Classical Communication With Displacement Receivers via Noisy Quantum Channels

Komunikacja klasyczna przez zaszumione kanały kwantowe z użyciem niekonwencjonalnych odbiorników koherentnych

Ludwig Kunz

PhD Thesis

University of Warsaw
Faculty of Physics
Poland

Abstract

In this dissertation we investigate the capabilities of a generalized version of the displacement receiver (DR) in the context of classical communication with coherent states over noisy channels. The receiver itself consists of a displacement operation followed by a photon-number resolving detector and enables flexible strategies to cope with different communication tasks at hand, e.g. different types of noise or encodings. The physical channel over which the states are transferred is modeled as a quantum channel and all scenarios are analyzed on the basis of the error probability as well as mutual information. When communication takes place over a lossless and noiseless channel with information encoded in a binary alphabet of coherent states, the optimal measurement is known to be the Dolinar receiver and the displacement receiver provides the same scaling of the error probability with the average energy of the alphabet.

Assuming an infinite photon-number resolution (PNR) the DR can interpolate between operating as a photon counter and a homodyne detector by controlling the displacement. Also, the Dolinar receiver can be understood as a series of DRs using the information provided by its predecessors. Therefore, we can use the theoretical concept of the DR as a basic model to understand advantages and disadvantages of the various receivers.

Imperfections in different parts of a DR have to be taken into account for a realistic theoretical model. We analyze the effects of non-unit visibility, dark counts, afterpulsing and limited detection efficiency on the DR. Non-unit visibility imposes the most severe restriction, but all of them can be mitigated by PNR and optimization of the communication strategy. Already current or near term technology allows for sub-shot noise detection and the DR is therefore an interesting candidate for optical communication.

Apart from standard classical communication the DR can also be employed to show superadditivity of accessible information. In the limit of vanishing average energy the DR approaches the optimal individual measurement. Based on previous proposals we show that there exist several setups for achieving a superadditive advantage of accessible information with the DR in the low power limit.

All of the applications rely on a near-optimal performance of the receiver. However, displacement receivers in particular are susceptible to noise in the channel that may render them useless. The particular impact depends highly on the type of noise in question.

First, we analyze a scenario where thermal noise is present in the communication channel. Here, the DR can be made more robust by increasing the PNR of the detector. Only for large amounts of thermal background photons the DR gets overwhelmed and conventional detection provides better results. In the photon-starved regime pulse-position modulation is a commonly employed type of encoding and also in this scenario the DR can offer an improvement in performance when thermal noise is present.

Another type of noise is phase noise which directly affects the phase variable of the state.

As a physical motivation, we develop a full quantum model of lossy state transmission through a nonlinear medium and show that quantum effects are indeed present. Due to the interplay between distributed losses and the nonlinearity a fundamental type of nonlinear phase noise arises where the strength of the phase noise is closely tied to the energy of the state. We analyze the effects of communication over such a nonlinear channel and the impact on the generation of squeezed states in a nonlinear Kerr medium. Furthermore we study how both linear and nonlinear phase noise impact binary alphabets and possible receivers. Conventionally implemented DRs are especially vulnerable to phase noise and loose their advantage over conventional receivers rather quickly with rising strength of the noise. However, the DR is able to retain its superior performance by optimizing its displacement at the same time as the alphabet. The particular strategy depends on the constraint placed on the energy of the alphabet and we consider a constraint on the average and maximum energy. Notably, the DR approaches the Helstrom bound for very strong phase noise.

Dedicated to
the memory of Darcy

Acknowledgements

During the course of my Phd studies many people helped and supported me for which I am eternally grateful. In particular, I want to give thanks to:

- my wife, who supported and stood by me in good and in hard times.
- my parents, who always supported me.
- Konrad Banaszek, my supervisor, whose guidance and discussions enabled me to expand my abilities and knowledge to new areas.
- Marcin Jarzyna, who had to share an office with me for all these years and was always ready for a discussion and to offer help.

List of Publications

The presented work is largely based on results that were previously published in the following publications:

- L. Kunz, M. G. A. Paris, and K. Banaszek.
  Noisy propagation of coherent states in a lossy Kerr medium.
  Journal of the Optical Society of America B, 35(2):214, jan 2018.
  doi: 10.1364/josab.35.000214.

- K. Banaszek, L. Kunz, M. Jarzyna, and M. Jachura.
  Approaching the ultimate capacity limit in deep-space optical communication.
  In Free-Space Laser Communications XXXI, page 109100A, San Francisco, CA, mar 2019. SPIE.
  doi: 10.1117/12.2506963.

- K. Banaszek, W. Zwolinski, L. Kunz, and M. Jarzyna.
  Photon efficiency limits in the presence of background noise.
  In Proceedings of the 2019 IEEE International Conference on Space Optical Systems and Applications (ICSOS), Portland, OR, USA, 2019.
  URL https://arxiv.org/abs/1911.06772.

- M. T. DiMario, L. Kunz, K. Banaszek, and F. E. Becerra.
  Optimized communication strategies with binary coherent states over phase noise channels.
  npj Quantum Information, 5(1):65, jul 2019.
  doi: 10.1038/s41534-019-0177-4.

- L. Kunz, M. T. DiMario, F. E. Becerra and K. Banaszek.
  Beating the standard quantum limit for binary constellations in the presence of phase noise.
  21st International Conference on Transparent Optical Networks (ICTON), 1-6, jul 2019
  doi: 10.1109/icton.2019.8840431

- L. Kunz, M. Jarzyna, W. Zwolinski, and K. Banaszek.
  Low-cost limit of classical communication with restricted quantum measurements.
  New Journal of Physics, 22(4): 043010, apr 2020.
  doi: 10.1088/1367-2630/ab7bd3.

- K. Banaszek, L. Kunz, M. Jachura, and M. Jarzyna.
  Quantum limits in optical communications.
  Journal of Lightwave Technology, 38(10):2741–2754, may 2020.
  doi: 10.1109/jlt.2020.2973890.

Contents

1 Introduction

The need for a fast exchange of information over long distances has always sparked the creation of new ideas. The invention of the telegraph and the first transatlantic telegraph line marked the dawn of a new era of fast global communication. Milestones like the telephone, radio-communication or the internet were significant advances to provide easy access to communication for everyone. To transmit information over very long distances through the atmosphere electromagnetic waves at radio-frequencies turned out to be well suited, since they are much less disturbed by atmospheric effects than other frequencies. However, their information transfer is limited by their low carrier frequency and therefore low bandwidth. Due to the unyielding growth in required capacity, optical communication technologies started to become more and more relevant. With the invention of the laser in the 1960s a suitable optical source arrived. The unrivaled high bandwidth and data rate of optical communication soon established them as the backbone of modern telecommunication and since the 1970s the worldwide optical fibre network saw a rapid growth. Despite big advances in technology it is soon to be expected that the demand overtakes the available capacity [Essiambre and Tkach 2012].

Another growing trend in optical communication is satellite communication. While traditionally radio-frequencies are used, their limitations become more and more apparent with increasing demand on high data rates. Optical frequencies offer a solution to these problems both in near-Earth communication as well as deep-space missions. With their high bandwidth and the possibility to limit diffraction losses by sending a low-divergence beam they promise to solve the capacity problem [Hemmati et al. 2011]. However, for a widespread acceptance the arising challenges have to be overcome first. Optical states require a high quality receiver and are very sensitive to imperfections on the employed antennas. Furthermore, optical frequencies are much more prone to atmospheric conditions than radio-frequencies. Despite these difficulties the feasibility of optical satellite communication already has been proven in a number of experiments [Biswas et al. 2006; Smith 2006; Smutny et al. 2008; Hemmati et al. 2011; Chorvalli et al. 2017; Günthner et al. 2017].

The fundamental building blocks necessary for successful communication are the preparation of signal states and the receiver to retrieve information from the signal after transmission. Already in the early development of optical communication it was recognized that quantum effects play an integral role in defining the limits of what is physically achievable [Gordon 1962].
On the one hand, non-classical states of light, such as Fock or squeezed states, can be used to push communication beyond the classical limits [Caves and Drummond 1994;

Yuen and Shapiro 1978]. However, coherent states, the quantum analog of a classical state, are widely used in optical telecommunication. They remain a good candidate for communication as they prove optimal or near-optimal in many scenarios and are readily produced by a laser [Kikuchi 2016]. Furthermore, coherent states are highly useful for quantum communication such as quantum key distribution [Silberhorn et al. 2002; Gisin et al. 2002; Grosshans et al. 2003; Arrazola and Lütkenhaus 2014].
On the other hand, quantum mechanics offers new prospects for the detection of light which go beyond the limitations of commonly used techniques such as homodyne detection [Shapiro 1985].

To explore such possibilities it is necessary to find a suitable figure of merit for quantifying the performance. Additionally, the effects imposed on the signal by traveling through a medium have to be understood and addressed to obtain the best possible result.
The information theoretical approach to communication came into existence with the seminal work of Shannon [1948]. He adopted the concept of entropy from thermodynamics to rigorously quantify classical information which sparked a whole new way of thinking about communication. Crucially, Shannon connected mutual information to the channel capacity and, most notably, showed contrary to the common understanding that for non-trivial channels there always exists a strategy such that the transmission takes place without error for long messages even over noisy communication channels. Based on these results capacities of several fundamental types of communication channels have been established, among them the additive white Gaussian noise channel whose capacity is given by the Shannon-Hartley theorem [Shannon 1949].
Despite its great success, the Shannon-Hartley theorem has its shortcomings. It only applies to channels that involve linear physical media causing additive noise. In particular with the advent of optical fibre communication, however, nonlinear channels became more and more common [Essiambre et al. 2010; Essiambre and Tkach 2012]. A single-mode optical fibre is a Kerr nonlinear medium causing a variety of nonlinear effects that are not covered by models established so far. Only decades after Shannon's work nonlinear effects started to be taken into account in various approaches [Tang 2001; Mitra and Stark 2001; Wegener et al. 2004; Ellis et al. 2010; Turitsyn and Turitsyn 2012; Temprana et al. 2015; Sorokina and Turitsyn 2014]. All show that the capacity scales drastically different with the available energy compared to the previous results on linear channels such that new limitations on the information transmission arise. A second shortcoming of the Shannon-Hartley theorem stems from a limited class of measurements considered in its derivation.
Even though the theory of quantum mechanics was developed many years before information theory and already utilized entropy as an information measure of quantum states [von Neumann 1932], the theory of transmitting classical information via quantum channels was only developed much later by Holevo [1973]. In his ground-breaking work Holevo worked out key-differences between classical and quantum communication theory. The latter allows for new types of measurements and imposes different limitation on what is possible as many effects can not be described classically. In particular, quantum mechanics gives rise to an effect called superadditivity of mutual information

[Holevo 1973; Sasaki et al. 1998]. Essentially, more information can be retrieved from a collective measurement performed on the output of many channel uses than from the same number of uses measured individually. However, finding concrete measurement schemes that achieve superadditivity remains highly elusive and only few are known [Guha 2011; Wilde et al. 2012].
Because of quantum effects like superadditivity, it proved to be challenging to extend many results of classical information theory to the realm of quantum mechanics. A quantum analog of Shannon's coding theorem was developed by Schumacher and Westmoreland [1997]; Holevo [1998] in what is known as the Holevo-Schumacher-Westmoreland theorem. Also Hausladen et al. [1996]; Holevo [1998] derived a highly useful upper bound on mutual information which takes all possible measurements into account.
Finding the classical capacities of quantum channels proved once again to be a challenge. Based on numerous important results [Yuen and Ozawa 1993; Caves and Drummond 1994; Holevo and Werner 2001], the capacity of a lossy bosonic channel was derived by Giovannetti et al. [2004]. And only in recent years a rigorous derivation of the capacity for the quantum analog to the additive white Gaussian noise channel was accomplished by Giovannetti et al. [2014]. The capacity can be achieved by a Gaussian ensemble of coherent states rendering them highly qualified candidates for nearly optimal optical communication.

Apart from the analysis of channel capacities much effort was devoted to particular communication scenarios and techniques to saturate the capacity. In order to actually achieve the absolute limit, retrieving the information encoded into optical states is a critical task. A long standing and widely used detection technique is the homodyne detector [Ip et al. 2008; Olivares et al. 2013; Wittmann et al. 2010] or its extension, the double homodyne detector [Ip et al. 2008; Shapiro 2009], which both saturate their respective version of the Shannon-Hartley limit. Also single photon diodes (SPD) [Martinez 2007; Ip et al. 2008] received much attention. However, the mentioned detectors do not constitute the optimal measurement apart from specific scenarios. In particular, both homodyne and double homodyne detection suffer from inevitable measurement noise, called shot-noise, and are subsequently confined to the classical limit. Quantum mechanics, in principle, allows for sub-shot-noise detection as it is only restricted by quantum fluctuations themselves.
For the discrimination of two states the minimum error probability and the corresponding optimal measurement were found by Helstrom [1976] in his seminal work. While his theory provides the framework to deduce the limits on discrimination, actual physical realizations remain highly elusive. One of the most well known Helstrom measurements is the Dolinar receiver Dolinar [1973] for the discrimination of binary coherent alphabets. It is based on a displacement operation, detection with a SPD and a fast-feedback loop to adjust the displacement based on detection events. Even though this scheme saturates the Helstrom limit and has been demonstrated in a proof-of-principle experiment Cook et al. [2007], it remains challenging to implement in a practical applications due to its technical complexity.
Other types of receivers with less stringent technical requirements have been investi-

gated. While not being able to attain the Helstrom or Holevo limit, they still provide a near optimal performance. First introduced was a simpler version of the Dolinar receiver by Kennedy [1973], the displacement receiver, which does not require any feedback.
Due to the dawn of quantum cryptography, an increasing demand in optical satellite communication and optical fibre communication has renewed the interest in quantum receivers going beyond the classical limit. The advent of photon number resolving detectors with very low dark count rates [Eisaman et al. 2011] sparked a new level of activity in the field of sub-shot-noise receivers. In recent years a lot of attention [Burenkov et al. 2021] was given to further develop the idea of a displacement receiver and demonstrate its capability with modern technologies [Wittmann et al. 2008; Takeoka and Sasaki 2008; Wittmann et al. 2010; Tsujino et al. 2010; 2011; DiMario and Becerra 2018; Shcherbatenko et al. 2020; Thekkadath et al. 2021]. The principle of a displacement receiver was also extended to discriminate between more than two coherent states [Izumi et al. 2012; 2013; DiMario et al. 2018; Sidhu et al. 2021].
Apart from the single-shot approach also Dolinar-like receivers are under investigation. Easing the requirement of fast-feedback to a more simple strategy was proposed by Bondurant [1993] and demonstrated by Jabir et al. [2020]. Also receivers based on Bayesian strategies are actively researched for discriminating a various number of states and encoding types [Becerra et al. 2011; Izumi et al. 2012; Li et al. 2013; Izumi et al. 2013; Becerra et al. 2013; 2015; Sych and Leuchs 2016; Ferdinand et al. 2017; Burenkov et al. 2018; 2020; Jabir et al. 2020; Izumi et al. 2021; Sidhu et al. 2021].

While promising a new level of detection, displacement receivers suffer from a high susceptibility to imperfections and noise on the signal. In particular, noise caused by phase diffusion is highly detrimental [Teklu et al. 2015; Trapani et al. 2015; Jarzyna et al. 2016], since a displacement receiver relies on accurately displacing the signal which requires a phase reference. Another limiting factor is thermal noise which is also present in fibre based communication [Essiambre et al. 2010], but especially relevant in the context of deep-space-communication [Leeb 1989]. Both types of noise represent a significant challenge to displacement receivers and can easily erase the advantage over conventional techniques.

Outline

The goal of this dissertation is to investigate the potential of a single-shot displacement receiver with photon number resolving capability in the context of classical communication with coherent states. We develop an extensive model which incorporates various imperfections in order to gain a realistic picture. The physical channel over which the states are transferred is modeled as a quantum channel in order to incorporate nonclassical effects. We will analyze a variety of communication scenarios including fibre based communication as well as photon starved satellite communication. All scenarios are analyzed on the basis of error probability as well as mutual information as figures of merit and compared to the corresponding limits.
In chapter 2 we will first introduce a number of basic concepts used throughout this

work. The relevant basics of quantum mechanics as well as classical electrodynamics are reviewed, before the fundamentals of quantum optics are introduced. Several important phase space representations are discussed, followed by various types of quantum states.

In chapter 3 the foundations for addressing communication scenarios are laid. The concept of a channel is introduced as well as the error probability and mutual information as figures of merit to quantify the performance. Furthermore, both classical and quantum capacities for classical information transmission are shown and their respective behavior in the limits of low and high available energy are discussed. Finally, several widely used types of encodings are introduced.

In chapter 4 several receivers are introduced, in particular, photon counting detectors, homodyne and double homodyne detection and the displacement receiver with photon number resolving capability. We will investigate generalizations of the displacement receiver and how they approach the ultimate limit imposed by Helstrom. Furthermore, we explore the close connection between the displacement receiver and commonly used coherent detection in the form of a homodyne detector. We take a look on how measurement noise arises in homodyne detection which gives rise to the shot-noise limit. Also, a theory of imperfections in the displacement receiver is developed and the effects on communication are discussed.
The remainder of chapter 4 is devoted to measurements and detection in the limit of vanishing energy. First, the limits for various receivers are discussed similar to [Banaszek et al. 2020]. Then, we develop a formalism which allows us to analyze the contribution of particular measurements to mutual information in the low energy limit based on [Kunz et al. 2020]. We derive bounds on mutual information for different types of coherent states for individual measurements, before utilizing the formalism to derive measurements, including their physical realizations, that achieve superadditivity of mutual information.

In chapter 5 we first introduce several well known models for channel noise based on the theory of open quantum systems. Then, as presented in [Kunz et al. 2018], we investigate nonlinear noise arising from the interplay between a Kerr medium and distributed losses. The model can be understood as a simple quantum model of a single mode optical glass fibre and we show that quantum fluctuations cause nonlinear phase noise. Furthermore, we discuss its effect on the generation of squeezed states in a Kerr nonlinear medium.

In chapter 6 we apply the derived models for channels noise and receivers to investigate various communication scenarios. First, a channel with thermal background noise is considered and its effect on sub-shot-noise detection with a displacement receiver. Also pulse position modulation in the photon-starved regime and the capability of photon-number resolved detection are considered based on the results of [Banaszek et al. 2019a;b].
Next, using the results of [Kunz et al. 2018] the effect of a Kerr nonlinearity on the information transfer in the phase variable of coherent states is discussed.

Finally, we consider detection strategies for phase noise channel, both linear and non-linear. In particular, we show that a great improvement in the performance of the displacement receiver is achieved by optimizing the state amplitudes and the displacement. For linear phase noise a proof-of-principle type of experiment was carried out in collaboration with M. DiMario and F. E. Becerra [DiMario et al. 2019].

2 Basic Concepts

2.1 Quantum Mechanics

In this section we briefly review the core aspects of quantum mechanics which are relevant to this work. A more complete picture can be found in one of many textbooks or reviews such as [Mandel and Wolf 1995; Chefles 2000; Barnett and Croke 2009; Sakurai and Napolitano 2017; Griffiths and Schroeter 2018].

2.1.1 Quantum States

At the very core of quantum mechanics stands the concept of a *quantum state* Ψ which contains all knowledge about the system in question. Formally, it is described by a vector from the Hilbert space, commonly denoted in the Dirac notation as $|\,\Psi\,\rangle$ [Dirac 1939]. The exact nature of the Hilbert space and its members is dictated by the specific system which it should describe. For instance, the Hilbert space of an electron spin is two-dimensional whereas the harmonic oscillator lives in an infinite dimensional Hilbert space. Each state can be expanded in terms of a basis $\{|\,\phi_i\,\rangle\}$, i.e.

$$|\,\Psi\,\rangle = \sum_i c_i\,|\,\phi_i\,\rangle\,, \tag{2.1}$$

where c_i are complex numbers that satisfy $\sum_i |c_i|^2 = 1$. The choice of a particular basis is arbitrary, which exemplifies that a quantum state is an abstract object which is not tied to a specific representation. While there is no need to specify the state in a basis, it is often useful to do so.
A shortcoming of the quantum state vector formalism is that only *pure states* are easily written down in this form. *Mixed states*, which are probabilistic mixtures of pure states $|\,\Psi_i\,\rangle$, can be described more conveniently by a *density operator* or *density matrix* $\hat{\varrho}$. In this formalism, a mixed state can be written as

$$\hat{\varrho} = \sum_i p_i\,|\,\Psi_i\,\rangle\,\langle\,\Psi_i\,|\,, \tag{2.2}$$

where p_i can be interpreted as the probability that the quantum state in mind is in the pure state $|\,\Psi_i\,\rangle$. Note, however, that the pure states $|\,\Psi_i\,\rangle$ can be non-orthogonal to each other which would render the interpretation problematic. The density matrix $\hat{\varrho}$ is a Hermitian $\left(\hat{\varrho} = \hat{\varrho}^\dagger\right)$, positive semidefinite $(\hat{\varrho} \geq 0)$ operator with unit trace

$$\mathrm{Tr}\,[\hat{\varrho}] = 1. \tag{2.3}$$

In later sections, several bases and quantum states are introduced that are particularly important in the context of optical communication.

2.1.2 Measurements

A crucial point in any physical theory is obtaining insight about a system by measuring appropriate physical quantities. In quantum mechanics any observable quantity is connected to an operator $\hat{\mathcal{O}}$ called *observable*. In terms of its eigenstates $|\, o_m \rangle$, the observable can be written as

$$\hat{\mathcal{O}} = \sum_m o_m \,|\, o_m \rangle \langle\, o_m \,| \,, \tag{2.4}$$

where o_m denotes the eigenvalue for the corresponding eigenstate $|\, o_m \rangle$. Equation (2.4) is also called the eigendecomposition of $\hat{\mathcal{O}}$. The set of eigenvalues $\{o_m\}$ defines the possible outcomes when measuring the observable. Since any physically measured value is real, the observable $\hat{\mathcal{O}}$ must be a Hermitian operator and its expectation value for state $\hat{\varrho}$ is given by

$$\left\langle\, \hat{\mathcal{O}} \,\right\rangle = \mathrm{Tr}\left[\hat{\mathcal{O}}\hat{\varrho}\right]. \tag{2.5}$$

Note that this expectation value is the expected average outcome when measuring identical copies of a state and not the expected outcome when the same state is measured repeatedly. According to standard theory of quantum measurements after the system in a state $\hat{\varrho}$ has been measured and the outcome o_m has been obtained, the state of the system changes to the eigenstate $|\, o_m \rangle$. This phenomenon is often referred to as state collapse [Heisenberg 1927; von Neumann 1932]. If then the system is measured again, one can only receive the same measurement result o_m.
Th eigendecomposition of eq. (2.4) naturally defines the operators $\hat{P}_m = |\, o_m \rangle \langle\, o_m \,|$ which are projections on the corresponding eigenstates. The set of operators $\left\{\hat{P}_m\right\}$ defines a *projective quantum measurement* and the probability of obtaining the outcome o_m is given by Born's rule [Born 1926]

$$p_m = \mathrm{Tr}\left[\hat{P}_m\hat{\varrho}\right]. \tag{2.6}$$

Not all quantum measurements, however, have to be projective. The most general measurement allowed by quantum mechanics is described by a *positive operator-valued measure (POVM)*. It is defined by a set of positive, semidefinite, Hermitian operators $\{\hat{\pi}_m\}$ which satisfy

$$\sum_m \hat{\pi}_m = \hat{\mathbb{1}}. \tag{2.7}$$

The probability of obtaining the outcome m is again given by Born's rule

$$p_m = \mathrm{Tr}\left[\hat{\pi}_m\hat{\varrho}\right]. \tag{2.8}$$

Due to eqs. (2.3) and (2.7) the probabilities p_m of eq. (2.8) sum up to $\sum_m p_m = 1$ as expected. Projective measurements are a special case of a POVM with $\hat{\pi}_m = \hat{P}_m$. Quantum mechanics guarantees that a POVM has a physical realization. However, a major drawback is that finding a physical realization of a given POVM is usually a nontrivial task and often no such realization is known. Nevertheless, the POVM formalism is particularly useful when looking for the optimal measurement to discriminate different quantum states as we will see in chapter 3.

2.1.3 Evolution

Before we dive into the dynamics of optical signals, let us briefly take a look on how the evolution of states is described within quantum mechanics. The time evolution of quantum state $|\,\Psi\,(t)\,\rangle$ of a closed system is given by the Schrödinger equation [Schrödinger 1926]

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\,|\,\Psi\,(t)\,\rangle = \hat{H}\,|\,\Psi\,(t)\,\rangle\,. \tag{2.9}$$

The Hamiltonian $\hat{H}$ is a Hermitian operator that is determined by the system at hand. Knowing the Hamiltonian, the evolution of the system is, in principle, completely characterized. The solutions to eq. (2.9) are time dependent and describe how states evolve in time. Their evolution is given by

$$|\,\Psi\,(t)\,\rangle = \hat{U}\,(t,t_0)\,|\,\Psi\,(t_0)\,\rangle\,, \tag{2.10}$$

where $\hat{U}\,(t,t_0)$ is the time evolution operator from time t_0 to t and can be written as $\hat{U}\,(t,t_0) = e^{-\frac{i\hat{H}(t-t_0)}{\hbar}}$ for time independent Hamiltonians. The operator $\hat{U}\,(t,t_0)$ is unitary, i.e. it satisfies

$$\hat{U}^{\dagger}\,(t,t_0)\,\hat{U}\,(t,t_0) = \hat{\mathbb{1}}, \tag{2.11}$$

which implies $\hat{U}^{-1}\,(t,t_0) = \hat{U}^{\dagger}\,(t,t_0)$. A unitary evolution is therefore reversible.
A particularly important set of solutions to eq. (2.9) are stationary states which are given by the eigenstates $\hat{H}\,|\,E\,\rangle = E\,|\,E\,\rangle$ of the Hamiltonian. These states are the energy eigenstates of the system and their corresponding eigenvalue E is the energy of the state. Their time evolution is given by a global phase factor $|\,E\,(t)\,\rangle = e^{-\frac{iEt}{\hbar}}\,|\,E\,\rangle$. As can be easily seen from eq. (2.5) the expectation value of any observable $\hat{\mathcal{O}}$ for such a state is time independent, since the global phase factors out. This fact is indeed one of the postulates of quantum mechanics which states that a quantum state is only defined up to a global phase [Dirac 1930; von Neumann 1932]. For these reasons the states $|\,E\,\rangle$ are called stationary and usually their time dependence is dropped.
The time evolution of density matrices of closed systems can be derived from eqs. (2.2) and (2.10) and is also given by the unitary time evolution operator as [Sidhu et al. 2021]

$$\hat{\varrho}\,(t) = \hat{U}\,(t,t_0)\,\hat{\varrho}\,(t_0)\,\hat{U}^{\dagger}\,(t,t_0)\,. \tag{2.12}$$

By differentiating eq. (2.12) with respect to time one arrives at the equation of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\varrho}\,(t) = -\frac{i}{\hbar}\left[\hat{H},\hat{\varrho}\,(t)\right], \tag{2.13}$$

which is also called the von Neumann equation [von Neumann 1932].
Both, the Schrödinger equation of eq. (2.9) and the von Neumann equation in eq. (2.13) are useful to describe closed systems, i.e. systems that do not interact with an environment. This, however, is often not a valid assumption. For instance, any light pulse traveling through a fibre continuously interacts with the fibre itself. Therefore, a different framework is needed to describe such open quantum systems efficiently, i.e. without including the evolution of the external environment in the calculations explicitly. Before writing down the equation of motions which describe open systems, let us briefly sketch how they arise. A detailed derivation can be found, for instance, in [Breuer and Petruccione 2007].
We start out by describing the evolution of the joint state of system and environment with the unitary evolution of eq. (2.12). Then the joint state is reduced to only describe the system itself by taking the partial trace [Breuer and Petruccione 2007] over the environment. This step leaves us with a density matrix that only considers the system state, but still includes the evolution caused by interactions between system and environment. The effect of this interaction is commonly referred to as noise. In the end, we arrive at the Markovian master equation [Sidhu et al. 2021]

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\varrho}\left(t\right) = \Phi\hat{\varrho}. \tag{2.14}$$

Note that $\hat{\varrho}$ is the density matrix describing only the system itself without the environment. The operator Φ acts again on operators and is therefore called a super-operator. Equation (2.14) can be seen as a generalized form of the von Neumann equation of eq. (2.13). For a memoryless evolution the most general form of the super-operator Φ in eq. (2.14) is given by the *Gorini-Kossakowski-Lindblad-Sudarshan* equation [Gorini et al. 1976; Lindblad 1976] which reads

$$\Phi\hat{\varrho} = -\frac{i}{\hbar}\left[\hat{H},\hat{\varrho}\right] + \sum_k \gamma_k \left(\hat{L}_k\hat{\varrho}\hat{L}_k^\dagger - \frac{1}{2}\hat{L}_k^\dagger\hat{L}_k\hat{\varrho} - \frac{1}{2}\hat{\varrho}\hat{L}_k^\dagger\hat{L}_k\right). \tag{2.15}$$

The operators $\hat{L}_k$ are called Lindblad operators and describe the interaction between the system and the environment and their coupling strength is given by γ_k. Usually, the evolution of an open quantum system is not unitary which is caused by the second term in eq. (2.15). An important consequence of this fact is that noisy evolution can not easily be reversed.
All equations of motions presented above are fundamentally describing the complete evolution of quantum states, but each of them is best suited for a different scenario. The von Neumann and Schrödinger equations are, in principle, equivalent as mixed states could be treated with the Schrödinger equation by considering each pure state of eq. (2.2) separately. The evolved mixed state then is given by a statistical mixture of the evolved pure states. This is closely connected to the fact that in quantum mechanics exist two inherently different kinds of probabilities. The probability of obtaining an outcome according to Born's rule of eq. (2.8) is fundamentally quantum. Even with complete knowledge of the system the probabilistic behavior can not be avoided and

quantum mechanics dictates the uncertainty of observables. However, the probabilities p_i arising in eq. (2.2) for mixed states are of statistical nature. They are connected to an imperfect knowledge and are not fundamental in nature.
Similarly, open quantum systems could be, in principle, completely described by the von Neumann or Schrödinger equations. This, however, would require a complete description of the environmental state which is often impossible or impractical. The Gorini-Kossakowski-Lindblad-Sudarshan equation offers a more accessible way to understanding open quantum systems as it allows to calculate the (noisy) evolution of open systems without requiring a detailed knowledge of the environment.

2.2 Classical Electrodynamics

In preparation for the quantum description of optical communication, we will briefly introduce the classical description of the evolution of light and concentrate one the electric field. A more complete treatment can be found, for instance, in the textbook of Mandel and Wolf [1995].
The evolution of an electromagnetic field in free space is described by the famous Maxwell equations [Maxwell 1865] which are given by

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \quad , \quad \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \tag{2.16}$$

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \quad , \quad \boldsymbol{\nabla} \cdot \boldsymbol{E} = 0 \tag{2.17}$$

where c denotes the speed of light in vacuum. For the following discussion, we assume that the electric field $\boldsymbol{E}$ consists of a single polarization and has a fixed propagation direction described by z. Then, the electric field can be described by a field $E(z,t)$ and separated into a positive and a negative frequency part such that it reads

$$E(z,t) = E_+(z,t) + E_-(z,t)\,. \tag{2.18}$$

The positive frequency part in terms of a plane wave expansion can be written as

$$E_+(z,t) = \int \mathrm{d}\omega \sqrt{\frac{\hbar\omega}{4\pi\varepsilon_0 A}}\tilde{\zeta}(\omega)\, e^{i(k(\omega)z-\omega t)}, \tag{2.19}$$

where ε_0 is the vacuum permittivity, A is the effective area of the transverse spatial mode and ω the angular frequency of the field. The form of the factor under the square root was chosen for later convenience. The corresponding negative frequency part is obtained by complex conjugation of $E_+(z,t)$, i.e. $E_-(z,t) = E_+^*(z,t)$. The complex amplitude function $\tilde{\zeta}(\omega)$ consists of contributions from many frequency modes and is given by

$$\tilde{\zeta}(\omega) = \sum_j \zeta_j \tilde{u}_j(\omega)\,, \tag{2.20}$$

where j denotes the mode and $\tilde{u}_j(\omega)$ the complex envelope function of the j-th mode in frequency space. The mode profiles are normalized and orthogonal between different modes such that

$$\int \mathrm{d}\omega\ \tilde{u}_j(\omega)\, \tilde{u}^*_{j'}(\omega) = \delta_{jj'} \tag{2.21}$$

holds with the Kronecker delta $\delta_{jj'}$. Substituting eq. (2.20) into eq. (2.19) and using the dispersion relation $k(\omega) = \frac{\omega}{c}$ of free space yields

$$E_+(z,t) = \sum_j \zeta_j \int \mathrm{d}\omega\ \sqrt{\frac{\hbar\omega}{4\pi\varepsilon_0 A}}\tilde{u}_j(\omega)\, e^{-i\omega\left(t-\frac{z}{c}\right)}. \tag{2.22}$$

We can further simplify this expression assuming a narrowband signal ω that varies slowly around a carrier frequency ω_c which gives us a good approximation of eq. (2.22) reading

$$E_+(z,t) \approx \sqrt{\frac{\hbar\omega_c}{4\pi\varepsilon_0 A}} \sum_j \zeta_j \int \mathrm{d}\omega\ \tilde{u}_j(\omega)\, e^{-i\omega\left(t-\frac{z}{c}\right)} \tag{2.23}$$

$$= \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}} \sum_j \zeta_j u_j\left(t - \frac{z}{c}\right), \tag{2.24}$$

where in the last step the Fourier transformation of the complex envelope $\tilde{u}_j(\omega)$ was used, i.e.

$$u_j\left(t - \frac{z}{c}\right) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\omega\ \tilde{u}_j(\omega)\, e^{-i\omega\left(t-\frac{z}{c}\right)}. \tag{2.25}$$

Finally, the full expression for the electric field of a narrowband signal reads

$$E(z,t) = \sum_j \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}} \left[\zeta_j u_j\left(t - \frac{z}{c}\right) + \zeta_j^* u_j^*\left(t - \frac{z}{c}\right)\right]. \tag{2.26}$$

Considering only a single mode at $z = 0$ the expressions simplify further and the electric field is given by

$$E_+(t) = \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}}\zeta u(t), \tag{2.27}$$

$$E(t) = \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}} \left[\zeta u(t) + \zeta^* u^*(t)\right]. \tag{2.28}$$

It is fully characterized by a complex amplitude ζ, a complex envelope $u(t)$ and a carrier frequency ω_c.
The energy H contained in the signal can be determined by the expression [Scully and Zubairy 1997]

$$H = 2\varepsilon_0 \int \mathrm{d}t \int \mathrm{d}^2\boldsymbol{r}\ \left|E_+(t)\right|^2 = \hbar\omega_c \left|\zeta\right|^2. \tag{2.29}$$

where in the last step eq. (2.27) was used. The factor $\hbar\omega_c$ expresses the energy of a single photon with frequency ω_c, whereas $|\zeta|^2$ has the interpretation of the expectation value for the number of photons contained in the signal. The energy is therefore independent of the complex envelope $u(t)$.
The latter fixes the shape in the time domain as well as the frequency spectrum. Since time t and frequency ω are conjugate variables, the frequency spectrum $\tilde{u}(\omega)$ and the time profile $u(t)$ are directly connected by means of a Fourier transformation given by

$$\tilde{u}(\omega) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}t\; e^{i\omega t} u(t)\,, \tag{2.30}$$

$$u(t) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\omega\; e^{-i\omega t} \tilde{u}(\omega)\,. \tag{2.31}$$

To gain some insight on this connection let us look at a couple of common pulse profiles. As a first example consider a signal with a Gaussian profile

$$u(t) = \frac{1}{\sqrt[4]{\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}, \tag{2.32}$$

where σ denotes the standard deviation. Then, the corresponding frequency spectrum is given by another Gaussian

$$\tilde{u}(\omega) = \sqrt[4]{\frac{\sigma^2}{\pi}} e^{-\frac{\sigma^2\omega^2}{2}} \tag{2.33}$$

with standard deviation σ^{-1}. Both the envelope of eq. (2.32) and the spectrum of eq. (2.33) are depicted in fig. 2.1(a) and (b), respectively. While both $u(t)$ and $\tilde{u}(\omega)$ remain Gaussian, a signal envelope more localized in time leads to a broader frequency spectrum and vice versa. Depending on the scenario at hand, it might be necessary to limit either temporal or spectral spread of the signal to a defined range. For instance, when sending a train of signals identifiable only by their position in time, a localized envelope is advantageous to minimize the overlap between neighboring signals. Contrary, a defined and bounded spectrum is needed when communicating via a communication network to minimize crosstalk. To achieve the latter the sinc profile, defined as

$$u(t) = \frac{\sin(\pi t)}{\pi t}, \tag{2.34}$$

is a commonly used envelope. Its corresponding spectrum is upper and lower bounded and given by

$$\tilde{u}(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}}, & \text{for } -\pi \le \omega \le \pi \\ 0, & \text{for } |\omega| > \pi \end{cases}. \tag{2.35}$$

Both, eqs. (2.34) and (2.35), are depicted in fig. 2.1(c) and (d). Here, the signal envelope is widely spread in the time domain, which usually leads to an non-zero overlap between consecutive signals, but the spectrum is rectangular and confined to a specified range.

