1}{2}\right), \tag{4.2.81}$$

with symplectic eigenvalues

$$\begin{aligned} \overline{\nu} &= \sqrt{\overline{\lambda}_x\,\overline{\lambda}_p}, \\ \nu^{\text{out}} &= \sqrt{\lambda_x^{\text{out}}\,\lambda_x^{\text{out}}}, \end{aligned} \tag{4.2.82}$$

where

$$\begin{aligned} \overline{\lambda}_x &= \lambda_x^{\text{in}} + N(1+\phi), \\ \overline{\lambda}_p &= \lambda_p^{\text{in}} + N(1-\phi) + \lambda_p^{\text{mod}}, \\ \lambda_{x,p}^{\text{out}} &= \lambda_{x,p}^{\text{in}} + N(1 \pm \phi) \end{aligned} \tag{4.2.83}$$

where $\lambda_x^{\text{in}}, \lambda_p^{\text{in}}$ denote the eigenvalues of the input (4.2.77), λ_p^{mod} the non zero (for $\eta < 1$) eigenvalue of (4.2.80) and N_x, N_p are as in (4.2.76). In figure 4.5 we plotted the eigenvalues resulting from a numerical optimization for (4.2.81). We see that above the threshold the overall output state, characterized by $\overline{\boldsymbol{\gamma}}$ is thermal, i.e. $\overline{\lambda}_x = \overline{\lambda}_p$ and the optimal transmission rate reads as in (4.2.71). In order to get a clearer view for the behavior of the

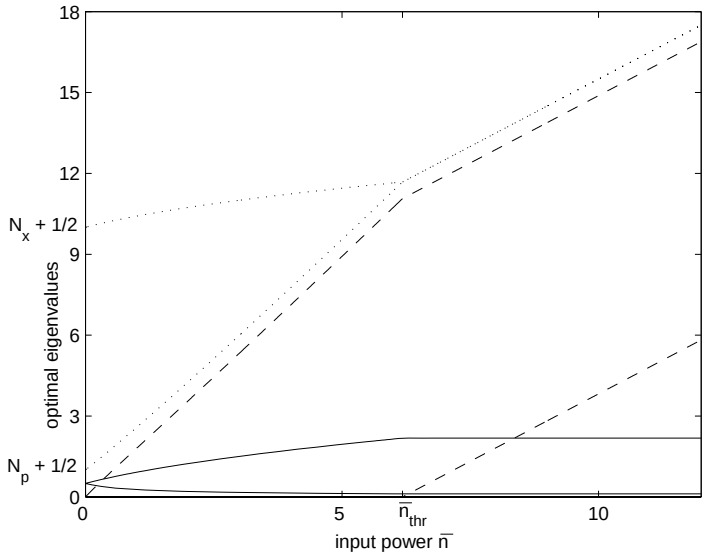

Figure 4.5: Optimal eigenvalues vs. input power $\overline{n}$. The upper and lower dotted line show $\overline{\lambda}_x, \overline{\lambda}_p$, the upper and lower dashed line show $\lambda_p^{\mathrm{mod}}, \lambda_x^{\mathrm{mod}}$, the upper and lower solid line show $\lambda_x^{\mathrm{in}}, \lambda_p^{\mathrm{in}}$. For $\overline{n} \geq \overline{n}_{\mathrm{thr}}$ the overall output state is a thermal state and the input sqeezing becomes constant. For $\overline{n} < \overline{n}_{\mathrm{thr}}$ the modulation in x is zero. For this plot we took $N = 5, \phi = 0.9$.

optimal squeezing in the region below the threshold we introduce a linear approximation for the overall output eigenvalues $\overline{\lambda}_x, \overline{\lambda}_p$ and investigate their differences to the numerical optimum. Knowing the definition of the threshold (4.2.75) we make the simple ansatz

$$\widetilde{\overline{\lambda}}_{x,p} = \underbrace{\overline{n} + \frac{1}{2} + N}_{\text{eigenvalue of thermal state}} \pm \underbrace{N\phi\left(\frac{\overline{n}_{\mathrm{thr}} - \overline{n}}{\overline{n}_{\mathrm{thr}}}\right)}_{\text{linear approximation}}. \tag{4.2.84}$$

With the latter and the purity condition (4.2.56) we determine the approximated input and modulation eigenvalues

$$\begin{aligned} \widetilde{\lambda_x^{\mathrm{in}}} &= \overline{n} + \frac{1}{2} + N\phi\left(1 + \frac{\overline{n}_{\mathrm{thr}} - \overline{n}}{\overline{n}_{\mathrm{thr}}}\right), \\ \widetilde{\lambda_p^{\mathrm{in}}} &= \frac{1}{4\widetilde{\lambda_x^{\mathrm{in}}}}, \\ \widetilde{\lambda_p^{\mathrm{mod}}} &= \widetilde{\overline{\lambda}}_p - \lambda_p^{\mathrm{in}} - N(1-\phi). \end{aligned} \tag{4.2.85}$$

The difference of the x-input eigenvalues $\lambda_x^{\mathrm{in}} - \widetilde{\lambda_x^{\mathrm{in}}}$ is plotted in figure 4.6. We observe a maximum (for each N) which increases and is shifted towards higher powers with increasing noise N. From this we can conclude that the optimal squeezing increases first (up to some

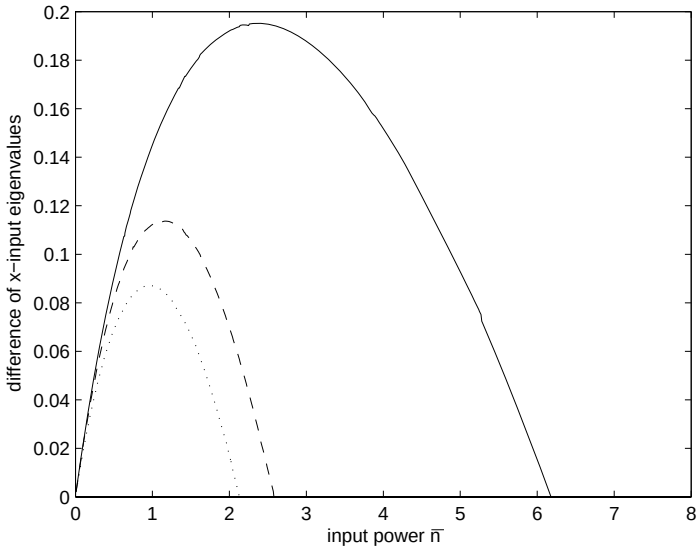

Figure 4.6: Difference of the x-input eigenvalues vs. input power $\overline{n}$. The dotted, dashed and solid line refer to $N = 0.5, 1, 5$ which attain maxima at $\overline{n} = 0.96, 1.1, 2.37$. All curves drop to zero at the threshold values. In addition we took always $\phi = 0.9$.

$\overline{n}$) stronger and then weaker than the linear approximation. In addition, we remark that the linear approximation seems to get worse for growing noise. The difference of the p-modulation $\widetilde{\lambda_p^{\mathrm{mod}}} - \lambda_p^{\mathrm{mod}}$ is plotted in figure 4.7 (as the input λ_x^{in} is higher the modulation λ_p^{mod} is lower). Since the modulation just uses the remaining energy from squeezing the plots look similar to the input difference. At last we compare the optimal transmission rates. In figure 4.8 we observe for each curve a maximum and a smooth drop zero (at each threshold $\overline{n}_{\mathrm{thr}}$). Although we may not matched the true capacity the approximation is close to the numerical optimum. With growing noise the maximal difference grows but the signal to noise ratio at the threshold (above which we know the true capacity) converges down to a constant. This simply because $\overline{n}_{\mathrm{thr}}$ is linear in N, i.e.

$$\overline{n}_{\mathrm{thr}} = \frac{1}{2}\left(\sqrt{\frac{1+\phi}{1-\phi}} + 2\phi N - 1\right). \tag{4.2.86}$$

Taking the limit

$$\lim_{N\to\infty} \frac{\overline{n}_{\mathrm{thr}}}{N} = \phi, \tag{4.2.87}$$

confirms that for each $\phi < 1$ the signal to noise ratio needed to achieve the capacity (4.2.71) is in the limit always lower than one. With decreasing noise the maximal difference gets smaller and therefore the linear approximation gets closer to the real capacity.

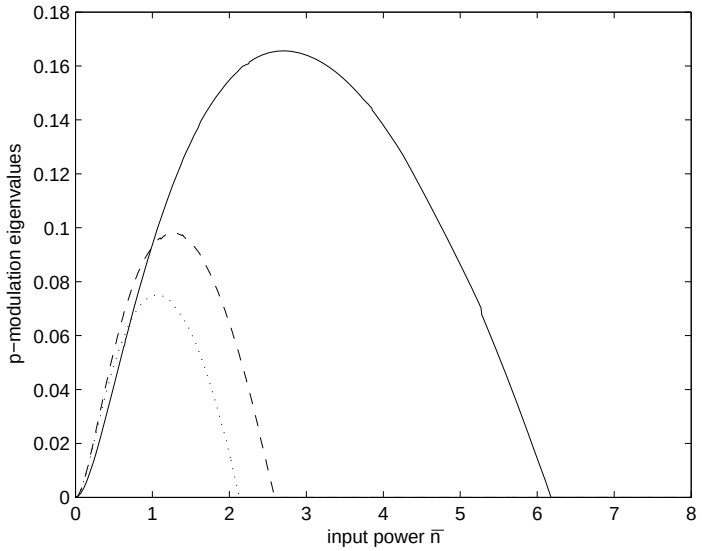

Figure 4.7: Difference of the p-modulation eigenvalues vs. input power $\overline{n}$. The dotted, dashed and solid line refer to $N = 0.5, 1, 5$ which attain maxima at $\overline{n} = 1.06, 1.27, 2.7$. All curves drop to zero at the threshold values. In addition we took always $\phi = 0.9$.

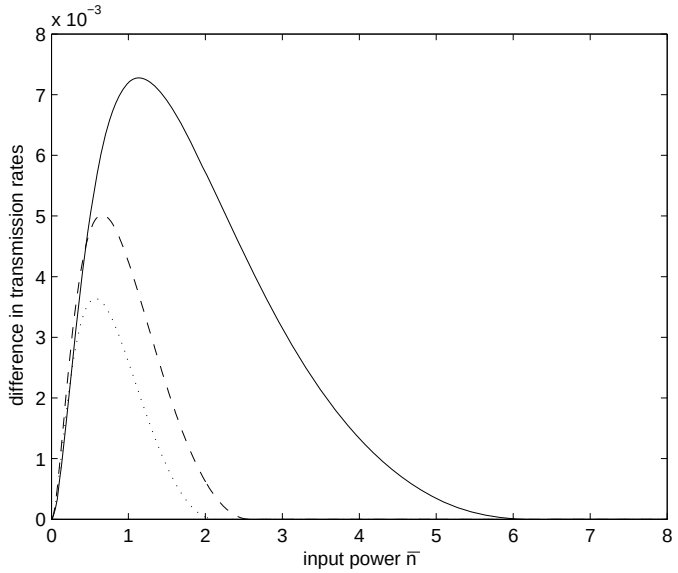

Figure 4.8: Difference in transmission rates $C - C_{\text{appr}}$ vs. input power $\overline{n}$. The dotted, dashed and solid line refer to $N = 0.5, 1, 5$ which attain maxima at $\overline{n} = 0.57, 0.66, 1.14$. All curves drop to zero at the threshold values. In addition we took always $\phi = 0.9$.

Bimodal channel with memory

We shortly review now an example of a bimodal memory channel, that is a channel as defined in (4.2.31) where the noise density function (4.2.29) has a 4 by 4 covariance matrix with off diagonal elements as already treated in [5], i.e.

$$\boldsymbol{\gamma}_{\text{env}} = \begin{pmatrix} N & -\phi N & 0 & 0 \\ -\phi N & N & 0 & 0 \\ 0 & 0 & N & \phi N \\ 0 & 0 & \phi N & N \end{pmatrix}, \tag{4.2.88}$$

with correlation coefficient $\phi \in [0,1]$. One observers that the $\hat{x}$-quadrature variances are anti-correlated, while the $\hat{p}$-quadrature variances are correlated. If $\phi = 0$ than (4.2.88) reduces to the memoryless case; if $\theta = 1$, than the channel has a full memory.
By applying a passive (trace preserving) symplectic transformation on the environment we can treat the problem of finding the capacity - with no loss of generality - in the basis of the noise. The diagonalized noise covariance matrix reads

$$\widetilde{\boldsymbol{\gamma}_{\text{env}}} = \begin{pmatrix} N(1+\phi) & 0 & 0 & 0 \\ 0 & N(1-\phi) & 0 & 0 \\ 0 & 0 & N(1-\phi) & 0 \\ 0 & 0 & 0 & N(1+\phi) \end{pmatrix}. \tag{4.2.89}$$

In the latter we see that the first mode suffers a phase dependent noise, high in x and weak in p, whereas the second mode suffers the same but swapped noise which is high in p and weak in x. The eigenvalue spectrum is therefore precisely the same as the one of two uncorrelated mono-modal phase dependent channels with indices interchanged in x and p. Thus, we can immediately conclude that above the threshold (4.2.75) we know that all solutions of the bimodal channel coincide with the phase dependent mono-modal channel: the best input and modulation read as in (4.2.68), (4.2.69) and the capacity reads as in (4.2.71). Similarly, below the threshold, the two-mode entangled state (as introduced in (4.2.3)) was used as the quantum input:

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2} \begin{pmatrix} \cosh 2r & -\sinh 2r & 0 & 0 \\ -\sinh 2r & \cosh 2r & 0 & 0 \\ 0 & 0 & \cosh 2r & \sinh 2r \\ 0 & 0 & \sinh 2r & \cosh 2r \end{pmatrix}, \tag{4.2.90}$$

with squeezing parameter r. We know that his input is useful, since it diagonalizes in the same basis as (4.2.90)[15]. The diagonal input reads

$$\widetilde{\gamma_{\text{in}}} = \frac{1}{2}\begin{pmatrix} e^{2r} & 0 & 0 & 0 \\ 0 & e^{-2r} & 0 & 0 \\ 0 & 0 & e^{-2r} & 0 \\ 0 & 0 & 0 & e^{2r} \end{pmatrix}, \tag{4.2.91}$$

Physically this diagonalization corresponds to a reversion of the beam splitter operation which entangled the two squeezed modes (see section 4.2.1) and resulted in the state with covariance matrix (4.2.90). We confirm that the eigenvalue spectrum of (4.2.90) is simply two times the spectrum of a squeezed input with interchanged x and p. Hence, our linear approximation for the mono-modal phase dependent channel and all results hold for the bimodal channel as well.