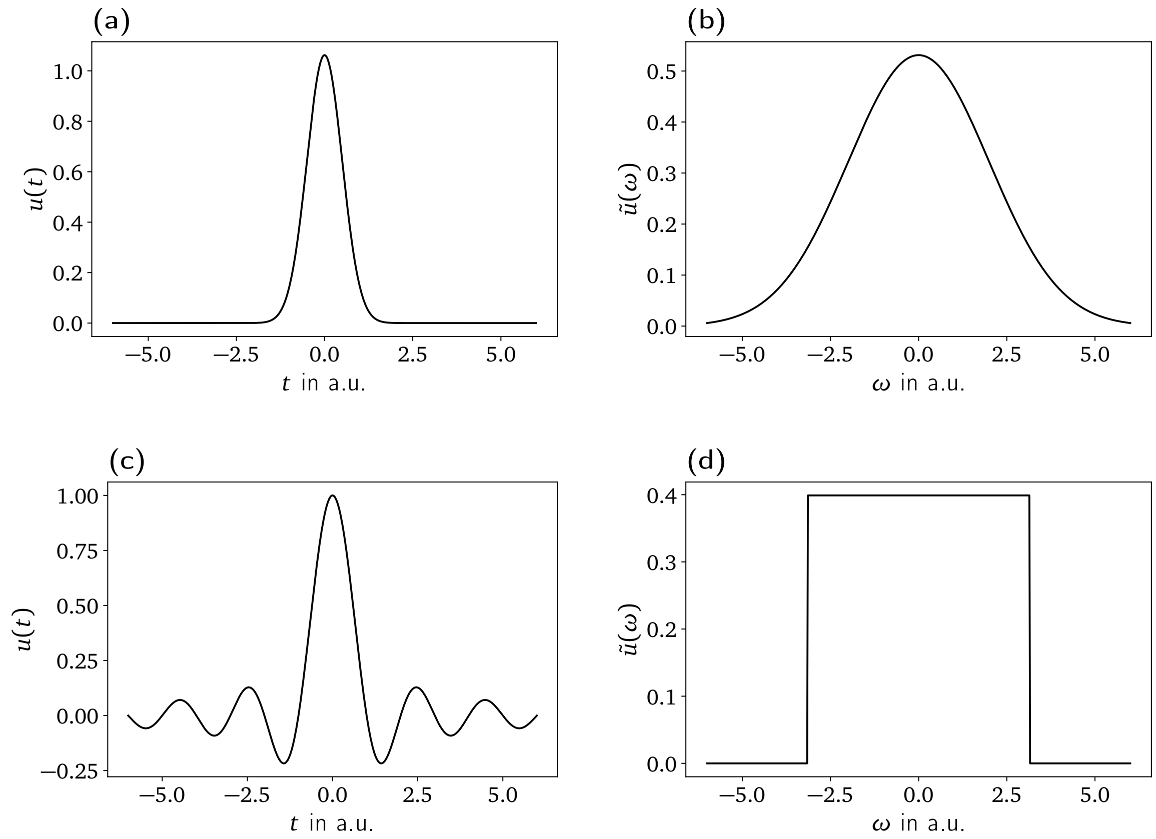


Figure 2.1: **(a)** Complex envelope $u(t)$ for the Gaussian profile of eq. (2.32) with standard deviation $\sigma = 0.5$. **(b)** The corresponding spectrum $\tilde{u}(\omega)$ of eq. (2.33). **(c)** Complex envelope $u(t)$ for the sinc profile of eq. (2.34). **(d)** The corresponding spectrum $\tilde{u}(\omega)$ of eq. (2.35).

2.3 Quantum Optics

In the previous section the classical evolution of electromagnetic waves was briefly introduced. Now we will turn to the quantum description of light with a focus on classical light in particular. This section is dedicated to review the very basics of single mode quantum optics and introduce the notion of a quantum optical phase space. A more complete treatment can be found in one of many excellent text books such as [Mandel and Wolf 1995; Scully and Zubairy 1997; Schleich 2001].

While in the classical treatment the conjugate variables ζ_j and ζ_j^* are given by complex numbers, in the quantum description they are replaced by conjugate non-Hermitian operators acting on the Hilbert space denoted as $\hat{a}_j$ and $\hat{a}_j^\dagger$. Therefore, the quantized version of the narrowband multimode electric field in eq. (2.26) reads

$$\hat{E}(z,t) = \sum_j \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}} \left[\hat{a}_j u_j\left(t - \frac{z}{c}\right) + \hat{a}_j^\dagger u_j^*\left(t - \frac{z}{c}\right)\right]. \tag{2.36}$$

Considering only a single mode at $z = 0$ reduces eq. (2.36) to

$$\hat{E}_{+}(t) = \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}}\hat{a}u(t), \tag{2.37}$$

$$\hat{E}(t) = \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}}\left[\hat{a}u(t) + \hat{a}^{\dagger}u^{*}(t)\right], \tag{2.38}$$

which corresponds to the classical field in eqs. (2.27) and (2.28). For the single mode approach to be valid it is required that all other modes remain unexcited. For instance, this assumption is valid for detection schemes that are only sensitive to a single mode. The operators $\hat{a}$ and $\hat{a}^{\dagger}$, also known as the annihilation and creation operators, are the ladder operators of a harmonic oscillator and satisfy the commutation relation

$$\left[\hat{a}, \hat{a}^{\dagger}\right] = 1. \tag{2.39}$$

The connection to the quantum harmonic oscillator becomes evident when considering the energy (Hamiltonian) of the electric field. The quantum analog of the classical energy in eq. (2.29) is given by

$$\hat{H} = \varepsilon_0 \int \mathrm{d}t \int \mathrm{d}^2\boldsymbol{r}\, \left[\hat{E}_{+}(t)\,\hat{E}_{+}^{\dagger}(t) + \hat{E}_{+}^{\dagger}(t)\,\hat{E}_{+}(t)\right] = \frac{\hbar\omega_c}{2}\left(\hat{a}\hat{a}^{\dagger} + \hat{a}^{\dagger}\hat{a}\right). \tag{2.40}$$

By using the commutation relation in eq. (2.39) we can easily see that the Hamiltonian of the electric field in eq. (2.40) is equivalent to the Hamiltonian of a harmonic oscillator which reads $\hat{H} = \hbar\omega\left(\hat{a}^{\dagger}\hat{a} + \frac{1}{2}\right)$. Consequently, the energy eigenstates $|\, E\,\rangle$ of the electric field are directly given by the eigenstates of a harmonic oscillator. In order to find the energy eigenstates it is convenient to introduce the number operator

$$\hat{n} = \hat{a}^{\dagger}\hat{a}, \tag{2.41}$$

which is Hermitian and has the eigenstates

$$\hat{n}\,|\, n\,\rangle = n\,|\, n\,\rangle, \tag{2.42}$$

where $|\, n\,\rangle$ is called a *number state* or a *Fock state*. In section 2.5 the properties of Fock states are discussed in more detail.

A striking difference between eq. (2.38) and the classical field is the existence of vacuum modes. While the expectation value of the electric field on the vacuum state, i.e. a Fock state that contains zero photons, vanishes

$$\left\langle 0 \,\middle|\, \hat{E}(t) \,\middle|\, 0 \right\rangle = 0, \tag{2.43}$$

the expectation value of the squared operator does not vanish, since

$$\left\langle 0 \,\middle|\, \hat{E}^2(t) \,\middle|\, 0 \right\rangle = \frac{\hbar\omega_c}{2\varepsilon_0 A}\left|u(t)\right|^2. \tag{2.44}$$

Note that this requires a detection scheme that measures only a single mode. The variance of the electric field operator is clearly nonzero, i.e. $\mathrm{Var}\left(\hat{E}\left(t\right)\right) \neq 0$, and as a consequence even the lowest energy eigenstate exhibits a phenomenon known as vacuum fluctuations. They are responsible for a number of quantum effects such as the Casimir effect [Casimir and Polder 1948] or the Lamb shift [Lamb and Retherford 1947]. With the help of Fock states, the energy contained in the electric field of eq. (2.40) can be conveniently expressed in terms of the number of photons such that

$$\hat{H}\,|\,n\,\rangle = \hbar\omega_c\left(n+\frac{1}{2}\right)|\,n\,\rangle\,, \tag{2.45}$$

where $\hbar\omega_c$ is the energy of a single photon with frequency ω_c. The ladder operators $\hat{a}$ and $\hat{a}^\dagger$ received their name due to the fact that they raise or lower the energy level of a harmonic oscillator by a single level. Their action on the number states is given by

$$\hat{a}\,|\,n\,\rangle = \sqrt{n}\,|\,n-1\,\rangle\,, \tag{2.46}$$
$$\hat{a}^\dagger\,|\,n\,\rangle = \sqrt{n+1}\,|\,n+1\,\rangle\,. \tag{2.47}$$

Apart from describing the energy levels of the electric field, Fock states are also very useful to represent general quantum states $\hat{\varrho}$. They satisfy the orthogonality condition

$$\langle\,n\,|\,m\,\rangle = \delta_{m,n}, \tag{2.48}$$

where $\delta_{m,n}$ denotes the Kronecker delta and form a complete orthogonal basis of the photonic Hilbert space, such that an arbitrary quantum optical state $\hat{\varrho}$ can be expressed as

$$\hat{\varrho} = \sum_{n,m} c_{n,m}\,|\,n\,\rangle\,\langle\,m\,|\,. \tag{2.49}$$

We will make use of this representation extensively in this thesis.
So far we only considered the energy eigenstates of the electric field. When considering only the positive frequency part of the electric field, which reads

$$\hat{E}_+\left(t\right) = \sqrt{\frac{\hbar\omega_c}{2\varepsilon_0 A}}\,\hat{a}u\left(t\right), \tag{2.50}$$

another type of eigenstate arises. It is directly defined by the eigenstate of the annihilation operator $\hat{a}$

$$\hat{a}\,|\,\zeta\,\rangle = \zeta\,|\,\zeta\,\rangle\,. \tag{2.51}$$

The eigenstate $|\,\zeta\,\rangle$ is called a coherent state and was first introduced by Glauber [1963]. They are of utmost importance to optical communication and in section 2.6 we will discuss the properties of coherent states in more detail. Just like Fock states, coherent states form a complete set in the Hilbert space and obey the completeness relation

$$\frac{1}{\pi}\int \mathrm{d}^2\zeta\;\,|\,\zeta\,\rangle\,\langle\,\zeta\,| = \hat{\mathbb{1}}. \tag{2.52}$$

Since two coherent states are non-orthogonal with an overlap of

$$\langle \zeta_1 \,|\, \zeta_2 \rangle = e^{-\frac{1}{2}\left(|\zeta_1|^2+|\zeta_2|^2-2\zeta_1^*\zeta_2\right)}, \tag{2.53}$$

$$|\langle \zeta_1 \,|\, \zeta_2 \rangle|^2 = e^{-|\zeta_1-\zeta_2|^2}, \tag{2.54}$$

the set is in fact overcomplete, which may seem to render them a rather awkward set to use as a basis. However, the set of coherent states allows for very useful insights as it is closely connected to phase space representations.

2.4 Phase Space Representations

In classical mechanics it is often convenient to represent the dynamics of a system in phase space. A phase space is spanned by two conjugate variables, usually the position x and momentum p. Each point in phase space represents a unique state of the system and the entirety of the space forms the set of all possible states. The time evolution of a system draws a path which is called *trajectory*. In this fashion the state and its time trajectory can be described pictorially.
The notion of a phase space can also be extended to quantum mechanics. In quantum optics the quantized version of the conjugate variables x and p are given by the operators [Schleich 2001]

$$\hat{x} = \frac{1}{\sqrt{2}}\left(\hat{a}+\hat{a}^\dagger\right), \tag{2.55}$$

$$\hat{p} = \frac{i}{\sqrt{2}}\left(\hat{a}^\dagger-\hat{a}\right), \tag{2.56}$$

which are also called *quadratures*. They span the quantum optical phase space and a general quadrature $\hat{q}_\theta$ with phase θ is obtained by a superposition of $\hat{x}$ and $\hat{p}$ such that it is given by

$$\hat{q}_\theta = \cos(\theta)\,\hat{x} + \sin(\theta)\,\hat{p} = \frac{1}{\sqrt{2}}\left(\hat{a}e^{-i\theta}+\hat{a}^\dagger e^{i\theta}\right). \tag{2.57}$$

The quadrature $\hat{x}$ is recovered by choosing the phase $\theta = 0$ and $\hat{p}$ by $\theta = \frac{\pi}{2}$. The eigenstate of the general quadrature operator is defined by the equation $\hat{q}_\theta \,|\, q_\theta \rangle = q_\theta \,|\, q_\theta \rangle$ and its overlap with a coherent state with amplitude ζ defined in eq. (2.51) is given by [Schleich 2001]

$$\langle \hat{q}_\theta \,|\, \zeta \rangle = \pi^{-\frac{1}{4}} \exp\left[-\frac{1}{2}q_\theta^2 - \frac{1}{2}\zeta^2 e^{-2i\theta} + \sqrt{2}q_\theta\zeta e^{-i\theta} - \frac{1}{2}|\zeta|^2\right]. \tag{2.58}$$

A couple of very useful expressions are given by the squared absolute values of eq. (2.58) for $\theta = 0$ and $\theta = \frac{\pi}{2}$ which read

$$|\langle x \,|\, \zeta \rangle|^2 = \frac{1}{\sqrt{\pi}} \exp\left[-\left(x-\sqrt{2}\mathrm{Re}\left[\zeta\right]\right)^2\right], \tag{2.59}$$

$$|\langle p \,|\, \zeta \rangle|^2 = \frac{1}{\sqrt{\pi}} \exp\left[-\left(p-\sqrt{2}\mathrm{Im}\left[\zeta\right]\right)^2\right]. \tag{2.60}$$

If we consider $\mathrm{Re}\left[\zeta\right]$ and $\mathrm{Im}\left[\zeta\right]$ as a set of quadratures to describe the phase space, then they are directly connected to quadrature values x and p via the relations [Schleich 2001]

$$x = \sqrt{2}\mathrm{Re}\left[\zeta\right], \tag{2.61}$$

$$p = \sqrt{2}\mathrm{Im}\left[\zeta\right]. \tag{2.62}$$

All phase space illustrations in the remainder of this thesis will make use of x and p as quadratures.
Even though the generalization from classical to quantum phase space is rather straightforward, there are important distinctions. Most importantly, the quadratures do not commute with each other, which means that they cannot be simultaneously measured and assigned well defined values. Therefore, it is impossible to characterize the state of a system by a single point in phase space. Instead, each state occupies a certain area in phase space and can be characterized by its shape which illustrates many properties. In general, the shape is subject to the time evolution of the system. However, it does not follow classical trajectories in phase space, but evolves more intricately. Only if the state retains its shape throughout the evolution, a classical trajectory might be recovered.
In the following, we will introduce two phase space representations. In section 2.4.1 the P-representation will be introduced and the Q-representation in section 2.4.2. Another notable representation is the Wigner function, which was the first representation to be introduced historically [Wigner 1932]. However, we will not make us of it in this thesis.

2.4.1 The Glauber-Sudarshan P-Representation

In 1963 Glauber and Sudarshan independently introduced a phase space representation of quantum states that uses coherent states $|\,\zeta\,\rangle$ as basis. In the Glauber-Sudarshan P-representation an arbitrary quantum state $\hat{\varrho}$ is described by a real function $P\left(\zeta\right)$ through

$$\hat{\varrho} = \int \mathrm{d}^2\zeta \; P\left(\zeta\right) |\,\zeta\,\rangle\,\langle\,\zeta\,|\,. \tag{2.63}$$

Since the density matrix has a unit trace, it follows immediately that

$$\int \mathrm{d}^2\zeta \; P\left(\zeta\right) = 1. \tag{2.64}$$

It might be tempting to think of $P\left(\zeta\right)$ as a probability density of the state $\hat{\varrho}$ in phase space. However, this is not the case. Nonclassical states exhibit a P function which has negative values or is more singular than a delta function. Furthermore, even when considering states for which $P\left(\zeta\right)$ is completely positive the conclusion that then the states are classical can be misleading as it is still possible for them to show quantum features. For example, two points in the phase space of $P\left(\zeta\right)$ are not mutually exclusive, since coherent states are non-orthogonal, i.e. two phase space points possess a non-zero overlap. However, mutual exclusivity is required to treat $P\left(\zeta\right)$ as a proper probability distribution as only in this case two points correspond to exclusive events. To illustrate this, let us consider a coherent state $|\,\zeta_0\,\rangle$ defined by

$$P_{\text{coh}}\left(\zeta\right) = \delta\left(\zeta - \zeta_0\right). \tag{2.65}$$

Two points in the phase space of $P\left(\zeta\right)$ posses the overlap written in eq. (2.53). For not too large values of the distance $|\zeta - \zeta_0|$, the overlap $\langle\,\zeta\,|\,\zeta_0\,\rangle$ is non zero and the point in phase space could also correspond to a different value of ζ. In this regime coherent states exhibit nonclassical features. For large values of $|\zeta - \zeta_0|$, the overlap $\langle\,\zeta\,|\,\zeta_0\,\rangle$ approximately vanishes, two points in phase space can be seen as mutually exclusive and $P\left(\zeta\right)$ can be treated as a probability distribution. Here, coherent states behave like classical states.

2.4.2 The Husimi-Q-Function

Another useful phase space representation was first introduced by Husimi [1940]. The Husimi-Q-function is defined as the expectation value of a projection of the density operator $\hat{\varrho}$ onto a coherent state $|\,\zeta\,\rangle$, i.e.

$$Q\left(\zeta\right) = \frac{1}{\pi}\,\langle\,\zeta\,|\,\hat{\varrho}\,|\,\zeta\,\rangle\,. \tag{2.66}$$

The Q-function is a quasi-probability distribution which has the nice properties of being non-negative and bounded, such that the relation

$$0 \leq Q\left(\zeta\right) \leq \pi \tag{2.67}$$

holds. Since the Husimi-Q-function is rather easy to calculate compared to other representations of quantum states it is particularly useful for the pictorial characterization of quantum states in phase space. For these reasons the Husimi-Q-function will be used to illustrate quantum states in phase space throughout this thesis.
It is noteworthy that the Q-function is closely connected to the P-function. By inserting eq. (2.63) into the Q-function of eq. (2.66) we obtain

$$Q\left(\zeta\right) = \frac{1}{\pi}\int \mathrm{d}^2\beta\; P\left(\beta\right)|\langle\,\zeta\,|\,\beta\,\rangle|^2 = \frac{1}{\pi}\int \mathrm{d}^2\beta\; P\left(\beta\right) e^{-|\zeta-\beta|^2}, \tag{2.68}$$

where in the last step eq. (2.54) was used. As can bee seen in the latter expression, the Q-function is obtained by a convolution of the P-function and a Gaussian centered at $\zeta - \beta$.

2.5 Fock States

Fock states constitute an important basis for quantum optics. In this section some of the main features of Fock states are reviewed.
Formally, as done in section 2.3, a Fock state $|\,n\,\rangle$ is introduced as the eigenstate of the

n-th energy level of a quantum harmonic oscillator, which corresponds to the energy levels of the electromagnetic field. Consequently, in quantum optics a Fock state $|\, n \,\rangle$ is a state containing exactly n-photons. Generating such an n-photon state is a non-trivial task and so far only multi-photon states up to $n = 6$ have been demonstrated [Hofheinz et al. 2008; Sayrin et al. 2011; Bimbard et al. 2010; Cooper et al. 2013; Yukawa et al. 2013]. Notable single-photon sources are quantum dots and color centers as well as non-linear processes such as parametric down-conversion or four-wave mixing [Eisaman et al. 2011].
The Q-function of a Fock state $|\, n \,\rangle$ is given by [Schleich 2001]

$$Q_{\text{Fock}}\left(\zeta\right) = \frac{e^{-|\zeta|^2}}{\pi} \frac{|\zeta|^{2n}}{n!}. \tag{2.69}$$

Examples for different photon numbers n are depicted in fig. 2.2. Since eq. (2.69) only depends on the absolute value of ζ, the Q-function of a Fock state is radially symmetric and describes a donut shape in phase space. Accordingly, Fock states are not able to carry any phase information, but only information about their energy.
A photon number state $|\, n \,\rangle$ is non-classical as is to be expected. Its P-function reads [Schleich 2001]

$$P_{\text{Fock}}\left(\zeta\right) = L_n \left[-\frac{1}{2} \frac{\partial^2}{\partial \zeta_p^2} - \frac{1}{2} \frac{\partial^2}{\partial \zeta_x^2} \right] \delta\left(\zeta\right), \tag{2.70}$$

where $L_n\left[x\right]$ is the n-th Laguerre polynomial and ζ_x (ζ_p) is the quadrature value of the quadrature $\hat{x}$ ($\hat{p}$). Equation (2.70) is highly singular as it involves a high order of partial derivatives of the delta function $\delta\left(\zeta\right)$ and therefore has no classical counterpart. For $n = 0$ the derivatives vanish and the state reduces to the vacuum state $|\, 0 \,\rangle$ with the P-function $P_{\text{vac}}\left(\zeta\right) = \delta\left(\zeta\right)$.

2.6 Coherent States

In this section we review some of the main features of a *coherent state*. It plays a central role in optical communication where it is utilized to carry classical information as it can be routinely produced by a well stabilized laser with a high repetition rate. Coherent states were first described by Glauber [1963] and are defined as eigenstates $|\, \zeta \,\rangle$ of the annihilation operator $\hat{a}$ by

$$\hat{a}\,|\, \zeta \,\rangle = \zeta\,|\, \zeta \,\rangle \tag{2.71}$$

as we have already seen in eq. (2.51). A coherent state is also the eigenstate of the positive frequency part of the electromagnetic field operator in eq. (2.50) with a well defined complex amplitude ζ and consequently, the direct analog of a classical electromagnetic wave. This bears the question whether a coherent state is actually simply a classical state. Indeed, it exhibits a number of classical features as it possesses a minimal uncertainty product, follows classical time trajectories and its electromagnetic field behaves

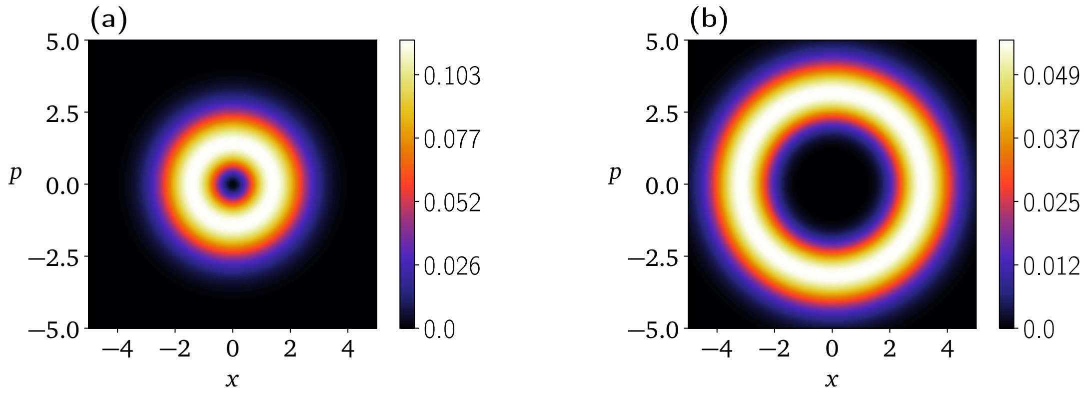


Figure 2.2: Husimi-Q-function of a n-photon state defined in eq. (2.42) with a photon number of **(a)** $n = 1$ and **(b)** $n = 5$. The photon number n determines the radius of the donut while the width of the donut is independent of n.

as classically as possible, since it can be described by a well defined complex amplitude. However, a coherent state is still inherently quantum and should not be understood as an actual classical state. For instance, classical states have a well defined amplitude and phase, whereas coherent states exhibit quantum fluctuations around their respective expectation values.
An explicit representation of a coherent state in the Fock basis is given by

$$|\,\zeta\,\rangle = \sum_{n=0}^{\infty} e^{-\frac{|\zeta|^2}{2}} \frac{\zeta^n}{\sqrt{n!}} \,|\,n\,\rangle\,. \tag{2.72}$$

The Q-function of state $|\,\zeta_0\,\rangle$ is given by

$$Q_{\mathrm{coh}}\left(\zeta\right) = \frac{1}{\pi} e^{-|\zeta-\zeta_0|^2} \tag{2.73}$$

and is shown in fig. 2.3. It is a Gaussian bell located at ζ_0. In contrast to Fock states, coherent states are able to carry information in both their amplitude and phase. The latter is defined via the polar representation of the complex amplitude which is given by

$$\zeta = |\zeta|\, e^{-i\phi}, \tag{2.74}$$

where $|\zeta|$ corresponds to the amplitude and ϕ to the phase. The variances of the quadratures defined in eqs. (2.55) and (2.56) for a coherent state can be determined using eq. (2.72) and read

$$\begin{aligned} \langle\,(\Delta x)^2\,\rangle &= \langle\,\zeta\,|\,\hat{x}^2\,|\,\zeta\,\rangle - (\langle\,\zeta\,|\,\hat{x}\,|\,\zeta\,\rangle)^2 = \frac{1}{2}, \\ \langle\,(\Delta p)^2\,\rangle &= \langle\,\zeta\,|\,\hat{p}^2\,|\,\zeta\,\rangle - (\langle\,\zeta\,|\,\hat{p}\,|\,\zeta\,\rangle)^2 = \frac{1}{2}. \end{aligned} \tag{2.75}$$

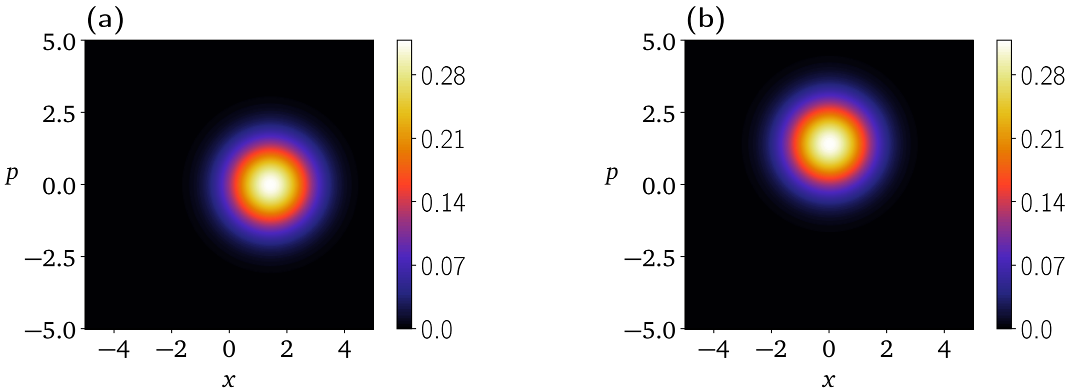


Figure 2.3: Husimi-Q-function of a coherent state given in eq. (2.73) with complex amplitude **(a)** $\zeta = 1$ and **(b)** $\zeta = i$. Both states contain the same energy $|\zeta|^2 = 1$.

Note that these are the same values as for the vacuum state, which is a special case of a coherent state with amplitude $\zeta = 0$. Therefore, the width of the Q-function is independent of the coherent state amplitude and all coherent states have exactly the same shape in phase space.

Often we will be interested in measuring the number of photons, i.e. the value of the photon number operator $\hat{n} = \hat{a}^\dagger \hat{a}$. The corresponding POVM is given by the set of projections $\{\hat{\pi}_n = |\, n\,\rangle \langle\, n\,|\}$ and the probability to measure n photons in a coherent state is given by the Poissonian distribution

$$p\,(n|\zeta) = |\langle\, n\,|\,\zeta\,\rangle|^2 = e^{-|\zeta|^2} \frac{|\zeta|^{2n}}{n!} \tag{2.76}$$

according to Born's rule in eq. (2.8) and the Fock representation in eq. (2.72). The average number of photons carried by a coherent state, i.e. the expectation value of $\hat{n}$, is equal to

$$\langle\,\zeta\,|\,\hat{n}\,|\,\zeta\,\rangle = e^{-\frac{|\zeta|^2}{2}} \langle\,\zeta\,| \sum_n \frac{\zeta^n}{\sqrt{n!}} \hat{n}\,|\,n\,\rangle = e^{-|\zeta|^2} \sum_n \frac{\left(|\zeta|^2\right)^n}{(n-1)!} = |\zeta|^2\,, \tag{2.77}$$

where in the second step the orthogonality condition for Fock states in eq. (2.48) was used. Consequently, the average number of photons contained in a coherent state is directly defined by its complex amplitude ζ.

A different, yet equivalent, approach to defining a coherent state is via the displacement operator

$$\hat{D}\,(\zeta) = e^{\zeta \hat{a}^\dagger - \zeta^* \hat{a}}, \tag{2.78}$$

that shifts a given state in phase space by a complex amplitude ζ. Its effect is schematically depicted in fig. 2.4. When applied to the vacuum state the result of the displacement

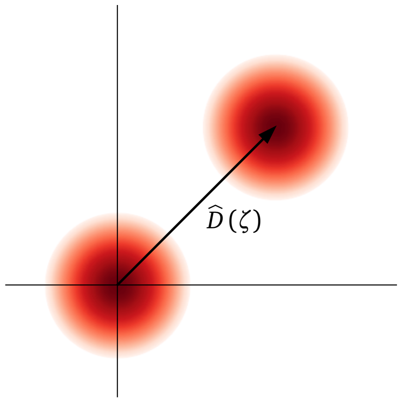


Figure 2.4: Schematic depiction of the displacement operator in eq. (2.78) acting on the vacuum state. The resulting state is a vacuum state displaced in phase space which is also known as a coherent state $|\,\zeta\,\rangle = \hat{D}\,(\zeta)\,|\,0\,\rangle$.

operation is a coherent state, i.e.

$$|\,\zeta\,\rangle = \hat{D}\,(\zeta)\,|\,0\,\rangle\,. \tag{2.79}$$

This fact exemplifies that all coherent states possess the same quadrature variances as the vacuum state. If the original state was a coherent state $|\,\zeta_0\,\rangle$, the displaced state remains coherent and is given by $|\,\zeta_0 + \zeta\,\rangle$ up to a global phase factor. As we will discuss in section 4.4 the displacement operation is highly useful in the context of the discrimination of coherent states.
So far we only discussed pure coherent states. A particular noisy and thus not pure type of coherent states are phase diffused coherent states. As seen above, a coherent state is characterized entirely by its complex amplitude which can be divided into two parts, an absolute value $|\zeta|$ and a phase ϕ as shown in eq. (2.74). Suppose now that the phase variable is not well defined. This can be caused by using a non-ideal laser source in the state preparation or random processes acting on the state during its transmission through a medium [Ip et al. 2008; Li 2009; Kikuchi 2016]. A simple model to incorporate such phase diffusion is given by a coherent state whose phase variable is blurred according to a Gaussian distribution. The density matrix is then given by

$$\hat{\varrho}_{\mathrm{pd}} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mathrm{d}\phi\; e^{-\frac{\phi^2}{2\sigma^2}}\,|\,e^{i\phi}\zeta\,\rangle\,\langle\,e^{i\phi}\zeta\,|\,, \tag{2.80}$$

where σ denotes the standard deviation of the Gaussian distribution and therefore the strength of the phase diffusion. The effect of phase diffusion in phase space is shown in fig. 2.5. It causes a blurring of the state on a circle whose radius is defined by the absolute value $|\zeta|$. The blurring is intensified with rising standard deviation σ.

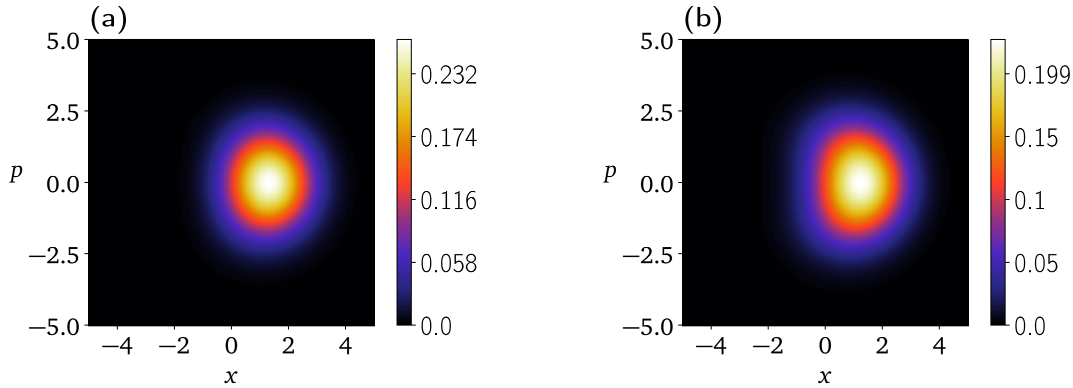


Figure 2.5: Husimi-Q-function of a phase diffused coherent state defined in eq. (2.80) with complex amplitude $\zeta = 1$ and phase diffusion of strength **(a)** $\sigma = 0.5$ and **(b)** $\sigma = 0.75$.

2.7 Squeezed States

Another useful type of quantum states for optical quantum communication are squeezed states which are characterized by their quadrature fluctuation properties. The fluctuations for one quadrature are suppressed below the vacuum level whereas the complementary quadrature fluctuations are increased. Unlike a coherent state, a squeezed state is highly non-classical. While not as readily accessible as coherent states, squeezed states can be generated in several ways. Most notably, they are produced via parametric down-conversion in a second-order nonlinear crystal or via the Kerr effect in an optical fibre [Andersen et al. 2016].

First investigated by Yuen [1976], a squeezed coherent state

$$|\, z, \zeta \,\rangle = \hat{S}\,(z)\, \hat{D}\,(\zeta)\,|\, 0 \,\rangle = \hat{S}\,(z)\,|\, \zeta \,\rangle \tag{2.81}$$

emerges by applying the unitary squeezing operator

$$\hat{S}\,(z) = \exp\left[\frac{1}{2}\left(z^* \hat{a}^2 - z\left(\hat{a}^\dagger\right)^2\right)\right] \tag{2.82}$$

to a coherent state. The complex squeezing parameter $z = re^{i\phi_z}$ defines the degree of squeezing $|z| = r$ and its direction ϕ_z. Let x_{ϕ_z} be the squeezed and p_{ϕ_z} the complementary quadrature. Then, the corresponding variances are given by

$$\left\langle\, (\Delta x_{\phi_z})^2 \,\right\rangle = \frac{1}{2} e^{-2r} \quad ; \quad \left\langle\, (\Delta p_{\phi_z})^2 \,\right\rangle = \frac{1}{2} e^{2r}. \tag{2.83}$$

The variance of one quadrature is now below the level of vacuum fluctuations while the other quadrature fluctuations are above it. It is only possible to suppress one quadrature variance at the cost of the other due to Heisenberg's uncertainty principle [Heisenberg

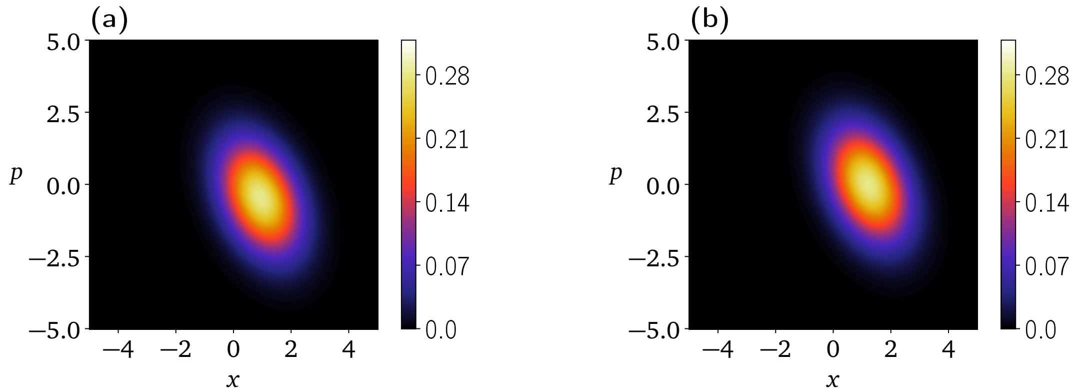


Figure 2.6: Husimi-Q-function of **(a)** a squeezed coherent state $|\, z, \zeta \,\rangle$ and **(b)** an ideal squeezed state $|\, \zeta, z \,\rangle$ both with mean photon number $\bar{n} = 1$ and squeezing parameter $z = 0.5e^{i\pi/4}$. While the ideal squeezed state retains the position of a corresponding coherent state the squeezed coherent state experiences an additional phase shift.

1927].

The P-function of a squeezed state is given by [Schleich 2001]

$$P_{\mathrm{sq}}(\beta) = \exp\left[\frac{1-r}{8r}\frac{\partial^2}{\partial\beta_x^2} - \frac{1-s}{8}\frac{\partial^2}{\partial\beta_p^2}\right]\delta(\beta-\zeta), \tag{2.84}$$

where β_x and ζ_x are the quadrature values in the squeezed direction x_{ϕ_z}, whereas β_p and ζ_p are the quadrature values of the conjugate quadrature p_{ϕ_z}. The P-function of a squeezed state involves an infinite order of derivatives of the delta function such that it is even more singular than the P-function of a Fock state (cf. eq. (2.70)). This demonstrates the non-classicality of squeezed states. The corresponding Q-function can be written as [Kim et al. 1989]

$$Q_{\mathrm{sq}}(\beta) = \frac{1}{\pi\cosh(r)}\exp\left[-2\frac{(\zeta_x-\beta_x)^2}{1+e^{2r}} - 2\frac{(\zeta_p-\beta_p)^2}{1+e^{-2r}}\right]. \tag{2.85}$$

In the Fock basis a squeezed coherent state is given by [Mandel and Wolf 1995]

$$|\, z, \zeta \,\rangle = \sum_{n=0}^{\infty}\sqrt{\frac{\left(e^{i\phi_z}\tanh(r)\right)^n}{2^n\cosh(r)\, n!}}\, H_n\left(\frac{\zeta}{\sqrt{2e^{i\phi_z}\sinh(r)\cosh(r)}}\right) e^{\frac{\zeta^2}{2}\left(e^{-i\phi_z}\tanh(r)-1\right)}\,|\, n \,\rangle\,, \tag{2.86}$$

where $H_n(x)$ are Hermite polynomials. A squeezed vacuum state can be retrieved from eq. (2.86) by setting $\zeta = 0$ which yields

$$|\, z, 0 \,\rangle = \sum_{n=0}^{\infty}\frac{\left(-e^{i\phi_z}\tanh(r)\right)^n}{\sqrt{2^n\cosh(r)\, n!}}\,|\, 2n \,\rangle\,, \tag{2.87}$$

since $H_n(0) = 0$ for odd n. Squeezed vacuum contains only even Fock states and consequently has a highly singular P-function.
The average number of photons

$$\langle z, \zeta \,|\, \hat{n} \,|\, z, \zeta \rangle = |\zeta|^2 + \sinh^2(r) \tag{2.88}$$

of a squeezed state defined in eq. (2.86) differs from a coherent state defined in eq. (2.77) as the squeezing operation adds energy to the state. When a coherent state is squeezed the resulting state does not simply retain its position in phase space, but also experiences a phase shift as illustrated in fig. 2.6. In order to produce a squeezed state that is centered at the same position as the corresponding coherent state, it is necessary to first squeeze the vacuum state and then apply a displacement operation. This produces the so-called ideal squeezed state [Caves 1981; Mandel and Wolf 1995]

$$|\, \zeta, z \rangle = \hat{D}(\zeta)\, \hat{S}(z) \,|\, 0 \rangle \,. \tag{2.89}$$

In fig. 2.6 both, the ideal squeezed state and the squeezed coherent state, are illustrated. The degree and direction of the squeezing are the same, but the squeezed coherent state is nevertheless slightly shifted in phase space. A connection between the two types of squeezed states can be made with the relation $|\, \zeta, z \rangle = \left|\, z, \tilde{\zeta} \right\rangle$ [Mandel and Wolf 1995], where

$$\tilde{\zeta} = \zeta \cosh(r) + \zeta^* e^{i\phi_z} \sinh(r) \,. \tag{2.90}$$

An explicit expression for the ideal squeezed state can be obtained by applying eq. (2.90) to eq. (2.86). Note that the ideal squeezed state and the squeezed coherent state have the same average energy written in eq. (2.88).

2.8 Thermalized States

When a state interacts with a heated environment the state itself becomes thermalized. Effectively, thermal environmental interactions incoherently add photons to the transmitted state. The strength of thermal noise can be characterized by the average number of thermal photons n_b added to a state. The distribution of these photons follow Bose-Einstein statistics and consequently, the density matrix $\hat{\varrho}_{\text{th}}$ of a thermal (or thermalized vacuum) state itself in the Fock basis reads [Helstrom 1976]

$$\hat{\varrho}_{\text{th}} = \frac{1}{n_b + 1} \sum_k \left(\frac{n_b}{n_b + 1} \right)^k |\, k \,\rangle \langle\, k \,| \,. \tag{2.91}$$

Clearly, a thermal state is characterized by an incoherent mixture of photon number states $|\, k \,\rangle$. The temperature T of the thermal state is connected to the average number of thermal photons n_b by the relation

$$T = \frac{2\pi\hbar f}{k_B \ln\left[\frac{n_b}{1+n_b}\right]}, \tag{2.92}$$

where k_B is the Boltzmann constant and f the considered frequency.
Naturally, not only the vacuum state is affected by thermal noise. For an initial state $\hat{\varrho}$ thermal noise effectively transforms the state according to

$$\hat{\varrho} \rightarrow \frac{1}{\pi n_b} \int \mathrm{d}^2\beta \, e^{-\frac{|\beta|^2}{n_b}} \hat{D}(\beta) \, \hat{\varrho} \hat{D}^\dagger(\beta) \, . \tag{2.93}$$

If a coherent state $|\,\zeta\,\rangle$ is thermalized, eq. (2.93) can be simplified with the help of eq. (2.79) and the state is formally given by a displaced thermal state, i.e.

$$\hat{\varrho}_{\mathrm{c,th}} = \hat{D}(\zeta) \, \hat{\varrho}_{\mathrm{th}} \hat{D}^\dagger(\zeta) \, . \tag{2.94}$$

Note that as long as the system is in thermal equilibrium, the state of eq. (2.94) is equivalent to the one obtained by first displacing the vacuum state and then adding thermal photons [Oz-Vogt et al. 1991]. The Q and P-functions of a thermalized coherent state are both Gaussians and given by [Schleich 2001]

$$Q_{\mathrm{c,th}}(\beta) = \frac{1}{\pi(n_b+1)} e^{-\frac{|\beta-\zeta|^2}{n_b+1}} \, , \tag{2.95}$$

$$P_{\mathrm{c,th}}(\beta) = \frac{1}{\pi n_b} e^{-\frac{|\beta-\zeta|}{n_b}} \, . \tag{2.96}$$

In fig. 2.7(a) and (b) examples of thermalized coherent states are shown. The effect of thermal noise is a broadening of the state in phase space whose strength depends on the number of background photons n_b. An element of the density matrix $\hat{\varrho}_{\mathrm{c,th}}$ in the Fock basis reads [Helstrom 1976]

$$\langle\, n \,|\, \hat{\varrho}_{\mathrm{c,th}} \,|\, m \,\rangle = \sqrt{\frac{n!}{m!}} \frac{n_b^m}{(1+n_b)^{m+1}} \left(\frac{\zeta^*}{n_b}\right)^{m-n} e^{-\frac{|\zeta|^2}{1+n_b}} L_n^{(m-n)}\left(-\frac{|\zeta|^2}{n_b(1+n_b)}\right) , \tag{2.97}$$

where $m \geq n$ and $L_n^{(m)}(x)$ denotes the generalized Laguerre polynomial. The matrix elements for $m < n$ are obtained via the relation $\langle\, n \,|\, \hat{\varrho}_{\mathrm{c,th}} \,|\, m \,\rangle = \langle\, m \,|\, \hat{\varrho}_{\mathrm{c,th}} \,|\, n \,\rangle^*$.
If squeezed states are affected by thermal noise, the resulting state is given by

$$\hat{\varrho}_{\mathrm{s,th}} = \hat{S}(z) \, \hat{\varrho}_{\mathrm{th}} \hat{S}^\dagger(z) \, , \tag{2.98}$$

where $\hat{S}(z)$ is the squeezing operator defined in eq. (2.82). Their P-function is equal to [Kim et al. 1989]

$$P_{\mathrm{sq,th}}(\beta) = \frac{\exp\left[-\frac{\beta_p e^{-r}}{\sinh(r)+n_b e^{r}} - \frac{\beta_x e^{r}}{-\sinh(r)+n_b e^{-r}}\right]}{\pi\sqrt{(\sinh(r)+n_b e^{r})(n_b e^{-r}-\sinh(r))}} . \tag{2.99}$$

The Q-function of a thermalized squeezed vacuum state is given by [Kim et al. 1989]

$$\begin{aligned} Q_{\mathrm{s,th}}(\beta) = & \frac{1}{\pi\sqrt{(n_b^2+2n_b)\cosh^2(r)+1}} \exp\left\{-\frac{\mathrm{Re}[\beta]^2}{2}\tanh(r) - |\beta|^2\right. \\ & \left. + \frac{n_b}{(n_b^2+2n_b)\cosh^2(r)+1}\left[(n_b+1)|\beta|^2 - n_b\frac{\mathrm{Re}[\beta]^2}{2}\tanh(r)\right]\right\} \end{aligned} \tag{2.100}$$

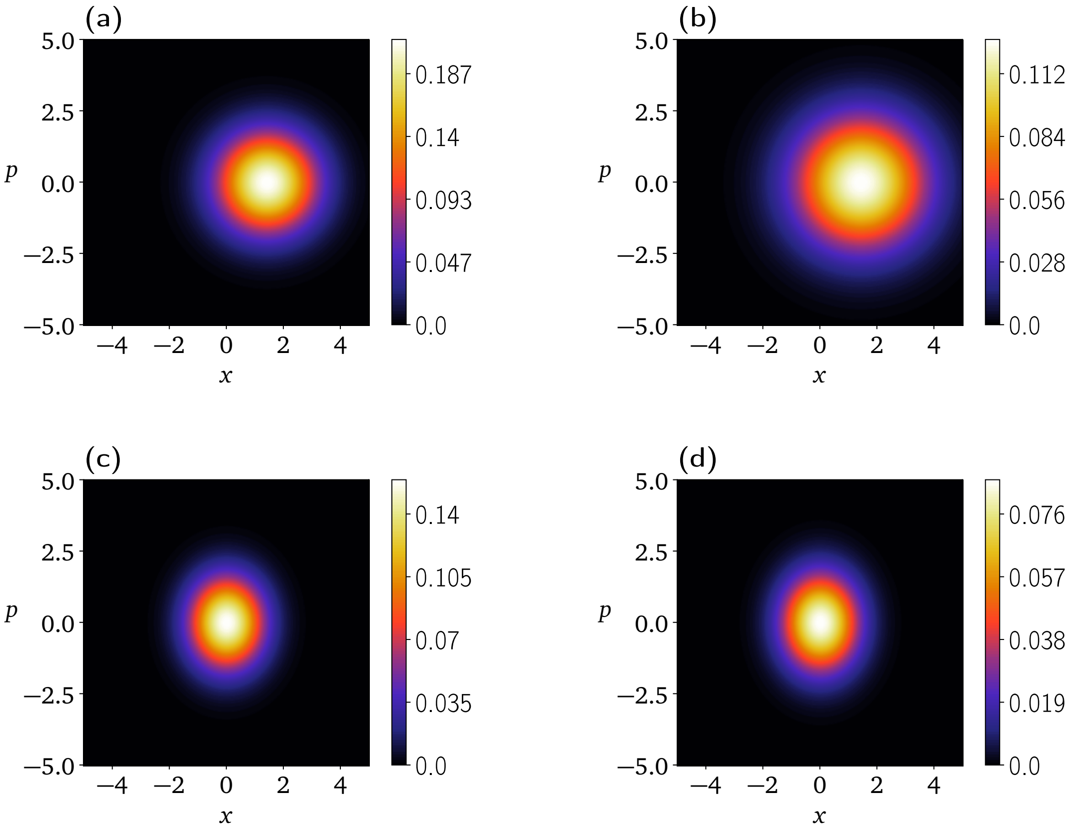


Figure 2.7: Husimi-Q-function of a thermalized **(a,b)** coherent or **(c,d)** squeezed state $\hat{\varrho}_{\mathrm{c,th}}$ with mean photon number $\bar{n} = 1$ and thermal background of **(a,c)** $n_b = 0.5$ and **(b,d)** $n_b = 1.5$ photons. The thermal background broadens the occupied area in phase space. Note that the used contour levels differ between the figures.

and depicted in fig. 2.7(c) and (d). Again, thermalization of a squeezed state broadens its occupied space in phase space similar as it did with coherent states. The broadening, however, is not equally strong in all directions as the squeezed quadrature is less affected. Thermal noise increases the energy of the state by adding the average number of thermal photons n_b. However, since the addition takes place incoherently the state looses coherence and becomes less pure. Even more, originally non-classical states evolve into more and more classical states.

3 Communication, Information and Capacities

Communication is omnipresent in our daily lives. Whether we are talking to another person, browsing the internet or sending a text message, all of these interactions depend on successfully transmitting messages between several parties. These communication acts take place in various ways, the mobile phone signal is broadcasted through the atmosphere, the TV receives its broadcast via cable and deep-space missions, like the Voyager satellites, transmit their signal through free space. All of these different scenarios can be broken down to several generic steps. First a message is prepared, then sent through a medium to the destination and finally, the receiving party detects the signal and attempts to decode the message.
In this chapter we review the basic concepts of communication. In section 3.1 the communication framework and the concept of a channel in particular are introduced. Section 3.2 establishes the error probability as a measure of performance for the detection of transmitted messages. Section 3.3 reviews a more general approach to quantifying performance with the help of entropy, mutual information or photon-information efficiency (PIE) and section 3.4 shows the capacities and limits of several important communication channels. Finally, section 3.5 presents several common ways to encoding messages.

3.1 Communication Channel

While (classical) communication can take various forms, all scenarios have several steps in common. A generic communication scenario with its building blocks is depicted in fig. 3.1. The term *communication channel* is somewhat ambiguous and commonly refers to state preparation, propagation and detection. It is entirely defined by the input and output symbols and their respective conditional probability distributions.
Before the actual transmission, the communication parties need to agree on a set of *symbols*, e.g. $\{0, 1, 2, \dots\}$, of which the messages will consist and the corresponding set of physical information carriers, the *alphabet* $\{\hat{\varrho}_i\}$. The latter are used to physically transmit the message from the sender to its destination point. Note that, in principle, there can also exists many destinations. However, this more general scenario will not be treated in this work.
The information carriers, also referred to as *input states*, are quantum states represented by density matrices $\hat{\varrho}_i$ and sent with their corresponding a-priori probability p_i. They form a full set such that $\sum_i p_i = 1$. Each state is attributed to one symbol, e.g. $0 \leftrightarrow \hat{\varrho}_0$ etc, such that after detecting a certain state the receiving party is able to translate the

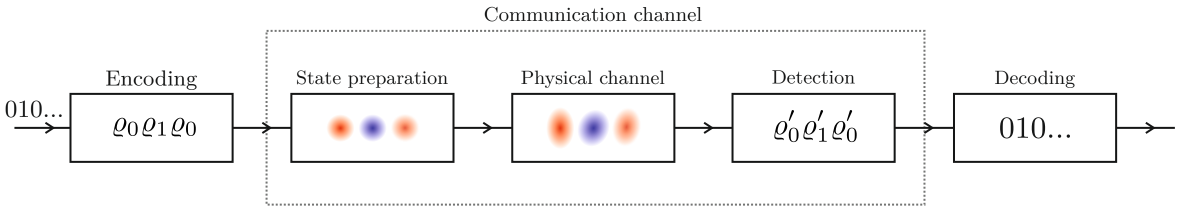


Figure 3.1: Schematic depiction of a generic communication scenario. First the message is encoded from a set of symbols into a sequence of physical information carrier states $\hat{\varrho}_i$, which are prepared by a transmitter and sent via a physical channel to their destination. The physical channel is noisy and alters the propagating states depending on the particular properties of the channel. At the destination the states $\hat{\varrho}'_i$ are detected by a receiver and an attempt is made to translate them back to the original message in symbols. In general, the detection is erroneous, since the states may not be perfectly distinguishable or they arrive distorted due to noise.

message back into symbols.

In optical communication the alphabet consists of states of light and coherent states are widely used for this purpose. They are produced by a laser at a given carrier frequency with high repetition rates and are good candidates for carrying information. Therefore, we will mostly consider alphabets consisting of coherent states, i.e. $\{\hat{\varrho}_i = |\,\zeta_i\,\rangle\,\langle\,\zeta_i\,|\}$, in this work.

After agreeing on a set of symbols and their corresponding states, the ground is set for the actual communication to take place. First, the sending party translates the message into a string of symbols, a procedure called *encoding*. Then for each symbol the corresponding state is prepared by a transmitter in order to sent them to their destination. Since the step of encoding the message into symbols and then into physical states is not of particular interest for this work, we will use the term “encoding” loosely also for the chosen set of physical states.

After a signal state is prepared, it is propagated over the *physical channel* connecting the transmitter and the receiver. In any realistic scenario the propagation is noisy and the states arrive distorted. The receiving party is now tasked to measure the arriving state $\hat{\varrho}'_i$ and to guess which input state $\hat{\varrho}_i$ actually was sent. Doing so, the detected states are translated back into symbols and finally the message is (hopefully) correctly recovered. This task is critical for successful communication at a high data rate. It is, however, inherently erroneous as it is impossible for the receiver to distinguish the states perfectly due to noise picked up in the channel, imperfections in the detector and fundamental limitations like quantum fluctuations or the non-orthogonality of coherent states.

There exist two fundamental ways of describing a channel. The first refers directly to the physical channel where a particular model for the signal propagation has to be found. These models naturally depend on the physical interaction between the propagating

signals and its environment, which can be treated classically or quantum. In atmospheric channels, for instance, these interactions are turbulences or scattering [Mahalov and Moustaoui 2010] whereas in optical fibre channels scattering and Kerr nonlinearities are central [Senior 2009].
The second way does not describe the particular interactions directly, but on the level of the communication channel and its symbols. Here, the symbols are treated as random variables and it suffices to define a statistical noise model between the input and the output. The particular model needs to be justified by an underlying physical noise model which is mimicked by the random process. A widely used statistical model is *additive white Gaussian noise (AWGN)* describing background noise that is present in most channels [Essiambre et al. 2010].
The noise models ultimately define the channel and its limits on the transmission of information. With unlimited resources, however, an arbitrary amount of information could be transmitted. Yet, in realistic scenarios resources are limited and in order to gain applicable insights these limitation must be taken into account. The most common constraint is a limitation on the available energy. One possible scenario is a transmitter that only has a certain amount of energy available within a time frame which can be distributed freely only within that frame. By using the number of photons as a measure of energy we can define the constraint via the *mean photon number* $\bar{n}$ of an alphabet of coherent states. It is given as the expectation value of the complex amplitude, i.e.

$$\bar{n} = \mathbb{E}\left[|\zeta|^2\right] = \mathbb{E}\left[\mathrm{Re}\left[\zeta\right]^2\right] + \mathbb{E}\left[\mathrm{Im}\left[\zeta\right]^2\right]. \tag{3.1}$$

For a discrete distribution the mean photon number reads

$$\bar{n} = \sum_i p_i \left|\zeta_i\right|^2. \tag{3.2}$$

Another possible type of constraint is an upper limit on the peak energy $n_{\max}$ of the transmitter. For coherent states this would mean that

$$n_{\max} \geq \left|\zeta_i\right|^2 \tag{3.3}$$

for all states in the alphabet.
Based on this framework it is now necessary to define suitable measures for the performance of a particular communication system or channel. In the following sections several approaches to quantify performance are reviewed. In section 3.2 the error probability, in section 3.3 entropy, information, mutual information and photon information efficiency are introduced. In section 3.5 several important types of encodings are presented and finally in section 3.4 the ultimate limits imposed on channels (their capacities) are discussed.

3.2 Error Probability

The most straightforward way to measure how well a message was received is to look at the probability of making a mistake when guessing the received state. This probability is

called the *error probability* p_{err}. While it is a rather easy to calculate quantity, the error probability suffers from a couple of drawbacks. Before turning to a formal definition we will consider the following example to showcase some of its features.
Let us assume that we are communicating with an alphabet consisting of two states, the vacuum state $\hat{\varrho}_0 = |\,0\,\rangle\,\langle\,0\,|$ and a coherent state $\hat{\varrho}_1 = |\,\zeta\,\rangle\,\langle\,\zeta\,|$, both sent with the probability $p_0 = p_1 = \frac{1}{2}$. For simplicity, let the transmission take place noiseless. Our task is now to identify the transmitted states. In such a procedure it is possible to make two errors: we can guess for $\hat{\varrho}_0$ when actually $\hat{\varrho}_1$ was sent and vice versa. A trivial guessing strategy is to randomly choose either of the states with their a-priori probability $\frac{1}{2}$. Then, the average error probability over many tries would be $p_{\mathrm{err}} = \frac{1}{2}$. This strategy of truly guessing the state defines the upper limit for p_{err} as we can always fall back on it regardless of any available measurement result. If the alphabet size would be different or the a-priori probabilities would change, the upper limit would change too. For instance, if $\hat{\varrho}_0$ is sent with $p_0 = \frac{1}{3}$ and $\hat{\varrho}_1$ sent with $p_1 = \frac{2}{3}$ the error probability would be $p_{\mathrm{err}} = \frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3} = \frac{4}{9}$. In short, the error probability has an upper bound which depends on the a-priori probabilities p_i.
A non-trivial strategy is to base the guess on obtained measurement results. One possible measurement is given by the POVM $\left\{|\,0\,\rangle\,\langle\,0\,|\,, \hat{\mathbb{1}} - |\,0\,\rangle\,\langle\,0\,|\right\}$ which enables us to test whether we received the vacuum state or not. Here, we never will make an error if the vacuum state was transmitted. However, if the coherent state $|\,\zeta\,\rangle$ was sent, we will be mistaken in some of the cases, since the coherent state has a non-zero overlap with the vacuum state $|\langle\,0\,|\,\zeta\,\rangle|^2 = e^{-|\zeta|^2}$. In this scenario the error probability is not symmetric which needs to be taken into account when quantifying the performance of the communication system.
To formalize the notion of error probability, we assume from now on that the state $\hat{\varrho}_i$ is associated with the symbol i such that $i \leftrightarrow \hat{\varrho}_i$. The probability of guessing the symbol j, when symbol i was sent, is given by the conditional probability $p\,(j|i)$. The overall probability of error is obtained by the sum of all $p\,(j|i)$ with $j \neq i$ weighted by the a-priori probability of the symbol i such that it reads

$$p_{\mathrm{err}} = \sum_i \sum_{j \neq i} p_i p\,(j|i) = 1 - \sum_i p_i p\,(i|i)\,, \tag{3.4}$$

where i defines the input and j the guessed output symbol. In the last step we used the fact that $\sum_j p\,(j|i) = 1$. The term $\sum_i p_i p\,(i|i)$ is also known as the average success probability. By taking into account the a-priori probabilities the error probability becomes symmetric. We can find an upper bound on the error probability by simply guessing the received symbol with the respective probabilities equal to the a-priori probabilities of the symbols, i.e. $p\,(i|i) = p_i$, such that

$$p_{\mathrm{err}} \leq 1 - \sum_i p_i^2. \tag{3.5}$$

For an alphabet of size M, where all states are sent with equal probabilities $p_i = \frac{1}{M}$, the upper bounds reduces to

$$p_{\mathrm{err}} \leq \frac{M-1}{M}. \tag{3.6}$$

This reveals a major down-side of the error probability as a measure for performance. Its upper limit depends on the a-priori probabilities and consequently, the error probability is only suited to compare the performance of different systems when using the same alphabet with the same a-priori probabilities.
Physically, the conditional probabilities $p\left(j|i\right)$ are obtained by measuring the output states $\hat{\varrho}'_i$ with a POVM $\{\hat{\pi}_i\}$, where the measurement operator $\hat{\pi}_i$ is associated with the outcome i. The error probability of eq. (3.4) then reads

$$p_{\mathrm{err}} = \sum_i \sum_{j\neq i} p_i \operatorname{Tr}\left[\hat{\varrho}'_i \hat{\pi}_j\right] = 1 - \sum_i p_i \operatorname{Tr}\left[\hat{\varrho}'_i \hat{\pi}_i\right] \tag{3.7}$$

according to the Born rule of eq. (2.8). Overall, the error probability is a simple to calculate and intuitive measure. However, it is necessary to define a specific POVM and decoding strategy in order to determine the error probability. One might therefore raise the question whether the chosen measurement is actually the optimal one. Below we will discuss this issue in detail in a few specific scenarios which are particularly relevant to this thesis.

3.2.1 Minimum Error Probability

In order to find the *minimum error probability* $p_{\min}$ we have to minimize eq. (3.7). The POVM achieving the minimal error probability must satisfy the necessary and sufficient conditions [Helstrom 1976; Barnett and Croke 2009; Chefles 2000]

$$\sum_i p_i \hat{\varrho}_i \hat{\pi}_i - p_j \hat{\varrho}_j \geq 0 \text{ for all } j, \tag{3.8}$$

$$\hat{\pi}_i \left(p_i \hat{\varrho}_i - p_j \hat{\varrho}_j\right) \hat{\pi}_j = 0 \text{ for all } i, j. \tag{3.9}$$

The first inequality stems from the requirement that the error probability is an extremum for the optimal POVM and the second condition ensures that it is a minimum rather than a maximum. Note that the latter condition can be derived from the first one combined with the fact that the involved operators are positive [Barnett and Croke 2009]. For problems involving only two states the minimum error probability is, in principle, known. Note, however, that only for particular situations a physical implementation of the minimum error measurement was found.

Binary decisions

The case of discriminating two states, $\hat{\varrho}_0$ and $\hat{\varrho}_1$, was first solved in the seminal work by Helstrom [1976]. Here, the POVM consists of two elements $\{\hat{\pi}_0; \hat{\pi}_1\}$ with $\hat{\pi}_1 = \hat{\mathbb{1}} - \hat{\pi}_0$

and the error probability of eq. (3.7) can be written as

$$p_{\text{err}} = p_0 \operatorname{Tr}[\hat{\pi}_1 \hat{\varrho}_0] + p_1 \operatorname{Tr}[\hat{\pi}_0 \hat{\varrho}_1] \tag{3.10}$$

$$= \frac{1}{2} \left\{ p_0 \operatorname{Tr}\left[\left(\hat{\mathbb{1}} - \hat{\pi}_0\right) \hat{\varrho}_0\right] + p_1 \operatorname{Tr}[\hat{\pi}_0 \hat{\varrho}_1] + p_0 \operatorname{Tr}[\hat{\pi}_1 \hat{\varrho}_0] + p_1 \operatorname{Tr}\left[\left(\hat{\mathbb{1}} - \hat{\pi}_1\right) \hat{\varrho}_1\right] \right\} \tag{3.11}$$

$$= \frac{1}{2} - \frac{1}{2} \left\{ \operatorname{Tr}\left[\hat{\pi}_0 \hat{A}\right] - \operatorname{Tr}\left[\hat{\pi}_1 \hat{A}\right] \right\}, \tag{3.12}$$

where in the last line we introduced the Hermitian operator

$$\hat{A} = p_0 \hat{\varrho}_0 - p_1 \hat{\varrho}_1. \tag{3.13}$$

The error probability becomes minimal, if the term $\operatorname{Tr}\left[\hat{\pi}_0 \hat{A}\right] - \operatorname{Tr}\left[\hat{\pi}_1 \hat{A}\right]$ is maximized. This is achieved by choosing a POVM such that $\hat{\pi}_0$ is a projection onto the subspace spanned by the eigenvectors corresponding to the positive eigenvalues of $\hat{A}$. Consequently, $\hat{\pi}_1$ has to be a projection onto the negative eigenvalue subspace. Considering the spectral eigendecomposition $\hat{A} = \sum_{i=1}^{n} \lambda_i \,|\, i \,\rangle \langle\, i \,|$ of $\hat{A}$, where $i \in \{1, \ldots, k\}$ denote the positive eigenvalues and $i \in \{k+1, \ldots, n\}$ the negative eigenvalues, we can write eq. (3.12) as

$$p_{\min} = \frac{1}{2} - \frac{1}{2} \left\{ \sum_{i=1}^{k} \lambda_i - \sum_{i=k+1}^{n} \lambda_i \right\} \tag{3.14}$$

$$= \frac{1}{2} - \frac{1}{2} \sum_{i=1}^{n} |\lambda_i|. \tag{3.15}$$

Therefore, the minimum error probability $p_{\min}$ depends on the sum of the absolute values of the eigenvalues of $\hat{A}$. It can be conveniently expressed in terms of the trace distance T between two quantum states $\hat{\varrho}$ and $\hat{\sigma}$, which is defined as

$$T(\hat{\varrho}, \hat{\sigma}) = \frac{1}{2} \operatorname{Tr}\left[\sqrt{(\hat{\varrho} - \hat{\sigma})^{\dagger} (\hat{\varrho} - \hat{\sigma})}\right]. \tag{3.16}$$

Given an eigendecomposition of the Hermitian operator $(\hat{\varrho} - \hat{\sigma}) = \sum_i \lambda_i \,|\, i \,\rangle \langle\, i \,|$, it is straightforward to see that the trace distance is given by

$$T(\hat{\varrho}, \hat{\sigma}) = \frac{1}{2} \sum_i |\lambda_i| \tag{3.17}$$

such that we can write the minimum error probability as [Hayashi 2006; Audenaert et al. 2007]

$$p_{\min} = \frac{1}{2} - T(p_0 \hat{\varrho}_0, p_1 \hat{\varrho}_1). \tag{3.18}$$

Interestingly, calculating the trace distance between two states corresponds to maximizing the classical total variation distance between the two probability distributions $\{\mathrm{Tr}[\hat{\pi}_i\hat{\varrho}]\}$ and $\{\mathrm{Tr}[\hat{\pi}_i\hat{\sigma}]\}$ created by the POVM $\{\hat{\pi}_i\}$, i.e. [Nielsen and Chuang 2009]

$$\begin{aligned} T(\hat{\varrho},\hat{\sigma}) &= \frac{1}{2}\max_{\{\hat{\pi}_i\}}\sum_i \left|\mathrm{Tr}[\hat{\pi}_i\hat{\varrho}] - \mathrm{Tr}[\hat{\pi}_i\hat{\sigma}]\right| \\ &= \frac{1}{2}\max_{\{\hat{\pi}_i\}}\sum_i \left|\mathrm{Tr}[\hat{\pi}_i(\hat{\varrho}-\hat{\sigma})]\right|. \end{aligned} \tag{3.19}$$

The appearance of the trace distance in eq. (3.18) exemplifies that the choice for the POVM projecting onto the eigenvalues is indeed the optimal one. The POVM for which the total variation distance attains the trace distance also constitutes the minimum error measurement.

Pure states

For two pure states $\hat{\varrho}_0 = |\,\Psi_0\,\rangle\langle\,\Psi_0\,|$ and $\hat{\varrho}_1 = |\,\Psi_1\,\rangle\langle\,\Psi_1\,|$ the minimum error probability and its POVM can be cast into a simpler form. Here, we can express the states in a generic orthonormal basis $|\,\pm\,\rangle$ as follows

$$|\,\Psi_0\,\rangle = \cos(\theta)\,|\,+\,\rangle + \sin(\theta)\,|\,-\,\rangle\,, \tag{3.20}$$

$$|\,\Psi_1\,\rangle = \cos(\theta)\,|\,+\,\rangle - \sin(\theta)\,|\,-\,\rangle\,, \tag{3.21}$$

where θ is defined by the overlap $\langle\,\Psi_0\,|\,\Psi_1\,\rangle = \cos(2\theta)$. Using this basis, the operator $\hat{A}$ defined in eq. (3.13) can be diagonalized and its eigenvalues and eigenvectors are

$$\lambda_\pm = \frac{1}{2}(p_0 - p_1 \pm \omega)\,, \tag{3.22}$$

$$|\,\lambda_\pm\,\rangle = \frac{1}{\sqrt{2}}\left[\sqrt{1\pm\upsilon}\,|\,+\,\rangle \pm \sqrt{1\mp\upsilon}\,|\,-\,\rangle\right] \tag{3.23}$$

with $\omega = \sqrt{1-4p_0p_1\cos^2(2\theta)}$ and $\upsilon = (p_0-p_1)\cos(2\theta)\,\omega^{-1}$. The optimal POVM is the projection onto $|\,\lambda_\pm\,\rangle$,

$$\hat{\pi}_0 = |\,\lambda_+\,\rangle\langle\,\lambda_+\,| \quad ; \quad \hat{\pi}_1 = |\,\lambda_-\,\rangle\langle\,\lambda_-\,|\,, \tag{3.24}$$

and the minimum error probability of eq. (3.12) can be cast into the form

$$p_{\min} = \frac{1}{2}(1-\omega) = \frac{1}{2}\left(1 - \sqrt{1-4p_0p_1\left|\langle\,\Psi_0\,|\,\Psi_1\,\rangle\right|^2}\right). \tag{3.25}$$

It is seen in eq. (3.25) that $p_{\min}$ depends on the a-priori probabilities and the overlap of the two pure states. The minimum error probability is closely related to the fidelity F between two pure states. For two arbitrary quantum states $\hat{\varrho}$ and $\hat{\sigma}$, fidelity is defined as [Nielsen and Chuang 2009]

$$F(\hat{\varrho},\hat{\sigma}) = \mathrm{Tr}\left[\sqrt{\sqrt{\hat{\varrho}}\hat{\sigma}\sqrt{\hat{\varrho}}}\right], \tag{3.26}$$

which reduces to $F(|\Psi_0\rangle, |\Psi_1\rangle) = |\langle\Psi_0|\Psi_1\rangle|$ for pure states. With this eq. (3.25) reads

$$p_{\min} = \frac{1}{2}\left(1 - \sqrt{1 - 4p_0p_1F(|\Psi_0\rangle, |\Psi_1\rangle)^2}\right) \tag{3.27}$$

$$= \frac{1}{2}\left(1 - \sqrt{1 - 4F(p_0|\Psi_0\rangle, p_1|\Psi_1\rangle)^2}\right). \tag{3.28}$$

This expression can also be derived directly from the general expression for the minimum error probability of eq. (3.18), since for pure states the trace distance $T(\hat{\varrho}_0, \hat{\varrho}_1)$ is equivalent to the fidelity $F(\hat{\varrho}_0, \hat{\varrho}_1)$ via the relation

$$T(|\Psi_0\rangle, |\Psi_1\rangle) = \sqrt{1 - F(|\Psi_0\rangle, |\Psi_1\rangle)^2}. \tag{3.29}$$

For general mixed states such a direct relation between the fidelity and the trace distance does not exist, with the exemption of the Fuchs–van de Graaf inequalities [Fuchs and van de Graaf 1999]. In the special case where both states are sent with the same probability $p_0 = p_1 = \frac{1}{2}$, the optimal POVM of eq. (3.24) is given by projections on the eigenstates

$$|\lambda_+\rangle = \frac{|+\rangle + |-\rangle}{\sqrt{2}} \quad ; \quad |\lambda_-\rangle = \frac{|+\rangle - |-\rangle}{\sqrt{2}}. \tag{3.30}$$

As shown in fig. 3.2 the two states are symmetrically located around the input states. For non-equal a-priori probabilities the eigenstates are rotated towards the more likely input state. In case of two orthogonal input states, the minimum error measurement is naturally given by projections onto the states themselves.

Multiple copies

So far we considered only binary decisions with a single copy available for the measurement. If instead N copies are available a better informed decision can be made. In principle, the minimum error probability is given by the expression

$$p_{\min}^N = \frac{1}{2} - T\left(p_0\hat{\varrho}_0^{\otimes N}, p_1\hat{\varrho}_1^{\otimes N}\right), \tag{3.31}$$

similar to eq. (3.18). An upper bound on the error probability is given by the Chernoff bound which states that [Audenaert et al. 2007; Calsamiglia et al. 2008]

$$p_{\min}^N \leq \min_{0\leq s\leq 1} p_0^s p_1^{1-s}\,\mathrm{Tr}\left[\hat{\varrho}_0^s\hat{\varrho}_1^{1-s}\right]^N. \tag{3.32}$$

Furthermore, the minimum error probability is proportional to

$$p_{\min}^N \sim e^{-NC_Q} \text{ with } C_Q = -\lim_{n\to\infty}\frac{\ln\left[p_{\min}^N\right]}{N} = -\ln\left[\min_{0\leq s\leq 1}\mathrm{Tr}\left[\hat{\varrho}_0^s\hat{\varrho}_1^{1-s}\right]\right], \tag{3.33}$$

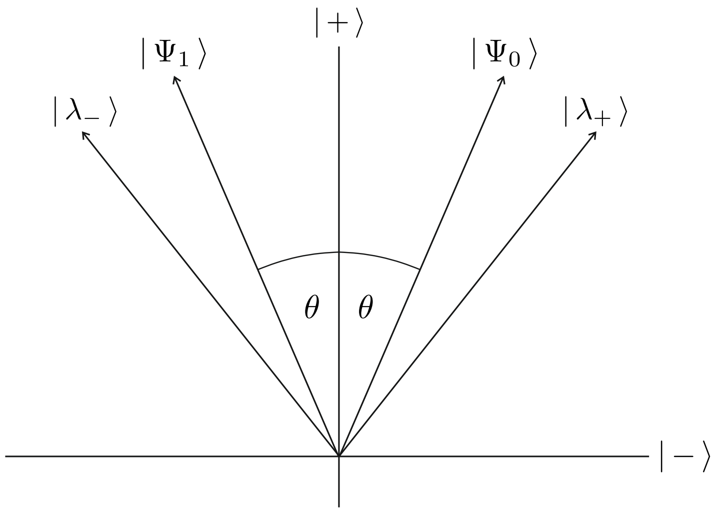


Figure 3.2: Illustration of the POVM (cf. eq. (3.30)) achieving the minimum error probability of eq. (3.25) for two equiprobable pure states $|\Psi_0\rangle$ and $|\Psi_1\rangle$. It is given by two projections on the states $|\lambda_\pm\rangle$ located symmetrically around the input states.

i.e. it decreases exponentially with the number of copies N. Note that eq. (3.33) only provides the leading order in the exponent. In general, other non-vanishing functions of N can appear [Audenaert et al. 2012; Li 2014]. However, to achieve the minimum error probability of eq. (3.32) collective measurements on all copies are needed in general. Since individual measurements only form a subset of collective ones this does not come as a surprise. Unfortunately, it is a notoriously hard task to find these collective POVMs and their physical implementation.
While the minimum error probability is an easy to calculate figure of merit for the performance of a particular communication channel it does not provide the ultimate limit on the information transfer. When calculating the error probability we already need to define a rulebook for decoding the message (i.e. the POVM) such that we know which physical detection event we interpret as which symbol. Of course it is desirable to quantify the performance in a way that does not impose this limitation and directly characterizes the amount of information that is transmitted. In section 3.3 the concept of mutual information is introduced which is able to quantify the channels successful transmissions in a more general fashion, where we do not need to define a particular decoding strategy.

3.3 Entropy and Mutual Information

Up to this point we only measured the communication performance by the probability of making an error, but ultimately we are interested in how much information can be sent regardless of errors. In order to quantify this we need to find an appropriate measure for the information content of an event. Certainly, the measure needs to fulfill a couple

of properties dictated by intuition.
First of all, a rare event should carry more information than a common event. Suppose, for instance, the weather is sunny virtually every day and only on a few days it is rainy. If the weather report confirms in the morning that the day will be sunny, little insight is gained in todays weather as this was expected anyway. If, however, the report predicts rainy weather for the day, a lot of information is gained, since it is an unexpected event.
Secondly, we would like information to be additive for completely independent events. Consider two separate and independent drawings of lottery numbers. If we learn the outcome of both, the amount of information gained should be the information content of one added to the other.
A function satisfying these desired properties is [Shannon 1948; MacKay 2003]

$$h\left(x\right)=\log_2\left[\frac{1}{p\left(x\right)}\right]=-\log_2\left[p\left(x\right)\right], \tag{3.34}$$

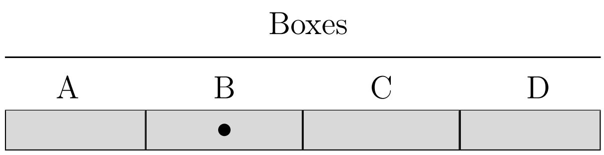
where $\log_2[p]$ is the logarithm to base 2 and x denotes the event that occurs with probability $p(x)$. Since we always associate the probability $p(x)$ with the event x we will use these notations interchangeably, i.e. $h(p(x)) = h(x)$. The quantity $h(x)$ is sometimes referred to as *Shannon information content* or *self-information* and is measured in bits if $\log_2$ is used. On some occasions it is more convenient to use the natural logarithm $\ln[x]$ to base e instead and quantify the information content in nats. Bits and nats can be directly converted into each other via the relation

$$1 \text{ nat} = \log_2(e) \text{ bits} \approx 1.44 \text{ bits}. \tag{3.35}$$

Equation (3.34) indeed satisfies the desired properties. Consider two independent events x and y. Their joint probability distribution is separable such that $p(x,y) = p(x)\,p(y)$ holds. Then, self-information is given by

$$\begin{aligned} h\left(x,y\right) &= \log_2\left[\frac{1}{p\left(x,y\right)}\right] = \log_2\left[\frac{1}{p\left(x\right)p\left(y\right)}\right] \\ &= \log_2\left[\frac{1}{p\left(x\right)}\right] + \log_2\left[\frac{1}{p\left(y\right)}\right] = h\left(x\right)+h\left(y\right), \end{aligned} \tag{3.36}$$

which shows clearly that the self-information of two independent events is additive.
Also, a rarer event carries more information than a more common one. Due to the definition of eq. (3.34) the condition $p(x) \leq p(y)$ directly implies $h(x) \geq h(y)$ such that the rarer event x indeed contains more information.
In order to further exemplify that the information content is well suited to describe information, consider a simple guessing game where four closed boxes, labelled A,B,C and D, are placed on a table. Beforehand, a single ball was put into one of the boxes. Our task is to guess which one contains a ball without any further prior knowledge. After an unsuccessful guess we are allowed to take another guess until we find the correct box. In table 3.1 the probabilities and the information content of possible series of guesses are shown for the exemplary case where the ball is in the second box. All other scenarios



Variant	Probability				Information content (bits)			
	A	B	C	D	A	B	C	D
1	0	$1/4$	0	0	0	2	0	0
2	$3/4$	$1/3$	0	0	0.415	1.585	0	0
3	$3/4$	$1/2$	$2/3$	0	0.415	1	0.585	0
4	$3/4$	0	$2/3$	$1/2$	0.415	0	0.585	1

Table 3.1: A game of guessing which box contains the ball. In the shown example the second box B contains the ball and marks the correct guess. For all possible guessing variants (up to permutations of wrong guesses) the probability for the event to occur and the corresponding information gain $\log_2\left[\frac{1}{p}\right]$ are shown.

are obtained by permutations of the boxes. With a probability of 25% the first guess will succeed and we learn the content of all four boxes which equals gaining 2 bits of information in a single guess. A wrong first guess gives us only 0.415 bits of information, since we only learn the content of a single box. With each guess the information content $h(x)$ grows until, finally, we reach 2 bits in total. A correct guess at any point reveals the remaining boxes and thus gives the missing bits of information.