Multimode memoryless channel

The case of a channel with n uncorrelated modes was already treated in [17][§3]. In this model, a classical information was transmitted via a displacement of a n-mode cavity state[16] The cavity was regarded as a quantum noise in contrary to the classical noise we defined for the channel (4.2.31). However, If we treat again the sum of the quantum input and the classical noise, i.e. the quantum output as an extended quantum noise we can adopt the results. In our terms the covariance matrix of the classical noise of *memoryless* channel reads

$$\boldsymbol{\gamma}_{\text{env}} = \text{diag}(N_1, N_2, ..., N_n; N_1, ..., N_n), \tag{4.2.92}$$

i.e. each channel j has a thermal noise

$$\boldsymbol{\gamma}_{\text{env},j} = \begin{pmatrix} N_j & 0 \\ 0 & N_j \end{pmatrix}. \tag{4.2.93}$$

For each individual channel we know therefore, that the optimal input is a coherent state and hence the total input is an ensemble of product states, that is

$$\rho^{\text{in}}_{\boldsymbol{\alpha}} = \bigotimes_j \rho^{\text{in}}_{\alpha_j},$$

where $\rho^{\text{in}}_{\alpha_j}$ is a coherent state as in (4.2.44), and $\boldsymbol{\alpha}$ denotes the classical information we would like to transmit. The quantum output is therefore just the sum of the covariance matrix of the vacuum and the given classical noise, i.e.

$$\boldsymbol{\gamma}_{\text{out}} = \mathbb{1}/2 + \boldsymbol{\gamma}_{\text{env}}. \tag{4.2.94}$$

[15] It was concluded in section 4.2.5 that the optimal input has to live in the same basis as the noise.
[16] In our notations we do not treat different frequencies ω_j for the modes as in [17][§3.].

with $\nu_j^{\text{out}} = 1/2 + N_j$. Thus, the full power can be used for modulation. The optimal modulation density function of the signal was determined to have the covariance matrix

$$\boldsymbol{\gamma}_{\text{mod}} = \text{diag}(m_1, ..., m_n; m_1, ..., m_n), \tag{4.2.95}$$

where

$$m_j = (\mu - \nu_j^{\text{out}})^+ = (\mu - N_j - \frac{1}{2})^+, \tag{4.2.96}$$

such that

$$\frac{1}{n}\sum_{j=1}^{n} m_j = \overline{n}, \tag{4.2.97}$$

with the maximum mean photon number constraint $\overline{n}$. This solution is the quantum analog to the classical "water-filling" described in section 3.1.3. The one shot capacity then reads

$$C_1 = \frac{1}{n}\sum_{j=1}^{n} (g(\mu) - g(N_j))^+, \tag{4.2.98}$$

with $(x)^+$ as in (3.1.23).

4.3 Bosonic memory channel with Gauss-Markov noise

We are now equipped with all necessary tools to treat the central part of this work: the bosonic memory channel with a Gauss-Markov noise. This channel is defined as in (4.2.31), where the covariance matrix for the noise reads

$$\boldsymbol{\gamma}_{\text{env}} = \begin{pmatrix} \boldsymbol{\Gamma}(\phi) & 0 \\ 0 & \boldsymbol{\Gamma}(-\phi) \end{pmatrix}, \tag{4.3.1}$$

where $\boldsymbol{\Gamma}$ is the covariance matrix resulting from the Gauss-Markov process specified in (2.2.10) with correlation parameter $0 \leq \phi < 1$. In the case where $\phi = 0$ the noise covariance matrix will be diagonal and the channel corresponds to a multimode memoryless channel. If we recall the definition of the Markovian correlated channel (4.2.19) we see that the displacement channel (4.2.31) with noise (4.3.1) indeed belongs to this class of channels (where integrate over density functions instead of summing over probability distributions), because the noise density function factorizes as

$$\begin{aligned} f_{\text{env}}(\beta_1, \beta_2, ..., \beta_n) &= \frac{1}{\pi^n \sqrt{\det(\boldsymbol{\gamma}_{\text{env}})}} \exp\left[-\frac{1}{2}\boldsymbol{\beta}^\dagger \boldsymbol{\gamma}_{\text{env}}^{-1} \boldsymbol{\beta}\right] \\ &= f_{\text{env}}(\beta_1)\, f_{\text{env}}(\beta_1|\beta_2) \cdots f_{\text{env}}(\beta_{n-1}|\beta_n), \end{aligned} \tag{4.3.2}$$

where $f_{\text{env}}(\beta_i|\beta_j)$ denote conditional quadvariate (two quadratures) Gaussian density functions. Thus the action of the bosonic channel with Gauss-Markov memory on the input state $\rho_{\boldsymbol{\alpha}}^{\text{in}}$ is given by

$$\begin{aligned} T^{(n)}[\rho_{\boldsymbol{\alpha}}^{\text{in}}] = \int d^2\beta_1 \int d^2\beta_2 ... \int d^2\beta_n\, f_{\text{env}}(\beta_1)\, f_{\text{env}}(\beta_1|\beta_2) \cdots f_{\text{env}}(\beta_{n-1}|\beta_n) \\ \times D(\beta_n) \otimes ... \otimes D(\beta_1)\, \rho_{\boldsymbol{\alpha}}^{\text{in}}\, D^\dagger(\beta_1) \otimes ... \otimes D^\dagger(\beta_n). \end{aligned} \tag{4.3.3}$$

We will rotate now the problem into the basis where the covariance matrix of the environment is diagonal. This can by done with no loss generality, since a rotation is a passive symplectic transformation and does not change the entropy of the system. Then, by same reasoning as in the mono-modal anisotropic case (see section 4.2.5) we impose the assumption that the optimal input covariance matrix has to diagonalize in the same basis as the noise. In the definition of the capacity (4.2.41), i.e.

$$C_n(T) = \frac{1}{n} \max_{\boldsymbol{\gamma}_{\text{in}}, \boldsymbol{\gamma}_{\text{mod}}} S(\overline{\rho}) - \int d^2\alpha\, f(\alpha)\, S(\rho_{\boldsymbol{\alpha}}^{\text{out}}) \tag{4.3.4}$$

the covariance matrix of the modulation only enters in the first term, (the integral in the second term is only taken over the normalized density function) and thus $S(\overline{\rho})$ can be maximized with respect to $\boldsymbol{\gamma}_{\text{mod}}$ independently. The subadditivity property of the entropy (4.2.9) for $\overline{\rho}$ reads:

$$S(\overline{\rho}) \leq S(\overline{\rho}_1) + S(\overline{\rho}_2) + ... + S(\overline{\rho}_n), \tag{4.3.5}$$

with equality if the covariance matrix associated to $\overline{\rho}$ is diagonal and with

$$\begin{aligned}
\overline{\rho}_1 &= \mathrm{Tr}_{\overline{\rho}_2,\ldots,\overline{\rho}_n}(\overline{\rho}) \\
&\ldots \\
\overline{\rho}_k &= \mathrm{Tr}_{\overline{\rho}_1,\ldots,\overline{\rho}_{k-1},\overline{\rho}_{k+1},\ldots\overline{\rho}_n}(\overline{\rho}) \\
&\ldots \\
\overline{\rho}_n &= \mathrm{Tr}_{\overline{\rho}_1,\ldots,\overline{\rho}_{n-1}}(\overline{\rho}),
\end{aligned} \tag{4.3.6}$$

where $\mathrm{Tr}_{\rho_A}(\rho_{AB})$ denotes the partial trace (4.2.10). Each of the states $\overline{\rho}_k$ corresponds to a reduced Gaussian state with a 2x2 covariance matrix composed by a sum of the corresponding input, noise and modulation. However, as all modulation cross terms do not contribute to the overall energy constraint (4.2.36) we are free to set them to zero and thus equality in (4.3.5) holds as all matrices are now diagonal. We therefore confirm that as in the mono modal case the covariance matrices of input, modulation and environment diagonalize in the same basis.
In the following we will focus on the asymptotic limit (channel uses $n \to \infty$) of the capacity of (4.3.3). In this limit we recall from (2.3.22) the eigenvalue spectrum of the $\hat{x}$ and $\hat{p}$ quadrature blocks of (4.3.1), that is

$$\begin{aligned}
\lambda_x^{\mathrm{env}}(x) &= N\,\frac{1-\phi^2}{1+\phi^2-2\phi\cos(x)}, \\
\lambda_p^{\mathrm{env}}(x) &= N\,\frac{1-\phi^2}{1+\phi^2+2\phi\cos(x)},
\end{aligned} \tag{4.3.7}$$

with the spectral parameter $x \in [0,\pi]$. In this limit we have an infinite number of parallel channels labeled by x which we use one time or equivalently an infinite number of uses of one channel with a noise spectrum varying with x.
In the following, we call the induced input power spectrum $\lambda(x)$, where then the saturated power constraint on the total system reads

$$\overline{n} = \frac{1}{\pi}\int_0^{\pi} dx\ \lambda(x). \tag{4.3.8}$$

4.3.1 Full thermal output

For each individual channel x we know that if its input power $\lambda(x)$ is sufficient the best overall output state is a thermal state, i.e.

$$\lambda_x^{\mathrm{in}}(x) + \lambda_x^{\mathrm{env}}(x) + \lambda_x^{\mathrm{mod}}(x) = \lambda_p^{\mathrm{in}}(x) + \lambda_p^{\mathrm{env}}(x) + \lambda_p^{\mathrm{mod}}(x). \tag{4.3.9}$$

As the noise over the channels is not constant we demand in addition that we have enough energy to fill this gap over all x such that the overall output state will be constant $\forall x$, i.e.

$$\overline{\lambda}_x(x) = \overline{\lambda}_p(x) = \mathrm{const}, \quad \forall x \tag{4.3.10}$$

At the end of the following sections we derive explicitly the input power which is needed such that the latter is fulfilled. Up to than we will adjust all spectra such that (4.3.10) is achieved.
If (4.3.9) is satisfied for a channel x then we know the best input of this channel which reads just as in the one mode case (4.2.68), but is now a function of x, i.e.

$$\lambda_{x,p}^{\text{in}}(x) = \frac{1}{2}\sqrt{\frac{\lambda_{x,p}^{\text{env}}(x)}{\lambda_{p,x}^{\text{env}}(x)}}. \tag{4.3.11}$$

Given this input we define in the asymptotic limit the fraction of power which is needed to squeeze (or entangle in the original basis) the input such that the anisotropy of the environment is matched. The fraction of a general squeezed state was defined for finite n in (4.2.37) by the normalized trace of the input covariance matrix and becomes therefore now a normalized integral over the total input spectrum, that is

$$\eta \equiv \frac{1}{2\pi\overline{n}}\int_0^\pi dx\ \left(\lambda_x^{\text{in}}(x) + \lambda_p^{\text{in}}(x)\right) - \frac{1}{2\overline{n}} = \frac{1}{\pi\overline{n}}\left(\int_0^\pi dx\ \lambda_x^{\text{in}}(x) - \frac{\pi}{2}\right), \tag{4.3.12}$$

where we subtracted the energy of the vacuum and where the simplification is due to the symmetry

$$\lambda_x^{\text{in}}(x) = \lambda_p^{\text{in}}(\pi - x). \tag{4.3.13}$$

Similar to the extension of the input we define the modulation as in the mono-modal case (4.2.69) but with x dependence, i.e.

$$\lambda_{x,p}^{\text{mod}}(x) = \mu - (\lambda_{x,p}^{\text{in}}(x) + \lambda_{x,p}^{\text{env}}(x)), \tag{4.3.14}$$

where μ is now a global waterfilling level over the whole $\hat{x}$,$\hat{p}$ spectrum, such that

$$\frac{1}{\pi}\int_0^\pi \lambda_{x,p}^{\text{mod}}(x) = (1-\eta)\overline{n}. \tag{4.3.15}$$

This waterfilling solution achieves the same total thermal output state for all channels. The fraction power used for modulation in $\hat{x}$ and $\hat{p}$ is identical because the output spectrum

$$\lambda_{x,p}^{\text{out}}(x) = \lambda_{x,p}^{\text{in}}(x) + \lambda_{x,p}^{\text{env}}(x), \tag{4.3.16}$$

also fulfills the symmetry

$$\lambda_x^{\text{out}}(x) = \lambda_p^{\text{out}}(\pi - x). \tag{4.3.17}$$

Thus, the waterfilling level μ is indeed the same for both quadrature spectra. If (4.3.11) and (4.3.14) hold then we fulfill by definition (4.3.9). We notice that the modulation defined in (4.3.14) is actually similar to the quantum water filling solution in the multimode memoryless case (4.2.96), whereas we do not introduce coherent states, but squeezed states as input. This is simply due to the fact that in the basis of the noise our system is a set of

parallel uncorrelated channels where for the modulation (i.e. the classical information we transmit) the classical water filling is optimal, provided that

$$\lambda_{x,p}^{\mathrm{mod}}(x) \geq 0, \forall x \tag{4.3.18}$$

holds. Then (4.3.10) will be fulfilled.
If we consider each of the channels individually, then we know that the allocated input power $\lambda(x)$ contains the input (where we subtract the vacuum) and the modulation, i.e.