3.3.1 Shannon Entropy

With the Shannon information content $h(x)$ we are able quantify the information contained in a particular event. However, most of the time we are interested in the outcome of a random process rather than a specific event that might occur. Luckily, the extension to this case is easily obtained.
Consider a random process X whose random variable x can take the value of any possible outcome, each obtained with its associated probability $p(x)$. The average information of a random process X is given by the expectation value of the information content $h(x)$ of all possible outcomes, i.e. it is defined as [Shannon 1948; MacKay 2003]

$$H(X) = \mathbb{E}[h(x)] = \sum_{x\in X} p(x)\log_2\left[\frac{1}{p(x)}\right] = -\sum_{x\in X} p(x)\log_2[p(x)] \tag{3.37}$$

and commonly referred to as the *Shannon entropy*. In the case of vanishing probability $p(x) = 0$ the convention $0\log_2[0] = 0$ holds. The Shannon entropy is measured in bits

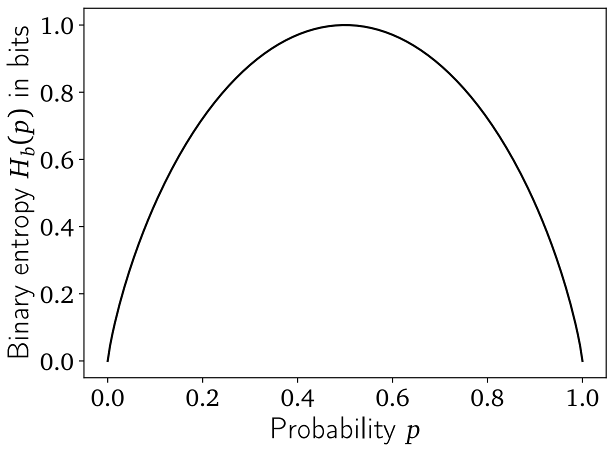


Figure 3.3: Binary entropy $H_b(p)$ of eq. (3.38) as a function of the probability p. The uncertainty is greatest if both outcomes are equiprobable.

or nats depending on the basis of the used logarithm. It is a non-negative quantity $H(X) \geq 0$ that reflects the uncertainty of a random process. The maximum value is obtained for a uniform probability distribution $p(x) = \frac{1}{N}$ where N denotes the number of outcomes. Contrary to the fact that a very rare event is the most informative, its contribution to the average information content is not necessarily significant. Consider a random process with two outcomes $\{x_1, x_2\}$ and their respective probabilities $\{p, 1-p\}$ which, for instance, is produced by a biased coin. The probability p denotes the success probability of the coin toss. Its entropy is called *binary entropy* $H_b(p)$ and given by

$$H_b(p) = -p \log_2[p] - (1-p) \log_2[1-p] \tag{3.38}$$

and shown in fig. 3.3. The most uncertain about whether we succeed is when $p = 0.5$, which also defines the maximum of the binary entropy. In other words, on average we gain the most information when tossing an unbiased coin. If the probability deviates to either side from 50%, the entropy decreases. If the random variable is continuous rather than discrete, the entropy can be defined analogously by replacing the sum with an integral as

$$H(X) = \int \mathrm{d}x \, p(x) \log_2 \left[\frac{1}{p(x)}\right]. \tag{3.39}$$

Another important quantity is the *joint entropy* of two random variables X and Y. It is defined as

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 [p(x,y)], \tag{3.40}$$

where $p(x,y)$ is the joint probability of x and y. For independent variables $p(x,y) = p(x)\,p(y)$ holds and it follows that the joint entropy is separable according to $H(X,Y) =$

$H(X) + H(Y)$. In fact, the joint entropy is subadditive, i.e.

$$H(X,Y) \leq H(X) + H(Y) \tag{3.41}$$

and equality is achieved only for independent variables. Subadditivity can be understood as the fact that looking at the joint state of two random variables can only reduce uncertainty, but never increase it.
We can also wonder about the uncertainty left in a random variable when gaining knowledge about another one. Consider the joint state of two variables X and Y with their joint entropy $H(X,Y)$. If one of the random variables is now measured, for the sake of concreteness Y, the amount of information gained is $H(Y)$. The uncertainty about X left is consequently given by

$$H(X|Y) = H(X,Y) - H(Y), \tag{3.42}$$

which defines the *conditional entropy* $H(X|Y)$. In terms of the conditional probability distribution $p(x|y)$, the conditional entropy reads

$$H(X|Y) = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2 [p(x|y)]. \tag{3.43}$$

If both random variables are identical $X = Y$, the conditional entropy vanishes $H(X|X) = 0$. In case they are independent, the conditional entropy reduces to the marginal entropy $H(X|Y) = H(X)$.

3.3.2 Mutual Information

With the various types of entropy defined above we are now in the position to consider how much information is conveyed through a communication channel. Let us denote the input by the random variable X and the output of the channel by Y. The amount of information gained is given by how much the uncertainty of the input X is reduced by measuring Y. In a perfect scenario the input would be recovered fully, i.e. $Y = X$, and the amount of information gained is equal to the entropy $H(X)$. Realistically, however, this is not the case. Due to noise or fundamental limitations the measured output will always be different from the input. The amount of information gained in any case is given by the *mutual information* $I(X;Y)$ which is defined as [MacKay 2003]

$$I(X;Y) = H(X) - H(X|Y) \tag{3.44}$$
$$= H(X) + H(Y) - H(X,Y), \tag{3.45}$$

where in the second line the definition of the joint entropy in eq. (3.42) was used. Since entropies are nonnegative, mutual information can never exceed the input entropy $H(X)$ due to eq. (3.44). Interestingly, eq. (3.45) implies that mutual information is symmetric in X and Y, i.e. $I(X;Y) = I(Y;X)$, and we gain the same amount of information about Y by measuring X. From the practical point of view it is often more useful to invoke symmetry and consider $I(X;Y) = H(Y) - H(Y|X)$ when calculating mutual information, since the conditional entropy $H(Y|X)$ is usually easier to express. In terms of probabilities mutual information can be written as

$$I(X;Y) = \sum_{x\in X} p(x) \sum_{y\in Y} p(y|x) \log_2 [p(y|x)] - \sum_{y\in Y} p(y) \log_2 [p(y)] \tag{3.46}$$

$$= \sum_{x,y} p(x,y) \log_2 [p(y|x)] - \sum_{x,y} p(x,y) \log_2 [p(y)] \tag{3.47}$$

$$= \sum_{x,y} p(x,y) \log_2 \left[\frac{p(x,y)}{p(x)\,p(y)}\right] \tag{3.48}$$

by using eqs. (3.37) and (3.43).
In order to showcase the usefulness of mutual information let us consider a communication scheme based on the example of table 3.1. Suppose the four boxes labelled A,B,C and D are used to transmit information. The boxes are connected and a ball is placed in one box to encode the symbol. Each box is used in 25% of the cases such that the input entropy is given by

$$H(X) = -\sum_{x\in X} p(x) \log_2 [p(x)] = \sum_{x\in X} \frac{1}{4} \log_2 [4] = 2. \tag{3.49}$$

Consequently, the maximum amount of information that can be conveyed is 2 bits. After placing the ball the boxes are sealed and transported to their destination where they are opened again. If the transport is flawless the ball arrives faithfully in the initial box and the conditional probabilities are $p(y|x) = \delta_{x,y}$. As a consequence the output entropy is

$$H(Y) = \sum_{y\in Y} \frac{1}{4} \log_2 [4] = 2 \tag{3.50}$$

and the conditional entropy $H(Y|X)$ vanishes as perfect knowledge about the input is gained, i.e. $X = Y$. Mutual information is therefore equal to $I(X;Y) = H(Y) = 2$ and indeed the full 2 bits are transmitted.
However, if the transport is faulty, mutual information will be less than 2 bits. For instance, suppose there is a chance of 25% that the ball moves to either of the adjacent boxes. The conditional probabilities for this scenario are written in table 3.2. The probability of obtaining a particular output remains $p(y) = \frac{1}{4}$ and the output entropy remains $H(Y) = 2$ as well. The conditional entropy, however, no longer vanishes since

$$H(Y|X) = -\sum_{x\in X} \frac{1}{4} \sum_{y\in Y} p(y|x) \log_2 [p(y|x)] = 0.723. \tag{3.51}$$

Therefore, mutual information for the flawed transportation is given by

$$I(X;Y) = H(Y) - H(Y|X) = 1.277 \tag{3.52}$$

Input x	Output y			
	A	B	C	D
A	$^3/_4$	$^1/_4$	0	0
B	$^1/_4$	$^1/_2$	$^1/_4$	0
C	0	$^1/_4$	$^1/_2$	$^1/_4$
D	0	0	$^1/_4$	$^3/_4$

Table 3.2: Conditional probabilities $p(y|x)$ for communicating by placing a ball into one of four boxes. Each of the boxes is used with a probability of 25%. During the transport there is a chance that the ball moves to one of the adjacent boxes.

and only 1.277 bits of information are transmitted on average.
Mutual information is a more general measure of performance than the error probability. It directly quantifies the amount of information that is conveyed from the input to the output. In principle, mutual information does not specify a decoding strategy whereas a defined decoding strategy is needed when calculating the error probability as discussed in section 3.2. In order to achieve the information transfer promised by mutual information the optimal strategy must be found first which is a highly non-trivial task [Moon 2005]. Only in some cases error probability and mutual information are equivalent quantities such as, for instance, when distinguishing binary alphabets of pure states [Tomassoni and Paris 2008].

3.3.3 Extension to Quantum Systems

The concept of entropy and mutual information can be extended to quantum systems. Remarkably, von Neumann introduced the quantum version of the entropy in 1932 first and following this Shannon adopted the concept to the classical theory in 1948. The *von Neumann entropy* of a quantum system X in state $\hat{\varrho}_X$ is defined as [von Neumann 1932]

$$S(\hat{\varrho}_X) = -\operatorname{Tr}\left[\hat{\varrho}_X \log_2\left[\hat{\varrho}_X\right]\right]. \tag{3.53}$$

For convenience we will drop the symbol $\hat{\varrho}$ and use the notation $S(\hat{\varrho}_X) = S(X)$ instead. Given the eigenvalues λ_i of the density matrix $\hat{\varrho}_X$, the entropy of eq. (3.53) can be expressed as

$$S(X) = -\sum_i \lambda_i \log_2\left[\lambda_i\right]. \tag{3.54}$$

Since density operators are positive semi-definite matrices with unit trace, the von Neumann entropy is non-negative and becomes zero if and only if the quantum state is pure. It is now straightforward to define the quantum version of the joint entropy of two systems X and Y as the entropy of the composite system $\hat{\varrho}_{X,Y}$ such that it reads

$$S(X,Y) = -\operatorname{Tr}\left[\hat{\varrho}_{X,Y}\log_2\left[\hat{\varrho}_{X,Y}\right]\right]. \tag{3.55}$$

For uncorrelated systems the joint entropy is again additive, i.e. $S(X,Y) = S(X) + S(Y)$. The quantum conditional entropy is defined as

$$S(X|Y) = S(X,Y) - S(Y) \tag{3.56}$$

following the classical approach of eq. (3.42). Note that, unlike the classical entropy, the quantum conditional entropy can become negative if two systems X and Y are entangled [Nielsen and Chuang 2009]. This also implies that it is possible to be more uncertain about the state of a system X than the joint state of X and Y.
Finally, the quantum mutual information is defined according to eq. (3.45) as

$$\begin{aligned} I_Q(X;Y) &= S(X) + S(Y) - S(X,Y) && (3.57)\\ &= S(X) - S(X|Y) = S(Y) - S(Y|X), && (3.58) \end{aligned}$$

where in the last line eq. (3.56) was used. Since the entropy for uncorrelated systems is additive, quantum mutual information $I_Q(X;Y)$ is unaffected by the addition of an uncorrelated system Z, i.e. $I_Q(X;Y) = I_Q(X;YZ)$. Also a couple of other intuitive properties hold. For example, measuring a system can never increase mutual information such that [Nielsen and Chuang 2009]

$$I_Q(X;Y) \geq I_Q(X';Y') \tag{3.59}$$

holds, where primes denote a quantum system after a measurement was performed. Similarly, discarding a system can only decrease mutual information, i.e. [Nielsen and Chuang 2009]

$$I_Q(X;YZ) \geq I_Q(X;Y). \tag{3.60}$$

In the following section 3.4 we will use mutual information and its properties to discuss the limits of optical communication.

3.4 Capacities and Limits

Mutual information $I(X;Y)$, as a measure for the transmitted information, is defined by the input and output probability distributions. It is important to note that the output distribution is fully determined by the input and the action of the channel, i.e. given a particular channel only the input distribution can be freely chosen. However, this is not entirely true from a physical perspective. Information theory and therefore mutual information sees the output distribution as given, since the measurement device is part of the channel. Naturally, we are free to choose the measurement which retrieves the

information from the physical information carrier and thus directly influences the output distribution. Having said this, we will discuss this question at a later point from the quantum mechanical perspective and first take the information theoretical approach.
In order to maximize the amount of information that can be conveyed via a particular channel, we need to find the optimal input distribution $p(x)$. The result of this maximization is the *capacity* C of a channel which is formally defined as [MacKay 2003]

$$C = \max_{p(x)} I(X;Y), \tag{3.61}$$

where the maximization is performed over all possible probability distributions $p(x)$. That this quantity indeed defines the maximum possible information transfer of a (memoryless) channel was proven by Shannon [1948] in his noisy-channel coding theorem. The latter guarantees that there exists a strategy to decode the message from the output that achieves the capacity C in the limit of a large number of channel uses. A channel use is defined as the act of sending a single symbol, which is then received at the output. In his seminal work Shannon only gave an outline of the proof and only much later the first rigorous proof was given by Feinstein [1954]. By now, many proofs can be found in textbooks such as [MacKay 2003; Cover and Thomas 2005]. Finding the capacity of a particular channel is a difficult task in general, since mutual information has to be maximized over all possible probability distributions including distributions where $p(x) = 0$ for some x. In the following, we will consider a few simple examples of known capacities.

Binary symmetric channel

The first example is the capacity of the *binary symmetric channel* (BSC) which is probably the simplest channel model possible. The binary symmetric channel consists of two input symbols $X = \{\mathrm{A}, \mathrm{B}\}$ that are sent with probabilities $\{p_\mathrm{A}, p_\mathrm{B} = 1 - p_\mathrm{A}\}$. Both are faithfully transmitted with probability $p(\mathrm{A}|\mathrm{A}) = p(\mathrm{B}|\mathrm{B}) = 1 - p$ and are falsely accounted for the other with probability $p(\mathrm{A}|\mathrm{B}) = p(\mathrm{B}|\mathrm{A}) = p$. An illustration of the binary symmetric channel is shown in fig. 3.4. Suppose a symbol was chosen and sent over the channel. Whether it arrives correctly is decided by a biased coin flip with the success probability $1 - p$ and consequently the conditional entropy equals the binary entropy $H(Y|X = x) = H_b(p)$ of eq. (3.38). Mutual information for the BSC is now given by

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) && (3.62)\\ &= H(Y) - \sum_{x\in X} p(x) H(Y|X = x) && (3.63)\\ &= H(Y) - H_b(p). && (3.64) \end{aligned}$$

The output entropy $H(Y)$ is defined by the output probabilities which read

$$\begin{aligned} p(Y = A) &= p_\mathrm{A}(1 - p) + (1 - p_\mathrm{A})\, p, && (3.65)\\ p(Y = B) &= p_\mathrm{A} p + (1 - p_\mathrm{A})(1 - p). && (3.66) \end{aligned}$$

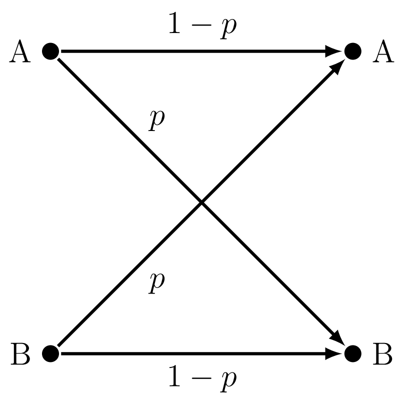


Figure 3.4: Sketch of a binary symmetric channel. Each symbol is correctly transmitted with probability $1-p$ and mistaken for the other symbol with probability p. The corresponding channel capacity is given by eq. (3.67).

Since the entropy $H(Y)$ is the entropy of a binary random variable that is at most 1 bit, we can bound eq. (3.64) from above and the capacity is given by [Cover and Thomas 2005]

$$C_{\mathrm{BSC}} = 1 - H_b(p) \tag{3.67}$$

in bits. In fig. 3.6 the capacity is shown as a function of the probability p. Since the binary entropy takes its maximum of 1 bit at $p = 0.5$ as shown in fig. 3.3, the capacity is minimal for this error probability. For $p > 0.5$ the capacity increases again until it reaches 1 bit at an error probability of $p = 1$. While this might seem odd at first glance, it makes perfect sense after a closer look. At $p = 0.5$ we are perfectly uncertain about the input, we might as well flip a fair coin to decide which symbol we received instead of performing a measurement. Naturally, it is impossible to infer any information in this case. If the error probability rises to $p \geq 0.5$, it is more likely that we received the wrong symbol instead of the correct one. Therefore, we simply switch the received symbol to the other symbol, which is equivalent to having an error probability of $1-p$. Due to its symmetric effect on all symbols, the channel capacity is symmetric too.

Z-channel

An example of an asymmetric channel is given by the *Z-channel*. Here, one symbol is transmitted correctly with unit probability and only the other symbol has a probability of p to flip during transmission. A sketch of the Z-channel is shown in fig. 3.5 and the conditional entropies are equal to

$$H(Y|\mathrm{A}) = -(1-p)\log_2[1-p] - p\log_2[p] = H_b(p), \tag{3.68}$$
$$H(Y|\mathrm{B}) = 1\log_2[1] = 0. \tag{3.69}$$

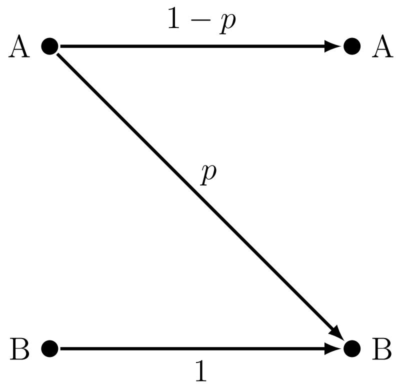


Figure 3.5: Sketch of a Z-channel. The symbol A is correctly transmitted with probability $1-p$ and mistaken for the other symbol with probability p. The symbol B is always transmitted faithfully. The corresponding channel capacity is given by eq. (3.77).

Let p_{A} be the a-priori probability of A. Then, the marginal probabilities of the two outcomes are given by

$$p\left(y=\mathrm{A}\right)=p_{\mathrm{A}}\left(1-p\right), \tag{3.70}$$

$$p\left(y=\mathrm{B}\right)=1-p_{\mathrm{A}}\left(1-p\right) \tag{3.71}$$

and consequently the outcome entropy $H\left(Y\right)$ is equal to

$$H\left(Y\right)=-p_{\mathrm{A}}\left(1-p\right)\log_2\left[p_{\mathrm{A}}\left(1-p\right)\right]-\left[1-p_{\mathrm{A}}\left(1-p\right)\right]\log_2\left[1-p_{\mathrm{A}}\left(1-p\right)\right] \tag{3.72}$$

$$=H_b\left(p_{\mathrm{A}}\left(1-p\right)\right). \tag{3.73}$$

Together with the conditional entropies of eq. (3.69) we can now write the mutual information of the Z-channel as

$$I\left(X;Y\right)=H_b\left(p_{\mathrm{A}}\left(1-p\right)\right)-p_{\mathrm{A}}H_b\left(p\right). \tag{3.74}$$

To maximize mutual information we need to calculate the derivative which reads

$$\frac{\mathrm{d}I\left(X;Y\right)}{\mathrm{d}p_{\mathrm{A}}}=-\left(1-p\right)\log_2\left[\frac{p_{\mathrm{A}}\left(1-p\right)}{1-p_{\mathrm{A}}\left(1-p\right)}\right]-H_b\left(p\right) \tag{3.75}$$

and consequently the optimal input probability distribution is given by

$$p_{\mathrm{A}}^*=\left[\left(1-p\right)\left(1+e^{\frac{H_b(p)}{1-p}}\right)\right]^{-1}. \tag{3.76}$$

With the optimal probability of eq. (3.76) the capacity of a Z-channel is given by

$$C_{\mathrm{Z}}\left(p\right)=H_b\left(p_{\mathrm{A}}^*\left(1-p\right)\right)-p_{\mathrm{A}}^*H_b\left(p\right) \tag{3.77}$$

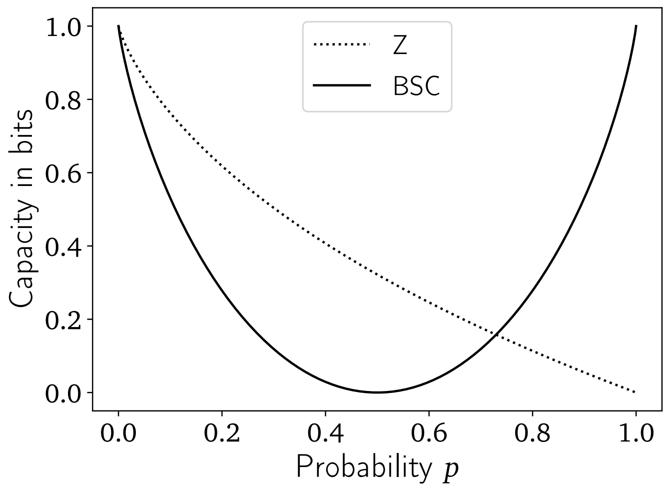


Figure 3.6: Capacity of the binary symmetric channel of eq. (3.67) and of the Z-channel of eq. (3.77) as a function of the probability p.

in bits and depicted in fig. 3.6. It is a monotonously decreasing function of probability p. Unlike for the binary symmetric channel the capacity C_Z does not recover for $p > 0.5$, since we are not able to use the symmetry of the error to our advantage. At $p = 1$ only the symbol B is obtained at the output rendering it impossible to infer the input. Nevertheless, the Z-channel has a larger capacity then the BSC except in the regime $p \gtrsim 0.73$ where the BSC obtains a higher capacity.

Gaussian channel

For channels involving only low dimensional input alphabets, the capacity can be obtained by straightforward optimization. For more complex channels it becomes more and more difficult and only few capacities are known. Luckily, the capacity of the arguably most important continuous variable channel is known. A *Gaussian channel* is characterized by a continuous input and output and noise which acts according to a Gaussian distribution, i.e. the conditional probability between input x and output y is given by

$$p\left(y|x\right) = \frac{1}{\sqrt{2\pi N}} \exp\left[-\frac{\left(x-y\right)^2}{2N}\right], \tag{3.78}$$

where N denotes the variance of the normal distribution. Sometimes this type of channel is also referred to as an *additive white Gaussian noise (AWGN)* channel. The practical importance of Gaussian channels is due to the central limit theorem which states that when adding a large number of independent random variables their joint distribution approaches a Gaussian distribution [MacKay 2003; Cover and Thomas 2005]. An example of this type of noise is thermal background noise (cf. section 2.8).

When analyzing the capacity of a Gaussian channel we face the problem that mutual information becomes infinite. Since the noise between neighboring symbols decays exponentially, we can choose any value of x as input that is far away from any other symbol and therefore all symbols are transmitted effectively noiseless. To be able to choose an arbitrary value of x, however, infinite energy is needed. To remedy this problem we need to introduce an appropriate constraint of which the most common is a constraint on the average energy $\mathbb{E}[x^2]$. Under this constraint $S \geq \mathbb{E}[x^2]$ the capacity becomes finite and was shown to be [Shannon 1949]

$$C_G = \frac{1}{2}\log_2\left[1 + \frac{S}{N}\right], \tag{3.79}$$

where the ratio $\frac{S}{N}$ is called *signal-to-noise ratio* [Essiambre et al. 2010]. Equation (3.79) is known as the *Shannon-Hartley theorem* and is attained by a Gaussian input distribution of zero mean and variance S. In the following we will consider two scenarios important in optical communication where information is encoded in one or both optical quadratures. First, suppose we use coherent states and only the quadrature x (cf. eq. (2.55)) of the optical phase space to encode information. Using the quadrature p instead would yield the same results. Since using only x implies $p = 0$, the signals energy calculates to

$$S = \mathbb{E}\left[x^2\right] = 2\mathbb{E}\left[\mathrm{Re}\left[\zeta\right]^2\right] = 2\bar{n} \tag{3.80}$$

with eqs. (2.61), (2.62) and (3.1). The noise consists of two contributions: AWGN, stemming from the transmission, and intrinsic noise coming from the measurement itself. The former is also known as *excess noise*. The variance of AWGN can also be characterized in terms of the number of photons n_b added to a transmitted state. The measurement noise is defined by the nature of the utilized measurement device on a fundamental level. While we will leave the actual device undefined at this point, a detection scheme that is able to measure any phase space quadrature, known as homodyne detector, will be introduced later on in section 4.2. If our classical detection performs at the fundamental limits defined by quantum mechanics, i.e. it measures at the level of the minimal variance also known as *shot noise* defined by eq. (2.75), the intrinsic noise variance is given by $\frac{1}{2}$. Since the variances of two Gaussian distributions are additive, the overall noise variance is given by

$$N = n_b + \frac{1}{2} \tag{3.81}$$

and together with eq. (3.80) the capacity according to the Shannon-Hartley theorem for a single-quadrature is given by

$$C_1 = \frac{1}{2}\log_2\left[1 + \frac{4\bar{n}}{2n_b + 1}\right]. \tag{3.82}$$

In the second scenario we encode not only in x, but also in the p quadrature and measure both. A physical device able to achieve such a measurement, the double homodyne detector, will be introduced in section 4.2. Since the available energy is now split between both quadratures, the energy for either of them is given by

$$\mathbb{E}\left[x^2\right] = \mathbb{E}\left[p^2\right] = 2\mathbb{E}\left[\mathrm{Re}\left[\zeta\right]^2\right] = 2\frac{\bar{n}}{2} = \bar{n}. \tag{3.83}$$

At the measurement stage the light is separated into two equal parts on which the quadratures x and p are measured respectively. Consequently, the signal energy arriving at the detectors is split into half again such that $S = \frac{\bar{n}}{2}$. Similarly, the contribution of the excess noise to the noise variance N is split in half whereas the measurement noise remains the same, i.e. the noise variance for a single quadrature reads

$$N = \frac{n_b}{2} + \frac{1}{2}. \tag{3.84}$$

Therefore, the Shannon capacity for one quadrature in this scheme is given by $C = \frac{1}{2}\log_2\left[1 + \frac{\bar{n}}{n_b+1}\right]$. The total capacity is given by the combined contributions of both quadratures x and p, which yields

$$C_2 = \log_2\left[1 + \frac{\bar{n}}{n_b + 1}\right]. \tag{3.85}$$

Both capacities C_1 and C_2 of eqs. (3.82) and (3.85) are shown in fig. 3.7 for the case of no excess noise $n_b = 0$. While for low mean photon numbers $\bar{n}$ it is beneficial to only encode in a single quadrature, C_2 yields more information than C_1 when more energy becomes available due to its superior scaling with $\bar{n}$.

3.4.1 Quantum Limits

So far we only investigated the limits on mutual information for measurements that can be fully described by a classical theory. Quantum mechanics, however, allows for more general measurements (POVMs) as well as quantum states and imposes different limitations on measurements (e.g. uncertainty principle) than the classical theory. In principle, one could expect that mutual information can be increased by invoking quantum features.

When allowing for general quantum measurements a notable feature of mutual information is *superadditivity*. Suppose instead of single states $\hat{\varrho}_x$, we send a product state $\hat{\varrho}^{\otimes} = \hat{\varrho}_{x_1} \otimes \hat{\varrho}_{x_2} \otimes \cdots \otimes \hat{\varrho}_{x_N}$, which corresponds to using the channel N times instead of a single use. Quantum mechanics allows us to measure the product state collectively, i.e. we can set up a single measurement that acts on all N received states simultaneously. The quantum mechanical capacity for mutual information is obtained by maximizing over all possible input states and POVMs, i.e.

$$\mathcal{C}_N = \max_{\{\hat{\varrho}^{\otimes}, \hat{\pi}_i\}} I\left(X; Y\right). \tag{3.86}$$

In general, the information $\mathcal{C}_N$ obtained via collective measurements can exceed the information of N single channel uses which is represented by the relation

$$\mathcal{C}_N \geq N\mathcal{C}_1 \tag{3.87}$$

and known as superadditivity [Sasaki et al. 1998]. It is a quantum feature that can not be obtained classically. Superadditivity requires memory effects such that the initially uncorrelated states $\hat{\varrho}_i$ become correlated. This correlation can also be achieved after the transmission in the measurement process [Guha 2011] and does not need to be a channel feature. In section 4.7 we will investigate superadditivity in the low energy limit and present detection schemes that are able to achieve it.
Maximization of the quantum capacity of eq. (3.86) is a notoriously hard task. In 1973 Holevo derived an upper bound on the mutual information that takes quantum mechanical effects into account and simplifies the expression dramatically. Consider a general quantum state $\hat{\varrho}_x$ that is drawn out of a set $X = \{\hat{\varrho}_x\}$ with probability p_x and sent to a destination point, where the outcome Y is obtained. For any physically possible measurement the mutual information is bounded by [Holevo 1973]

$$I\left(X;Y\right) \leq S\left(\hat{\varrho}\right) - \sum_x p_x S\left(\hat{\varrho}_x\right), \tag{3.88}$$

where $\hat{\varrho} = \sum_x p_x \hat{\varrho}_x$ is the average received state and $S\left(\hat{\varrho}\right)$ the von Neumann entropy defined in eq. (3.53). Note that the Holevo bound of eq. (3.88) takes superadditive effects into account. Due to its outstanding usefulness the quantity defining the upper bound,

$$\chi = S\left(\hat{\varrho}\right) - \sum_x p_x S\left(\hat{\varrho}_x\right), \tag{3.89}$$

received its own name and is referred to as *Holevo quantity* or *Holevo information*. While the Holevo bound is tight [Hausladen et al. 1996; Schumacher and Westmoreland 1997; Holevo 1998], it remains highly elusive to find a practically feasible measurement that saturates it as the proofs do not give a constructive way to derive them.
Holevo's idea to prove the bound is simple, yet beautiful. Suppose P describes a quantum system which is used to prepare the state of a quantum system Q. The system Q is now sent to the destination, where it is measured with the help of a quantum system M. After the quantum state of Q was prepared, the mutual information between the preparation system P and the messenger system Q is given by $I\left(P;Q\right)$. Since in the beginning the measurement system M is entirely uncorrelated with those two systems, we can easily extend the mutual information to the joint expression $I\left(P;Q\right) = I\left(P;QM\right)$. Now Q is sent to M and the state is measured. Since a measurement can never increase mutual information, the relation $I\left(P;QM\right) \geq I\left(P';Q'M'\right)$ holds, where the primes denote systems after the measurement operation was applied. In the end we discard system Q'. Since discarding can never increase mutual information, we can write down $I\left(P';Q'M'\right) \geq I\left(P';M'\right)$. After combining these relations, we arrive at

$$I\left(P;Q\right) = I\left(P;QM\right) \geq I\left(P';Q'M'\right) \geq I\left(P';M'\right), \tag{3.90}$$

which essentially states that the mutual information between sender P and receiver M never exceeds the mutual information between sender P and prepared messenger system Q. After formalizing the left and right-hand side of eq. (3.90) one arrives at the Holevo bound of eq. (3.88). A detailed proof can be found, for instance, in the textbook of Nielsen and Chuang [2009].
To derive the ultimate quantum limit on classical information, the Holevo quantity χ has to be maximized over all input states $\hat{\varrho}_x$ and their probabilities p_x which defines the *Holevo capacity* as

$$\mathcal{C} = \max_{p_x, \hat{\varrho}_x} \chi. \tag{3.91}$$

Note that, in general, the maximization has to be carried out over all possible input states, including entangled states, in the limit of infinite channel uses and normalized to a single channel use [Hastings 2009]. While already very early ideas existed how to beat the classical capacity [Gordon 1962], a proof for the Holevo capacity of an optical channel was achieved only much later.
For a noisy optical channel this was first achieved in the case where only losses are present. In quantum mechanics losses can not be described as straightforwardly as classically, where simply a fraction of the energy is lost. The simplest idea to model losses in a quantum way is to consider a beamsplitter that reflects part of the light into a second mode which is lost. However, this model proves to be to simplistic on some occasions as it assumes that the loss takes place at a specific point in time and space. Often, particularly in optical communication, losses are distributed over a large time or space interval and also other effects may take place as well while traveling. To model an appropriate loss mechanism for this scenario one has to resort to a quantum master equation that describes the interactions between the state of light and a reservoir describing the environment (cf. eq. (2.15)). Most importantly, coherent states remain coherent under losses and only their amplitude is damped. Let τ be the energy transmission factor that describes the fraction of photons that are lost such that we can denote the average photon number after losses as

$$\bar{n}_\tau = \tau \bar{n}. \tag{3.92}$$

The capacity of a lossy bosonic quantum channel was derived by Giovannetti et al. [2004] and reads

$$\mathcal{C}_{\mathrm{LB}} = g\left(\bar{n}_\tau\right), \tag{3.93}$$

where the function $g\left(x\right)$ is defined as

$$g\left(x\right) = \left(x+1\right) \ln\left[x+1\right] - x \ln\left[x\right] \tag{3.94}$$

and the basis of the logarithm defines the unit used to express information. Here nats are used. In fig. 3.7 the capacity $\mathcal{C}_{\mathrm{LB}}$ of eq. (3.93) is compared to the classical capacities of eqs. (3.82) and (3.85). It lies considerably above the Shannon capacities for all $\bar{n}_\tau$.

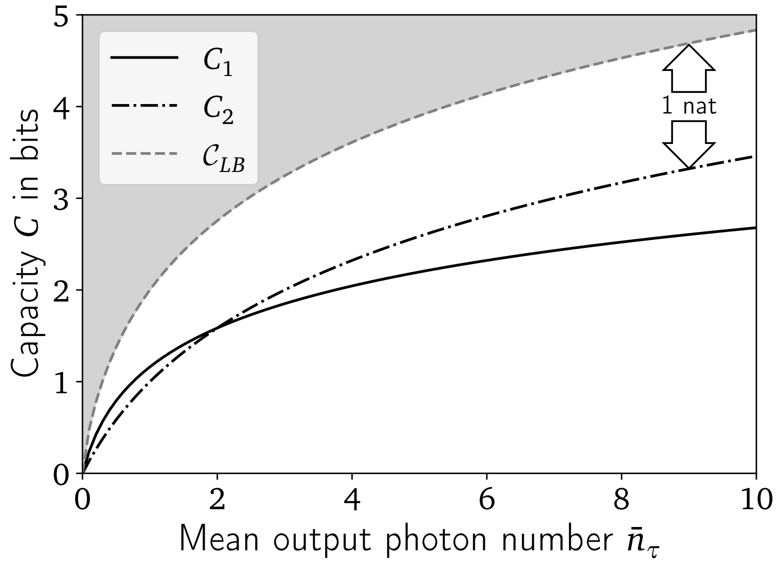


Figure 3.7: The single-quadrature and two-quadrature capacities C_1 and C_2 of eqs. (3.82) and (3.85) as a function of the mean output photon number $\bar{n}_\tau$. The Holevo capacity $\mathcal{C}_{\text{LB}}$ of eq. (3.93) defines the upper bound on mutual information attainable with quantum mechanics. In the regime of large mean photon number $\bar{n}_\tau \gg 1$, the gap between the quantum limit and the capacity dictated by classical electrodynamics reaches 1 nat $\approx$ 1.44 bits of information.

The capacity $\mathcal{C}_{\text{LB}}$ can be achieved in the lossless case $\tau = 1$ by an ensemble of Fock states with a thermal probability distribution [Yuen and Ozawa 1993; Caves and Drummond 1994; Lukanowski and Jarzyna 2021]. Even though for this ensemble a physical realization of the optimal measurement is known, the practical importance of communication with Fock states is very limited, since generating photon number states remains very challenging (cf. section 2.5). For other values of $\tau \neq 1$ the capacity of eq. (3.93) can be achieved by an ensemble of coherent states with Gaussian probability distribution, which are much more readily available than Fock states (cf. section 2.6). Unfortunately, however, a corresponding practical measurement that saturates the capacity bound remains unknown.

The quantum version of an AWGN (cf. eq. (3.78)) is given by a phase insensitive bosonic Gaussian channel. It describes all interactions of an environment with the system that lets a state of initially Gaussian character remain Gaussian. A more detailed description of Gaussian states or channels can be found, for instance, in the review of Weedbrook et al. [2012]. An effective description of such a quantum channel is thermalization of the state as given in eq. (2.93).