$$\lambda(x) = \frac{1}{2}(\lambda_x^{\mathrm{in}}(x) + \lambda_p^{\mathrm{in}}(x) + \lambda_x^{\mathrm{mod}}(x) + \lambda_p^{\mathrm{mod}}(x)) - \frac{1}{2}, \tag{4.3.19}$$

where here $\lambda_{x,p}^{\mathrm{in}}(x)$ and $\lambda_{x,p}^{\mathrm{mod}}(x)$ are assumed as in (4.3.11), (4.3.14), respectively. Recalling the definitions in the mono-modal case, we know that if

$$\lambda(x) \geq \overline{n}_{\mathrm{thr}}(x), \quad \forall x \tag{4.3.20}$$

where $\overline{n}_{\mathrm{thr}}(x)$ is now a function of x for each channel, then all channels are in a thermal state with same temperature over all x. $\overline{n}_{\mathrm{thr}}$ reads equivalently to the mono-modal definition, but as we imposed in (4.2.75) the condition that the $\hat{x}$ noise is always greater or equal to the $\hat{p}$ noise we have to switch the quadratures at $x = \pi/2$ since above that position the $\hat{p}$ noise is greater than the $\hat{x}$ noise. Therefore we write

$$\overline{n}_{\mathrm{thr}}(x) = \frac{1}{2}\begin{cases} \left(\sqrt{\frac{\lambda_x^{\mathrm{env}}(x)}{\lambda_p^{\mathrm{env}}(x)}} + \lambda_x^{\mathrm{env}}(x) - \lambda_p^{\mathrm{env}}(x) - 1\right), & 0 < x \leq \pi/2 \\ \left(\sqrt{\frac{\lambda_p^{\mathrm{env}}(x)}{\lambda_x^{\mathrm{env}}(x)}} + \lambda_p^{\mathrm{env}}(x) - \lambda_x^{\mathrm{env}}(x) - 1\right), & \pi/2 < x \leq \pi. \end{cases} \tag{4.3.21}$$

In figure 4.9 we plotted the quadrature noises as well as the threshold for certain N, ϕ. Since all quantities of the allocated power $\lambda(x)$ spectrum are given by the environment except the water filling level μ, we will now specify the condition $\overline{n}$ such that the resulting μ leads to a positive modulation spectrum. This condition was already found in the treatment of the classical pendant in (3.2.9) and reads in the quantum case as

$$\begin{aligned} &\pi(1-\eta)\overline{n} \geq \pi\lambda_x^{\mathrm{out}}(0) - \int_0^\pi dx\ \lambda_x^{\mathrm{out}}(x) \\ \Leftrightarrow\ &(1-\eta)\overline{n} \geq \frac{1+\phi}{1-\phi}\left(N + \frac{1}{2}\right) - \underbrace{\frac{1}{\pi}\int_0^\pi dx\ \lambda_x^{\mathrm{in}}(x)}_{=\eta\overline{n}+\frac{1}{2}} - \underbrace{\frac{1}{\pi}\int_0^\pi dx\ \lambda_x^{\mathrm{env}}(x)}_{=N} \\ \Leftrightarrow\ &\overline{n} \geq N\left(\frac{1+\phi}{1-\phi} - 1\right) + \frac{1}{2}\left(\frac{1+\phi}{1-\phi} - 1\right). \end{aligned} \tag{4.3.22}$$

Therefore, we define the quantity which is needed to modulate the full spectrum, that is

$$\overline{n}_{\mathrm{s}}(\phi, N) \equiv N\left(\frac{1+\phi}{1-\phi} - 1\right) + \frac{1}{2}\left(\frac{1+\phi}{1-\phi} - 1\right), \tag{4.3.23}$$

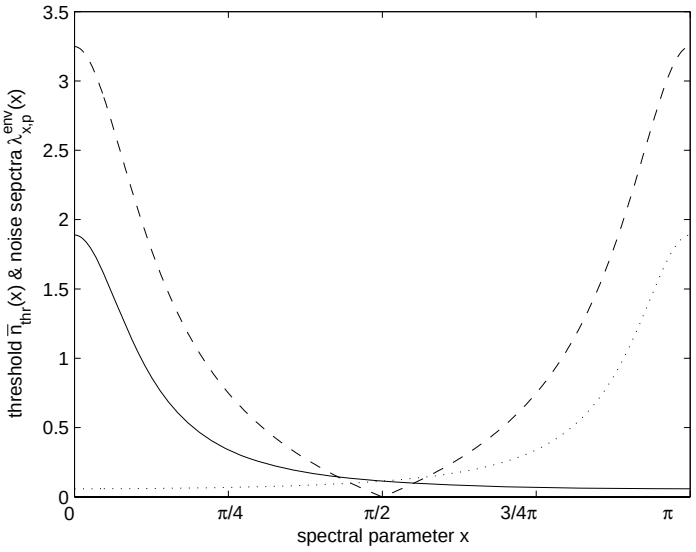

Figure 4.9: Threshold function $\overline{n}_{\rm thr}(x)$ and noise eigenvalue functions $\lambda_x^{\rm env}(x), \lambda_p^{\rm env}(x)$ vs. spectral parameter x. The dashed line shows $\overline{n}_{\rm thr}(x)$. The solid and dotted line the noise spectrum of the $\hat{x}$ and $\hat{p}$ quadrature $\lambda_x^{\rm env}(x), \lambda_p^{\rm env}(x)$ which are mirror symmetrical around $x = \pi/2$. For this plot we took $N = 1/3, \phi = 0.7$.

as a function of the correlation strength ϕ. The assumption in the first line of (4.3.22) could have been expressed in terms of $\lambda_p^{\rm out}(x)$ as well, since (4.3.17) holds. In the memoryless case $\phi = 0$ we confirm that the input power which is needed to have a full positive modulation spectrum $\overline{n}_{\rm s}(\phi, N)$ is anything above zero, because in this case the channel is $\forall x$ the same thermal channel and therefore we will not use any energy to prepare the input state[17]. If $\phi \to 1$ then $\overline{n}_{\rm s}(\phi, N)$ will tend to infinity since the highest eigenvalue $\lambda_x^{\rm out}(0)$ tends to infinity as well, making it impossible to modulate the whole spectrum for finite input power. Interestingly, if we recall the definition of the power $P_{\rm s}$ which is sufficient to modulate the whole spectrum in the classical case, (see (3.2.9)) we receive the simple connection with the quantum expression $\overline{n}_{\rm s}(\phi, N)$ by

$$\overline{n}_{\rm s}(\phi, N) = P_{\rm s}(\phi) + \frac{1}{2}\left(\frac{1+\phi}{1-\phi} - 1\right). \tag{4.3.24}$$

In the quantum case we have an extra term which is diverging with $\phi \to 1$ which is the highest eigenvalue of the quantum input, i.e. $\lambda_x^{\rm in}(0) = \lambda_p^{\rm in}(\pi)$. A more detailed comparison with the classical solution is presented in section 4.3.1.
For the quantum water filling level we obtain an expression similar to the mono-modal

[17] Coherent states are conjectured to be optimal and just convey the vacuum energy

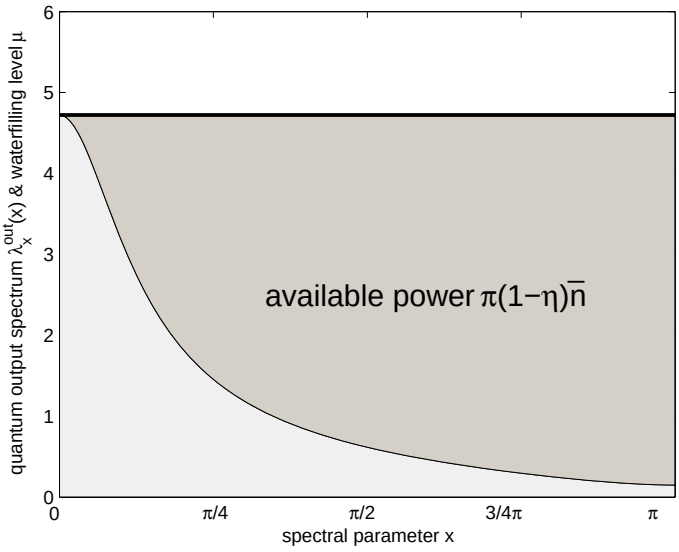

Figure 4.10: Output spectrum $\lambda_x^{\text{out}}(x)$ (decreasing curve), water filling level μ (bold line) vs. spectral parameter x. The dark gray area shows the remaining distributed power $\pi(1-\eta)\overline{n}$. We took $\overline{n} = \overline{n}_{\text{s}}(\phi, N) = 3.89$, for $N = 1/3, \phi = 0.7$, such that the spectrum is precisely completely filled.

case (4.2.70), i.e.

$$\begin{aligned} \mu &= (1-\eta)\overline{n} + \frac{1}{\pi}\int_0^\pi dx\ \lambda_x^{\text{out}}(x) \\ &= \overline{n} + N + \frac{1}{2}, \end{aligned} \tag{4.3.25}$$

which is again just the total average power of the system. Assuming (4.3.22) is fulfilled we know with (4.3.25) now indeed the overall output state or in other words the output temperature $\forall x$ which is the constant in (4.3.10), i.e.

$$\overline{\lambda}_x(x) = \overline{\lambda}_p(x) = \mu = \overline{n} + N + \frac{1}{2}. \tag{4.3.26}$$

In figure 4.10 we plotted the "water filling-picture" (with $\overline{n} = \overline{n}_{\text{s}}(\phi, N)$) similar to the classical case (see section 3.2), whereas now the quantum output spectrum $\lambda_x^{\text{out}}(x)$ ($\lambda_p^{\text{out}}(x)$, respectively) plays the role of the noise spectrum and the filled area leads to the classical modulation spectrum instead of the classical input. The resulting $\hat{x}, \hat{p}$-spectra and the total power spectrum are shown in 4.11. The threshold $\overline{n}_{\text{thr}}(x)$, the resulting modulation and power spectrum $\lambda(x)$ are displayed in figure 4.12. With (4.3.26) we will now give an explicit expression for the Capacity of the channel with $\overline{n} \geq \overline{n}_{\text{s}}(\phi, N)$. The Capacity for finite uses n of the bosonic channel was defined in (4.2.41), but since we are in the limit

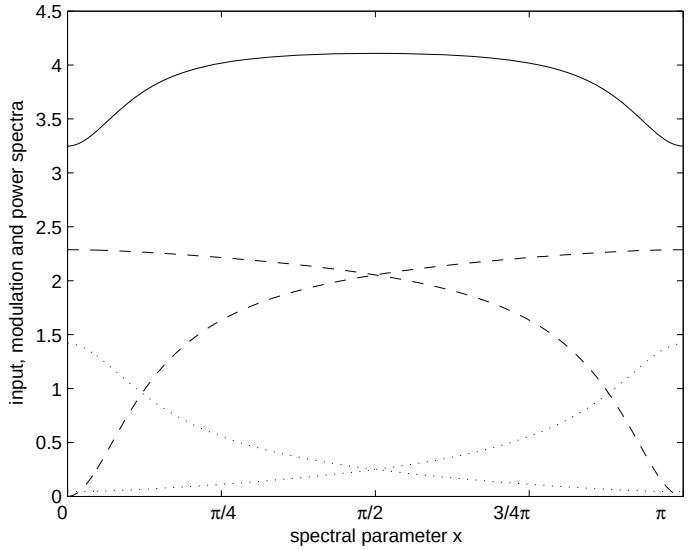

Figure 4.11: Input, modulation and power spectrum vs. spectral parameter x. The decreasing, increasing dotted curves correspond to $\frac{1}{2}\lambda_x^{\rm in}(x), \frac{1}{2}\lambda_p^{\rm in}(x)$, the increasing, decreasing dashed curves correspond to $\frac{1}{2}\lambda_x^{\rm mod}(x), \frac{1}{2}\lambda_p^{\rm mod}(x)$ and the solid line shows $\lambda(x)$. For this plot we took $N = 1/3, \phi = 0.7, \overline{n} = \overline{n}_{\rm s}(\phi, N) = 3.89$.

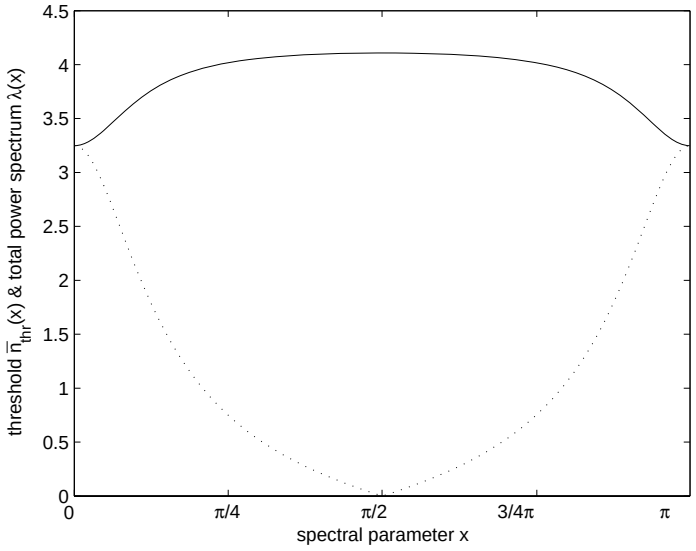

Figure 4.12: Threshold function $\overline{n}_{\rm thr}(x)$ (dotted curve) and power spectrum $\lambda(x)$ (solid curve) vs. spectral parameter x. For $\lambda(x)$ we took $\overline{n} = \overline{n}_{\rm s}(\phi, N) = 3.89$, for $N = 1/3, \phi = 0.7$.

$n \to \infty$, we are only considering the asymptotic expression (4.2.26) of the finite capacity, where the resulting expression is the limit of Riemann sums, i.e.

$$C_{\mathrm{s}}^{(\mathrm{qu})} = \lim_{n\to\infty} C_{n,s}^{(\mathrm{qu})} = \frac{1}{\pi}\int_0^\pi dx\; g\left(\overline{\nu}(x) - \frac{1}{2}\right) - g\left(\nu^{\mathrm{out}}(x) - \frac{1}{2}\right), \tag{4.3.27}$$

where the symplectic eigenvalue functions $\overline{\nu}(x), \nu^{\mathrm{out}}(x)$ read

$$\begin{aligned} \overline{\nu}(x) &= \sqrt{\overline{\lambda}_x(x)\,\overline{\lambda}_p(x)}, \\ \nu^{\mathrm{out}}(x) &= \sqrt{\lambda_x^{\mathrm{out}}(x)\,\lambda_p^{\mathrm{out}}(x)}. \end{aligned} \tag{4.3.28}$$

Inserting in the latter the result (4.3.26), (4.3.16) and using (4.3.27) we obtain the final expression for the optimal transmission rate above the threshold, which is now indeed the capacity of the system in this domain, that is

$$C_{\mathrm{s}}^{(\mathrm{qu})} = \frac{1}{\pi}\int_0^\pi dx\; g\left(\overline{\nu}(x) - \frac{1}{2}\right) - g\left(\nu^{\mathrm{out}}(x) - \frac{1}{2}\right), \quad \overline{n} \geq \overline{n}_{\mathrm{s}}(\phi, N) \tag{4.3.29}$$

$$\begin{aligned} &= g\left(\mu - \frac{1}{2}\right) - g\left(\sqrt{\left(\frac{1}{2}\sqrt{\frac{\lambda_x^{\mathrm{env}}(x)}{\lambda_p^{\mathrm{env}}(x)}} + \lambda_x^{\mathrm{env}}(x)\right)\left(\frac{1}{2}\sqrt{\frac{\lambda_p^{\mathrm{env}}(x)}{\lambda_x^{\mathrm{env}}(x)}} + \lambda_p^{\mathrm{env}}(x)\right)} - \frac{1}{2}\right) \\ &= g(\overline{n} + N) - \frac{1}{\pi}\int_0^\pi dx\; g\left(\sqrt{\lambda_x^{\mathrm{env}}(x)\,\lambda_p^{\mathrm{env}}(x)}\right), \quad \overline{n} \geq \overline{n}_{\mathrm{s}}(\phi, N) \\ &= g(\overline{n} + N) - \frac{1}{\pi}\int_0^\pi dx\; g\left(\frac{N(1-\phi^2)}{\sqrt{(1+\phi^2)^2 - 4\,\phi^2\,\cos^2 x}}\right), \quad \overline{n} \geq \overline{n}_{\mathrm{s}}(\phi, N). \end{aligned} \tag{4.3.30}$$

where $0 \leq \phi < 1$. The first line in (4.3.30) is, as expected, equivalent to the one-mode result (4.2.71), where we now have to integrate over the noise spectrum. In the memoryless case $\phi = 0$ we recover that the channel is no longer anisotropic but a thermal channel with same noise for the whole spectrum and thus, the capacity (4.3.30) coincides with the one shot capacity of the thermal channel (4.2.50).
Although we will not be able to express $C_{\mathrm{s}}^{(\mathrm{qu})}$ completly in terms of elementary functions we can treat its limit behavior.

Limit of full correlations

At first we investigate the limit of full correlations, i.e. $\phi \to 1$. As the first term in (4.3.30) is indepent of ϕ we will just take the limit of the integral term (where we drop the constant

N for now):

$$\lim_{\phi\to 1}\int_0^\pi dx\; g\left(\frac{1-\phi^2}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right)=\int_0^\pi dx\;\lim_{\phi\to 1} g\left(\frac{1-\phi^2}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right), \tag{4.3.31}$$

The latter holds, since we do not integrate over ϕ. We see that

$$\lim_{\phi\to 1} g\left(\frac{1-\phi^2}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right)\to 0,\quad 0<x<\pi, \tag{4.3.32}$$

since $g(x)$ is monotonously decreasing with decreasing x. Only at the limits of the integral we find (for arbitrary $0\leq\phi<1$)

$$g\left(\frac{(1-\phi^2)}{\sqrt{(1+\phi^2)^2-4\,\phi^2}}\right)=\lim_{\phi\to 1} g\left(\frac{1-\phi^2}{\sqrt{(1-\phi^2)^2}}\right)=1,\quad x=\{0,\pi\}. \tag{4.3.33}$$

By (4.3.32) we know that the integral in (4.3.31) will vanish everywhere exact for an infinitesimal small contribution at the boundaries. As these can be neglected we conclude that

$$\int_0^\pi dx\;\lim_{\phi\to 1} g\left(\frac{1-\phi^2}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right)=0, \tag{4.3.34}$$

and that thus the capacity (above threshold) for full correlation stays finite, namely

$$\lim_{\phi\to 1} C_{\mathrm{s}}^{(\mathrm{qu})}=g(\overline{n}+N). \tag{4.3.35}$$

If we now fix $\overline{n}=\overline{n}_{\mathrm{s}}(\phi,N)+1$ to ensure that the input power always suffices we can observe the limit behavior of the first and second term in (4.3.30) in figure 4.13. In figure 4.14 the capacity (4.3.30) is plotted for different noises versus the correlation parameter ϕ, where we fixed again the input power $\overline{n}=\overline{n}_{\mathrm{s}}(\phi,N)+1$. At $\phi=0$ the capacity of the system with highest noise N is the lowest and vice versa. Due to the behaviour of the integral noise terms (here for $N=0.01,1,10$), which all tend to zero, for some ϕ the input power (given by $\overline{n}_{\mathrm{s}}(\phi,N)$) of the noisiest system clearly overcomes the others. In the classical case, due to the divergence of the noise part all capacities tend to infinity and thus the noisiest part always has the lowest capacity (see 3.5.

Classical limit

In the following we recover the classical result, i.e. the capacity of the classical Gaussian channel from the expression of the capacity of the bosonic channel. In the classical limit we increase the input power and the noise while the signal to noise ratio is kept constant. Recalling the expression for the capacity of the bosonic channel (4.3.30) and the classical

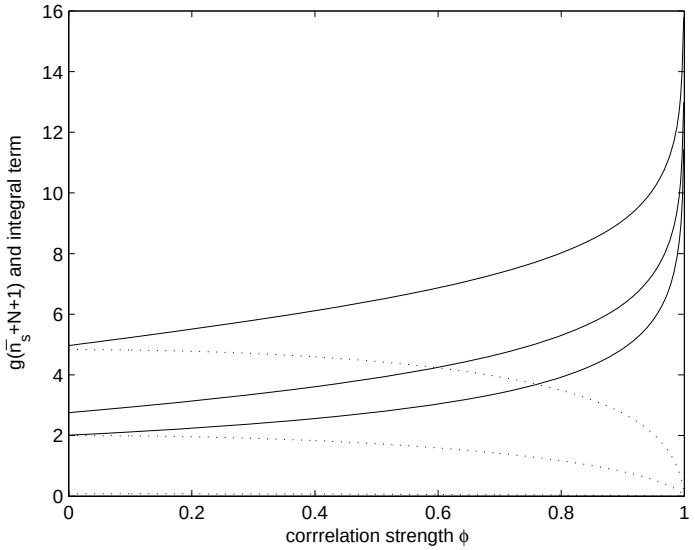

Figure 4.13: $g(\overline{n}_{\rm s}(\phi,N)+1+N)$ and integral term of (4.3.30) vs. correlation strength ϕ. The three dotted curves correspond from bottom to top to the integral term with $N = 0.01, 1, 10$. The three solid curves correspond from bottom to top to $g(\overline{n}_{\rm s}(\phi,N)+1+N)$ with $N = 0.01, 1, 10$.

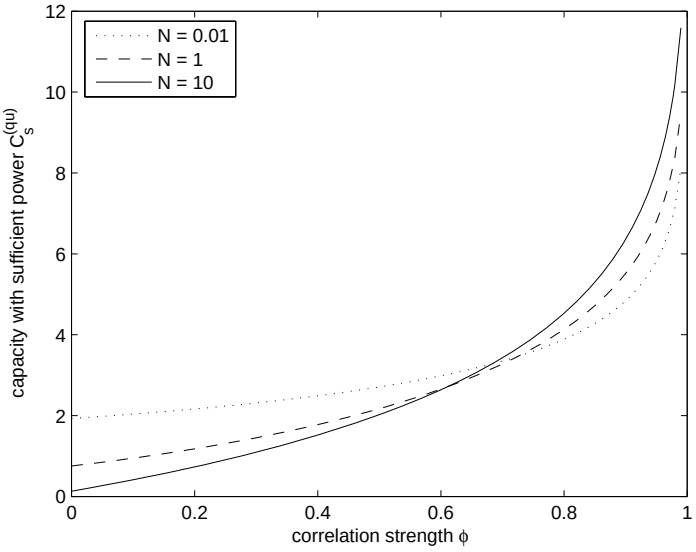

Figure 4.14: Capacity $C_{\rm s}^{({\rm qu})}$ vs. correlation strength ϕ. The dotted, dashed and solid curve correspond to $N = 0.01, 1, 10$. For all cureves we took $\overline{n} = \overline{n}_{\rm s}(\phi,N)+1$.

limit behavior of $g(x)$ (4.2.51) we obtain the classical limit terms, that is

$$\begin{aligned}\lim_{\overline{n},N\to\infty,\overline{n}/N=c} g(\overline{n}+N) - \frac{1}{\pi}\int_0^\pi dx\; g\left(\frac{N(1-\phi^2)}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right)\\ -\log(\overline{n}+N) - \frac{1}{\pi}\int_0^\pi dx\;\log\left(\frac{N(1-\phi^2)}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right) = 0,\end{aligned} \tag{4.3.36}$$

where the integral term can be simplified:

$$\frac{1}{\pi}\int_0^\pi dx\;\log\left(\frac{N(1-\phi^2)}{\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}}\right) = \log(N(1-\phi^2)) + \frac{1}{\pi}\int_0^\pi dx\;\log\left(\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}\right). \tag{4.3.37}$$

(4.3.36) is almost the desired the result. We solve the remaining integral as follows:

$$\begin{aligned}\frac{1}{\pi}\int_0^\pi dx\;\log\left(\sqrt{(1+\phi^2)^2-4\,\phi^2\,\cos^2 x}\right) &= \frac{1}{2\pi}\int_0^\pi dx\;\left(\log\left(1-\frac{4\phi^2}{(1+\phi^2)^2}\cos^2 x\right)\right) + \frac{\pi}{2}\log((1+\phi^2)^2)\\ &= \log\left(\frac{1}{2}\left(1+\sqrt{1-\frac{4\phi^2}{(1+\phi^2)^2}}\right)\right) + \log(1+\phi^2)\\ &= \log\left(1+\phi^2+\sqrt{(1+\phi^2)^2\left(1-\frac{4\phi^2}{(1+\phi^2)^2}\right)}\right) - 1\\ &= \log(1+\phi^2+\sqrt{1-2\phi^2+\phi^4}) - 1\\ &= \log(2) - 1 = 0,\end{aligned} \tag{4.3.38}$$

where the integral in the first line was found in [11](4.226). Thus, we conclude that

$$\lim_{\overline{n},N\to\infty,\,\overline{n}/N=c} C_{\mathrm{s}}^{(\mathrm{qu})} = C_{\mathrm{s}}^{(\mathrm{cl})} = \log\left(\frac{1}{1-\phi^2}\left(1+\frac{\overline{n}}{N}\right)\right), \tag{4.3.39}$$

where $C_{\mathrm{s}}^{(\mathrm{qu})}$ is defined as in (4.3.30), c is a constant and $C_{\mathrm{s}}^{(\mathrm{cl})}$ the capacity of the classical Gaussian channel (with Gauss-Markov noise) with sufficient input power (3.2.13). As the $g(x)$-function is always smaller than $\log x$ we will reach this limit from below and hence, for small $\overline{n}, N$ we can in addition introduce the inequality

$$C_{\mathrm{s}}^{(\mathrm{qu})} < C_{\mathrm{s}}^{(\mathrm{cl})}. \tag{4.3.40}$$

This is the expected result, because for the bosonic channel we will always (except in the classical limit) invest a certain amount of the input power to prepare the input entangled state, which we need not invest in the classical case. In the classical limit the fraction of power used for entanglement vanishs to zero and therefore the capacities coincide. In figure

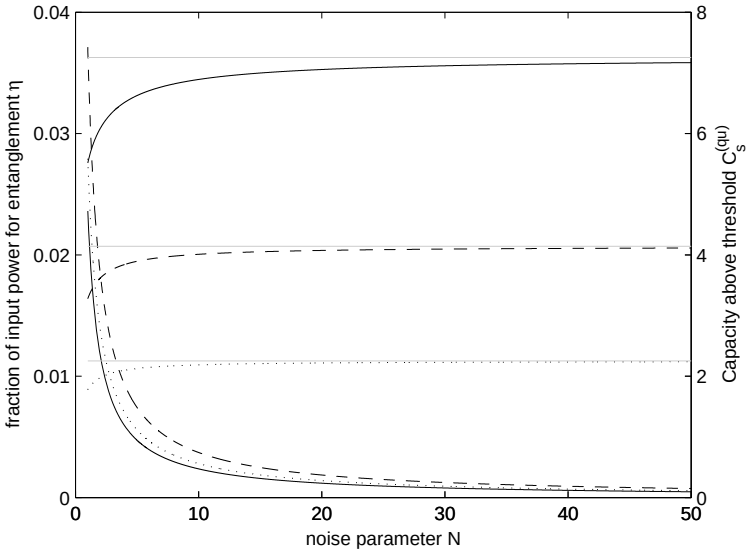

Figure 4.15: Fraction of input power used for entanglement η and Capacity $C_{\mathrm{s}}^{(\mathrm{qu})}$ vs. noise parameter N. The increasing curves correspond to the capacities, the decreasing to η. The gray bars correspond to the classical limit of each capacity. For all graphs the dotted, dashed and solid line correspond to $\phi = 0.4, 0.7, 0.9$. For this plot we took kept $\overline{n}/N = \overline{n}_{\mathrm{s}}(\phi, N = 1), \forall N \geq 1$.

4.15 we plotted the capacity $C_{\mathrm{s}}^{(\mathrm{qu})}$ and fraction η versus the noise parameter N. For each curve we kept a constant signal to noise ratio which we defined as follows: since the signal to noise ratio $\overline{n}_{\mathrm{s}}/N$ is monotonically decreasing with increasing N towards a constant (see (4.3.23)) we simply set

$$\frac{\overline{n}}{N} = \frac{\overline{n}_{\mathrm{s}}(\phi, N=1)}{1} = \overline{n}_{\mathrm{s}}(\phi, N=1), \tag{4.3.41}$$

for each ϕ, and therefore are sure that $\overline{n} \geq \overline{n}_{\mathrm{s}}, \forall N \geq 1$. The capacities all tend to their classical limit (4.3.39), while all fractions of power used to prepare the input are decreasing towards zero. Interestingly, for the introduced SNR $\eta(\phi = 0.7) > \eta(\phi = 0.9)$. This is due to the fact that $\overline{n}_{\mathrm{s}}(\phi = 0.7, 1) \ll \overline{n}_{\mathrm{s}}(\phi = 0.9, 1)$ and that hence, although we need more input to match the higher correlation its cost with respect to the total power $\overline{n}$ is lower.