Only in recent years the quantum capacity of an AWGN channel was proven to be

$$\mathcal{C}_{\text{AWGN}} = g\left(\bar{n}_\tau + n_b\right) - g\left(n_b\right) \tag{3.95}$$

by Giovannetti et al. [2014]. For no excess noise $n_b = 0$, the capacity $\mathcal{C}_{\text{AWGN}}$ reduces to the capacity of a lossy bosonic channel in eq. (3.93). Interestingly, $\mathcal{C}_{\text{AWGN}}$ can be achieved by a Gaussian ensemble of coherent states just like for the lossy bosonic channel. Equation (3.95) is the direct quantum counterpart to the Shannon capacity C_G of eq. (3.79). Note that, unlike C_G, the quantum capacity $\mathcal{C}_{\text{AWGN}}$ does not depend on the signal to noise ratio, but on the noise and signal strength separately.

3.4.2 Large Energy Limit

In order to better understand the relation between the quantum and classical capacity, it is instructive to perform an expansion of the former in the large energy regime. For this purpose we first expand the function $g(x)$ of eq. (3.94) in powers of x^{-1} for large $x \gg 1$, which up to the first order takes the form

$$g(x) \approx 1 + \ln[1 + x] + \frac{1}{2x}. \tag{3.96}$$

With this the capacity of a lossy bosonic channel in eq. (3.93) can be expanded in $\bar{n}_\tau^{-1}$ for large $\bar{n}_\tau \gg 1$, such that it reads

$$\begin{aligned} \mathcal{C}_{\text{LB}} &\approx \ln[1 + \bar{n}_\tau] + 1 + \frac{1}{2\bar{n}_\tau} &(3.97)\\ &= C_2 + 1 + \frac{1}{2\bar{n}_\tau}. &(3.98) \end{aligned}$$

Note that here the capacities are calculated in nats. Therefore, quantum mechanics offers a 1 nat advantage with respect to the classical capacity in the limit of large energy $\bar{n}_\tau \gg 1$ and a loss only channel (cf. fig. 3.7). For non-negligible excess noise $n_b \gg 1$, however, the situation is different. Using the same expansion of eq. (3.96) for the AWGN capacity of eq. (3.95) yields

$$\begin{aligned} \mathcal{C}_{\text{AWGN}} &\approx \ln[1 + \bar{n}_\tau + n_b] + 1 + \frac{1}{2(\bar{n}_\tau + n_b)} - \ln[1 + n_b] - 1 - \frac{1}{2n_b} &(3.99)\\ &= C_2 + \frac{1}{2(\bar{n}_\tau + n_b)} - \frac{1}{2n_b}. &(3.100) \end{aligned}$$

For large noise $n_b \gg 1$ the latter two terms vanish and the quantum capacity coincides with the classical capacity C_2. This suggest that a quantum advantage can only be obtained in scenarios with low excess noise, since otherwise AWGN shrinks the gap until the classical and the quantum capacity coincide.

3.4.3 Low Energy Limit

If the energy available for communication is very limited $\bar{n}_\tau \ll 1$, both mutual information and the capacity become very small. This regime is also known as *photon starved communication* and is a common scenario in deep-space communication [Hemmati 2006;

Biswas et al. 2017; Sodnik et al. 2017; Boroson 2018]. Here, it is often more instructive to quantify the performance not as the total amount of information transmitted, but as information transmitted per photon. Following this approach, the *photon information efficiency (PIE)* is defined as

$$\mathrm{PIE} = \frac{I\left(X;Y\right)}{\bar{n}_\tau}, \tag{3.101}$$

with the mean output photon number $\bar{n}_\tau$ of eq. (3.92). Equivalently we can define the PIE for capacities as

$$\mathrm{PIE} = \frac{C}{\bar{n}_\tau}. \tag{3.102}$$

While the same symbol PIE is used for both versions, the type of PIE used will be clear from its context.
Let us first investigate the loss-only scenario. The Shannon PIEs for the capacities of eqs. (3.82) and (3.85) expanded at $\bar{n}_\tau = 0$ take the form

$$\mathrm{PIE}_1 \approx 2 - 4\bar{n}_\tau \xrightarrow[\bar{n}_\tau \to 0]{} 2, \tag{3.103}$$

$$\mathrm{PIE}_2 \approx 1 - \frac{\bar{n}_\tau}{2} \xrightarrow[\bar{n}_\tau \to 0]{} 1. \tag{3.104}$$

Note that PIE for C_1 reaches 2 nats whereas using two-quadrature PIE only yields 1 nat of information. Intuitively, shot noise is overwhelming for small $\bar{n}_\tau$ and since using both quadratures effectively doubles the contribution of shot noise (cf. derivation of eq. (3.85)), the two quadrature measurement provides only half of the information attained by PIE_{C_1} in this regime.
Similarly, we can expand the PIE for the lossy bosonic channel of eq. (3.93). Up to the first order in $\bar{n}_\tau$ this yields

$$\mathrm{PIE}_{\mathrm{LB}} \approx \ln\left[\frac{1}{\bar{n}_\tau}\right] + 1 \xrightarrow[\bar{n}_\tau \to 0]{} \infty, \tag{3.105}$$

i.e. the quantum PIE diverges for vanishing signal energy. At first glance, this result seems like a paradox. An infinite PIE suggests that a single photon is able to carry an arbitrary amount of information and therefore it is possible to communicate an arbitrary amount of information with the usage of infinitesimally small amounts of energy. However, a closer look reveals that this scenario can never occur in reality. While the capacity does not specify the measurement, it is nevertheless necessary to measure in a noiseless way. This means that a physical realization needs to operate at a zero temperature which is impossible due to the third law of thermodynamics. Since for any other temperature some thermal noise will be present, the limit of eq. (3.105) is non-physical. An alternative way of stating the above is that in order to achieve infinite PIE for a lossy bosonic channel, the measurement system requires an infinite amount of energy [Ding et al. 2019]. As we will see in the following, the quantum PIE does not diverge

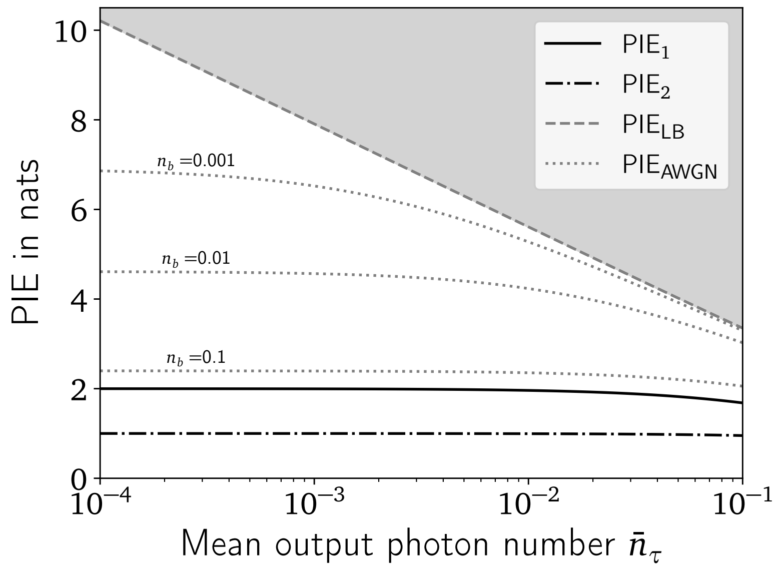


Figure 3.8: The classical Shannon PIEs of eqs. (3.103) and (3.104) are shown as a function of the mean output photon number $\bar{n}_\tau$. The quantum limit defined by eq. (3.108) is shown for several values of the excess noise photon number n_b. While in the noiseless scenario the quantum limit offers a different scaling with $\bar{n}_\tau$, it reaches a plateau when excess noise is present.

if background noise is taken into account. Assuming a finite n_b the Shannon capacities can be similarly expanded for $\bar{n}_\tau \ll 1$. Up to the first order in $\bar{n}_\tau$ this yields

$$\mathrm{PIE}_1 \approx \frac{2}{1+2n_b} - \frac{4\bar{n}_\tau}{(1+2n_b)^2} \xrightarrow[\bar{n}_\tau \to 0]{} \frac{2}{1+2n_b} \tag{3.106}$$

for the one-quadrature capacity of eq. (3.82) and

$$\mathrm{PIE}_2 \approx \frac{1}{1+n_b} - \frac{\bar{n}_\tau}{2\,(1+n_b)^2} \xrightarrow[\bar{n}_\tau \to 0]{} \frac{1}{1+n_b} \tag{3.107}$$

for the two-quadrature capacity of eq. (3.85). Unsurprisingly, the information per photon decreases with rising excess noise n_b. Expanding the corresponding quantum PIE for AWGN according to eq. (3.95) yields

$$\mathrm{PIE}_{\mathrm{AWGN}} \approx \ln\left[1+\frac{1}{n_b}\right] - \frac{\bar{n}_\tau}{2n_b+2n_b^2} \xrightarrow[\bar{n}_\tau \to 0]{} \ln\left[1+\frac{1}{n_b}\right]. \tag{3.108}$$

With non-zero background noise $n_b \neq 0$ taken into account, the PIE tends to a finite value and the paradox of eq. (3.105) is resolved.
In fig. 3.8 the noiseless PIEs of eqs. (3.103) and (3.104) are shown alongside PIE$_{\mathrm{AWGN}}$ for several values of n_b. In the noiseless case, the gap to the classical PIE grows with decreasing $\bar{n}_\tau$. If excess noise is present, the PIE$_{\mathrm{AWGN}}$ reaches a plateau at the value dictated by eq. (3.108).

3.5 Encodings

In order to achieve the capacities derived in the previous section 3.4 ensembles of quantum states with a continuous probability distribution are needed. In practice it is often more convenient to use discrete encodings to ease the technical requirements. Coherent states are most commonly used in optical communication scenarios, since their preparation is less challenging and can be done at high repetition rates (cf. section 2.6). In the following, we will briefly introduce a few discrete encoding schemes based on coherent states that are useful for the further discussion. This list is by no means complete and many more encodings do exist [Haykin 2001; Senior 2009; Essiambre et al. 2010]. Additionally, POVMs which are commonly used to detect different encodings are discussed, whereas physical realizations of these POVMs will be discussed separately in chapter 4.

3.5.1 On-Off Keying (OOK)

Suppose we want to perform binary communication, i.e. only based on two symbols, 0 and 1, out of which messages are build up. The simplest way to encode this binary set of symbols into physical information carriers is *on-off keying (OOK)* modulation. The vacuum state $|\,0\,\rangle$ is used to transmit the symbol 0 and a coherent state $|\,\zeta\,\rangle$ for symbol 1. Hence, the information is encoded into the absolute value of the amplitude $|\zeta|$. The average energy at our disposal for state preparation is given by the mean photon number $\bar{n}$, the probability to send 0 is denoted by p_0 and the probability to send 1 by $p_1 = 1 - p_0$. Without the loss of generality, we assume that both states are located on the real axis of the phase space. Any other choice can be transformed into this scenario by an appropriate rotation of the phase space. Since the energy required to send the vacuum state is zero, the coherent state contains all energy such that its amplitude is given by

$$\zeta = \sqrt{\frac{\bar{n}}{1 - p_0}}. \tag{3.109}$$

An illustration of OOK in phase space is shown in fig. 3.9. A basic strategy to distinguish both states is a projection onto the vacuum state. Essentially, the arriving energy is measured and the vacuum state is assumed only if no energy was registered at all. Such a measurement can be physically performed by counting the arriving photons, which will be discussed in more detail in section 4.1. The POVM for projecting onto the vacuum state is given by $\{\hat{\pi}_i\}$ with the elements given by

$$\hat{\pi}_0 = |\,0\,\rangle\langle\,0\,|\,, \tag{3.110}$$

$$\hat{\pi}_1 = \hat{\mathbb{1}} - |\,0\,\rangle\langle\,0\,|\,. \tag{3.111}$$

The conditional probabilities $p\,(y|x)$ for obtaining an outcome are given by Born's rule of eq. (2.8) and written concretely in table 3.3. The symbol 0 will always be identified correctly while symbol 1 will be mistaken for 0 in some instances due to the non-zero transition probability $p\,(0|1)$. Physically, the reason is the non-zero overlap $|\langle\,0\,|\,\zeta\,\rangle|^2 =$

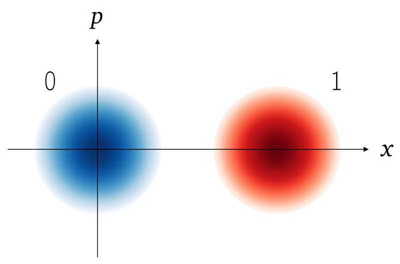


Figure 3.9: Possible states of the OOK alphabet illustrated in phase space. The left (blue) state is the vacuum state corresponding to symbol 0 and the right (red) a coherent state corresponding to 1.

Input x	Output y	
	0	1
0	1	0
1	$e^{-\lvert\zeta\rvert^2}$	$1-e^{-\lvert\zeta\rvert^2}$

Table 3.3: Conditional probabilities $p\,(y|x)$ for OOK when using the POVM written in eqs. (3.110) and (3.111) that projects onto the vacuum state .

$e^{-|\zeta|^2}$ between the vacuum state and a coherent state. While the vacuum state will always be correctly identified, since it is orthogonal to $\hat{\pi}_1$, a coherent state can be falsely classified as a vacuum state. Consequently, the average error probability is given by

$$p_{\text{err}}^{\text{OOK}} = (1-p_0)\,e^{-|\zeta|^2} = (1-p_0)\,e^{-\frac{\bar{n}}{1-p_0}}. \tag{3.112}$$

Note that for very large available energy $\bar{n} \gg 1$, the error probability tends to zero. To evaluate OOK in terms of mutual information it is helpful to look at the structure of the conditional probabilities. One symbol is transmitted faithfully, while the second one can flip to the first symbol. This behavior forms a Z-channel (cf. fig. 3.5) and mutual information is therefore given by eq. (3.74) with $p_{\text{A}} = 1-p_0$ and $p = p\,(0|1) = \exp\left[-\frac{\bar{n}}{1-p_0}\right]$. The scenario for $p_0 = \frac{1}{2}$ is shown in fig. 3.12. The maximum of 1 bit is reached in the limit of $\bar{n} \gg 1$. The full capacity of a Z-channel is reached for the probability given by eq. (3.76), which defines the optimum value for p_0. Note that the transition probability $p = p\,(0|1)$ in eq. (3.76) depends implicitly on p_0.

3.5.2 Binary Phase Shift Keying (BPSK)

A second approach to encode binary symbols is *binary phase shift keying (BPSK)*. While in OOK the information was encoded into the absolute value of the amplitude, in BPSK

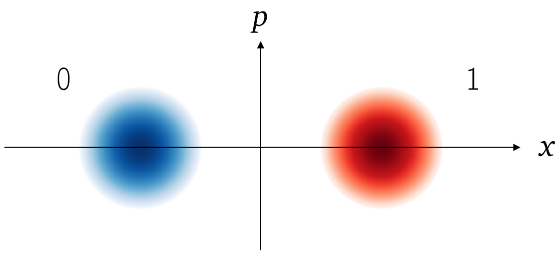


Figure 3.10: BPSK alphabet schematically depicted in the optical phase space. Both states are coherent states and are located symmetrically around the origin. The left (blue) coherent state is used for encoding the symbol 0 whereas the right (red) state is used for symbol 1.

the phase variable is used instead. For this purpose two coherent states with the same amplitude, but opposite phases are used. For concreteness, let us assume that the symbol 0 corresponds to state $|-\zeta\rangle$ and 1 is encoded in $|\zeta\rangle$. The amplitude is defined by the mean photon number and given by $\zeta = \sqrt{\bar{n}}$. A phase space illustration of BPSK is shown in fig. 3.10. The two states are located symmetrically around the origin on the x axis, due to their opposite phases. Therefore, they can be distinguished by constructing a POVM that measures the x quadrature, which corresponds to a POVM consisting of

$$\hat{\pi}_0 = \int_{-\infty}^{0} \mathrm{d}x \, |x\rangle\langle x| , \tag{3.113}$$

$$\hat{\pi}_1 = \int_{0}^{\infty} \mathrm{d}x \, |x\rangle\langle x| . \tag{3.114}$$

We can decide which state was received based on the following rule: for $x < 0$ we decide that symbol 0 was measured whereas for $x \geq 0$ 1 is assumed. The conditional probabilities obtained by this type of decision are determined with the help of Born's rule in eq. (2.8) and eq. (2.59). The obtained expressions are written in table 3.4. Note that the conditional probabilities are symmetric in the sense that both symbols have the same probability to flip to the other one. The average error probability is therefore given by

$$p_{\mathrm{err}}^{\mathrm{BPSK}} = \frac{1}{2}\left[1 - \mathrm{erf}\left(\sqrt{2\bar{n}}\right)\right] . \tag{3.115}$$

BPSK forms a binary symmetric channel due to the symmetry of the conditional probabilities and its mutual information is given by eq. (3.64).
The capacity of a binary symmetric channel of eq. (3.67) is achieved, when the marginal output entropy $H(Y)$ reaches 1 bit. This is the case for equiprobable input states, i.e.

Input x	Output y	
	0	1
0	$\frac{1}{2}\left[1+\mathrm{erf}\left(\sqrt{2\bar{n}}\right)\right]$	$\frac{1}{2}\left[1-\mathrm{erf}\left(\sqrt{2\bar{n}}\right)\right]$
1	$\frac{1}{2}\left[1-\mathrm{erf}\left(\sqrt{2\bar{n}}\right)\right]$	$\frac{1}{2}\left[1+\mathrm{erf}\left(\sqrt{2\bar{n}}\right)\right]$

Table 3.4: Conditional probabilities $p\,(y|x)$ for BPSK located on the real axis in phase space when measuring the corresponding quadrature x. $\mathrm{erf}\,(x)$ denotes the error function.

when their a-priori probabilities are $p_0 = p_1 = \frac{1}{2}$. Unlike for OOK, the optimal a-priori probabilities for BPSK are independent of the error probability $p_{\mathrm{err}}^{\mathrm{BPSK}}$. In fig. 3.12 mutual information of BPSK is compared to OOK, the Holevo limit of eq. (3.93) and the single-quadrature Shannon limit of eq. (3.82). It is seen that BPSK achieves higher mutual information than OOK, since it provides a larger separation of the alphabet states in phase space for the same energy cost. While the distance between states for OOK is $\sqrt{2\bar{n}}$, BPSK provides a distance of $2\sqrt{\bar{n}}$ which gives a superior state distinguishability and translates directly to larger mutual information. However, BPSK only gets close to the Shannon limit C_1 for small $\bar{n} \ll 1$. The Shannon limit itself is attained by a Gaussian distribution of coherent states which becomes well approximated by just two states for low $\bar{n}$.

3.5.3 Quadrature Phase Shift Keying (QPSK)

A non-binary type of encoding is *quadrature phase shift keying (QPSK)*. Similar to BPSK the information is encoded only in the phase variable of coherent states, but with four different phases such that up to 2 bits can be transmitted per channel use. Let us represent the symbols by a binary code such that the set of symbols is given by $\{00, 10, 11, 01\}$ with their respective a-priori probabilities $\{p_{00}, p_{10}, p_{11}, p_{01}\}$. The states used for transmitting the symbols are coherent states $|\,\zeta_j\,\rangle$ with their complex amplitudes defined by

$$\zeta_j = \sqrt{\bar{n}}e^{i\theta_j} = \sqrt{\bar{n}}\left[\cos\left(\theta_j\right) + i\sin\left(\theta_j\right)\right], \tag{3.116}$$

where $\bar{n}$ is the mean photon number and θ_j are the different phases for the symbols given by $\left\{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\right\}$. A phase space illustration of a QPSK alphabet is shown in fig. 3.11. One possible way to distinguish these four states is to measure two quadratures simultaneously such that the quadrant A_{00}, A_{01}, A_{10} or A_{11} of the phase space can be determined. This can be achieved by first splitting the transmitted state on a 50:50 beamsplitter into two modes and measure them separately. A more detailed discussion on a possible realization can be found in section 4.3. Overall this defines a POVM with

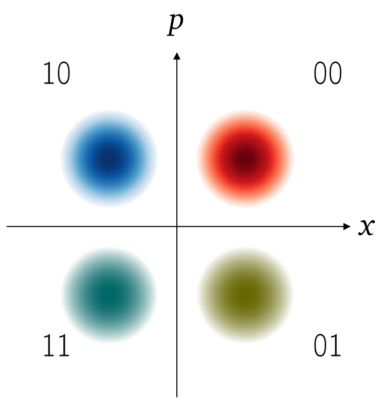


Figure 3.11: The four states of a QPSK alphabet illustrated in phase space. All states are coherent states with $|\zeta|^2 = \bar{n}$ which are located symmetrically around the origin due to their different phases.

elements given by

$$\hat{\pi}_{00} = \frac{1}{\pi}\int_{A_{00}} \mathrm{d}^2\beta \; |\,\beta\,\rangle\langle\,\beta\,|\,, \tag{3.117}$$

$$\hat{\pi}_{01} = \frac{1}{\pi}\int_{A_{01}} \mathrm{d}^2\beta \; |\,\beta\,\rangle\langle\,\beta\,|\,, \tag{3.118}$$

$$\hat{\pi}_{10} = \frac{1}{\pi}\int_{A_{10}} \mathrm{d}^2\beta \; |\,\beta\,\rangle\langle\,\beta\,|\,, \tag{3.119}$$

$$\hat{\pi}_{11} = \frac{1}{\pi}\int_{A_{11}} \mathrm{d}^2\beta \; |\,\beta\,\rangle\langle\,\beta\,|\,. \tag{3.120}$$

The corresponding conditional probabilities are written in table 3.5, where the expressions are obtained by Born's rule combined with the overlaps of eqs. (2.59) and (2.60). The arising probability

$$p_f = \frac{1}{2}\left[1 - \operatorname{erf}\left(\sqrt{\frac{\bar{n}}{2}}\right)\right] \tag{3.121}$$

can be interpreted as the probability that a single bit of the binary symbol code is flipped, e.g. the transitions $00 \to 01$ or $10 \to 00$. The average error probability is equal to

$$p_{\mathrm{err}}^{\mathrm{QPSK}} = 1 - \sum_x p_x \left(1 - p_f\right)^2 = 1 - \left(1 - p_f\right)^2 \tag{3.122}$$

$$= \frac{3}{4}\left[1 - \operatorname{erf}\left(\sqrt{\frac{\bar{n}}{2}}\right)\right]. \tag{3.123}$$

Input x	Output y			
	00	01	10	11
00	$(1-p_f)^2$	$(1-p_f)\,p_f$	$p_f\,(1-p_f)$	p_f^2
01	$(1-p_f)\,p_f$	$(1-p_f)^2$	p_f^2	$p_f\,(1-p_f)$
10	$p_f\,(1-p_f)$	p_f^2	$(1-p_f)^2$	$(1-p_f)\,p_f$
11	p_f^2	$p_f\,(1-p_f)$	$(1-p_f)\,p_f$	$(1-p_f)^2$

Table 3.5: Conditional probabilities $p\,(y|x)$ for QPSK when measuring both phase space quadratures x and p simultaneously. The bit flip probability p_f is defined as $p_f = \frac{1}{2}\left[1 - \operatorname{erf}\left(\sqrt{\frac{\bar{n}}{2}}\right)\right]$, where $\operatorname{erf}(x)$ denotes the error function.

Since the error function erf is monotonically decreasing with $\bar{n}$, the maximum error probability of QPSK is given by 75%.
To calculate mutual information, the output entropy $H\,(Y)$ and the conditional entropy $H\,(Y|X)$ have to be determined based on table 3.5. The conditional entropy is given by

$$H\,(Y|X) = 2\left\{-p_f^2 \log_2\left[p_f\right] - (1-p_f)^2 \log_2\left[1-p_f\right] - p_f\,(1-p_f)\log_2\left[p_f\,(1-p_f)\right]\right\} \tag{3.124}$$

$$= 2\left\{p_f \log_2\left[p_f\right] - (1-p_f)\log_2\left[1-p_f\right]\right\} \tag{3.125}$$

$$= 2H_b\,(p_f) \tag{3.126}$$

with the binary entropy of eq. (3.38). Now, mutual information can be written as

$$I_{\text{QPSK}} = H\,(Y) - 2H_b\,(p_f)\,. \tag{3.127}$$

For equiprobable input states, i.e. $p_x = \frac{1}{4}$, mutual information takes its maximum value. In this scenario the output probabilities are given by $p\,(y) = \frac{1}{4}$ and therefore, the marginal output entropy is equal to $H\,(Y) = 2$ in bits. Consequently, mutual information in bits is given by

$$I_{\text{QPSK}} = 2\left[1 - H_b\,(p_f)\right], \tag{3.128}$$

which looks exactly like twice the mutual information for BPSK. However, the bit flip probability p_f depends only on $\frac{\bar{n}}{2}$ whereas the dependency is given by $2\bar{n}$ for BPSK. In fig. 3.12 the mutual information of QPSK is compared to the Holevo limit of eq. (3.93) and the two-quadrature Shannon limit of eq. (3.82). In all regimes remains a gap between the mutual information of QPSK and the corresponding Shannon limit C_2.

3.5.4 Pulse Position Modulation (PPM)

A different type of encoding which is particularly useful for the communication with very limited energy $\bar{n} \ll 1$ is given by *pulse position modulation (PPM)*. Unlike for

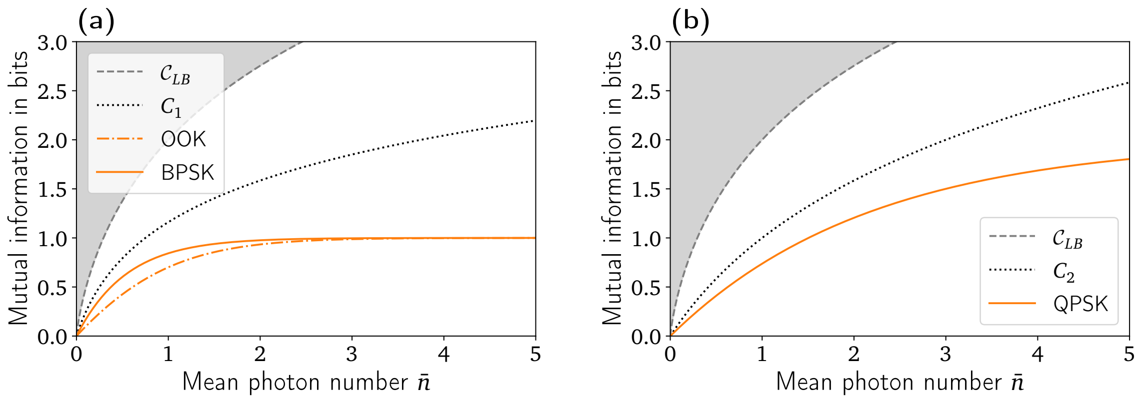


Figure 3.12: **(a)** Mutual information of binary alphabets OOK and BPSK compared to the capacity of a bosonic channel $\mathcal{C}_{\mathrm{LB}}$ for $\tau = 1$ and the one-quadrature Shannon capacity C_1. The a-priori probability distribution of OOK and BPSK is assumed to be flat with $p_0 = p_1 = \frac{1}{2}$. Note that this is the optimal choice for BPSK, while suboptimal for OOK. An optimal p_0 for OOK would yield higher mutual information, but does not exceed mutual information of BPSK. **(b)** Mutual information of QPSK compared to the capacity of a lossless bosonic channel $\mathcal{C}_{\mathrm{LB}}$ and the two-quadrature Shannon capacity C_2. The QPSK states are assumed to be equiprobable with $p_{ij} = \frac{1}{4}\ \forall\, i, j$.

the previously discussed encodings, PPM does not transmit information within a single channel use. Instead, each channel use is considered a time bin and M consecutive bins are grouped together into a PPM frame. The length M of a PPM frame is also referred to as the PPM order. Note that here $\bar{n}$ denotes the mean photon number per channel use such that for the entire frame the energy of $M\bar{n}$ is available. Instead of preparing a state for each time bin, the energy is collected and a single strong signal state is created. Information is encoded by placing the coherent state $\left|\sqrt{M\bar{n}}\right\rangle$ in a single time bin within the PPM frame.

For instance, consider PPM where a frame consists of 4 time bins. An illustration of the four possible symbols $\{0, 1, 2, 3\}$ is shown in fig. 3.13. Placing a state into the first bin corresponds to the symbol 1, in the second bin to symbol 2, etc. For the sake of simplicity, it is assumed that the coherent state stays perfectly contained within its time bin and does not leak into other bins.

After transmission to the destination the time bin containing the state is identified and the symbol recovered. Similarly to OOK, this can be achieved by a projection on the vacuum state in each time bin defining the same POVM as written in eqs. (3.110) and (3.111). It is possible that no signal is detected in the entire PPM frame due to the non-zero overlap between a coherent and the vacuum state $\left\langle 0 \middle| \sqrt{M\bar{n}} \right\rangle = e^{-\frac{M\bar{n}}{2}}$. This erasure event, denoted by e, conveys no information and is the only source of error.

In total for PPM of order M, the set of possible input symbols is given by $X =$

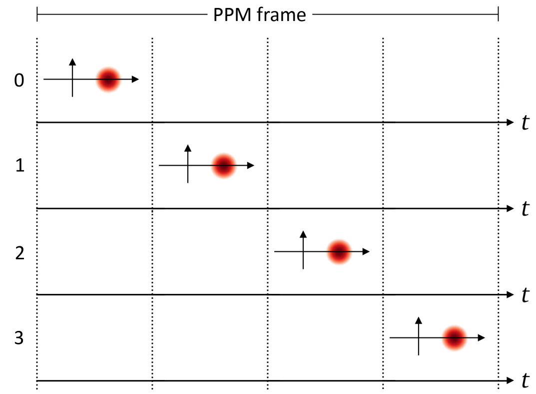


Figure 3.13: Illustration of a PPM alphabet of order $M = 4$. Depending on which time bin contains the coherent state, a particular symbols out of the set $\{0, 1, 2, 3\}$ is transmitted.

$\{1, \ldots, M-1, M\}$ whereas the output symbols are $Y = \{\mathrm{e}, 1, \ldots, M-1, M\}$. The conditional probabilities $p\left(y|x\right)$ for PPM are specified by eqs. (2.8), (3.110) and (3.111) and written comprehensively in table 3.6. The overall error probability is given by the probability for an erasure event, i.e.

$$p_{\mathrm{err}}^{\mathrm{PPM}} = e^{-M\bar{n}}. \tag{3.129}$$

In principle, the error probability could be made arbitrary small by increasing the PPM order M. However, this involves two problems.

First of all, the time until an actual symbol is transmitted increases together with the length of a PPM frame. In other words, arbitrary large PPM orders require an arbitrary long waiting time at the receiving end to actually identify the symbol.

The second problem is that by increasing the PPM order M the amplitude of the coherent state $\left|\sqrt{M\bar{n}}\right\rangle$ increases as well. Preparing such a state requires a very large peak to average energy ratio which can become problematic. A possible solution to circumvent this issue will be investigated in section 4.7.

The conditional entropy can be calculated from the conditional probabilities of table 3.6 and reads

$$H\left(Y|X\right) = -\left(1 - e^{-M\bar{n}}\right)\log_2\left[1 - e^{-M\bar{n}}\right] - e^{-M\bar{n}}\log_2\left[e^{-M\bar{n}}\right]. \tag{3.130}$$

The probabilities to obtain a particular outcome y are given by

$$p\left(Y = e\right) = e^{-M\bar{n}}, \tag{3.131}$$

$$p\left(Y \neq e\right) = \frac{1}{M}\left(1 - e^{-M\bar{n}}\right) \tag{3.132}$$

and the outcome entropy can be written as

$$H\left(Y\right) = -e^{-M\bar{n}}\log_2\left[e^{-M\bar{n}}\right] - \left(1 - e^{-M\bar{n}}\right)\log_2\left[\frac{1}{M}\left(1 - e^{-M\bar{n}}\right)\right]. \tag{3.133}$$

Input x	Output y					
	e	1	2	$\dots$	$M-1$	M
1	$e^{-M\bar{n}}$	$1-e^{-M\bar{n}}$	0	$\dots$	0	0
2	$e^{-M\bar{n}}$	0	$1-e^{-M\bar{n}}$		0	0
$\vdots$		$\vdots$			$\vdots$	
$M-1$	$e^{-M\bar{n}}$	0	0	$\dots$	$1-e^{-M\bar{n}}$	0
M	$e^{-M\bar{n}}$	0	0		0	$1-e^{-M\bar{n}}$

Table 3.6: Conditional probabilities $p\,(y|x)$ for PPM when the measurement consists of a projecting onto the vacuum state.

Finally, mutual information normalized to a single channel use takes the form

$$I_{\mathrm{PPM}}\,(X;Y) = \left(1-e^{-M\bar{n}}\right)\frac{\log_2{[M]}}{M}. \tag{3.134}$$

Note that this expression essentially consists of two factors, the maximum possible information per time bin $\frac{\log_2[M]}{M}$ and the average success probability $1-e^{-M\bar{n}}$. Unlike the error probability, mutual information does not grow continuously with the PPM order M, but exhibits a maximum value. Since PPM is mainly used under very limited average energy $\bar{n}$, it is convenient to consider PIE instead. The PIE for PPM reads

$$\mathrm{PIE}_{\mathrm{PPM}} = \frac{I_{\mathrm{PPM}}\,(X;Y)}{\bar{n}} = \left(1-e^{-M\bar{n}}\right)\frac{\log_2{[M]}}{M\bar{n}}. \tag{3.135}$$

The optimal PPM order can be determined for a finite but small energy $\bar{n} \ll 1$. By expanding the erasure probability up to the second order in $\bar{n}$ the PIE can be approximated by

$$\mathrm{PIE}_{\mathrm{PPM}} = \frac{I_{\mathrm{PPM}}\,(X;Y)}{\bar{n}} = \left(1-\frac{M\bar{n}}{2}\right)\log_2{[M]}\,. \tag{3.136}$$

In the limit $\bar{n} \to 0$ PIE approaches $\log_2{[M]}$. Based on eq. (3.136) the optimal PPM order M^* can be obtained and reads [Jarzyna et al. 2015]

$$M^* = \frac{2}{\bar{n}}W\left[\frac{2e}{\bar{n}}\right]^{-1}, \tag{3.137}$$

where $W\,[x]$ denotes the Lambert function. Note that this approximation assumes that M is continuous while in reality it can only take integer values. However, this is not an issue, since the corresponding PIE only changes slowly in the vicinity of M^*. In fig. 3.14 the PIE of PPM is compared to the single-quadrature Shannon PIE of eq. (3.106) and the PIE of a bosonic channel of eq. (3.105). While PPM comes much closer to the ultimate quantum limit than the other considered encodings, still a gap remains.

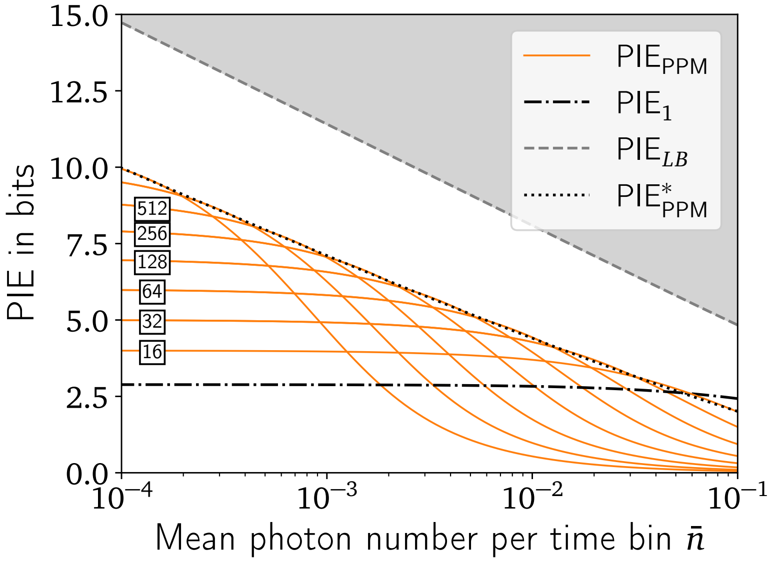


Figure 3.14: PIE per channel use of PPM for several orders M specified in boxes compared to the limit of a bosonic channel $\mathrm{PIE_{LB}}$ for $\tau = 1$ and the one-quadrature Shannon limit $\mathrm{PIE_1}$. The PIE for the approximate optimal PPM order M^* is denoted as PIE^*.

4 Receivers

In this chapter we will take a look at physical realizations of measurements that distinguish coherent state alphabets. The respective expressions for mutual information and the error probability are obtained and subsequently compared to the best possible measurement defined by either the Helstrom or Holevo limit.
First, commonly used receivers are reviewed, in particular photon counting in section 4.1, the homodyne receiver in section 4.2 and the double homodyne receiver in section 4.3. Both, the homodyne and the double homodyne detector, are fundamentally limited by the shot noise limit. In order to achieve a detection at a sub shot noise level a different type of receiver is needed. In section 4.4 the displacement receiver is introduced. It stands at the heart of this thesis and its capability of distinguishing states is discussed. In section 4.5 the behavior of the previously introduced receives is discussed in the photon starved regime. In section 4.6 we present structured receivers as a solution to the problem of high peak powers in PPM and finally, in section 4.7 a measurement scheme based on collective detection will be introduced that is able to achieve superadditivity of mutual information in the regime of low energy.

4.1 Photon Counting Detector

Probably the most fundamental idea for the detection of light is to measure the energy contained in a state. By using a POVM that consists of the projections $\{\hat{\pi}_n = |\, n\,\rangle\,\langle\, n\,|\}$ onto the energy eigenbasis of the quantum harmonic oscillator (cf. section 2.3), the energy of quantum states is measured which is particularly useful if the information is encoded solely into the amplitude of the states, e.g. for OOK or PPM. Since the photon number states $|\, n\,\rangle$ form an orthogonal basis, they can be, in principle, distinguished perfectly. Consequently, the ideal process of counting photons is inherently noiseless and serves as an important tool to beat classical detection schemes.
Physically counting photons can be achieved, e.g. by an avalanche photo diode that exploit the photoelectric effect. An incident photon creates an electric signal by creating an electron-hole pair [Hadfield 2009]. Depending on the used bias voltage, the signal current is either proportional to the number of incident photons or simply marks the incidence of a single photon by triggering an avalanche of electrons.
Other important types of photon counting devices are transition-edge sensors [Lita et al. 2008; Fukuda et al. 2011] that exploit a change in the resistance due to temperature changes caused by the photon absorption or superconducting nanowires [Marsili et al. 2013; Yamashita et al. 2013] which detect the breaking of Cooper pairs.
Regardless of the specific device, the incident photon creates a detection event which is

registered including a time stamp. Realistically, photon counters are limited by their *photon number resolution* (PNR), which specifies the largest number of photons the detector is able to resolve. For example, a photon counter with PNR of 2 is able to distinguish the events corresponding to the registration of 0 photons, 1 photon or 2+ photons. A special class of photon counters are those with a PNR of 1, called *single-photon detectors* (SPD), that can only distinguish between no photons at all or at least one. This type of detector is also referred to as Geiger-mode detector or on-off detector. Assuming a coherent state $|\,\zeta\,\rangle$ as incident state, the output probability distribution produced by a perfect photon detector is defined through Born's rule together with eq. (2.72) and reads

$$p\,(n|\zeta) = \mathrm{Tr}\left[\hat{\pi}_n\,|\,\zeta\,\rangle\,\langle\,\zeta\,|\right] = e^{-|\zeta|^2}\frac{|\zeta|^{2n}}{n!}. \tag{4.1}$$

It is a Poissonian distribution centered at the energy of the coherent state $|\zeta|^2$. Also for imperfect detectors with non-unit detection efficiency this distribution is obtained as it only causes a rescaling of the amplitude of a coherent state (cf. section 4.4.4).