Simplification of the integral term

We try now to simplify the integral term in (4.3.30). By defining

$$\begin{aligned} a &\equiv N(1-\rho^2) \\ b(x) &\equiv (1+\phi^2)\sqrt{1 - \frac{4\phi^2}{(1+\phi^2)^2}\cos^2 x}, \end{aligned} \tag{4.3.42}$$

we can rewrite

$$\int_0^\pi dx\; g\left(\frac{a}{b(x)}\right) = \int_0^\pi dx\;\left[\log\left(a + b(x)\right)\left(\frac{a}{b(x)} + 1\right) - \frac{a}{b(x)}\log a - \log b(x)\right]. \tag{4.3.43}$$

Secondly, we calculate the third term in the integral, namely

$$\begin{aligned} \int_0^\pi dx\; \frac{a}{b(x)}\log a &= N\left(\frac{1-\phi^2}{1+\phi^2}\right)\log\left(N(1-\phi^2)\right)\int_0^\pi dx\; \frac{1}{\sqrt{1 - \frac{4\phi^2}{(1+\phi^2)^2}\cos^2 x}} \\ &= 2N\left(\frac{1-\phi^2}{1+\phi^2}\right)\log\left(N(1-\phi^2)\right) K\left(\frac{2\phi}{1+\phi}\right), \end{aligned} \tag{4.3.44}$$

where $K(k)$ is the complete elliptic integral[18] of the first kind defined as

$$K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2\,\sin^2\theta}}, \tag{4.3.45}$$

which cannot be expressed further by elementary functions. Given the fact that in (4.3.43) another term includes a complete elliptic integral and an additional logarithm function we conclude that this is as far as the capacity expression can be simplified.

[18] Elliptic integrals are needed to calculate for instance the circumference of an ellipse.

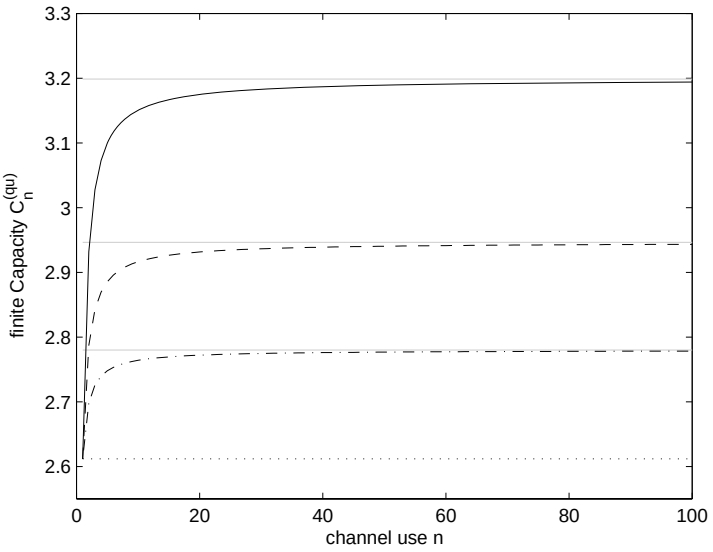

Figure 4.16: Capacity for finite uses $C^{(\mathrm{qu})}_{n,\mathrm{s}}$ and asymptotic capacity $C^{(\mathrm{qu})}_{\mathrm{s}}$ vs. number of channel uses n. The dotted, dashed, dashed-dotted and solid curve correspond to $\phi = 0, 0.4, 0.55, 0.7$. For each ϕ the corresponding asymptotic capacity is shown by a gray bar. For all plots we took $N = 1$ and $\overline{n} = 7.5 > \overline{n}_s(\phi = 0.7)$.

From finite to infinite uses

Let us have a look at the behavior of the finite Capacity of the channel (where $\overline{n} \geq \overline{n}_{\mathrm{s}}(\phi, N)$) with increasing channel uses (or equivalently dimension n of the system). We defined this expression most generally in (4.2.41), but since we are now in the domain where we know the optimal input and modulation for all eigenvalues we can simply the expression to

$$C^{(\mathrm{qu})}_{n,s} = \frac{1}{n}\sum_{j=1}^{n} g\left(\overline{\nu}_j - \frac{1}{2}\right) - g\left(\nu_j^{\mathrm{out}} - \frac{1}{2}\right), \tag{4.3.46}$$

where $\{\overline{\nu}_j\}, \{\nu_j^{\mathrm{out}}\}$ are now again finite symplectic eigenvalue spectra with

$$\begin{aligned} \nu_j^{\mathrm{out}} &= \sqrt{(\lambda^{\mathrm{in}}_{x,j} + \lambda^{\mathrm{env}}_{x,j})(\lambda^{\mathrm{in}}_{p,j} + \lambda^{\mathrm{env}}_{p,j})}, \\ \overline{\nu}_j &= \sqrt{(\lambda^{\mathrm{in}}_{x,j} + \lambda^{\mathrm{env}}_{x,j} + \lambda^{\mathrm{mod}}_{x,j})(\lambda^{\mathrm{in}}_{p,j} + \lambda^{\mathrm{env}}_{p,j} + \lambda^{\mathrm{mod}}_{p,j})}, \end{aligned} \tag{4.3.47}$$

where all definitions of the finite eigenvalue spectra read as in the continuous case but with the discrete index j. However, as we do not know the eigenvalue spectrum of the noise covariance matrix (4.3.1), we have to calculate for each dimension n of the system the spectra numerically. In figure 4.16 we plotted the capacity for finite n (4.3.46) versus the channel uses n for fixed N, various correlation strength ϕ and a fixed input power $\overline{n}$ which is greater than $\overline{n}_{\mathrm{s}}(\phi, N)$ of the strongest ϕ. In addition we denoted by gray bars for

each ϕ the asymptotic capacity (4.3.30). Clearly, in the plotted region, we confirm that the asymptotic expression is indeed an upper limit[19]. The evolution of the transmission rate with respect to the uses n can be further understood if we decompose the finite capacity (4.3.46) and compare it with the asymptotic function (over which we integrate). Therefore, we introduce

$$\begin{aligned} \Delta g_j &= g\left(\overline{\nu}_j - \frac{1}{2}\right) - g\left(\nu_j^{\text{out}} - \frac{1}{2}\right), \\ \Delta g(x) &= g\left(\overline{\nu}(x) - \frac{1}{2}\right) - g\left(\nu^{\text{out}}(x) - \frac{1}{2}\right). \end{aligned} \tag{4.3.48}$$

In order to compare both functions in the region $[0, \pi]$ we define the plot positions for Δg_j by $\pi/n(j - 1/2)$. Since the finite rate (4.3.46) is normalized by n and the asymptotic rate (4.3.30) by π we multiply each Δg_j by π/n and then can compare in one bar / line plot the contributions for the finite and asymptotic rate. This is realized in figure 4.17 for different channel uses n, where $\phi = 0.7, N = 1/3, \overline{n} = 8$. In these plots one observes the convergent behavior of the finite capacity with respect to the asymptotic limit. We observe that the sum of the finite $\Delta g_j \, \pi/n$ is tending with increasing channel uses n more and more towards the asymptotic area (the area below $\Delta g(x)$).

[19] We tested numerically that for higher n this asymptotic bound is indeed not exceeded.

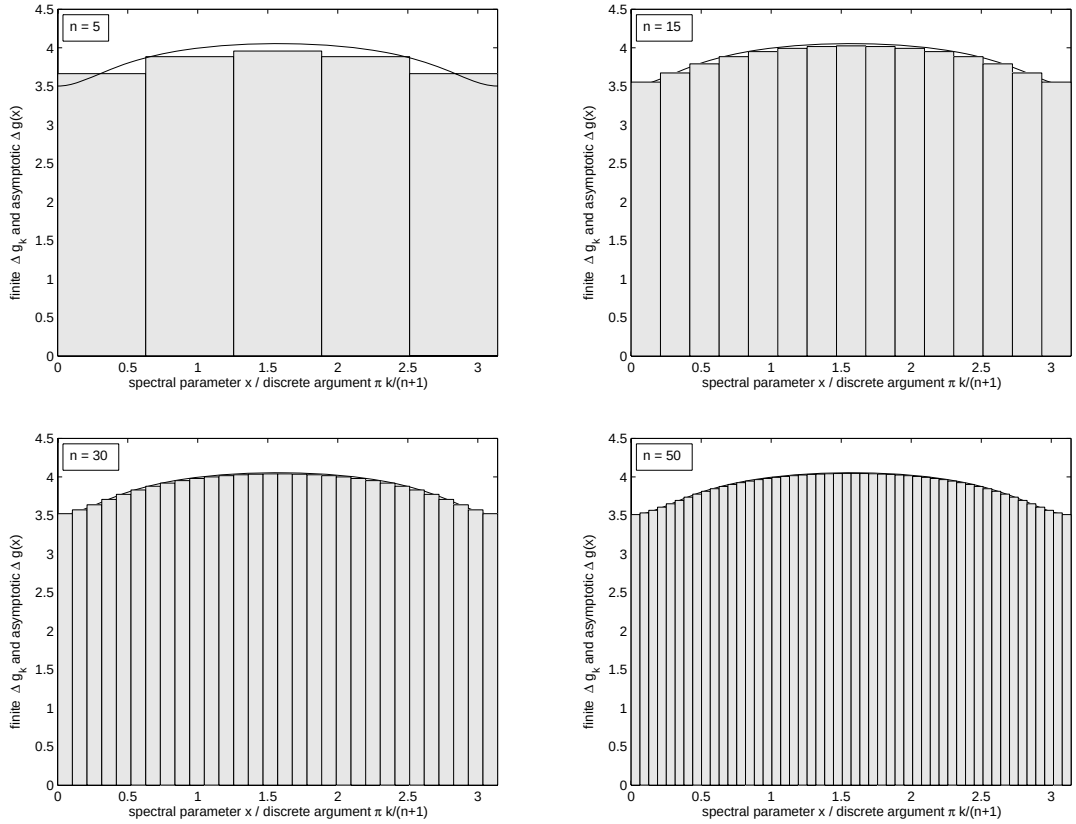

Figure 4.17: Rate contributions $\Delta g_j \cdot \pi/n$ (bars) and $\Delta g(x)$ (line) vs. parameter x, $\pi k/(n+1)$.

4.3.2 Partly thermal output

As we characterized the condition on $\overline{n}$ precisely we know now when we will not have sufficient input power, such that the overall output states are all in one thermal state, i.e. if

$$\overline{n} < \overline{n}_{\mathrm{s}}(\phi, N). \tag{4.3.49}$$

We will introduce in the following three ideas of distributing the given total power if the latter holds. Afterwards we will make several comparison between the models.

Ansatz 1: Dropping channels

One way of keeping the property of a total constant thermal state is by ignoring the nosiest channels. This means that we will shrink our spectrum up to the region in which we know that the input power will be exactly sufficient to fit the squeezing anisotropy and at the same time modulate all channels in this region, i.e.

$$\overline{\lambda}_x(x) = \overline{\lambda}_p(x) = \mathrm{const}, \quad \forall x \in [\kappa, \pi - \kappa]. \tag{4.3.50}$$

This condition can be rewritten in terms of the threshold, where we now take again $\lambda^{\mathrm{in}}_{x,p}(x), \lambda^{\mathrm{mod}}_{x,p}(x)$ as in (4.3.11), (4.3.14), respectively, that is

$$\lambda_\kappa(x) = \overline{n}_{\mathrm{thr}}(x), \quad \forall x \in [\kappa, \pi - \kappa]. \tag{4.3.51}$$

With this new spectral domain we define the fraction of power used for matching the noise asymmetry analog to (4.3.12), i.e.

$$\eta_\kappa \equiv \frac{1}{(\pi - 2\kappa)\overline{n}} \left(\int_\kappa^{\pi-\kappa} dx\ \lambda^{\mathrm{in}}_x(x) - \frac{\pi - 2\kappa}{2} \right). \tag{4.3.52}$$

The equation for the mean power constraint then reads

$$\begin{aligned} (\pi - 2\kappa)(1 - \eta)\overline{n} &= (\pi - 2\kappa)\lambda^{\mathrm{out}}_x(\kappa) - \int_\kappa^{\pi-\kappa} dx\ \lambda^{\mathrm{out}}_x(x) \\ \Leftrightarrow \overline{n} &= \lambda^{\mathrm{out}}_x(\kappa) - \frac{1}{\pi - 2\kappa} \int_\kappa^{\pi-\kappa} dx\ \lambda^{\mathrm{env}}_x(x) - \frac{1}{2}. \end{aligned} \tag{4.3.53}$$

Since we know that the modulation will "just fill" the output spectra $\lambda^{\mathrm{out}}_{x,p}(x)$, i.e. that $\lambda^{\mathrm{mod}}_x(\kappa) = \lambda^{\mathrm{mod}}_p(\pi - \kappa) = 0$, the water filling level is simply determined by

$$\mu_\kappa = \lambda^{\mathrm{out}}_x(\kappa) = \overline{n} + \frac{1}{\pi - 2\kappa} \int_\kappa^{\pi-\kappa} dx\ \lambda^{\mathrm{env}}_x(x) + \frac{1}{2}. \tag{4.3.54}$$

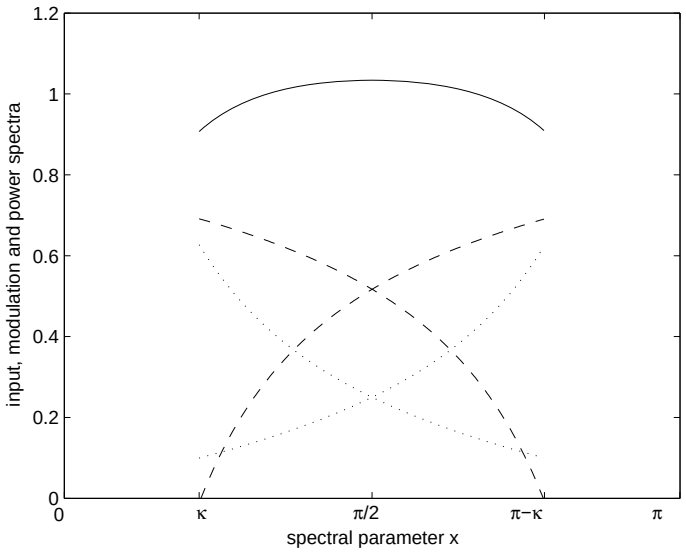

Figure 4.18: Input, modulation and power spectrum vs. spectral parameter x. The decreasing, increasing dotted curves correspond to $\frac{1}{2}\lambda_x^{\text{in}}(x), \frac{1}{2}\lambda_p^{\text{in}}(x)$, the increasing, decreasing dashed curves correspond to $\frac{1}{2}\lambda_x^{\text{mod}}(x), \frac{1}{2}\lambda_p^{\text{mod}}(x)$ and the solid line shows $P_\kappa(x)$. For this plot we took $N = 1/3, \phi = 0.7, \overline{n} = 1$.

As in the previous section all overall output eigenvalues are just determined by the water-filling level, i.e.

$$\overline{\lambda}_x(x) = \overline{\lambda}_p(x) = \mu_\kappa, \quad \forall x \in [\kappa, \pi - \kappa]. \tag{4.3.55}$$

The different spectra are shown in figure 4.18. Using (4.3.27) we can write out the transmission rate[20], i.e.

$$R_\kappa = \frac{1}{\pi - 2\kappa} \int_\kappa^\pi dx \; g\left(\overline{\nu}(x) - \frac{1}{2}\right) - g\left(\nu^{\text{out}}(x) - \frac{1}{2}\right), \tag{4.3.56}$$

where the symplectic eigenvalue functions $\overline{\nu}(x), \nu^{\text{out}}(x)$ are defined as in (4.3.28). Unfortunatly we have no further way of simplifying expression (4.3.56), which means that the solution will be purely numerical.

Ansatz 2: Hybrid solution

An additional possibility way of optimizing the transmission rate is to choose a part of the spectrum where we know that we can achieve a thermal output for each channel, apply the appropriated squeezing and an overall spectral modulation (whereas the output temperature of each channel will be different) and to apply the linear approximation analog

[20] We no longer call this rate "capacity" as the optimal transmission rate is unknown in this region

to the mono-modal linear approximation with the remaining power which is distributed uniformly among the channels. We split the spectrum at position $x = \beta$, which is found by solving

$$\overline{n} = \overline{n}_{\text{thr}}(x = \beta), \tag{4.3.57}$$

for the given N, ϕ. We know that in the region $x \in [\beta, \pi - \beta]$ each channel could be made thermal, however the energy will not be enough to modulate the complete spectrum. For a calculated β we define the input by

$$\lambda_x^{\text{in}}(x) = \frac{1}{2} \begin{cases} \left(\sqrt{\frac{\lambda_x^{\text{env}}(x)}{\lambda_p^{\text{env}}(x)}} \right), \ \beta \leq x \leq \pi - \beta \\ \left(2\overline{n} + 1 + \Delta_{p,x}^{\text{env}}(x) + \Delta_{x,p}^{\text{env}}(x) \left(\frac{\overline{n}_{\text{thr}}(x) - \overline{n}}{\overline{n}} \right) \right), \ \text{otherwise} \end{cases} \tag{4.3.58}$$

with $\lambda_{x,p}^{\text{env}}(x)$ as in (4.3.7) and with $\Delta_{x,p}^{\text{env}}(x) = \lambda_x^{\text{env}}(x) - \lambda_p^{\text{env}}(x)$. The fraction of energy used to prepare the input reads similar to the case with full thermal output, i.e.

$$\eta_{\text{hyb}} = \frac{1}{\pi \overline{n}} \left(\int_0^\pi dx \ \lambda_x^{\text{in}}(x) \right) - \frac{1}{2\overline{n}}. \tag{4.3.59}$$

If we follow the same steps as in (4.3.22) we see that for given mean input power $\overline{n}$ indeed

$$\overline{n} < \lambda_x^{\text{out}}(0) - N - \frac{1}{2}, \tag{4.3.60}$$

where $\lambda_x^{\text{out}}(x) = \lambda_x^{\text{in}}(x) + \lambda_x^{\text{env}}(x)$, which means that the remaining energy will not be sufficient to modulate the whole spectrum. Therefore we apply a modified water filling (analog to the classical solution in section 3.2), because the output spectrum is monotonously decreasing (increasing for $\hat{p}$), as well as for the Markovian model. We define

$$\begin{aligned} \lambda_x^{\text{mod}}(x) &= \Theta(x - \alpha_{\text{hyb}})(\mu_{\text{hyb}} - \lambda_x^{\text{out}}(x)) \\ \lambda_p^{\text{mod}}(x) &= \Theta(\pi - \alpha_{\text{hyb}} - x)(\mu_{\text{hyb}} - \lambda_p^{\text{out}}(x)), \end{aligned} \tag{4.3.61}$$

where $\alpha_{\text{hyb}}, \mu_{\text{hyb}}$ are determined such that

$$\frac{1}{\pi} \int_0^\pi dx \ \lambda_{x,p}^{\text{mod}}(x) = (1 - \eta_{\text{hyb}})\overline{n}. \tag{4.3.62}$$

The transmission rate is then again given by

$$R_{\text{hyb}} = \frac{1}{\pi} \int_0^\pi dx \ g\left(\overline{\nu}(x) - \frac{1}{2} \right) - g\left(\nu^{\text{out}}(x) - \frac{1}{2} \right), \tag{4.3.63}$$

with symplectic eigenvalue spectra as in (4.3.28).

Ansatz 3: Change of input

The third and last idea contains a different spectrum for the input. We know from the previous investigation that an optimal input has to diagonalize in the same basis as the noise Such an input, i.e. a multimode finite squeezed state, was defined in [24], that is

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}\begin{pmatrix} e^{r\Omega} & 0 \\ 0 & e^{-r\Omega} \end{pmatrix}, \tag{4.3.64}$$

where r denotes the squeezing parameter and the generator Ω reads

$$\Omega_{n\times n} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.3.65}$$

Since we know that the inverse of the environment covariance matrix will be a three diagonal matrix in the asymptotic limit $n \to \infty$ (see (2.3.12)), we state that

$$\lim_{n\to\infty}[\boldsymbol{\gamma}_{\text{env}}^{-1}, \Omega] = 0 = \lim_{n\to\infty}[\boldsymbol{\gamma}_{\text{env}}, \Omega], \tag{4.3.66}$$

where $[A, B] = AB - BA$ is the *commutator*[21] of two matrices A, B and where $\boldsymbol{\gamma}_{\text{env}}$ reads as defined in (4.3.1). As Ω is the generator of $\boldsymbol{\gamma}_{\text{in}}$ we know that therefore $\boldsymbol{\gamma}_{\text{in}}$ and $\boldsymbol{\gamma}_{\text{env}}$ commute as well. In the section of Toeplitz matrices we quoted a spectrum of a general three diagonal matrix (see 2.3.15) which we use to immediately write out the spectrum of (4.3.65) in the limit of $n \to \infty$, i.e.

$$\lambda^{(\Omega)}(x) = 2\cos(x), \quad x \in [0, \pi]. \tag{4.3.67}$$

With the latter we determine the spectrum of the input, that is

$$\lambda^{\text{in}}_{x,p}(x) = e^{\pm 2r\cos(x)}. \tag{4.3.68}$$

where the plus sign corresponds to the $\hat{x}$-quadrature, the minus sign to the $\hat{p}$-quadrature. Since the input is again generated by the cosine we remark again the symmetry

$$\lambda^{\text{in}}_{x}(x) = \lambda^{\text{in}}_{p}(\pi - x). \tag{4.3.69}$$

The fraction of energy used for entanglement (for given squeezing r) is again determine by the integral over the input spectrum, i.e.

$$\eta_{\text{exp}}(r) = \frac{1}{\pi\overline{n}}\left(\int_0^{\pi} dx\, e^{2r\,\cos(x)} - \frac{\pi}{2}\right) = \frac{1}{2\overline{n}}(I_0(2r) - 1), \tag{4.3.70}$$

[21] Two matrices A, B which *commute*, i.e. [A,B] = 0, diagonalize in the same basis.

where $I_0(x)$ denotes the modified Bessel function of first kind. Since we know that for a given $\overline{n} < \overline{n}_{\rm s}(\phi, N)$ we will not be able to modulate the full spectrum we define the modulation analog to the input in the classical case, i.e.

$$\lambda_x^{\rm mod}(x) = \Theta(x-\alpha)(\mu_r - (\lambda_{x,p}^{\rm in}(x) + \lambda_{x,p}^{\rm env}(x))), \tag{4.3.71}$$

where α, μ_r are determined such that

$$\frac{1}{\pi}\int_0^\pi \lambda_{x,p}^{\rm mod}(x) = (1-\eta_r)\overline{n}. \tag{4.3.72}$$

The optimal transmission rate for this input has to be determined by a maximization over the squeezing r or equivalently the fraction of entanglement η, that is

$$R_{\rm exp} = \max_{\eta\in[0,1]} \frac{1}{\pi}\int_0^\pi dx\; g\left(\overline{\nu}(x) - \frac{1}{2}\right) - g\left(\nu^{\rm out}(x) - \frac{1}{2}\right), \tag{4.3.73}$$

where the symplectic eigenvalue functions $\overline{\nu}(x), \nu^{\rm out}(x)$ read as in (4.3.28).

Best ansatz?

In the following we will try to sort out the best of the three presented ideas. Firstly, we compare all rates for a fixed input power $\overline{n}$ and noise N versus the correlation parameter ϕ. The latter is plotted in figure 4.19 for three different N. All curves correspond to the difference $R_\kappa - R_{\rm hyb}, R_\kappa - R_{\rm exp}$, respectively. We do not show the difference in rates $R_{\rm hyb} - R_{\rm exp}$ because it was more or less fluctuating around zero with some numerical error, thus we could not get useful information out of it. We confirm in the region of low correlations that all differences in $R_\kappa - R_{\rm hyb}$ and $R_\kappa - R_{\rm exp}$ are zero which simply corresponds to $\overline{n} \geq \overline{n}_{\rm s}(\phi, N)$ where all at least the rates using ansatz 1 and 2 coincide with the true capacity found in section 4.3.1. R_κ dominates the other two rates in the limit of high correlations for all N, whereas at $\phi \approx 1$ all rates seem to coincide. In the region of full correlations the differences become negligible which is due to the dominance of the noise (we assume a constant input power). For lower correlations R_κ seems to loose its benefit with respect to the other two with increasing noise N.
In a second comparison we plot in figure 4.20 the rate differences versus input power $\overline{n}$ for fixed correlation $\phi = 0.7$ and noise $N = 1/3$ (where for these parameters R_κ was dominating in figure 4.19). Clearly, we realize that it is not at all obvious to impose a "good" ad-hoc solution to the problem. However, we can conclude that a general property of the transmission rate (or the capacity) seems to be its robustness, since the difference in all rates is below a tenths of a bit for the given noise regions.

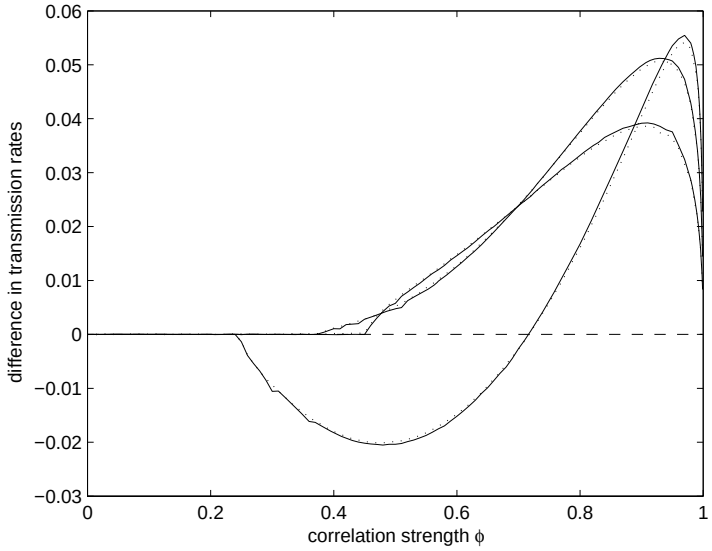

Figure 4.19: Difference in transmission rates vs. correlation parameter ϕ. The solid curves correspond to the difference $R_\kappa - R_{\rm hyb}$, the dotted curve to $R_\kappa - R_{\rm exp}$. The difference in $R_{\rm hyb} - R_{\rm exp}$ is not shown. The dashed curves shows the zero line. The curves with highest to lowest maximum correspond to $N = 1, 1/3, 0.1$. For all plots we fixed $\overline{n} = 1$, the maximum correlation value was $\phi = 0.999$.

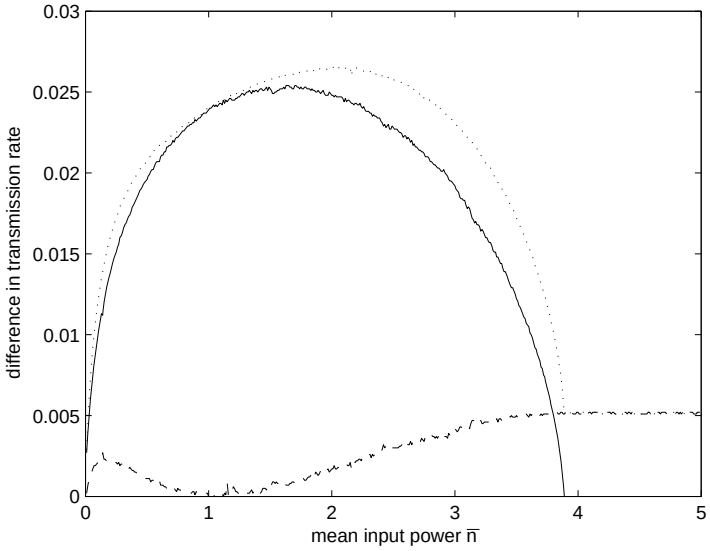

Figure 4.20: Difference in transmission rates vs. mean input power $\overline{n}$. The solid curve correspond to the difference $R_\kappa - R_{\rm hyb}$, the dotted curve to $R_\kappa - R_{\rm exp}$, the dashed curve to $R_{\rm hyb} - R_{\rm exp}$. For all plot we took $N = 1/3, \phi = 0.7$. The zero/constant drop is at $\overline{n} = \overline{n}_{\rm s}(\phi, N) = 3.89$.

Chapter 5

Conclusions and Outlook

In this work we discussed the classical Gaussian additive channel and a bosonic quantum memory channel with Gauss-Markov noise with input power constraint.

At first, we derived and characterized a covariance matrix which corresponds to such a noise process and we concluded that the matrix belongs to the class of Toeplitz matrices. As a result we concluded that a Gauss-Markov covariance matrix generates a stationary process, that is a process whose variance does not change when shifted in time.