Detecting OOK

Because the photon detector is able to perform a projection onto the vacuum state, it is an ideal candidate for the detection of OOK or PPM (cf. sections 3.5.1 and 3.5.4). In the following, we will discuss the detection of OOK. A phase space representation of an OOK alphabet for $\bar{n} = 1$ is shown in fig. 4.1 alongside the corresponding output probability distribution $p\left(n|\sqrt{2\bar{n}}\right)$ given by eq. (4.1) for a photon detector with a PNR of 5. To make a decision which state was received a threshold value k_n is chosen. For any event $n \leq k_n$ the vacuum state is assumed, whereas for $n > k_n$ one chooses a coherent state. In the noiseless scenario the optimal threshold value is given by $k_n = 0$, since the vacuum state will never trigger a detection event. Therefore, a PNR of 1 is sufficient to obtain optimal results.

The average error probability p_{err} for detecting OOK by counting photons is given by

$$p_{\mathrm{err}}^{\mathrm{OOK}} = \frac{1}{2}\left\{1 - \sum_{n=0}^{k_n}\left[p\,(n|0) - p\left(n|\sqrt{2\bar{n}}\right)\right]\right\}. \tag{4.2}$$

For a threshold of $k_n = 0$ only the probability $p\,(0|2\bar{n})$ that no photon is detected such that the error probability which is given by

$$p_{\mathrm{err}}^{\mathrm{OOK}} = \frac{1}{2}e^{-2\bar{n}} \tag{4.3}$$

and coincides with the value for the POVM considered in eq. (3.112). In fig. 4.2 the error probability for several values of k_n is compared to the minimum error probability defined by the Helstrom limit of eq. (3.25). Photon counting provides a similar scaling with the mean photon number as the Helstrom limit, but a constant gap between the two remains. Choosing a suboptimal threshold value $k_n > 0$ causes the error probability

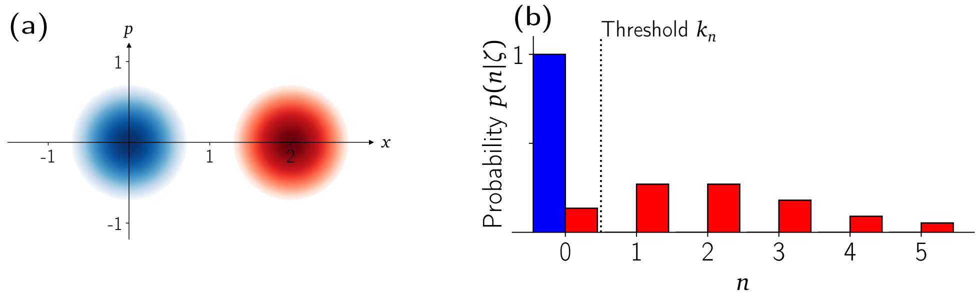


Figure 4.1: **(a)** Schematic depiction of an OOK alphabet in phase space for $\bar{n} = 1$. **(b)** The probability distribution obtained by a photon counting detector with a PNR of 5 defined in eq. (4.1). The decision threshold k_n that achieves the minimum error probability is marked by a dashed line. If no photon is registered the vacuum state $|\,0\,\rangle$ is assumed, while for any detection event the state $\left|\,\sqrt{2\bar{n}}\,\right\rangle$ was received.

to rise dramatically.
Apart from the error probability we can also consider mutual information as a figure of merit. For any threshold value, the expression for mutual information of OOK reduces to the mutual information of a Z-channel given in eq. (3.74). In fig. 4.2 mutual information of OOK is compared to the ultimate quantum limit for detecting OOK which is given by the Holevo bound of eq. (3.88). Counting photons leaves a considerable gap to the Holevo bound which assumes the optimal quantum measurement such that photon counting can not be considered as the optimal measurement.

4.2 Homodyne Detector

In order to distinguish alphabets for which information is encoded into the phase of coherent states, e.g. BPSK, using a photon counting detector does not suffice as it only reveals information about the amplitude. To also gain information about the phase a more intricate setup is needed. As we have seen in section 3.5.2 a POVM that is able to distinguish BPSK consists of projections onto a particular phase space quadrature. In the following a physical realization of such a POVM is introduced which is known as a *homodyne detector*.
A schematic setup of a homodyne detector is shown in fig. 4.3. The incoming signal is interfered with a laser at a 50:50 beamsplitter. The light of the laser comes in form of a coherent state and is usually referred to as the *local oscillator (LO)*. Since a splitting ratio of 50:50 is used, the particular setup is also referred to as balanced homodyne detector. Both output modes of the beamsplitter are measured by photon counting

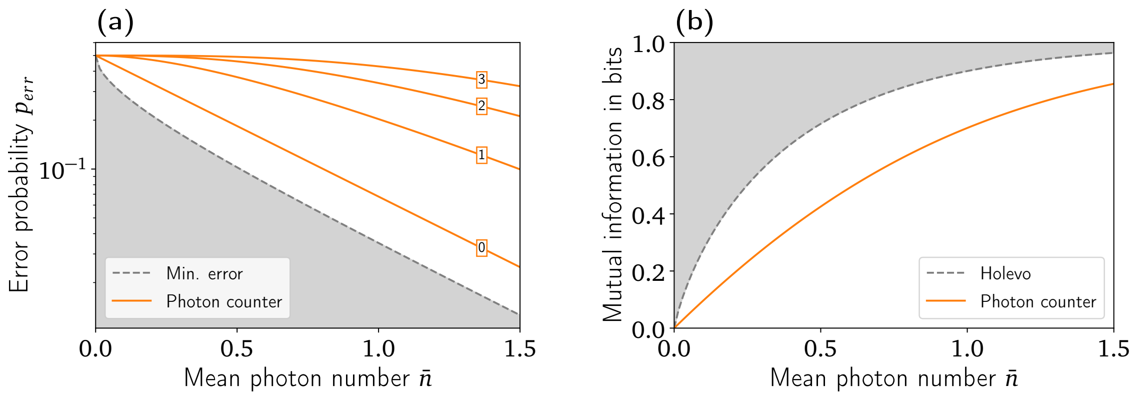


Figure 4.2: **(a)** Error probability p_{err} and **(b)** mutual information I for OOK with $p_0 = 0.5$, when detected by a photon counter of PNR 5. The corresponding threshold value k_n for calculating the error probability is denoted in boxes.

detectors and the resulting photon numbers are subtracted from each other. As we will see, the photon number difference is proportional to the quadrature component of the incident state. The particular phase of the measured quadrature can be controlled via the phase of the local oscillator. A more general derivation than presented below will be given in section 4.4.2 or can be found in many textbooks such as [Scully and Zubairy 1997; Schleich 2001].

Let $\hat{a}$ be the annihilation operator of the incident signal mode and $\hat{b}$ the annihilation operator of the LO mode. Then the two output modes of the beamsplitter are given by

$$\hat{c} = \frac{1}{\sqrt{2}}\left(\hat{a} + i\hat{b}\right), \tag{4.4}$$

$$\hat{d} = \frac{1}{\sqrt{2}}\left(\hat{b} + i\hat{a}\right). \tag{4.5}$$

Both of these modes are measured by a photon counter described by the respective number operator $\hat{c}^\dagger\hat{c}$ or $\hat{d}^\dagger\hat{d}$ such that the subtracted events are given by the operator

$$\hat{d}^\dagger\hat{d} - \hat{c}^\dagger\hat{c} = -i\left(\hat{a}^\dagger\hat{b} - \hat{b}^\dagger\hat{a}\right). \tag{4.6}$$

The expectation value of the obtained statistics is given by [Bandilla 1990; Scully and Zubairy 1997]

$$\left\langle \hat{d}^\dagger\hat{d} - \hat{c}^\dagger\hat{c} \right\rangle = \sqrt{2\left\langle \hat{b}^\dagger\hat{b} \right\rangle}\left\langle \hat{q}_\theta \right\rangle, \tag{4.7}$$

where $\hat{q}_\theta = \frac{1}{\sqrt{2}}\left(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}\right)$ is the quadrature component of the signal mode and θ is the phase difference between the incoming state and the LO. Note that we assumed an LO in form of a coherent state. The balanced homodyne detector therefore can

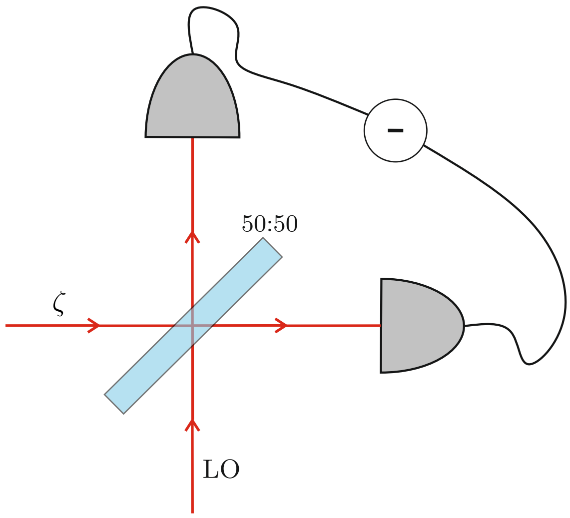


Figure 4.3: A schematic depiction of a balanced homodyne detector. The incident light is interfered with a strong local oscillator (LO) at a 50:50 beam splitter. Both output modes are measured by a photon counter and the obtained results are subtracted. The subtracted statistics are proportional to the quadrature $\hat{q}_\theta$, where the phase θ can be controlled via the phase of the LO.

measure a single arbitrary quadrature of the incident state only rescaled by a factor which depends on the LO. In the following discussion we will neglect this rescaling to keep the expressions concise.
Given a coherent state $|\,\zeta\,\rangle$ as input signal the probability distribution $p\,(q_\theta|\zeta)$ obtained by a balanced homodyne detector is defined by Born's rule of eq. (2.6) together with the overlap $|\langle\, q_\theta\,|\,\zeta\,\rangle|^2$ of eq. (2.58). It is a Gaussian distribution which reads

$$p\,(q_\theta|\zeta) = \frac{1}{\sqrt{\pi}} \exp\left\{-\left[q_\theta - \sqrt{2}\mathrm{Re}\,(\zeta)\cos\,(\theta) - \sqrt{2}\mathrm{Im}\,(\zeta)\sin\,(\theta)\right]^2\right\}. \tag{4.8}$$

For a phase of $\theta = 0$ the quadrature corresponds to the position x and for $\theta = \frac{\pi}{2}$ to the momentum p and the respective probability distributions are given by

$$p\,(x|\zeta) = \frac{1}{\sqrt{\pi}} \exp\left\{-\left[x - \sqrt{2}\mathrm{Re}\,(\zeta)\right]^2\right\}, \tag{4.9}$$

$$p\,(p|\zeta) = \frac{1}{\sqrt{\pi}} \exp\left\{-\left[p - \sqrt{2}\mathrm{Im}\,(\zeta)\right]^2\right\}. \tag{4.10}$$

Contrary to photon counting this measurement is phase-sensitive, since $\mathrm{Re}\,(\zeta) = |\zeta|\cos\,(\phi)$ and $\mathrm{Im}\,(\zeta) = |\zeta|\sin\,(\phi)$, where ϕ is the phase of the coherent state defined by $\zeta = |\zeta|\, e^{i\phi}$.

Detecting BPSK

Consider a BPSK alphabet (cf. section 3.5.2) located on the real axis x in phase space with a-priori probabilities p_0 and p_1. A straightforward way of detecting a binary alpha-

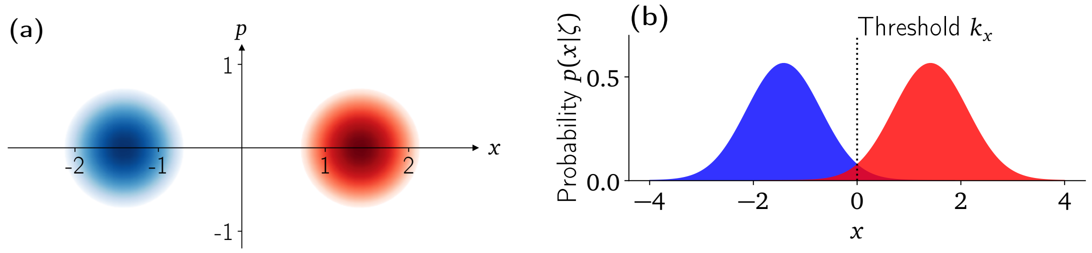


Figure 4.4: **(a)** The schematic depiction of a BPSK alphabet in phase space for $\bar{n} = 1$. **(b)** The corresponding probability distribution $p(x|\zeta)$ according to eq. (4.9). The optimal decision threshold k_x is marked.

bet is to define a threshold value k_x. For measurement results $x \leq k_x$ we assume that state $|\,\zeta_0\,\rangle$ was received, while for $x > k_x$ the state $|\,\zeta_1\,\rangle$ is guessed. In fig. 4.4 a BPSK alphabet is depicted in phase space alongside the corresponding output probability distribution $p(x|\zeta)$ when detected by a homodyne receiver. The conditional probabilities are given by

$$p(0|\zeta_i) = \int_{-\infty}^{k_x} \mathrm{d}x \;\; p(x|\zeta_i)\,, \tag{4.11}$$

$$p(1|\zeta_i) = \int_{k_x}^{\infty} \mathrm{d}x \;\; p(x|\zeta_i) \tag{4.12}$$

and depend on the threshold k_x. Consequently, the error probability is equal to

$$p_{\mathrm{err}}^{\mathrm{bin}}(\zeta_0, \zeta_1) = p_1 p(0|\zeta_1) + p_0 p(1|\zeta_0)\,. \tag{4.13}$$

In fig. 4.5 the error probability is shown for several thresholds k_x as a function of the mean photon number $\bar{n}$ and compared to the corresponding minimum error probability. Choosing the correct threshold is critical for obtaining a low error probability. The optimal decision threshold is given by $k_x = 0$ as is to be expected, since the states are located symmetrically around $x = 0$. It can also be obtained by the Neyman-Pearson criterion [Neyman and Pearson 1933; Helstrom 1976]. For the optimal threshold the conditional probabilities and error probability reduce to the expression derived in section 3.5.2 and the homodyne detector performs the POVM described by eqs. (3.113) and (3.114). However, homodyne detection does not offer the optimal scaling with $\bar{n}$ defined by the minimum error probability which suggests that homodyne detection is suboptimal.
Next, we consider mutual information of the homodyne detector. Up to now we used a decision rule which takes the actual output probability distribution $p(x|\zeta_i)$ defined by

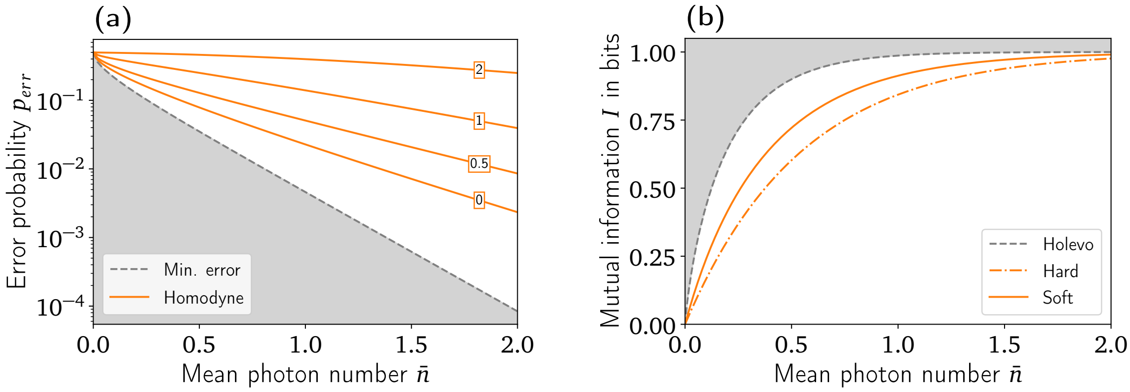


Figure 4.5: **(a)** Error probability p_{err} for BPSK detected with a homodyne detector. The chosen decision threshold k_x is indicated in boxes. The homodyne receiver offers a suboptimal scaling with the mean photon number $\bar{n}$. **(b)** Mutual information for soft and hard decision when detecting BPSK. The ultimate quantum limit is defined by the Holevo bound of eq. (3.88) determined for the BPSK alphabet.

the detector and determines a guess for the received symbol based on the value of x. In this way a new conditional probability distribution $p(j|\zeta_i)$ is formed, where j denotes the guessed received symbol. By defining such a rule we make what is called a *hard decision* and with eqs. (4.11) and (4.12) mutual information reads

$$I_{\text{hard}} = \sum_{i,j} p_i p(j|\zeta_i) \log_2 [p(j|\zeta_i)] - \sum_j p(j) \log_2 [p(j)] \tag{4.14}$$

with the outcome probabilities

$$p(j) = p_0 p(j|\zeta_0) + p_1 p(j|\zeta_1) \,. \tag{4.15}$$

Other than the error probability mutual information can also be calculated directly from the output distribution $p(x, \zeta_i)$, which is then known as *soft decision*. Here, we base mutual information on the actual output rather then the assumed symbols and consequently it is calculated as

$$I_{\text{soft}} = \sum_i \int \mathrm{d}x \; p_i p(x|\zeta_i) \log_2 [p(x|\zeta_i)] - \int \mathrm{d}x \; p(x) \log_2 [p(x)] \tag{4.16}$$

with the outcome probability

$$p(x) = p_0 p(x|\zeta_0) + p_1 p(x|\zeta_1) \,. \tag{4.17}$$

For equiprobable input states with $p_0 = p_1 = \frac{1}{2}$, eq. (4.16) can be written explicitly as

$$I_{\text{soft,eq}} = 4\bar{n} \log_2 [e] - \frac{e^{-2\bar{n}}}{\sqrt{\pi}} \int \mathrm{d}x \; e^{-x^2} \cosh\left(2x\sqrt{2\bar{n}}\right) \log_2 \left[\cosh\left(2x\sqrt{2\bar{n}}\right)\right] . \tag{4.18}$$

The difference between hard and soft decision homodyne receivers can be crucial, especially for low mean photon numbers $\bar{n}$. Figure 4.5 shows mutual information for soft and hard decision homodyne detection of a BPSK alphabet. Soft decision improves the accessible information significantly and always yields a larger or equal amount of information than hard decision. The direct output probability distribution $p(x|\zeta)$ always contains at least as much information as the probability distribution after any decision rule that can be defined. However, soft decision mutual information gives no hint on how to retrieve this information and an appropriate decoding strategy needs to be found first. This highly non-trivial problem gives rise to the field of error correction codes. More details can be found, for instance, in the textbook of Moon [2005].
For both decision types, soft and hard, a considerable gap to the ultimate Holevo limit remains. Only for large mean photon numbers $\bar{n} \gg 1$ the homodyne detector approaches the optimal measurement.

4.3 Double Homodyne Detector

Previously, we considered binary alphabets where it is sufficient to measure only one quadrature in phase space. Of course it is also possible to employ larger alphabets such as QPSK (cf. section 3.5.3), where the information is encoded in both quadratures. A single homodyne detector is not sufficient any more as it is only able to detect one quadrature at a time. However, there is a straightforward way of generalizing the homodyne scheme to detect two quadratures simultaneously. The signal can be measured by splitting the incoming state on a 50:50 beamsplitter into two modes which are both detected by an independent homodyne detection setup as depicted in fig. 4.6. This type of detector is called *double homodyne detector*. The output probability distribution is the joint distribution of both homodyne detectors and based on eqs. (4.9) and (4.10) it reads

$$p(x, p|\zeta) = \frac{1}{\pi} \exp\left\{-\left[x - \mathrm{Re}(\zeta)\right]^2 - \left[p - \mathrm{Im}(\zeta)\right]^2\right\}, \tag{4.19}$$

where $\zeta \to \frac{\zeta}{\sqrt{2}}$ was used to incorporate the preceding 50:50 beamsplitter.

Detecting QPSK

We can define a hard decision rule for the double homodyne detector similarly as for the homodyne detector, but for the detection of QPSK based on the distribution in eq. (4.19). In principle, two decision thresholds, one for each quadrature, have to be defined. For simplicity we will assume that both threshold values are located at the origin, i.e. $k_x = k_p = 0$, since these are the optimal values for the corresponding homodyne detector. For these thresholds the double homodyne detector effectively performs the POVM given by eqs. (3.117) to (3.120) and the conditional probabilities are given by table 3.5. The error probability is given by eq. (3.123) which reads

$$p_{\mathrm{err}}^{\mathrm{QPSK}} = \frac{3}{4}\left[1 - \mathrm{erf}\left(\sqrt{\frac{\bar{n}}{2}}\right)\right]. \tag{4.20}$$

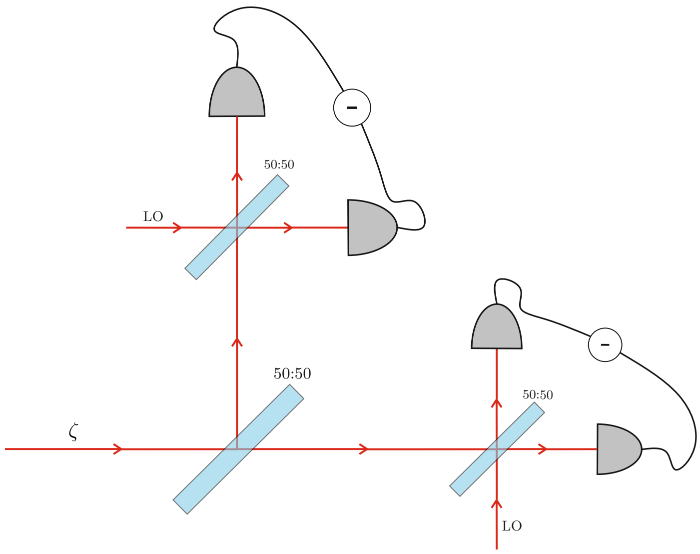


Figure 4.6: A schematic depiction of a double homodyne detector. The incident light is split into two modes at a 50:50 beamsplitter and both modes are measured by an independent homodyne detector. Each homodyne setup is able to measure an arbitrary phase space quadrature such that two quadratures are measured simultaneously, for instance x and p.

In fig. 4.7 the error probability of detecting QPSK symbols with a double homodyne detector is compared to the corresponding Helstrom limit. Since it is non-trivial to determine the minimum error probability for discriminating four states, the Helstrom limit is approximated by a particular measurement, the square root measurement [Holevo 1979; Hughston et al. 1993; Hausladen and Wootters 1994]. For symmetrical states this approximation is quite accurate [Kato et al. 1999; Becerra et al. 2011]. Similar to the homodyne receiver the double homodyne receiver offers a suboptimal scaling with the mean photon number and only gets close to the minimum error measurement for $\bar{n} \gg 1$. Apart from the error probability, we can again consider mutual information as a figure of merit. The hard decision mutual information can be calculated based on the conditional probabilities of table 3.5. Soft decision mutual information is based on the probability distribution $p\left(x,p|\zeta\right)$ of eq. (4.19) and is given by

$$I_{\text{soft}} = \sum_i \int \mathrm{d}x \int \mathrm{d}p \; p_i \, p\left(x,p|\zeta_i\right) \log_2\left[p\left(x,p|\zeta_i\right)\right] - \int \mathrm{d}x \int \mathrm{d}p \; p\left(x,p\right) \log_2\left[p\left(x,p\right)\right]. \tag{4.21}$$

Both types of mutual information are shown in fig. 4.7 in comparison with the Holevo limit. For the double homodyne detector it is very beneficial to employ a soft decision rule as much more information can be extracted from the raw output probability distribution than for a hard decision. Nevertheless, a considerable gap to the ultimate quantum limit remains and only for very high mean photon number $\bar{n}$ the double homodyne detector approaches the optimal measurement for QPSK.

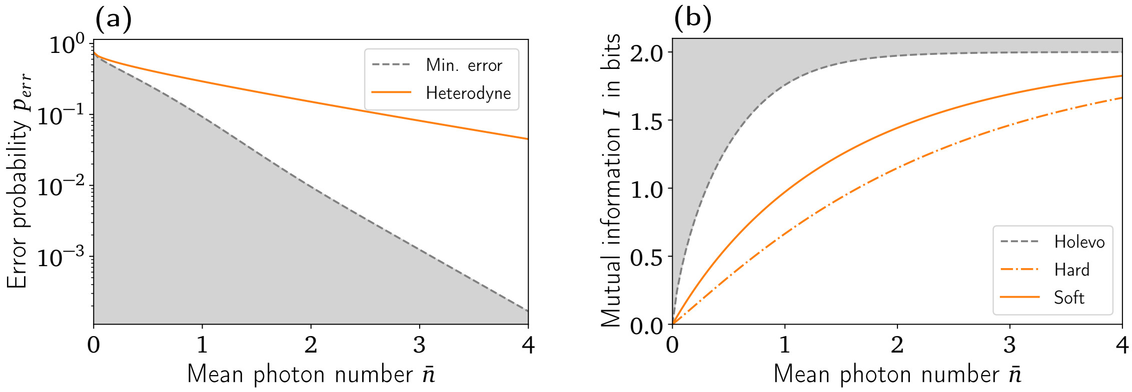


Figure 4.7: **(a)** Error probability p_{err} for QPSK detected with a double homodyne detector. The double homodyne receiver offers a suboptimal scaling with the mean photon number $\bar{n}$ compared to the minimum error probability. **(b)** Mutual information for soft and hard decision heterodyning when detecting QPSK. The ultimate quantum limit is defined by the Holevo bound of eq. (3.88) determined for the QPSK alphabet.

4.4 Displacement Receiver

As we have seen in the previous sections, for binary alphabets a photon counting detector is very useful for detecting OOK, while a homodyne receiver does well for BPSK alphabets. When compared to the ultimate limit, whether it is the Helstrom limit or the Holevo capacity, both fall short. This begs the question if there is an alternative detection scheme that can reach the ultimate limit or at least come closer to it.
Coherent detection, for example homodyne detection, is able to resolve alphabets where information is encoded in the phase of the state, e.g. BPSK. When operating under a constraint on the mean photon number $\bar{n}$, the BPSK alphabet has an inherent advantage over OOK as it possesses a larger state separation in phase space, which translates directly to a better distinguishability. Homodyne detection, however, is inherently limited by the fundamental quantum fluctuations of light known as the shot noise of a receiver. Photon counting itself does not have this limitation as it performs a projection onto an orthogonal energy basis. In order to use the advantages of a BPSK alphabet and overcome the shot noise limitation of a homodyne detector a *displacement receiver* (DR) can be used, which was first introduced by Kennedy [1973]. A sketch of a displacement receiver is shown in fig. 4.8. Here, the incoming state is first displaced by a complex amplitude β by applying the displacement operator $\hat{D}(\beta)$ defined in eq. (2.78). Physically, this is achieved by interfering the signal state with a local oscillator at a highly transmissive beamsplitter. After applying the displacement, the coherent state has a complex amplitude of $\zeta + \beta$ and is detected by a photon counting detector.
To illustrate the operational principle of the displacement receiver, consider a BPSK alphabet, i.e. the discrimination of two states $\left|-\sqrt{\bar{n}}\right\rangle$ and $\left|\sqrt{\bar{n}}\right\rangle$. The displacement

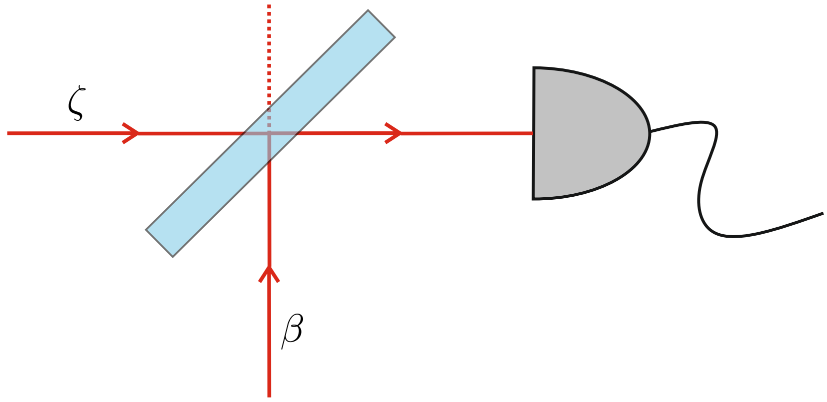


Figure 4.8: A schematic depiction of a displacement receiver. The incident state is displaced by a complex amplitude β. This is achieved by interfering the incoming light with a local oscillator (LO) at a beam splitter with almost unit transmission. The incoming state after the displacement operation has an amplitude of $\zeta + \beta$ and is measured by a photon counting detector.

amplitude is chosen as $\beta = \sqrt{\bar{n}}$ such that the state after the displacement is either the vacuum state $|\, 0\, \rangle$ or the coherent state $\left|\, 2\sqrt{\bar{n}}\, \right\rangle$. If a photon is registered at the photon counter we know state $\left|\, 2\sqrt{\bar{n}}\, \right\rangle$ was received whereas for no detection event the vacuum state is assumed.

The states before and after the displacement operation together with the corresponding probability distribution $p\,(n|\zeta)$ of eq. (4.1) obtained by the photon counter are shown in fig. 4.9. After applying the displacement, the states essentially constitute an OOK alphabet which now can be distinguished by a photon counting detector as the probability distributions are well distinguishable. It is important to note that the OOK alphabet obtained after the displacement has a twice larger energy compared to a directly encoded OOK alphabet.

The distinguishability of the displaced states heavily depends on the displacement amplitude β and, in general, the optimal value needs to be determined through optimization. In the following, we will assume a generic binary coherent alphabet with states $|\, \zeta_0\, \rangle$ and $|\, \zeta_1\, \rangle$ and their respective a-priori probabilities p_0 and p_1.

First, we consider the error probability as a figure of merit. To that end it is necessary to define a decision threshold k_n for the detected number of photons. For $n \leq k_n$ it is assumed that state $|\, \zeta_0\, \rangle$ was measured, and for $n > k_n$ state $|\, \zeta_1\, \rangle$. Then the error probability is given by

$$p_{\text{err}}^{DR} = p_0 \sum_{n=0}^{k_n} p\,(n|\zeta_0 + \beta) + p_1 \left(1 - \sum_{n=0}^{k_n} p\,(n|\zeta_1 + \beta) \right). \tag{4.22}$$

Now the error probability p_{err}^{DR} is minimized with respect to the displacement β for a fixed threshold k_n. This minimization is repeated for all possible values of k_n until the optimal threshold is found eventually. In fig. 4.10 the error probability of the displace-

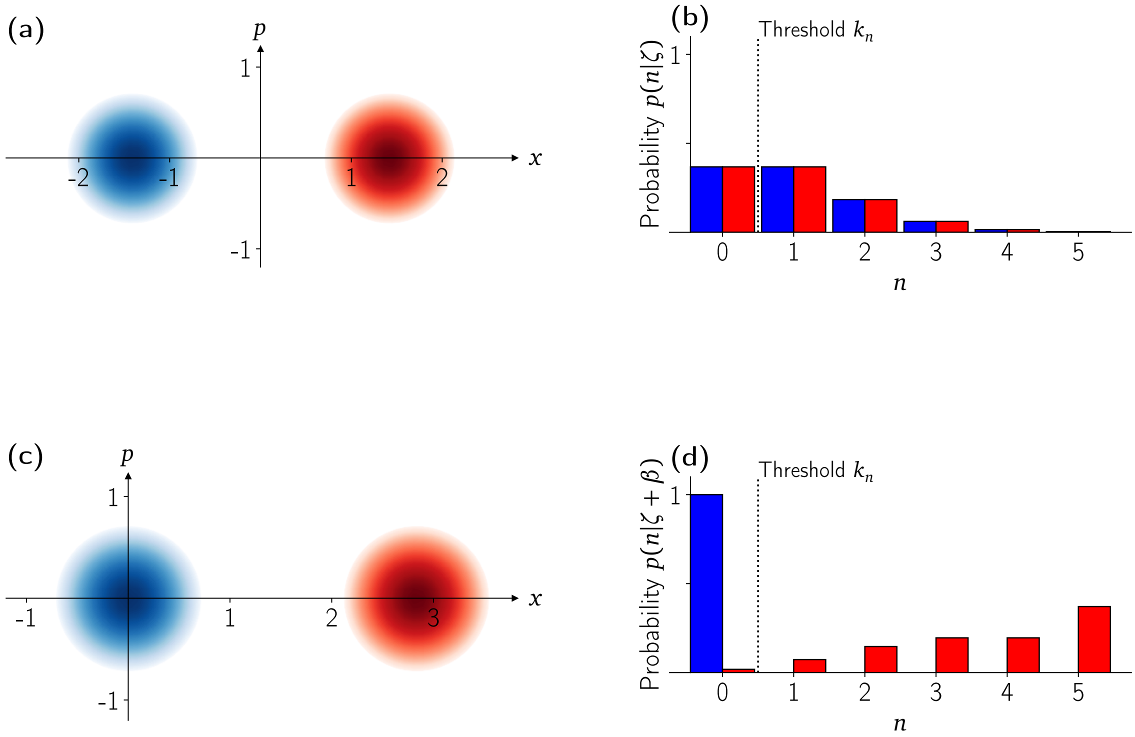


Figure 4.9: **(a)** The BPSK alphabet for $\bar{n} = 1$ before displacement depicted in phase space and **(b)** the corresponding probability distribution when detected by a photon counting detector with a PNR of 5. **(c)** The BPSK alphabet for $\bar{n} = 1$ after applying a displacement of $\beta = \sqrt{\bar{n}}$ depicted in phase space and **(d)** the corresponding probability distribution for detector limited to a PNR of 5. The optimal decision threshold k_n is marked with a dashed line.

ment receiver is shown for OOK and BPSK. If one of the states is simply displaced to the vacuum state, the displacement receiver achieves a similar scaling with the mean photon number $\bar{n}$ as the minimum error measurement. Yet, a much lower error probability is achieved by optimizing the displacement, in particular for small values of $\bar{n}$. Homodyne detection is beaten for all mean photon numbers and thus detection at a sub-shot noise level is achieved.

Soft decision mutual information for the displacement receiver can be calculated based on the probability distribution $p(n|\zeta_i + \beta)$ given by eq. (4.1). In fig. 4.10 mutual information for OOK and BPSK is compared to the corresponding Holevo limit. Here, the displacement receiver offers a similar performance as a homodyne detector, but for larger values of $\bar{n}$ gives a slight advantage. Optimizing the displacement β increases mutual information significantly. Nevertheless, a considerable gap to the optimal measurement defined by the Holevo limit remains in all cases.

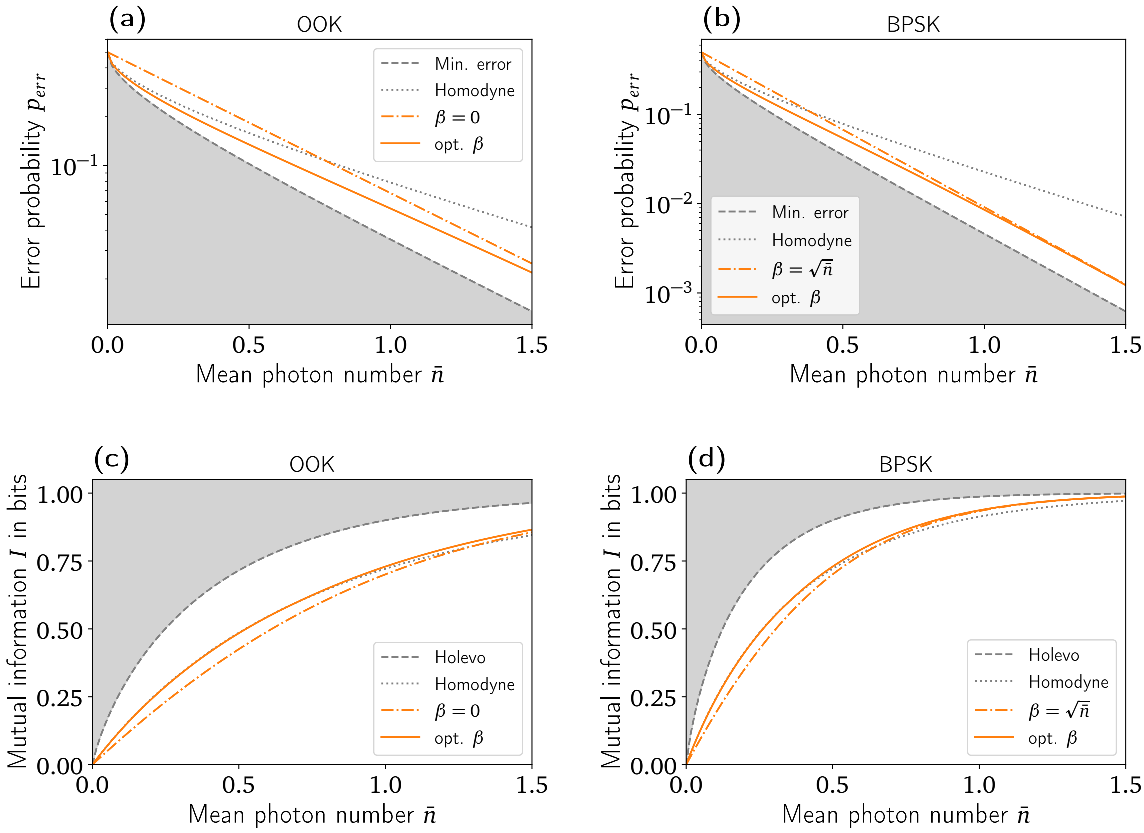


Figure 4.10: Error probability p_{err} for **(a)** OOK and **(b)** BPSK detected by a displacement receiver compared to the minimum error measurement. Optimizing the displacement β is crucial to obtain a low error probability. Mutual information I for soft decision for **(c)** OOK and **(d)** BPSK compared to the Holevo limit. Optimizing the displacement amplitude β increases mutual information significantly.