The capacity of a classical Gaussian channel with the derived noise model was then determined in the full input power domain, given by the well-known water filling solution. In addition, we showed that the capacity is diverging in the limit of full noise correlations.

In the quantum part we treated at first the bosonic mono-modal channel with anisotropic noise. We imposed by physical reasoning the assumption that the best input covariance matrix diagonalizes in the same basis as the noise covariance matrix. Furthermore, we showed that the quantum input has to be squeezed in a way such that the anisotropy of the noise is matched in order to achieve the optimal transmission rate. Moreover, we determined a power threshold above which the best overall modulated output is a thermal state. We showed that this is achieved by a water filling solution for the classical modulation with respect to both quadrature outputs. We saw that as a consequence the quadrature which experiences the stronger noise will no longer be modulated when the input power is below the threshold. An analytical solution could be provided above the threshold, whereas in the low power regime we introduced an approximative linear solution and showed that the difference to the numerically obtained capacity is less than a percent for the assumed noise strengths.

The results of the one-mode case were then extended to the bosonic memory channel with Gauss-Markov noise. At first, we realized that the problem of finding the capacity could be treated in the basis where the noise covariance matrix is diagonal and which corresponds to an uncorrelated environment. Afterwards, we extended the results of the one-mode case, where similarly the optimal input was assumed to be diagonal (and based on that the modulation as well) in the same basis as the noise covariance matrix. For an input power which is above a certain threshold, we could show that, analogous to the one dimensional case, the input of each a channel requires a certain degree of squeezing

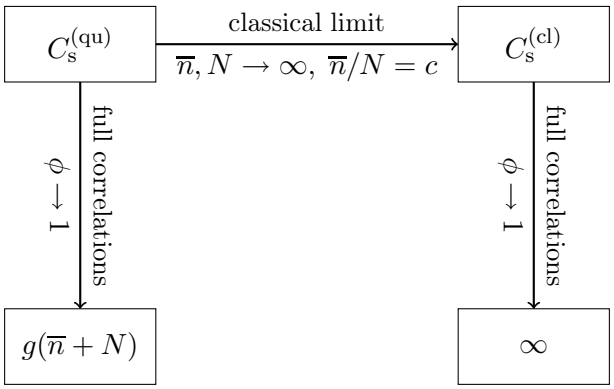

Figure 5.1: Limit relations between the capacity of the bosonic channel and the classical Gaussian channel (both with Gauss-Markov memory).

matching the anisotropy of its noise. Hence, when the noise is translated back to its correlated form we see that a certain degree of entanglement between the input modes is necessary to achieve the capacity. The best modulation of the input signal was furthermore determined by the classical water filling solution, whereas in the quantum case the output spectrum played the role of the environment of the classical case.
In the limit of full correlation we discovered in contrary to the classical case a convergent behavior of the capacity.

However, in this limit the cost of power which is required to prepare the quantum input state is divergent and hence, the ratio of (classical) modulation power to noise does not diverge, as in the classical case. In the limit of high input power and noise with constant signal to noise ratio (i.e. the classical limit) we showed that the capacity (above threshold) tends to the capacity of the classical Gaussian channel (the limit behavior of the capacity in the quantum and classical case is summarized in figure 5.1). Furthermore, we concluded that the capacity of the bosonic channel is in general always smaller than the capacity of the classical Gaussian channel. This is again due to the cost of power for the preparation of the entangled input state.
In the low power regime an analytical solution could not be derived. However, several approximative approaches were presented and compared.

The techniques we developed in this work were in addition applied to a non-Markovian model where the capacity was determined above a certain power threshold.

For both cases, the mono-modal bosonic and the infinite dimensional bosonic channel, an exact solution of the capacity in the low input power regime is missing. Once the one-dimensional case is fully understood the extension to the asymptotic limit is an interesting problem for future studies.

Appendix A

Entropy of Gaussian states

According to [16][V.§1] the Gaussian density operator ρ_0, representing the thermal state of the harmonic oscillator (with zero mean) has the spectral decomposition

$$\rho_0 = \frac{1}{N+1}\sum_{m=0}^{\infty}\left(\frac{N}{N+1}\right)^m |m\rangle\langle m|, \tag{A.0.1}$$

where $\{|m\rangle\}, m = 0, 1, ..., \infty$ stands for the Fock basis as introduced in section (4.1.3). By inserting (A.0.1) into (4.2.6), that is

$$\begin{aligned}
-\mathrm{Tr}\left(\rho\log_2\rho\right) &= -\mathrm{Tr}\left[\frac{1}{N+1}\sum_{m=0}^{\infty}\left(\frac{N}{N+1}\right)^m |m\rangle\langle m|\log_2\left(\frac{1}{N+1}\sum_{m=0}^{\infty}\left(\frac{N}{N+1}\right)^m |m\rangle\langle m|\right)\right] \\
&= -\mathrm{Tr}\left[\frac{1}{N+1}\sum_{m=0}^{\infty}\left(\frac{N}{N+1}\right)^m \log_2\frac{1}{N+1}\left(\frac{N}{N+1}\right)^m |m\rangle\langle m|\right] \\
&= -\frac{1}{N+1}\sum_{k,m=0}^{\infty}\left(\frac{N}{N+1}\right)^m \log_2\frac{1}{N+1}\left(\frac{N}{N+1}\right)^m \underbrace{\langle k|m\rangle}_{=\delta_{km}}\underbrace{\langle m|k\rangle}_{=\delta_{mk}} \\
&= -\frac{1}{N+1}\log_2\left(\frac{1}{N+1}\right)\underbrace{\sum_{m=0}^{\infty}\left(\frac{N}{N+1}\right)^m}_{=N+1} \\
&\quad -\frac{1}{N+1}\log_2\left(\frac{N}{N+1}\right)\underbrace{\sum_{m=0}^{\infty} m\left(\frac{N}{N+1}\right)^m}_{=N(N+1)} \\
&= (N+1)\log_2(N+1) - N\log(N),
\end{aligned} \tag{A.0.2}$$

we conclude that

$$S(\rho) = g(N), \tag{A.0.3}$$

where

$$g(x) = \begin{cases} (x+1)\log_2(x+1) - x\,\log_2(x), & x > 0 \\ 0 & x = 0 \end{cases} \tag{A.0.4}$$

is a continuous concave monotonously increasing function for $x > 0$.
In addition we state from [16][V.§1], that the operator (A.0.1) has the maximum entropy among all density operators ρ satisfying

$$\mathrm{Tr}\left(\rho \hat{a}^\dagger \hat{a}\right) \leq N. \tag{A.0.5}$$

Since $\hat{a}^\dagger\,\hat{a}$ is just the number operator the quantity N is just the number of photons of the system.
For a general n-mode Gaussian state with covariance matrix $\boldsymbol{\gamma}$ the eigenvalue equation

$$\det\left(\boldsymbol{\gamma} - i\nu\boldsymbol{J}\right) = 0, \tag{A.0.6}$$

where $\boldsymbol{J}$ denotes the symplectic matrix (4.1.24), is solved by a set $\pm\{\nu_j\}$ which we call in the following the *symplectic eigenvalues* of $\boldsymbol{\gamma}$. Hence, if $\widetilde{\boldsymbol{\gamma}}$ is the matrix $\boldsymbol{\gamma}$ in diagonal form, then we find that

$$\widetilde{\boldsymbol{\gamma}} = \mathrm{diag}(\nu_1, ..., \nu_n; \nu_1, ...\nu_n). \tag{A.0.7}$$

For each matrix

$$\boldsymbol{\gamma}_j = \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_j \end{pmatrix}, \tag{A.0.8}$$

we the relation between the symplectic eigenvalue ν_j and the number of photons N_j reads

$$\nu_j = N_j + \frac{1}{2},$$

with $N_j \geq 0$ and $\nu_j \geq 1/2$ (due to the Heisenberg uncertainty principle (4.1.27)). If we assume now a n-mode thermal state ρ_{th}, i.e. a state with a covariance matrix $\widetilde{\boldsymbol{\gamma}}$ we find with the help of the subadditivity property of the entropy (4.2.9) the relation:

$$S(\rho_{\mathrm{th}}) = S(\rho_{\mathrm{th},1}) + S(\rho_{\mathrm{th},2}) + ... + S(\rho_{\mathrm{th,n}}), \tag{A.0.9}$$

where $\rho_{\mathrm{th}} = \rho_{\mathrm{th},1} \otimes \rho_{\mathrm{th},2} \otimes ... \otimes \rho_{\mathrm{th,n}}$. Therefore the total entropy of ρ_{th} reads:

$$S(\rho_{\mathrm{th}}) = \sum_{j=1}^{n} g(N_j) = \sum_{j=1}^{n} g\left(|\nu_j| - \frac{1}{2}\right), \tag{A.0.10}$$

where the sum of the latter runs just from 1 to n (with the abs-function) and not $2n$ since all ν_j's are doubly degenerate.
Now we extend that result to a general Gaussian state with a possibly non-diagonal covariance matrix $\boldsymbol{\gamma}$. We define the function

$$G(x^2) = g\left(x - \frac{1}{2}\right) \tag{A.0.11}$$

and rewrite the second equation of (A.0.10) as

$$\sum_{j=1}^{n} g\left(|\nu_j| - \frac{1}{2}\right) = \sum_{j=1}^{n} G(\nu_j^2) = \frac{1}{2}\mathrm{Tr}\left(G(-(\boldsymbol{J}^{-1}\widetilde{\boldsymbol{\gamma}})^2)\right).$$

By applying passive (trace preserving) symplectic transformations S we may write

$$\widetilde{\boldsymbol{\gamma}} = S\boldsymbol{\gamma}S^{\mathsf{T}}.$$

Furthermore, we state that

$$\boldsymbol{J}^{-1}\widetilde{\boldsymbol{\gamma}} = (S\boldsymbol{J}S^{\mathsf{T}})^{-1}S\boldsymbol{\gamma}S^{\mathsf{T}} = (S^{\mathsf{T}})^{-1}(\boldsymbol{J}^{-1}\boldsymbol{\gamma})S^{\mathsf{T}},$$

and

$$G(-(\boldsymbol{J}^{-1}\widetilde{\boldsymbol{\gamma}})^2) = (S^{\mathsf{T}})^{-1}G(-(\boldsymbol{J}^{-1}\boldsymbol{\gamma})^2)S^{\mathsf{T}}.$$

Thus, due to cyclic permutations in the trace, we obtain the final expression for the entropy of the general Gaussian state ρ with covariance matrix $\boldsymbol{\gamma}$ with symplectic spectrum $\{\nu_j\}$, that is

$$S(\rho) = \frac{1}{2}\mathrm{Tr}\left(G(-(\boldsymbol{J}^{-1}\boldsymbol{\gamma})^2)\right) = \sum_{j=1}^{n} g\left(|\nu_j| - \frac{1}{2}\right). \tag{A.0.12}$$

We conclude that passive symplectic transformations applied on the covariance matrix do not change the entropy of the state.

Appendix B

Lossy channel with non-Markovian memory

We will investigate now a lossy bosonic channel with Non-Markovian model, which was introduced in [24] and apply the same steps as for the Markovian model such as the concept of quantum water filling to find the true capacity above a certain threshold. The covariance matrix for the environment is given by

$$\gamma_{\text{env}} = \frac{1}{2}\begin{pmatrix} e^{w\Omega} & 0 \\ 0 & e^{-w\Omega} \end{pmatrix}, \tag{B.0.1}$$

where $w \in \mathbb{R}$, Ω as in (4.3.65). Thus, as the inverse of this matrix is not three-diagonal the generated stochastic process cannot be a Markov process. The input reads as in (4.3.64), i.e.

$$\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}\begin{pmatrix} e^{r\Omega} & 0 \\ 0 & e^{-r\Omega} \end{pmatrix}, \tag{B.0.2}$$

with squeezing parameter r. In addition the proposed modulation is defined by

$$\gamma_{\text{mod}} = \frac{n(1-\eta_{\text{exp}}(r))\overline{n}}{\text{Tr}(\exp(y\Omega))}\begin{pmatrix} e^{y\Omega} & 0 \\ 0 & e^{-y\Omega} \end{pmatrix}, \tag{B.0.3}$$

with the unconstrained correlation strength $y \in \mathbb{R}$ and the fraction of energy used for entanglement by (B.0.2) $\eta_{\text{exp}}(r)$, which will be defined shortly. Here, in addition a beam splitter interaction with the environment is introduced. Denoting the beam splitter transmittivity by χ, the covariance matrices than read

$$\begin{aligned} \boldsymbol{\gamma}_{\text{out}} &= \chi\boldsymbol{\gamma}_{\text{in}} + (1-\chi)\boldsymbol{\gamma}_{\text{env}}, \\ \overline{\boldsymbol{\gamma}} &= \boldsymbol{\gamma}_{\text{out}} + \chi\boldsymbol{\gamma}_{\text{mod}}. \end{aligned} \tag{B.0.4}$$

In [24] a numerical optimization over the squeezing parameter r and classical correlation parameter y was performed. However, we already know that a water filling modulation is optimal in the whole power regime. As for the Gauss-Markov model which is discussed in this work we will hence also be able to find the true capacity above a certain threshold

value of the input power. We follow now the same steps as in section 4.3.1. Let us assume that the given input power is sufficient such that $\forall x$ the input will be squeezed "noise-like" and that there is enough power left, such that the full system can be modulated. At first, we state the input (B.0.2) is actually optimal for $r = w$ (above the threshold), since its spectrum is than identical to the one of the noise, i.e.

$$\lambda^{\text{in}}_{x,p}(x) = \frac{1}{2} e^{\pm 2w \cos(x)} = \lambda^{\text{env}}_{x,p}(x), \tag{B.0.5}$$

with plus for the $\hat{x}$-quadrature, minus for the $\hat{p}$-quadrature. We remark that due to the beam splitter interaction in the region above the threshold the spectrum of the output is, as well as the input, identical to the environment:

$$\lambda^{\text{out}}_{x,p}(x) = \chi \lambda^{\text{in}}_{x,p}(x) + (1-\chi)\lambda^{\text{env}}_{x,p}(x) = \lambda^{\text{env}}_{x,p}(x). \tag{B.0.6}$$

The fraction of energy used for entanglement for given w reads

$$\eta_{\text{exp}}(w) = \frac{1}{2\overline{n}}(I_0(2w) - 1), \tag{B.0.7}$$

which is identical to (4.3.70) where we already used such an input. If we apply again the water filling solution for the modulation, i.e.

$$\lambda^{\text{mod}}_{x,p}(x) = \mu - \lambda^{\text{out}}_{x,p}(x) = \mu - \lambda^{\text{env}}_{x,p}(x) \tag{B.0.8}$$

with

$$\frac{1}{\pi} \int_0^\pi dx \; \lambda^{\text{mod}}_{x,p} = (1 - \eta_{\text{exp}}(w))\overline{n}, \tag{B.0.9}$$

we retrieve an inequality similar to the one in (4.3.22), which needs to be fulfilled in order to modulate a the full spectrum, whereas now the beam splitter transmittivity χ is taken into account:

$$\begin{aligned}
&\chi\pi(1-\eta_{\text{exp}}(w))\overline{n} \geq \pi\lambda^{\text{out}}_x(0) - \int_0^\pi dx \; \lambda^{\text{out}}_x(x) \\
\Leftrightarrow\; &\chi\overline{n} - \chi\eta_{\text{exp}}(w)\overline{n} \geq \frac{1}{2}e^{2w} - \chi \underbrace{\frac{1}{\pi}\int_0^\pi dx \; \lambda^{\text{in}}_x(x)}_{=\eta_{\text{exp}}(w)\overline{n}+\frac{1}{2}} - (1-\chi) \underbrace{\frac{1}{\pi}\int_0^\pi dx \; \lambda^{\text{env}}_x(x)}_{=\frac{1}{2}I_0(2w)} \\
\Leftrightarrow\; &\overline{n} \geq \frac{1}{2}\frac{1}{\chi}\left(e^{2w} - (1-\chi)I_0(2w)\right) - \frac{1}{2} \equiv \overline{n}_{\text{exp,s}}(w).
\end{aligned} \tag{B.0.10}$$

where $I_0(x)$ denotes the modified Bessel function of first kind. With $\overline{n}_{\text{exp,s}}$ we characterized for the given model analog to $\overline{n}_{\text{s}}(\phi)$ the input power which is needed to modulate the

full spectrum, assuming we match for the whole input spectrum the noise anisotropy. Furthermore, we can derive an explicit expression for the water filling level, i.e.

$$\begin{aligned}\mu &= \chi(1-\eta_{\text{exp}}(w))\overline{n} + \frac{1}{\pi}\int_0^\pi dx\ \lambda_x^{\text{out}}(x) \\ &= \chi(\overline{n}+\frac{1}{2}) + \frac{1-\chi}{2}I_0(2w),\end{aligned} \tag{B.0.11}$$

which reads equivalent to (4.3.25). If $\overline{n} \geq \overline{n}_{\text{s,exp}}$ We assume that the modulated output of all channels is identical, i.e.

$$\overline{\lambda}_x(x) = \overline{\lambda}_p(x) = \mu, \quad \forall x. \tag{B.0.12}$$

Hence, using (4.3.27) we find the capacity of the system, if $\overline{n} \geq \overline{n}_{\text{exp,s}}(w)$, that is

$$\begin{aligned}C_{\text{exp,s}} &= \frac{1}{\pi}\int_0^\pi dx\ g(\overline{\nu}(x)-\frac{1}{2}) - g(\nu^{\text{out}}(x)-\frac{1}{2}), \quad \overline{n} \geq \overline{n}_{\text{exp,s}}(w) \\ &= g(\mu-\frac{1}{2}) - \frac{1}{\pi}\int_0^\pi dx\ g\left(\sqrt{\frac{1}{2}e^{2w\cos x}\frac{1}{2}e^{-2w\cos x}} - \frac{1}{2}\right), \quad \overline{n} \geq \overline{n}_{\text{exp,s}}(w) \\ &= g\left(\chi(\overline{n}+\frac{1}{2}) + \frac{1-\chi}{2}I_0(2w) - \frac{1}{2}\right), \quad \overline{n} \geq \overline{n}_{\text{exp,s}}(w),\end{aligned} \tag{B.0.13}$$

Interestingly, due to the beam splitter interaction the second term which contains only environment terms completely disappears and we arrive therefore at a full analytical expression (unlike the integral in (4.3.30)).
Below the threshold a numerical optimization over the squeezing parameter r with the input spectrum

$$\lambda_{x,p}^{\text{in}} = \frac{1}{2}e^{\pm 2r\cos(x)}, \tag{B.0.14}$$

is performed, with plus sign for the $\hat{x}$-quadrature and minus sign for the $\hat{p}$-quadrature. The fraction of power used for entanglement reads as in (4.3.70) but is now a function of r instead of w, i.e. $\eta_{\text{exp}}(r)$. As we know that we can no longer modulate the full spectrum we define the modulation by

$$\begin{aligned}\lambda_x^{\text{mod}}(x) &= \Theta(x-\alpha)(\mu - \chi\lambda_x^{\text{in}}(x) - (1-\chi)\lambda_x^{\text{env}}(x)), \\ \lambda_p^{\text{mod}}(x) &= \Theta(\pi-x-\alpha)(\mu - \chi\lambda_p^{\text{in}}(c) - (1-\chi)\lambda_p^{\text{env}}(x)),\end{aligned} \tag{B.0.15}$$

where $\mu = \chi\lambda_x^{\text{in}}(\alpha) + (1-\chi)\lambda_x^{\text{env}}(\alpha)$ and α is determined such that

$$\frac{1}{\pi}\int_0^\pi dx\ \lambda_{x,p}^{\text{mod}}(x) = (1-\eta_{\text{exp}}(r))\overline{n}. \tag{B.0.16}$$

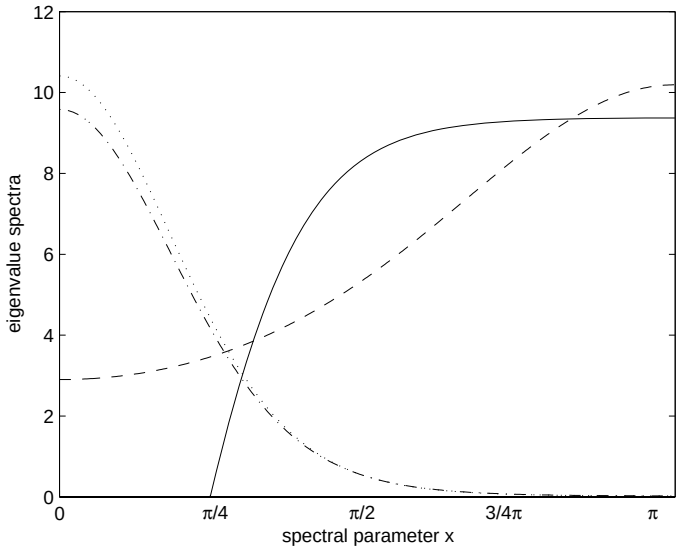

Figure B.1: Input and modulation eigenvalue spectra using either the modulation (B.0.3) (dashed-dotted and the dashed line) or the quantum water filling method (dotted and the solid line) vs. the spectral parameter x. All curves show spectra of the x quadrature. The optimal squeezing with water filling modulation is $r^* = 1.52$. The optimal squeezing and correlation factor with modulation (B.0.3) read $(r^*, y^*) = (1.48, -0.31)$. Here we took $\overline{n} = 8, w = 1.6, \chi = 0.7$.

The capacity then reads

$$C_{\text{exp,is}} = \max_{\eta_{\text{exp}}(r)\in[0,1]} \frac{1}{\pi} \int_0^\pi dx \; g(\overline{\nu}(x) - \frac{1}{2}) - g(\nu^{\text{out}}(x) - \frac{1}{2}), \quad \overline{n} < \overline{n}_{\text{exp,s}}, \tag{B.0.17}$$

where the symplectic eigenvalue functions $\overline{\nu}(x), \nu^{\text{out}}(x)$ read as in (4.3.28).

In figure B.1 we plot the input and modulation spectra which result from a numerical optimization over r, y using modulation (B.0.3) and a numerical optimization over r (or $\eta_{\text{exp}}(r)$) using water filling for modulation in an area where $\overline{n} < \overline{n}_{\text{exp,s}}(w)$. The numerical optimum (which corresponds to the numerical capacity) uses slightly more squeezing than when using modulation (B.0.3). The improvement of the transmission rate using the water filling modulation is demonstrated in figure B.2, where the optimal transmission rate for both ways of modulation is shown versus the memory parameter w. The dotted line shows the theoretical curve (up to $\overline{n} = \overline{n}_{\text{exp,s}}$) which does not connect precisely with the solid curve (which denotes the optimization with water filling modulation) which may be due to numerical errors.

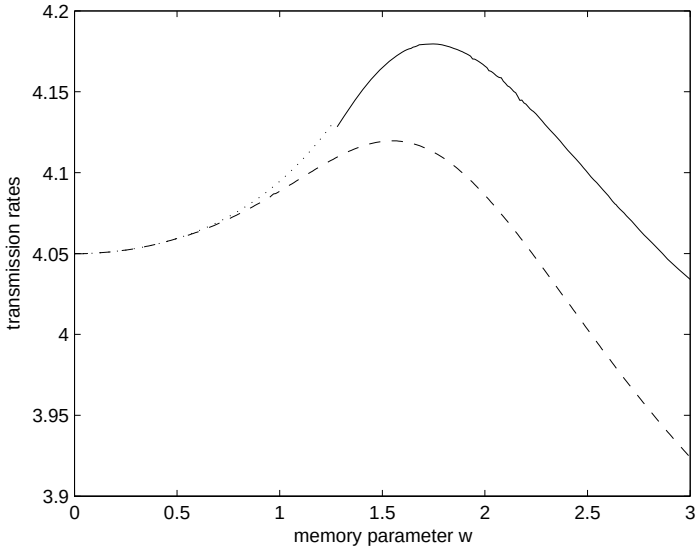

Figure B.2: Optimal transmission rates vs. noise parameter s. The dotted curve (up to $\overline{n} = \overline{n}_{exp,s}(s)$ shows the true capacity $C_{\text{exp,s}}$. The dashed line denotes the transmission rate using modulation (B.0.3), where the solid line shows the numerical capacity $C_{\text{exp,is}}$ (where $\overline{n} < \overline{n}_{exp,s}(s)$) using the quantum water filling method (4.3.71). Here we took $\overline{n} = 8, \chi = 0.7$.

Bibliography

[1] Arvind, B. Dutta, and R. Simon. The real symplectic groups in quantum mechanics and optics. *arXiv:quant-ph/9509002v3*, 2005.

[2] C. H. Benett and G. Brassard. Proceedings of ieee international conference. In *Quantum cryptography: Public-key distribution and coin tossing*, pages 175–179, 1984.

[3] S. L. Braunstein and P. v. Loock. Quantum information with continuous variables. *Rev. Mod. Phys.*, 77:519, 2005.

[4] A. Böttcher and S. Grudsky. *Spectral Properties of Banded Toeplitz Matrices*. Siam, 2005. ISBN 0-89871-599-7.

[5] N. J. Cerf, J. Clavareau, J. Roland, and C. Macchiavello. Quantum entanglement enhances the capacity of bosonic channels with memory. *Phys. Rev. A*, 72:1–5, 2005.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, inc., 1991. ISBN 0-471-06259-6.

[7] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777, 1935.

[8] S. N. Ethier and T. G. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics, Utah, USA, Wisconsin, USA, 1986. ISBN 0-471-08186-8. 156 pp.

[9] R. Garcia-Patron and N. J. Cerf. Unconditional optimality of gaussian attacks against continuous-variable quantum key distribution. *Phys. Rev. Lett.*, 97:190503, 2006.

[10] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Phys. Rev. Lett.*, 92(2), 2004.

[11] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 1980. ISBN 0-12-294760-6.

[12] M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2:155–239, 2006.

[13] F. Grosshans and P. Grangier. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.*, 88:057902, 2002.

[14] A. Gut. *An Intermidiate Course in Probability*. Springer Verlag, Uppsala, Sweden, 1995. ISBN 0-387-94507-5. 122 pp.

[15] T. Hiroshima. Additivity and multiplicativity properties of some gaussian channels for gaussian inputs. *Phys. Rev. A*, 73:012330, 2006.

[16] A. S. Holevo. Coding theorems for quantum channels. *arXiv:quant-ph/9809023v1*, 1998.

[17] A. S. Holevo, M. Sohma, and O. Hirota. Capacity of quantum gaussian channels. *Phys. Rev. A*, 59:1820–1828, 1999.

[18] M. Keyl. Fundamentals of quantum information theory. *arXiv:quant-ph/0202122v1*, 2002.

[19] D. Kretschmann and R. Werner. Quantum channels with memory. *Phys. Rev. A*, 72: 1–19, 2005.

[20] M. Lévy, N. J. Cerf, and G. V. Assche. Quantum distribution of gaussian keys using squeezed states. *Phys. Rev. A*, 63:052311, 2001.

[21] C. Macchiavello and G. Palma. Entanglement-enhanced information transmission over a quantum channel with correlated noise. *Phys. Rev. A*, 65:1–4, 2002.

[22] S. Mancini. Models for quantum memory channels. *Journal of Physics: Conference Series*, 36:121–125, 2006.

[23] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2007. ISBN 978-0-521-63503-5.

[24] O. Pilyavets, V. Zborovskii, and S. Mancini. Lossy bosonic quantum channel with non-markovian memory. *Phys. Rev. A*, 77:1–8, 2008.

[25] E. Schrödinger. Discussion of probability relations between separated systems. *Proceedings of the Cambridge Philosophical Society*, 31:555–563, 1935.

[26] B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997.

[27] P. Shor. Proc. 35nd annual symposium on foundations of computer science. In *Algorithms for quantum computation: Discrete logarithms and factoring*, pages 124–134. IEEE Computer Society Press, 1994.

[28] R. Ursin et al. Entanglement-based quantum communication over 144 km. *Nature Physics*, 3:481–486, 2007.

[29] D. Walls and G. Millburn. *Quantum Optics*. Springer, 1995. ISBN 3-540-58831-0.

[30] C. Weedbrook et al. Quantum cryptography without switching. *Phys. Rev. Lett.*, 93: 170504, 2004.

[31] S. Wiesner. Conjugate coding. *ACM SIGACT News*, 15:78–88, 1983.

[32] Wikipedia.org. General number field sieve, dec 2008.

[33] B. J. Yen and J. H. Shapiro. Multiple-access bosonic communications. *Phys. Rev. A*, 72:062312, 2005.