Note that, in principle, the displacement receiver is not only able to resolve OOK or BPSK, but any binary coherent alphabet. While there is no advantage to do so in the noiseless scenario, it is essential when communicating under phase noise as will be discussed in section 6.3.
Since the displacement receiver offers sub-shot noise detection, it continues to be an active field of research. Various experiments show the feasibility of the approach, also in fibre-based implementations at telecom wavelength or with modern sensors allowing for a high detection bandwidth [Wittmann et al. 2008; Takeoka and Sasaki 2008; Wittmann et al. 2010; Tsujino et al. 2010; 2011; DiMario and Becerra 2018; DiMario et al. 2019; Shcherbatenko et al. 2020; Thekkadath et al. 2021].
In the following sections we will investigate several aspects of the displacement receiver in more detail. First, in section 4.4.1 we will discuss how the gap to the minimum error measurement for binary alphabets can be closed by the addition of feedback. In sec-

tion 4.4.2 the relationship between the homodyne detector and the displacement receiver is investigated whereas section 4.4.3 is dedicated to the problem of detecting non-binary alphabets, e.g. QPSK, with the displacement receiver. Finally, in section 4.4.4 possible imperfections in realistic implementations of the displacement receiver and their impact on the performance are discussed.

4.4.1 Approaching the Limit

None of the so far discussed receivers actually reaches the Helstrom limit. While the POVMs for the minimum error measurement are known, their physical realization often is not. Nevertheless, for discriminating binary coherent states an implementation was found by Dolinar [1973]. It is based on the displacement receiver, but utilizes a real-time feedback loop to adjust the displacement based on the so far observed detection events. As we will see in this section, the gap in performance between a displacement receiver and the Dolinar receiver can be closed by a multi-copy version of the displacement receiver, where subsequent measurements are adapted to previously obtained results.
Let us start out by considering a very simple idea to improve a displacement receiver with PNR $= 1$. We divide the incoming coherent state $|\,\zeta_{0,1}\,\rangle$ into a series of M states with weaker amplitudes $\frac{\zeta_{0,1}}{\sqrt{M}}$. Physically, this can be achieved by splitting the original pulse into M time bins and treat each time bin as a separate copy. Each of these weaker copies is measured separately and for all measurements the same displacement amplitude β is used. For the sake of simplicity we assume $\beta = -\frac{\zeta_0}{\sqrt{M}}$, such that state $|\,\zeta_0\,\rangle$ is displaced to the vacuum state. If we detect a photon for any copy, we know state $|\,\zeta_1\,\rangle$ was received and for no photons at all we assume it was $|\,\zeta_0\,\rangle$. The only source of error is when $|\,\zeta_1\,\rangle$ does not produce any detection event and consequently, the error probability for M measurements is equal to $p_{\text{err}} = \exp\left(-\frac{4|\zeta_1|^2}{M}\right)^M = \exp\left(-2\,|\zeta_1|^2\right)$. Not surprisingly the benefit of having M-copies is completely mitigated by their weaker amplitudes, such that this scheme does not provide any improvement. In fact, it has the same error probability as a single-copy displacement receiver. Note that this does not change even if the displacement amplitude used in all stages is freely optimized.
Since a static M-copy measurement clearly does not improve the error probability, it is necessary to turn to an adaptive measurement scheme. By adding feedforward between subsequent measurements each stage can be adapted according to the current knowledge. This approach is commonly referred to as Bayesian updating.
Each measurement is performed by a displacement receiver, who can either detect one or zero photons, such that possible outcomes are $e_m \in \{0, 1\}$. We will also refer to these detection events as (detector) click or no click. The string of results known after measurement m is denoted as $\mathbf{e}_m$. After an outcome was obtained, the displacement β_m for the following stage is optimized based on the current knowledge. Formally, this is achieved by using the posterior probability, that is the probability of having received state $|\,\zeta_i\,\rangle$ given the results $\mathbf{e}_{m-1}$, to update the prior probability for measurement m. The prior probability $P^{(m)}\left(\zeta_i\right)$ of state $|\,\zeta_i\,\rangle$ for the m-th measurement can be determined

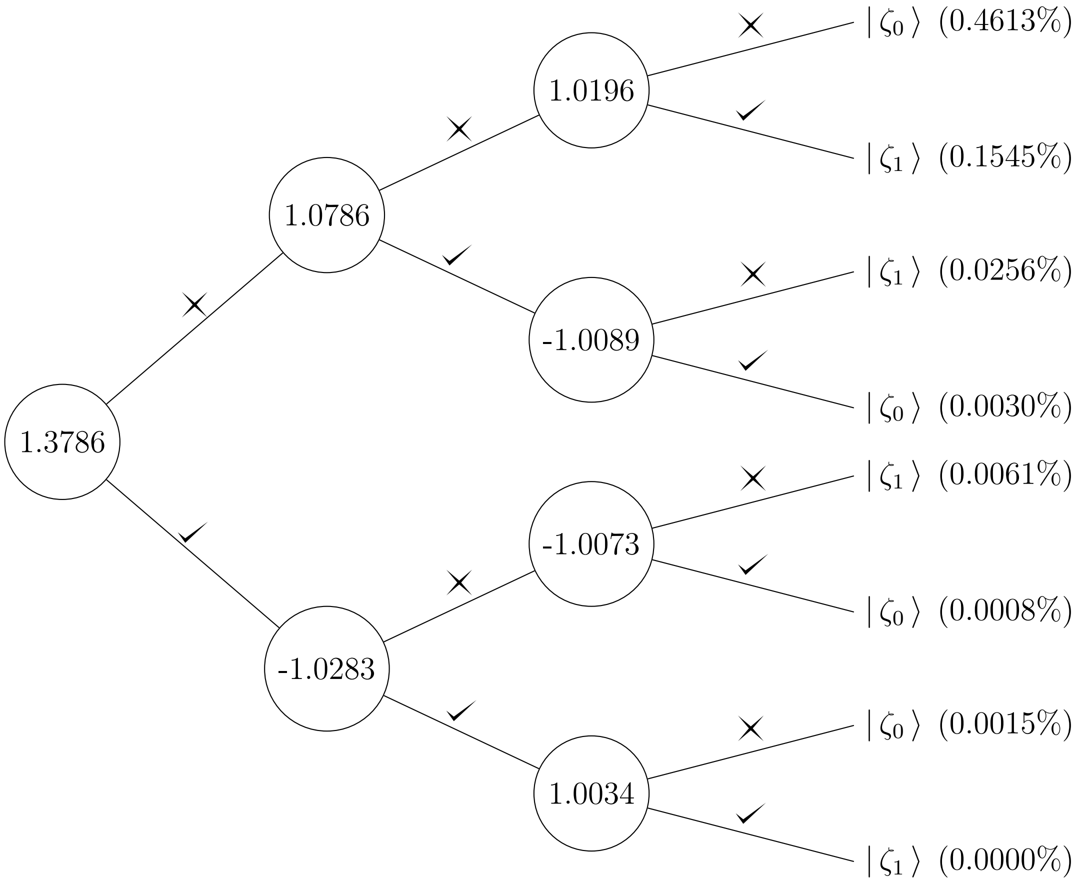


Figure 4.11: Decision tree for a 3-copy measurement scheme ($M = 3$) with a displacement receiver detecting a coherent BPSK alphabet under the constraint $\bar{n} = 1$. The displacement amplitude is optimized for each measurement stage by minimizing eq. (4.22) with the corresponding priors of eq. (4.23). The optimal displacement is depicted in a circle at each node as a fraction of the state amplitude, i.e. the shown value is calculated as $\frac{\sqrt{M}\beta}{\zeta_1} = \sqrt{3}\beta$. A fraction of 1 corresponds to displacing $|\zeta_0\rangle$ to the vacuum state. A positive sign of the displacement fraction corresponds to test for $|\zeta_0\rangle$, while a negative sign tests for $|\zeta_1\rangle$. The symbol ✓(✕) denotes that a (no) photon was detected in the previous measurement. The leafs show the guessed state for the corresponding string of detection events and the shown percentage denotes the weighted (and rounded) error probability for the particular guess. The overall error probability is $p_{\text{err}} \approx 0.6514\%$. Note that this value was calculated from the exact results and then was rounded. The corresponding error probability for an optimized single-shot displacement receiver is $p_{\text{err}} = 0.8561\%$ and the minimum error probability is $p_{\min} = 0.4600\%$.

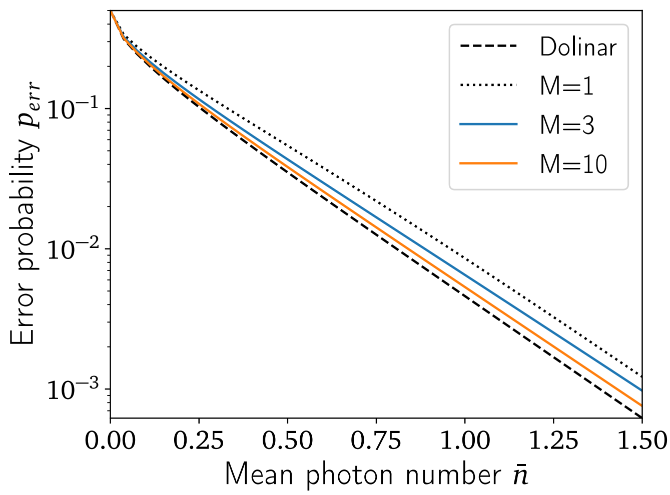


Figure 4.12: The error probability of the adaptive M-copy detection scheme with a displacement receiver as a function of the mean photon number $\bar{n}$ for a BPSK alphabet. The incoming state is divided into M-copies which are measured separately. Each subsequent measurement depends on the outcomes of previous measurements. The minimum error probability is reached with the Dolinar receiver.

according to Bayes's rule by

$$P^{(m)}\left(\zeta_i\right) = \frac{P^{(m-1)}\left(\zeta_i\right) P\left(\mathbf{e}_m|\zeta_i\right)}{P\left(\mathbf{e}_m\right)}, \tag{4.23}$$

where $P\left(\mathbf{e}_m|\zeta_i\right)$ is the probability of obtaining outcome e_m given state $|\,\zeta_i\,\rangle$ after the previous events $\mathbf{e}_{m-1}$. The marginal probability for outcome e_m is given by

$$P\left(\mathbf{e}_m\right) = P^{(m-1)}\left(\zeta_0\right) P\left(\mathbf{e}_m|\zeta_0\right) + P^{(m-1)}\left(\zeta_1\right) P\left(\mathbf{e}_m|\zeta_1\right). \tag{4.24}$$

The optimal displacement amplitude β_m for each measurement is determined by minimizing eq. (4.22) with the prior probabilities $p_i = P^{(m)}\left(\zeta_i\right)$ and state amplitudes $\frac{\zeta_i}{M}$. Figure 4.11 shows the decision tree for measuring a coherent BPSK alphabet for $\bar{n} = 1$ divided into $M = 3$ adaptive stages. The overall error probability is $p_{\text{err}} = 0.6514\%$ compared to 0.8561% for a single-copy measurement.
At each node the more likely state should be displaced close to vacuum. Whenever a photon is detected, our current guess seems to be wrong and the other state becomes more likely. In other words, a detector click tells us we guessed wrong, while no click means we guessed correctly. Therefore, it becomes necessary to change the displacement for the following measurement after a click, such that the other state will be displaced close to vacuum.
The contribution of a particular string of detection events to the error probability varies heavily. In general, each detector click lowers the error probability of the corresponding

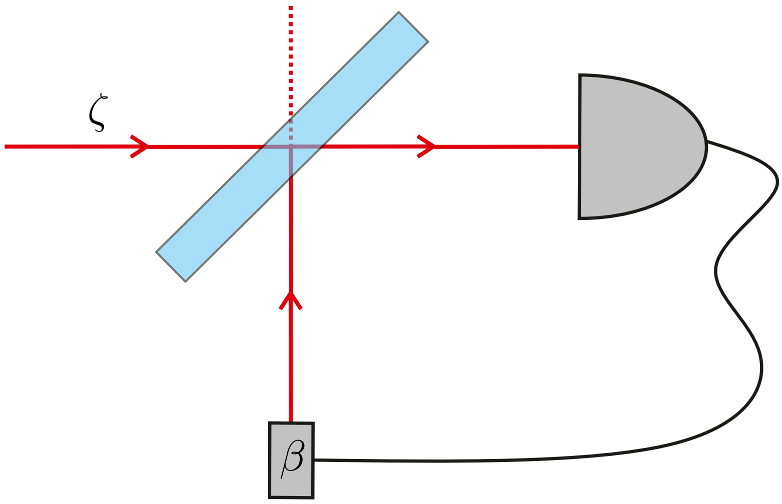


Figure 4.13: A schematic depiction of the Dolinar receiver. The incident state is displaced by a complex amplitude β. This is achieved by interfering the incoming light with a local oscillator (LO) at a beam splitter power transmission of almost 1. The displaced state with amplitude $\zeta + \beta$ is measured by a photon counter. Depending on the obtained result, feedback is sent to the displacement stage and the displacement amplitude β is adjusted to the current knowledge in real time.

branch more than the opposite event. A detected photon tells us with great confidence that our current guess is wrong, whereas the absence of a photon confirms our current believe. However, in the latter case the uncertainty is much greater as the further displaced state has a non-zero probability of producing a click. Due to the nature of Poissonian statistics this probability is larger than receiving a click from the state displaced close to vacuum. Consequently, the largest error stems from measuring no photons at all.
Lastly, clicks obtained earlier in the procedure contribute more to lowering the error probability than events observed in later stages as they help to focus the measurements in subsequent stages. This is also visible in the shrinking displacement amplitude, which indicates narrowing knowledge about the received state.
The error probabilities obtained for the displacement receiver for several values of M are shown in fig. 4.12. With increasing number of copies M the error probability approaches the minimum error probability. In order to saturate the Helstrom bound it is necessary to increase $M \to \infty$, i.e. go to a continuous version of the adaptive displacement receiver [Assalini et al. 2011]. This is achieved by placing a real-time feedback loop between the measurement and displacement stage, which amounts to the Dolinar receiver [Dolinar 1973]. A schematic of the latter is shown in fig. 4.13. Whenever the detector clicks, feedback is sent to the displacement stage to adjust it appropriately. The optimal displacement is given by two symmetric time-dependent solutions for the envelopes of the displacement field, such that each of them corresponds to test for one of the states. After each detector click the displacement is switched to the other solution. The Dolinar receiver requires very fast feedback which has, so far, limited its applica-

tions to proof-of-principle experiments [Cook et al. 2007]. Technically less challenging is the M-copy displacement receiver as it requires only feed-forward, as was demonstrated by Sych and Leuchs [2016].

Finding physical implementations for reaching the Helstrom limit remains an elusive task as the Helstrom formalism does not provide a constructive way to derive them and can require collective measurements for distinguishing multiple copies of a state. Interestingly, for the discrimination of pure quantum states collective measurements are not needed as the limit can be attained by a single-copy measurement [Jarzyna and Kolodynski 2020]. In fact, when discriminating pure states, collective measurements are equivalent to adaptive M-copy measurements which require only local operations and classical communication (LOCC) and can be implemented by Bayesian updating [Acín et al. 2005]. Another way of reaching the Helstrom limit was recently proposed and utilizes non-demolition measurements based on atom-light interaction [Han et al. 2018]. Reaching the Holevo bound remains a much more challenging task and so far has not been demonstrated. Only in particular regimes the Holevo bound can be saturated with known receivers, such as PNR measurements in the low power limit [Jarzyna 2019] or double homodyne detection in the high power regime [Giovannetti et al. 2004]. The latter, however, only approaches the Holevo limit asymptotically in the sense that the ratio $I_{\text{soft}}/\mathcal{C}_{\text{LB}} \to 1$ only leaving a constant gap of 1 nat.

4.4.2 A Universal Coherent Receiver

Usually, displacement and homodyne receivers are seen as different types of detectors. A reason for this is, for instance, that a homodyne receiver is limited by shot-noise, whereas a displacement receiver can perform at a sub-shot noise level. Furthermore, for coherent states a homodyne detector obtains Gaussian statistics originating from the measured quadrature, while the displacement receiver outputs Poissonian statistics. In the following section, we want to explore the connection between these two receivers and we will see that they can be understood as different operational modes of the same setup. Also, we will discuss the origin of shot noise, which is often seen as fundamental intrinsic noise, but actually is introduced into the system.

Let us consider a coherent receiver depicted schematically in fig. 4.14. The incoming signal pulse is interfered with a local oscillator (LO) at a beamsplitter with power transmission T. The output mode $\hat{c}$ is given by

$$\hat{c} = \sqrt{T}\hat{a} + i\sqrt{1-T}\hat{b}, \tag{4.25}$$

where $\hat{a}$ is the annihilation operator of the signal state mode and $\hat{b}$ of the LO mode. The annihilation operator of the output mode $\hat{c}$ is measured by a photon counting detector with unlimited PNR and the average number of detected photons is given by [Bandilla

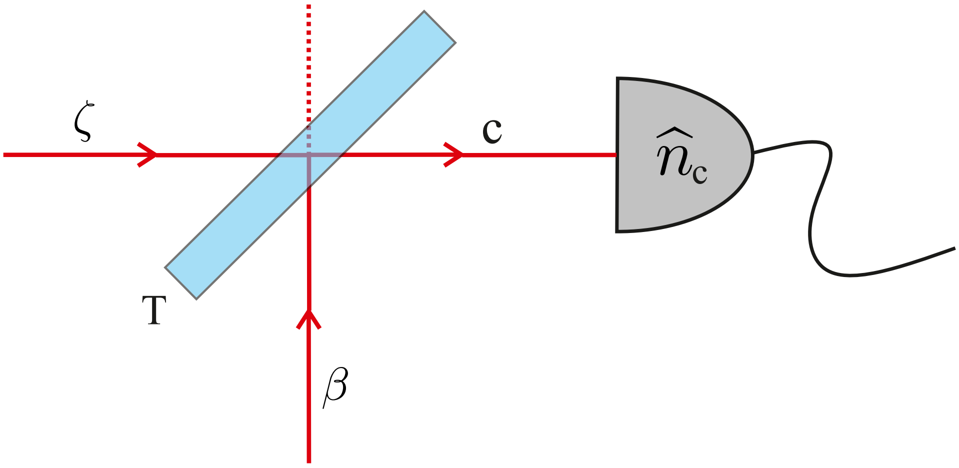


Figure 4.14: A schematic depiction of a coherent receiver. The incident light is interfered with a local oscillator with complex amplitude β at a beam splitter with power transmission T. The output mode c is measured by a photon counter.

1990; Scully and Zubairy 1997]

$$\hat{n}_c = \hat{c}^\dagger \hat{c} = T\hat{a}^\dagger \hat{a} + (1-T)\,\hat{b}^\dagger \hat{b} + i\sqrt{T\,(1-T)}\left(\hat{a}^\dagger \hat{b} - \hat{b}^\dagger \hat{a}\right), \tag{4.26}$$

$$\Rightarrow \langle\, \hat{n}_c \,\rangle = T\,\langle\, \hat{a}^\dagger \hat{a} \,\rangle + (1-T)\left\langle\, \hat{b}^\dagger \hat{b} \,\right\rangle - \sqrt{2T\,(1-T)}\sqrt{\left\langle\, \hat{b}^\dagger \hat{b} \,\right\rangle}\,\langle\, \hat{q}_\theta \,\rangle\,, \tag{4.27}$$

where $\hat{q}_\theta = \frac{1}{\sqrt{2}}\left(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}\right)$ is the quadrature operator of the signal mode and θ is the phase difference between the incoming state and the LO. The variance of $\hat{n}_c$ can be calculated directly from eq. (4.26) and reads

$$\mathrm{Var}\,(\hat{n}_c) = T^2 \mathrm{Var}\,(\hat{a}^\dagger \hat{a}) + (1-T)^2\,\mathrm{Var}\left(\hat{b}^\dagger \hat{b}\right) + 2T\,(1-T)\,\mathrm{Var}\left(\hat{b}^\dagger \hat{b}\right)\mathrm{Var}\,(\hat{q}_\theta)\,. \tag{4.28}$$

The obtained statistics for $\hat{n}_c$ in eqs. (4.27) and (4.28) consist of three terms. The first two are proportional to the photon statistics of the incoming signal state and the LO respectively. The third part is proportional to the quadrature $\hat{q}_\theta$ of the signal state. One can choose a particular quadrature to measure by controlling the phase of the LO. Note that a phase jump of $\frac{\pi}{2}$ occurs when the LO is reflected at the beamsplitter. For the sake of simplicity we assume from here on out, that the LO phase is chosen such that the quadrature operator corresponds to the position operator $\hat{x} = \frac{1}{\sqrt{2}}\left(\hat{a} + \hat{a}^\dagger\right)$. Additionally, we assume that the LO comes in form of a coherent state with real-valued amplitude β. The corresponding expectation value and variance are equal to

$$\left\langle\, \hat{b}^\dagger \hat{b} \,\right\rangle = \beta^2, \tag{4.29}$$

$$\mathrm{Var}\left(\hat{b}^\dagger \hat{b}\right) = \beta^2. \tag{4.30}$$

With these assumptions, the eqs. (4.27) and (4.28) simplify to

$$\langle n_c \rangle = T \left\langle \hat{a}^\dagger \hat{a} \right\rangle + (1-T)\,\beta^2 - \sqrt{2T\,(1-T)}\beta \left\langle \hat{x} \right\rangle, \tag{4.31}$$
$$\mathrm{Var}\,(\hat{n}_c) = T^2 \mathrm{Var}\left(\hat{a}^\dagger \hat{a}\right) + (1-T)^2\,\beta^2 + 2T\,(1-T)\,\beta^2 \mathrm{Var}\,(\hat{x})\,. \tag{4.32}$$

For the considered detection scheme, the transmission T of the beamsplitter is close to unity ($T \approx 1$) in order to retain most of the incoming signal. The LO amplitude β is chosen such that the reflected fraction $\sqrt{1-T}\beta$ remains finite. For the further discussion it is therefore convenient to introduce the effective LO amplitude as $\beta_{\text{eff}} = \sqrt{1-T}\beta$. Now, the output statistics simplify to

$$\langle n_c \rangle = T \left\langle \hat{a}^\dagger \hat{a} \right\rangle + \beta_{\text{eff}}^2 - \sqrt{2T}\beta_{\text{eff}} \left\langle \hat{x} \right\rangle, \tag{4.33}$$
$$\mathrm{Var}\,(\hat{n}_c) = T \mathrm{Var}\left(\hat{a}^\dagger \hat{a}\right) + (1-T)\,\beta_{\text{eff}}^2 + 2T\beta_{\text{eff}}^2 \mathrm{Var}\,(\hat{x})\,. \tag{4.34}$$

We see that the eventually performed measurement is defined by the value of the effective LO amplitude β_{eff}.

Homodyne detection

Let us consider an operational mode for the coherent receiver in which a very strong LO $\left(\beta_{\text{eff}} \gg \left\langle \hat{a}^\dagger \hat{a} \right\rangle\right)$ is used. This mode corresponds to homodyne detection. The terms $T \left\langle \hat{a}^\dagger \hat{a} \right\rangle$ can be neglected due to the large displacement amplitude and we obtain

$$\langle n_c \rangle = \beta_{\text{eff}}^2 - \sqrt{2T}\beta_{\text{eff}} \left\langle \hat{x} \right\rangle, \tag{4.35}$$
$$\mathrm{Var}\,(\hat{n}_c) = (1-T)\,\beta_{\text{eff}}^2 + 2T\beta_{\text{eff}}^2 \mathrm{Var}\,(\hat{x})\,. \tag{4.36}$$

The first term of the expectation value in eq. (4.35), β_{eff}^2, is known and can be easily subtracted from the results. Then, the only remaining term is proportional to the quadrature $\hat{x}$ of the incoming state, which is the quantity effectively measured. For a coherent state as input, this means Gaussian statistics are obtained.
The variance in eq. (4.36) has two contributions, the first one depends only on the LO, known as the LO shot noise, and a second one containing the quadrature shot noise. Consequently, homodyne detection is limited by the shot noise and not the signal noise. While this limitation can not be overcome when using a strong LO, there exists an alternative homodyne detection scheme, the balanced homodyne detection as discussed in section 4.2. It has a lower overall noise as it suppresses the LO shot noise. However, the quadratures shot noise remains.

Displacement receiver

The second operation mode of the universal coherent receiver utilizes a LO with an effective amplitude approximately as large as the signal amplitude, i.e. $\beta_{\text{eff}}^2 \approx T\zeta^2$. This mode corresponds to operating as a displacement detector. If we also assume a coherent state as input, the output statistics of eqs. (4.33) and (4.34) read

$$\langle n_c \rangle = 2\zeta^2 \pm 2\zeta^2, \tag{4.37}$$
$$\mathrm{Var}\,(\hat{n}_c) = 2\zeta^2, \tag{4.38}$$

where $T \approx 1$, the quadrature expectation value $\langle \hat{x} \rangle = \sqrt{2}\zeta$ and the variance $\mathrm{Var}(\hat{x}) = 0.5$ of a coherent state was used. The expectation value depends on the sign of the LO amplitude. Here, the measurement is only limited by the signal noise contrary to homodyne detection, where the LO noise dominates. Therefore, shot noise has to be understood as a limitation stemming from the usage of a very strong LO and not as a fundamental limit on the measurement of light. In fact, in the considered scheme shot noise gets gradually introduced with growing LO power. This is illustrated by the following reasoning. For a coherent state as input state and $T \approx 1$, the general variance of eq. (4.34) can be written as

$$\mathrm{Var}(\hat{n}_c) = \zeta^2 + \beta_{\text{eff}}^2. \tag{4.39}$$

We see that both signal and LO contribute equally to the variance. For the displacement receiver, both parts remain as the LO energy is roughly equal to the signal energy. If the LO strength is gradually increased, the LO noise β_{eff}^2 contributes more and more to the overall noise, until it completely overwhelms the signal noise. This is the regime where the detection is limited to the shot noise level. Note that in principle the shot noise can be reduced to a sub coherent level when measuring different types of states. Squeezed states, for instance, can lower the quadrature noise further and therefore the detection noise also gets reduced.
In summary, we see that the displacement receiver and homodyne detection are closely connected. Each of them corresponds to a different regime of the LO strength of the universal coherent receiver. The displacement receiver is able to operate as a homodyne receiver with increasing displacement. For very large displacement (and unlimited PNR), both homodyne and displacement detectors, measure a quadrature of the state. Intuitively, the discussed behavior of the coherent receiver can be understood as a consequence of the fact that Poissonian and Gaussian statistics are approximately equal in the limit of very large mean values.

4.4.3 Detecting Larger Alphabets

So far we only investigated the performance of a displacement receiver when detecting binary coherent alphabets. However, standard communication strategies often employ non-binary alphabets, for instance, QPSK. Particularly, when the average energy $\bar{n}$ is large it becomes beneficial to extend the alphabet to a larger number of states. In contrast to coherent receivers, like double homodyne detection, the displacement receiver is not suited to discriminate more than two states in a single shot. Its operation principle is to exclude one particular state, which is naturally not sufficient to discriminate non-binary alphabets.
To detect higher-dimensional alphabets with a displacement receiver in a single shot, it becomes necessary to split the signal into several copies with lower energy. Each copy is then detected by a displacement receiver in independent ports. A schematic depiction of the setup is shown in fig. 4.15. In the following we investigate the capability of such a detection scheme with M ports to detect alphabets consisting of up to 4 coherent states.

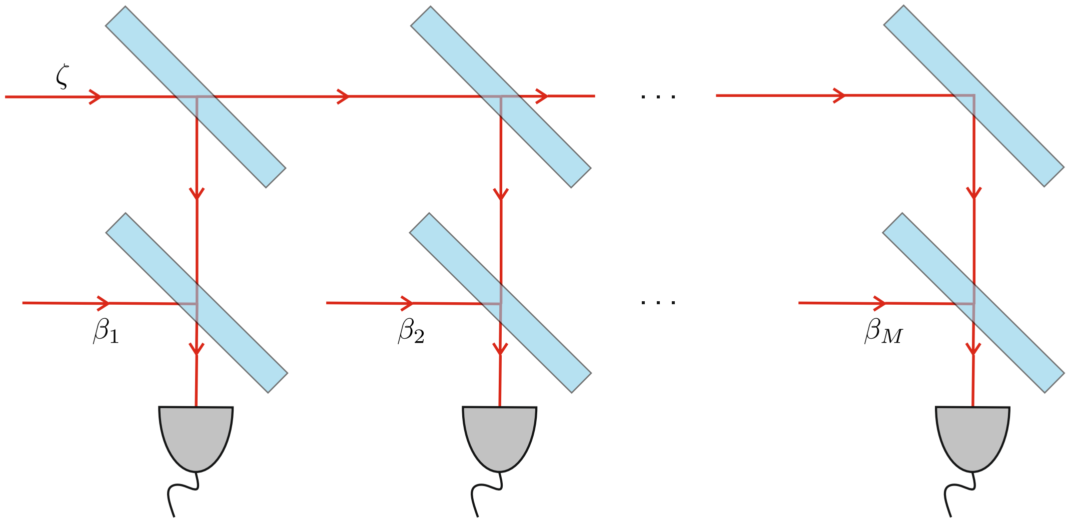


Figure 4.15: A schematic depiction of a receiver for detecting higher order alphabets. The incident state ζ is split up into M ports by a series of beamsplitters with variable transmission. In each port a displacement receiver is detecting the copy in the respective port.

Each port receives a copy of the signal with only a fraction υ_m of the signals total energy, such that $\sum_m^M \upsilon_m = 1$. Physically, this can be achieved by a series of beamsplitters with appropriate transmissivities. Each copy is displaced by an amplitude β_m and consecutively detected by a photon counter. The energy fractions υ_m and the displacements β_m for each port have to be optimized globally. Note that it is equivalent to using only a single displacement receiver which switches its displacement at certain times t_m. Each time frame corresponds to a single copy in one of the M ports. This, however, requires that the signal is stretched over a certain time which can be accomplished, for instance, by appropriate delay lines.
For the following optimization we fix the displacement direction (specified by its phase) in the m-th port to be $e^{im\pi/2}$, such that only the absolute value of the displacement $|\beta_m|$ has to be optimized. The displacement amplitude for port i itself is given by the relation $\beta_m = |\beta_m| e^{im\pi/2}$. To quantify the performance mutual information for both soft and hard decision is used. Error probability itself is not suited for a comparison as it is only comparable for schemes with the same alphabet size. The overall goal of this optimization is to find the optimal communication scheme and therefore both, BPSK and QPSK, are investigated in order to determine which one is optimal. For detectors with a PNR of 1 mutual information is shown in fig. 4.16 alongside the optimal parameters. Here, the displacement receiver provides a constant advantage over conventional measurements. It is optimal to use BPSK for low mean photon numbers $\bar{n}$ such that a single displacement receiver is sufficient to obtain optimal mutual information. For higher mean photon numbers $\bar{n}$ it becomes necessary to increase the number of ports M, since QPSK becomes superior. The optimum is reached for $M = 4$, where each port excludes one of the QPSK states and receives 25% of the signals energy. However, the

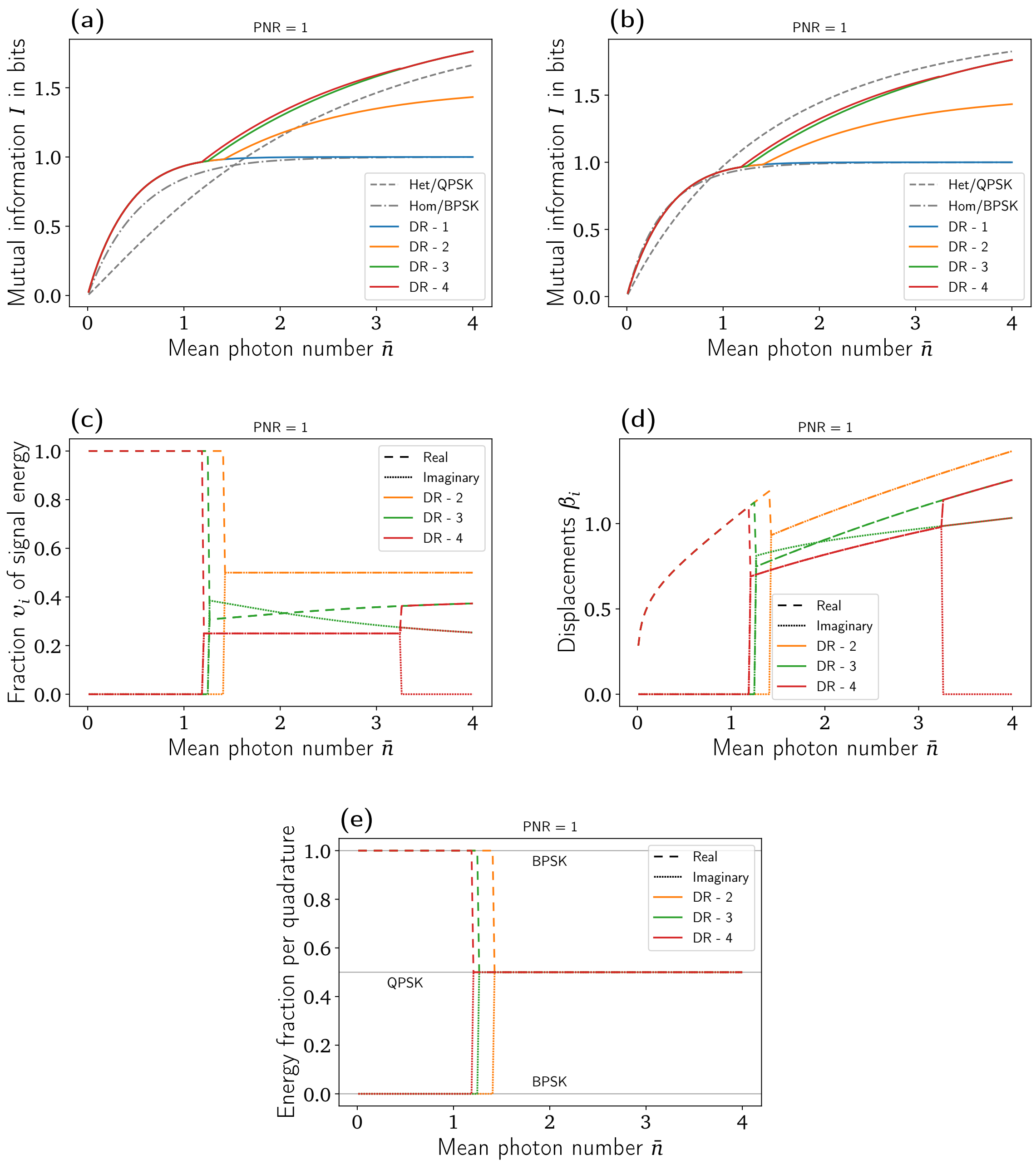


Figure 4.16: Mutual information I for the M-port detection scheme and alphabets under a **(a)** hard and **(b)** soft decision rule for detectors with PNR 1. The optimized parameters are **(c)** the energy fraction v_m per port and **(d)** the displacements β_m. The energy fraction per quadrature shown in **(e)** displays whether BPSK or QPSK should be used. The optimal parameters are the same for hard and soft decision.

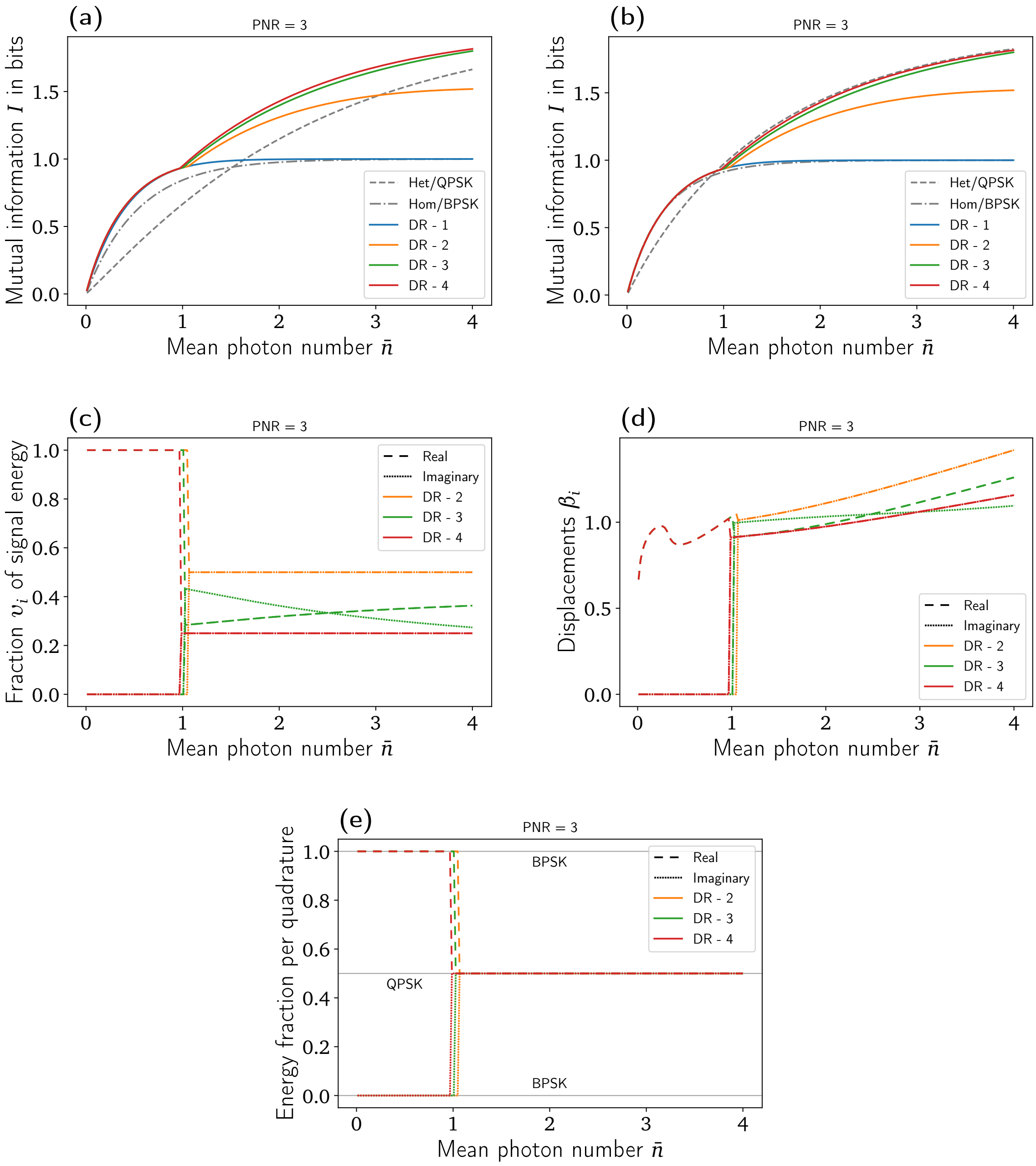


Figure 4.17: Mutual information I for the M-port detection scheme and alphabets under **(a)** hard and **(b)** soft decision rule for detectors with PNR 3. The optimized parameters are **(c)** the energy fraction v_m per port and **(d)** the displacements β_m. The energy fraction per quadrature shown in **(e)** displays whether BPSK or QPSK should be used. The optimal parameters are the same for hard and soft decision.

advantage over the 3-port scheme is minimal and vanishes with rising $\bar{n}$. In this regime it is best to direct the majority of the signals energy to ports displacing on the real axis and only a small fraction to the remaining port displacing on the imaginary axis. Interestingly, the optimal amount of energy for detecting the real and imaginary quadrature depends on the mean photon number $\bar{n}$ and is not equally distributed between the quadratures.
When soft decision is used the displacement receiver does not change its performance compared to hard decision decoding, while homodyne and double homodyne detection are both able to retrieve much more information from the raw probability distributions. As a result, double homodyne detection remains superior for larger mean photon numbers $\bar{n}$. When increasing the PNR of the photon counters, the displacement receiver approaches the performance of double homodyne detection as is shown in fig. 4.17. Here, the 4-port detection scheme remains the optimal choice also for higher values of $\bar{n}$ and its mutual information is comparable to soft decision double homodyne detection. The single shot approach to detect larger alphabets was demonstrated experimentally in several works [Izumi et al. 2012; 2013; DiMario et al. 2018; Sidhu et al. 2021].
Apart from the considered one-shot detection schemes, adaptive schemes can greatly enhance the performance of a displacement receiver for larger alphabets. Here, only a single detector is used and the displacement is adapted according to the obtained measurement results. A near-optimal sub-shot-noise receiver with a rather simple feedback strategy was proposed by Bondurant [1993] and demonstrated by Jabir et al. [2020]. More powerful feedback strategies involve Bayesian updating in real time. These Bayesian schemes were demonstrated in several experiments with either feed-forward or feedback for many different alphabet sizes and still are an active field of research [Becerra et al. 2011; Izumi et al. 2012; Li et al. 2013; Izumi et al. 2013; Becerra et al. 2013; 2015; Ferdinand et al. 2017; Izumi et al. 2020; 2021; Sidhu et al. 2021].
Another possibility of extending the alphabet size is to not only utilize different phases (phase shift keying) for encoding information, but additionally use different frequencies as was proposed by Burenkov et al. [2018] and recently demonstrated [Burenkov et al. 2020; Jabir et al. 2020]. A drawback of adaptive procedures is that they can be computationally expensive and need to be performed in real time for optimal results.

4.4.4 Imperfections

The displacement receiver relies on accurately displacing the received signal and measuring the photon number distribution afterwards. Already small errors in this system can distort the results significantly and reduce the performance. To get a realistic picture of the detector behavior it is necessary to take various imperfections into account. In the following, we present models for several different imperfections and their impact on the displacement receiver.

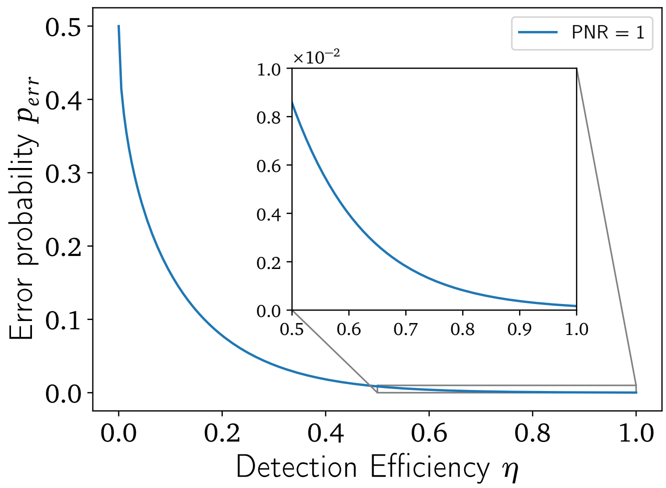


Figure 4.18: The error probability p_{err} for a displacement receiver detecting a BPSK alphabet with $\bar{n} = 2$ as a function of the detection efficiency η.

Detection efficiency

Detection efficiency η describes the percentage of photons that are present, but dot not create a detection event at the photon counter. These photons are lost and the effect can be modeled as a perfect detector which is preceded by a beam splitter with transmission η, where the second outgoing port is inaccessible [Gardiner and Zoller 2004]. Formally, the corresponding transmission η is given by the ratio of the number of generated photoelectrons to the number of incident photons. When detecting coherent states an imperfect detection efficiency is equivalent to amplitude attenuation. Therefore, the probability of measuring n photons is reduced to

$$p\left(n|\zeta,\eta\right) = e^{-\eta|\zeta|^2}\frac{\left(\eta\left|\zeta\right|^2\right)^n}{n!}. \tag{4.40}$$

The error probability for detecting a BPSK alphabet with $\bar{n} = 2$ is shown in fig. 4.18 as a function of the detection efficiency. The error probability rises constantly with decreasing detection efficiency as the communication task at hand is effectively rescaled to a lower $\bar{n}$. Since detection efficiency only rescales the amplitude of the signal, increasing PNR does not change the error probability.

Dark counts

Even in the complete absence of light, photon detectors sometimes register an event due to thermal instabilities. These additional counts are called *dark counts* and are usually characterized by a Poissonian distribution. Assuming that n_{dark} dark counts per signal are registered on average, the probability of detecting n photons for a coherent state is

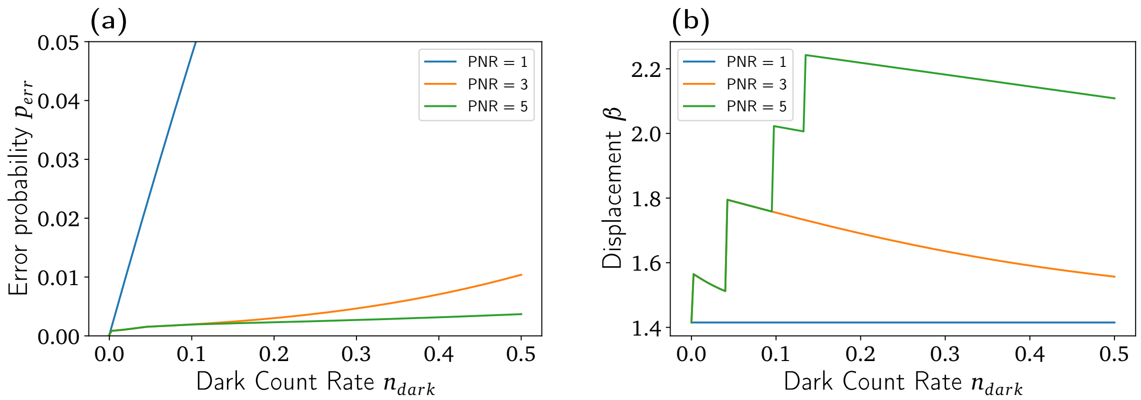


Figure 4.19: The **(a)** error probability p_{err} and **(b)** optimal displacement for a displacement receiver detecting a BPSK alphabet with $\bar{n} = 2$ as a function of the average dark count rate n_{dark}.

given by

$$p\left(n|\zeta, n_{\text{dark}}\right) = e^{-|\zeta|^2 - n_{\text{dark}}} \frac{\left(|\zeta|^2 + n_{\text{dark}}\right)^n}{n!}. \tag{4.41}$$

The error probability for detecting a BPSK alphabet accompanied by dark counts is shown in fig. 4.19 alongside the optimal displacement. A displacement receiver without PNR is very susceptible to dark counts and its error probability rises quickly to large values with the dark counts rate n_{dark}. Contrarily, the displacement receiver can tolerate much higher values of n_{dark} without loosing much performance, if a PNR detector is used. To compensate for the dark count rate, however, it is necessary to increase the displacement amplitude β. The discrete jumps in β are due to the discrete nature of the photon count statistics and correspond to a change in the decision threshold.

Mode mismatch / visibility

An ideal displacement operation relies on coherently combining the LO with the signal state at a highly transmissive beam splitter. If the signal mode does not fully match the LO mode, the modes are not mixed fully coherent and the output will contain two modes. The first of them contains a combination of the signal and the LO mode, whereas the second one only a residue LO mode. If both the LO and the signal are coherent states, the output statistics are given as [Banaszek et al. 1999]

$$p\left(n|\zeta, \Upsilon\right) = e^{-|\zeta + \Upsilon\beta|^2 - \left(1 - \Upsilon^2\right)|\beta|^2} \frac{\left[|\zeta + \Upsilon\beta|^2 + \left(1 - \Upsilon^2\right)|\beta|^2\right]^n}{n!}, \tag{4.42}$$

where the dimensionless parameter Υ describes the mode overlap. The parameter Υ is directly related to the visibility ν by

$$\Upsilon = \frac{\nu}{2 - \nu}. \tag{4.43}$$

The visibility ν itself is defined as the ratio between the average amplitude and intensity of the measured signal [Mandel and Wolf 1995]. It measures the fringe contrast and is directly related to the coherence between the two overlapping signals and therefore to the mode mismatch. In fig. 4.20 the optimized error probability p_{err} is shown alongside the optimal displacements β and amplitudes ζ. Without PNR the displacement receiver is very prone to mode mismatch and the error probability rises quickly with deteriorating mode mismatch. In fact, even for visibilities close to unity OOK quickly becomes the superior encoding. With higher PNR the displacement receiver is able to retain good performance even for lower values of ν. Apart from increasing PNR it is also possible to improve the performance by simultaneously optimizing the state amplitudes ζ_i. Below a certain value of the visibility ν, depending on the PNR and mean photon number $\bar{n}$, it becomes beneficial to use OOK instead of BPSK where no displacement or PNR is needed to distinguish the states.

Afterpulsing

The imperfections of the receiver also depend on the type of the used photon counter. In avalanche photon-diodes, specifically, an additional avalanche can be triggered when the detector comes out of its dead time. If this afterpulsing effect occurs, an additional event is created without another photon being present. Afterpulsing can also occur in sequence, since one event can cause another one. A single afterpulsing event has a probability p_{ap} to take place. Therefore, the obtained statistics without afterpulsing need to be modified and only the probability of not detecting any photons $p\,(0|\zeta)$ remains unchanged. The probability of detecting one photon $p\,(1|\zeta)$ is lowered by the factor $(1-p_{\text{ap}})$, since the fraction p_{ap} registers an additional event and needs to be added to $p\,(2|\zeta)$, where again afterpulsing can occur. By repeating the procedure iteratively the following probabilities are obtained

$$p\,(0|\zeta) \longrightarrow p\,(0|\zeta)\,, \tag{4.44}$$
$$p\,(1|\zeta) \longrightarrow p\,(1|\zeta)\,(1-p_{\text{ap}})\,, \tag{4.45}$$
$$p\,(2|\zeta) \longrightarrow [p\,(2|\zeta) + p\,(1|\zeta)\,p_{\text{ap}}]\,(1-p_{\text{ap}})\,, \tag{4.46}$$
$$p\,(3|\zeta) \longrightarrow \left[p\,(3|\zeta) + p\,(2|\zeta)\,p_{\text{ap}} + p\,(1|\zeta)\,p_{\text{ap}}^2\right](1-p_{\text{ap}}) \tag{4.47}$$
$$\vdots$$

The probabilities for $n>0$ can be written in a closed form as

$$p\,(n|\zeta,p_{\text{ap}}) = (1-p_{\text{ap}})\sum_{i=1}^{n} p\,(i|\zeta)\,p_{\text{ap}}^{n-i}. \tag{4.48}$$

Naturally, afterpulsing has an impact on the performance only if PNR >1 is needed for the task, which is the case, for instance, when dark counts occur. In fig. 4.21 the error probability for distinguishing a BPSK alphabet is shown for $n_{\text{dark}} = 0.3$ as a function of the afterpulsing probability p_{ap}. For the optimal discrimination of BPSK alphabets with

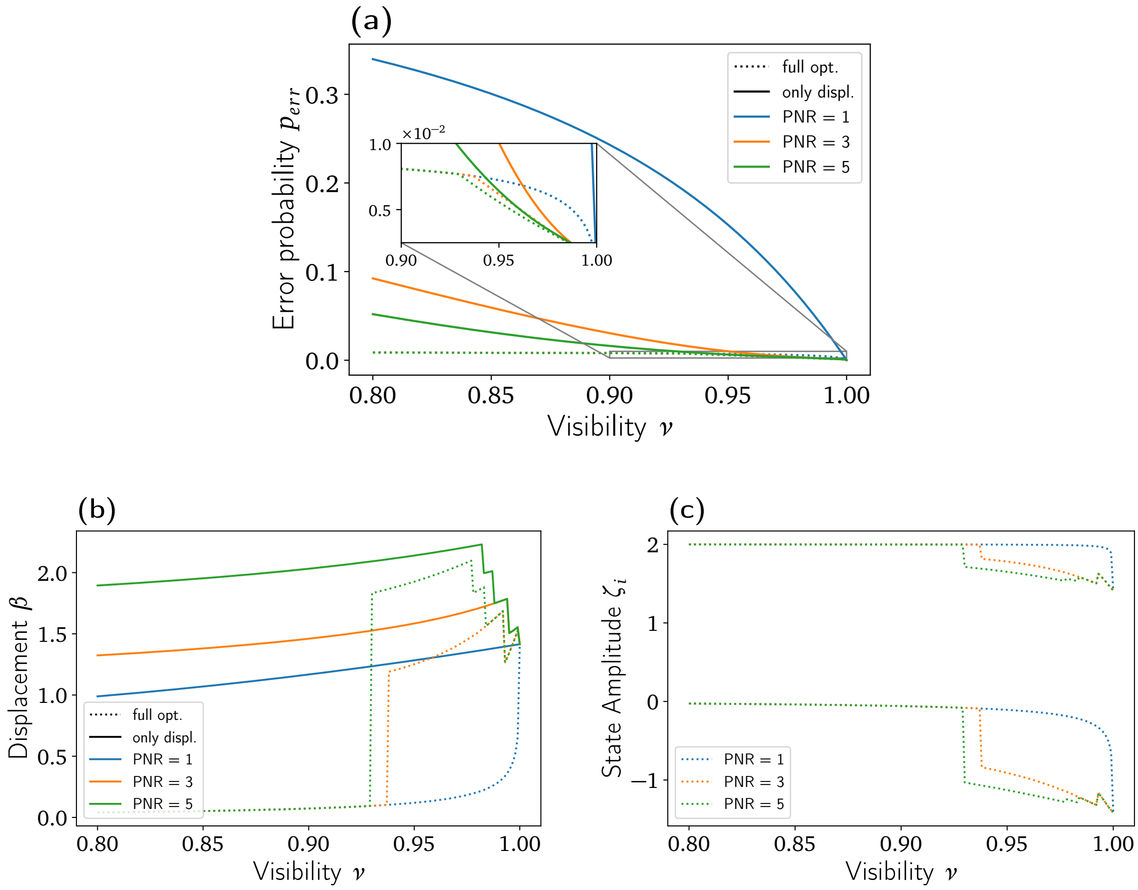


Figure 4.20: The **(a)** error probability p_{err}, **(b)** optimal displacement and **(c)** optimal alphabet for a displacement receiver detecting a BPSK alphabet with $\bar{n} = 2$ as a function of the visibility ν. The solid lines show the error probability when the alphabet is fixed to BPSK and only the displacement amplitude β is optimized, whereas for the dotted lines both the amplitudes ζ_i and the displacement β were optimized. Discrete jumps in the amplitude and the displacement are due to a change in the decision threshold.

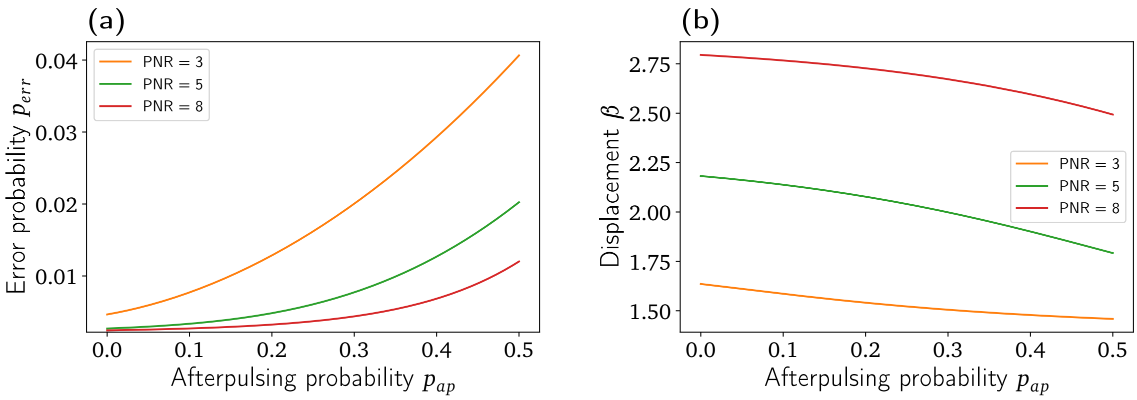


Figure 4.21: **(a)** The error probability p_{err} and **(b)** optimal displacement for an imperfect displacement receiver with dark count rate $n_{\mathrm{dark}} = 0.3$ detecting a BPSK alphabet with $\bar{n} = 2$ as a function of the afterpulsing probability p_{ap}.

dark counts, PNR is needed (cf. fig. 4.19) and afterpulsing further increases the error probability p_{err}. The deteriorating effect is more pronounced the lower the PNR of the photon detector is. With growing afterpulsing probability p_{ap} the optimal displacement amplitude β decreases as it becomes favorable to produce less detection events.

4.5 Low Energy Limit

In the regime where the available energy is very limited $\bar{n} \ll 1$, dubbed the photon starved regime, PIE becomes the favorable figure of merit to quantify communication. Due to the energy constraint the arising limiting factors are different from outside the photon starved regime and the channel capacities admit to particular limits, which were already discussed in section 3.4.3. Most notably, the PIE corresponding to the capacity C_1 with single quadrature detection, eqs. (3.82) and (3.103), approaches its maximal value of 2 nats for $\bar{n} \to 0$.
In this section we will discuss the behavior of the previously introduced detection techniques in the photon starved regime and how their maximum PIE compares to the corresponding capacity. Since it is advantageous to use only a single quadrature to encode information in the photon-starved regime as indicated by eqs. (3.103) and (3.104), we limit the discussion to just this scenario.
In order to analyze the PIE of different receivers we need to choose a particular alphabet. As we encode only in a single quadrature and operate in the limit $\bar{n} \ll 1$, a BPSK alphabet is the natural choice as it provides the optimal state separation of two coherent states with a given average energy $\bar{n}$.
According to the Helstrom limit the optimal receiver for a binary coherent alphabet is the Dolinar receiver. Therefore, the respective PIE can be determined from the minimum error probability in eq. (3.12) and the mutual information of a binary symmetric

channel in eq. (3.67). Expanding the resulting expression up to the first order in $\bar{n}$ yields

$$\mathrm{PIE}_{\mathrm{Dol}} \approx 2 - \frac{8}{3}\bar{n}, \tag{4.49}$$

where the information content is expressed in nats. Since we are only considering a binary set of pure states, the minimum error measurement given by the Dolinar receiver also achieves the optimal mutual information for non-collective measurements [Tomassoni and Paris 2008]. Consequently, the Dolinar receiver is the optimal individual measurement and defines the maximum PIE of 2 nats attainable without the use of collective measurements. This agrees with the single-quadrature Shannon limit of eq. (3.103), however, the first order of the expression for the Dolinar receiver provides a more favorable scaling with $\bar{n}$.

Interestingly, the limit of 2 nats is achieved also by various other detectors that are suboptimal outside the low energy regime. The first of these detectors is the homodyne detector. Its mutual information for soft decision decoding is given by eq. (4.18). As before for the Dolinar receiver, the corresponding PIE admits to a power series expansion in $\bar{n}$ which up to the first order reads

$$\mathrm{PIE}_{\mathrm{Hom}} \approx 2 - 4\bar{n} \tag{4.50}$$

in nats. Note that this expression is exactly equal to the Shannon-Hartley limit of eq. (3.103). Since the Shannon-Hartley limit is achieved by a Gaussian distribution of coherent states, the equality illustrates that the Gaussian distribution is well approximated by just two such states if very little energy is at hand.

The second receiver in our discussion that reaches the PIE value of 2 nats is the displacement receiver. Its PIE can be calculated based on the general expression for mutual information in eq. (3.46) together with the probabilities of eq. (4.1). However, to obtain the optimal mutual information the displacement has to be optimized such that the PIE for the displacement receiver can be calculated only numerically.

In fig. 4.22 all three PIEs discussed above are shown as a function of the mean photon number $\bar{n}$. Both, the homodyne detector and displacement receiver, approach the optimal PIE of 2 nats in the limit $\bar{n} \to 0$. Consequently, both can be seen as an optimal detection strategy in the photon starved regime for coherent alphabets encoded in a single-quadrature.

Ultimately, however, PIE of a binary coherent alphabet is not limited to 2 nats. The Holevo limit, as shown in fig. 3.8 allows for much higher PIEs and the discussed detection strategy remain suboptimal from the perspective of quantum mechanics. A known solution that gets us closer to the Holevo limit is given by PPM as is shown in fig. 3.14. While direct detection is sufficient as a detection method, PPM requires large peak powers to generate the necessary states (cf. section 3.5.4). Furthermore, it does not operate with a BPSK alphabet such that the exact nature of its connection to the limit of 2 nats remains unclear. As we will see in the following both these problems can be addressed and PPM can be seen as a strategy based on states from a BPSK alphabet and collective measurements. Note that PPM is not the optimal strategy as it can be regarded as a

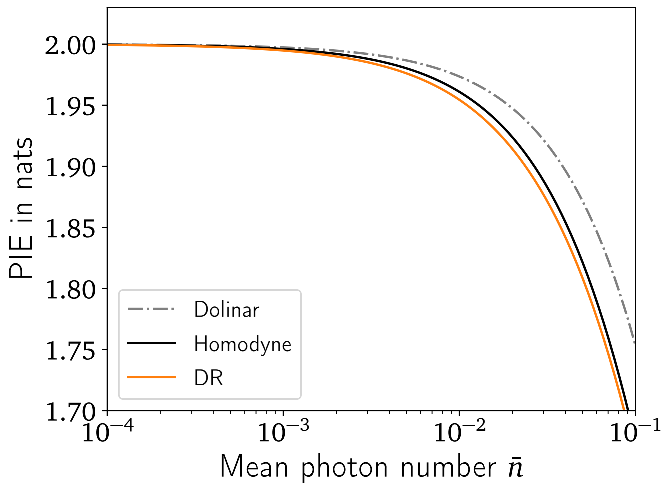


Figure 4.22: PIE of a homodyne, Dolinar and displacement receivers for detecting a BPSK alphabet. All detection schemes approach 2 nats, the limit for separable measurements.

protocol based on a restricted subclass of generalized OOK alphabets. The latter can provide higher PIE and the optimal alphabet can be found [Jarzyna et al. 2015; Chung et al. 2016; Jarzyna and Banaszek 2017].

4.6 Structured Receiver

First, we address the problem of large peak powers necessary for PPM. The solution requires only a uniform energy distribution in time at the cost of a more intricate detection setup. Essentially, all time bins of a PPM frame are filled with weak coherent states from a BPSK alphabet which at the receiving end are gathered into a particular time bin by a so-called structured receiver [Guha 2011]. The position of a particular time bin that contains all the energy depends on the phases of the input states such that effectively, PPM symbols can be translated into the phases of BPSK symbols.
The operational principle of a structured receiver is based on Hadamard matrices H_m, which are symmetric square matrices containing only entries with values ± 1. They are constructed in such a way that all columns are orthogonal to each other. This property holds for rows as well [Moon 2005]. This construction implies the relation

$$H_m H_m^T = m\mathbb{1}_m, \tag{4.51}$$

which holds for all Hadamard matrices, where $\mathbb{1}_m$ denotes the $m \times m$ identity matrix. Hadamard matrices are known to exist for $m = 1, 2$ or multiples of 4 and selected other integers.
A consequence of eq. (4.51) is the following: if a Hadamard matrix acts on a vector given by any of its columns, the result contains only a single non-zero entry at the position

corresponding to the column. For instance, consider the Hadamard matrix for $m = 4$ given by

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \tag{4.52}$$

Acting on its first or second column yields

$$H_4 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad , \quad H_4 \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 4 \\ 0 \end{pmatrix}, \tag{4.53}$$

whereas acting on the third or fourth column yields

$$H_4 \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 4 \\ 0 \end{pmatrix} \quad , \quad H_4 \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4 \end{pmatrix}. \tag{4.54}$$

By constructing a unitary operation from a Hadamard matrix this feature can be utilized to build a structured receiver.
Consider a series of m coherent states $\left| \Psi_{\vec{j}} \right\rangle = |\pm\zeta\rangle \otimes \cdots \otimes |\pm\zeta\rangle$ where each mode corresponds to a particular time bin. The phases ± 1 are chosen such that the vector $\vec{j} = (\pm 1, \ldots, \pm 1)^T$ containing all phases corresponds to a column of the corresponding Hadamard matrix H_m. A state $\left| \Psi_{\vec{j}} \right\rangle$ constructed in this way is called a Hadamard word. Now we define the unitary $\hat{H}_m$ by

$$\hat{H}_m = \frac{1}{\sqrt{m}} H_m. \tag{4.55}$$

If such operation is applied to a Hadamard word $\left| \Psi_{\vec{j}} \right\rangle$, all the energy is collected into a single time bin due to eq. (4.51). In this fashion, PPM words are obtained by applying the unitary of eq. (4.55) even though only Hadamard words have been transmitted. As Hadamard words consist of states drawn from a BPSK alphabet, their energy distribution in time is entirely flat which eases the requirements for their preparation.
Physically, the necessary unitary $\hat{H}_m$ can be implemented by a series of beamsplitters with either passive or active control [Guha 2011; Banaszek and Jachura 2017]. In the simplest scenario with $m = 2$, the structured receiver consists of a single 50:50 beamsplitter, where one input mode experiences an additional phase shift of π compared to the other input mode. The output modes of the beamsplitter are given by

$$\hat{c} = \frac{1}{\sqrt{2}} \left(\hat{a} + \hat{b} \right), \tag{4.56}$$

$$\hat{d} = \frac{1}{\sqrt{2}} \left(\hat{a} - \hat{b} \right) \tag{4.57}$$

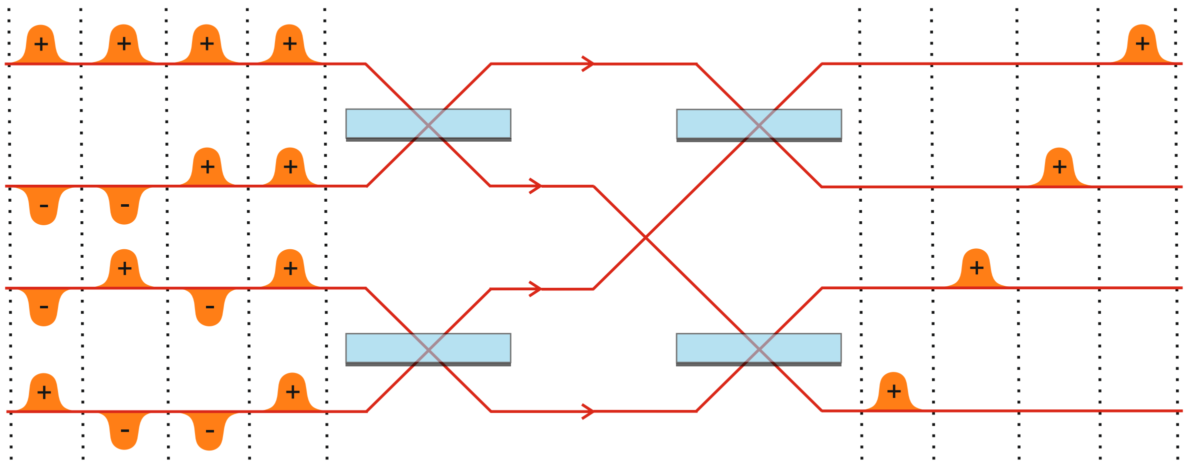


Figure 4.23: Example of a structured receiver for $m = 4$. The incoming states are interfered in a series of 50:50 beamsplitters defined by eqs. (4.56) and (4.57). The incoming words are transformed such that only one output port contains all energy.

and the expectation values for the mean photon number in each respective mode are given by

$$\left\langle \zeta, \zeta \,\middle|\, \hat{c}^\dagger \hat{c} \,\middle|\, \zeta, \zeta \right\rangle = 2\zeta^2 \quad , \quad \left\langle \zeta, \zeta \,\middle|\, \hat{d}^\dagger \hat{d} \,\middle|\, \zeta, \zeta \right\rangle = 0, \tag{4.58}$$

$$\left\langle \zeta, -\zeta \,\middle|\, \hat{c}^\dagger \hat{c} \,\middle|\, \zeta, -\zeta \right\rangle = 0 \quad , \quad \left\langle \zeta, -\zeta \,\middle|\, \hat{d}^\dagger \hat{d} \,\middle|\, \zeta, -\zeta \right\rangle = 2\zeta^2. \tag{4.59}$$

Evidently, each of the output modes contains all energy of a particular Hadamard word. While for $m = 2$ a single beamsplitter suffices, for $m > 2$ a series of beamsplitters is necessary to achieve the required transformation. In fig. 4.23 a structured receiver for $m = 4$ is depicted. In this scenario four beamsplitters of the type defined by eqs. (4.56) and (4.57) are needed. Again a single output mode contains all the energy for the respective Hadamard word. Different words are distinguished by measuring which output mode contains energy which can be done with SPDs. Since we interfere all BPSK states on a series of beamsplitters, the resulting measurement is an example of a collective measurement.

4.7 Measurements in the Low Energy Limit

As we have seen in the previous section, a collective measurement on BPSK symbols can be achieved relatively simply by a series of beamsplitters. Finding collective measurements is of great interest as they can provide a larger amount of mutual information than measurements based on the detection of individual symbols. They can achieve superadditivity of mutual information as defined in eq. (3.87). The main problem with attaining superadditivity is to find corresponding collective measurements and their physical

realizations. In the following, we will present a systematic way to analyze mutual information for coherent states and measurements including their physical realizations that achieve superadditivity in the low energy limit. Note that, while the expressions below are derived with coherent states in mind, the formalism can be applied to any type of quantum state that can be expanded in a non-negative cost parameter ζ in the limit $\zeta \to 0$. Furthermore all expressions are presented in nats.
Let $\hat{\varrho}_x(\zeta)$ be a coherent state appearing with a-priori probability p_x. The cost of a coherent state is given by the squared absolute value of its complex amplitude $|\zeta|^2$. To retrieve information from the state a POVM $\{\hat{\pi}_y\}$ is performed, where y denotes the outcome of the measurement. According to Born's rule of eq. (2.6) the POVM defines the conditional probabilities

$$p_\zeta(y|x) = \mathrm{Tr}\left[\hat{\pi}_y \hat{\varrho}_x\right]. \tag{4.60}$$

To subsequently analyze mutual information it will be very useful to decompose the full expression into a sum of contributions from different measurement results y, i.e.

$$I(X;Y) = \sum_y I_y, \tag{4.61}$$

where the individual contributions are given by

$$I_y = \sum_x p_x p_\zeta(y|x) \ln\left[\frac{p_\zeta(y|x)}{p_\zeta(y)}\right] \tag{4.62}$$

due to eq. (3.46). To obtain more insight into the low energy regime we will now expand the state $\hat{\varrho}_x$ around $\zeta = 0$ up to the second order such that we arrive at the power series

$$\hat{\varrho}_x \approx \hat{\varrho}^{(0)} + \zeta\hat{\varrho}_x^{(1)} + \zeta^2\hat{\varrho}_x^{(2)}, \tag{4.63}$$

where $\hat{\varrho}^{(0)} = |\,0\,\rangle\langle\,0\,|$ denotes the vacuum state and

$$\hat{\varrho}_x^{(1)} = \left.\frac{\partial\hat{\varrho}_x}{\partial\zeta}\right|_{\zeta=0}, \qquad \hat{\varrho}_x^{(2)} = \left.\frac{1}{2}\frac{\partial^2\hat{\varrho}_x}{\partial\zeta^2}\right|_{\zeta=0} \tag{4.64}$$

the first and second order contributions respectively. This power series for the state implies a corresponding expansion for the probabilities. They admit to a power series given by

$$p_\zeta(y|x) \approx p^{(0)}(y) + \zeta p^{(1)}(y|x) + \zeta^2 p^{(2)}(y|x), \tag{4.65}$$

where a particular contribution is defined by $p^{(k)}(y|x) = \mathrm{Tr}\left[\hat{\pi}_y\hat{\varrho}_x^{(k)}\right]$. Similarly, expanding the marginal probabilities yields

$$p_\zeta(y) \approx p^{(0)}(y) + \zeta p^{(1)}(y) + \zeta^2 p^{(2)}(y) \tag{4.66}$$

with $p^{(k)}(y) = \sum_x p^{(k)}(y|x)$. Note that the marginal probabilities can be defined equivalently by $p^{(k)}(y) = \mathrm{Tr}\left[\hat{\pi}_y \hat{\varrho}^{(k)}_{\mathrm{ens}}\right]$, where

$$\hat{\varrho}^{(k)}_{\mathrm{ens}} = \sum_x p_x \hat{\varrho}^{(k)}_x. \tag{4.67}$$

To further analyze mutual information we will divide the POVM elements $\hat{\pi}_y$ into two groups depending on whether an outcome y can be generated by the vacuum state. This distinction ties naturally to different types of measurement devices.
The first group, denoted by $\mathcal{M}_\perp$, contains measurements orthogonal to the vacuum state, i.e. measurements that satisfy $\mathrm{Tr}\left[\hat{\pi}_y \hat{\varrho}^{(0)}\right] = 0$. The second group, denoted by $\mathcal{M}$, consists of all remaining measurements which satisfy $\mathrm{Tr}\left[\hat{\pi}_y \hat{\varrho}^{(0)}\right] > 0$. In the optical scenario with $\hat{\varrho}^{(0)} = |\,0\,\rangle\langle\,0\,|$, for instance, a photon counting detector is a realization of an orthogonal measurement, since its corresponding POVM is given by $\{|\,n\,\rangle\langle\,n\,|\}_n$ and all elements with $k > 0$ naturally belong to $\mathcal{M}_\perp$. Physical realization of a type-$\mathcal{M}$ measurement are the displacement receiver or homodyne detector. The classification of a homodyne detector, however, is problematic in a strict sense. Its output quadrature takes a continuous form and with increasing value of the quadrature the probability $\mathrm{Tr}\left[\hat{\pi}_y \hat{\varrho}^{(0)}\right]$ becomes arbitrary small.
The overall mutual information can now be written as

$$I = \sum_{y\in\mathcal{M}_\perp} I_y + \sum_{y\in\mathcal{M}} I_y \tag{4.68}$$

and the contribution from each measurement type will be analyzed separately.
First, we will analyze the contribution of type-$\mathcal{M}$ measurements. By expanding the logarithm appearing in eq. (4.62) into a power series, one obtains

$$I_y \approx \frac{\zeta^2}{2p^{(0)}(y)} \sum_x p_x \left[p^{(1)}(y|x) - p^{(1)}(y)\right]^2, \tag{4.69}$$

where we can also write

$$p^{(1)}(y|x) - p^{(1)}(y) = \mathrm{Tr}\left[\hat{\pi}_y \left(\hat{\varrho}^{(1)}_x - \hat{\varrho}^{(1)}_{\mathrm{ens}}\right)\right]. \tag{4.70}$$

With the help of techniques commonly used in quantum estimation theory [Helstrom 1976; Holevo 1982; Braunstein and Caves 1994], we can transform this expression into a more useful form. The *symmetric logarithmic derivative (SLD)* at $\zeta = 0$ of the density operators of the individual as well as the ensemble states are implicitly defined by the respective equations

$$\hat{\varrho}^{(1)}_x = \left.\frac{\partial \hat{\varrho}_x}{\partial \zeta}\right|_{\zeta=0} = \frac{1}{2}\left(\hat{L}_x \hat{\varrho}^{(0)} + \hat{\varrho}^{(0)} \hat{L}_x\right), \tag{4.71}$$

$$\hat{\varrho}^{(1)}_{\mathrm{ens}} = \left.\frac{\partial \hat{\varrho}_{\mathrm{ens}}}{\partial \zeta}\right|_{\zeta=0} = \frac{1}{2}\left(\hat{L}_{\mathrm{ens}} \hat{\varrho}^{(0)} + \hat{\varrho}^{(0)} \hat{L}_{\mathrm{ens}}\right). \tag{4.72}$$

We assume that the SLDs are well defined for all considered physical scenarios [Liu et al. 2016; Šafránek 2017; Czajkowski et al. 2017]. Inserting both eqs. (4.71) and (4.72) into eq. (4.70) yields

$$p^{(1)}(y|x) - p^{(1)}(y) = \frac{1}{2}\,\mathrm{Tr}\left[\hat{\pi}_y\left(\hat{D}_x\hat{\varrho}^{(0)} - \hat{\varrho}^{(0)}\hat{D}_x\right)\right], \tag{4.73}$$

where

$$\hat{D}_x = \hat{L}_x - \hat{L}_{\mathrm{ens}}. \tag{4.74}$$

Applying the Schwartz inequality

$$\left|\mathrm{Tr}\left[\hat{A}\hat{B}\right]\right|^2 \leq \mathrm{Tr}\left[\hat{A}^\dagger\hat{A}\right]\mathrm{Tr}\left[\hat{B}^\dagger\hat{B}\right] \tag{4.75}$$

to the operators

$$\hat{A} = \frac{1}{\sqrt{p^{(0)}(y)}}\sqrt{\hat{\varrho}^{(0)}}\sqrt{\hat{\pi}_y} \quad , \quad \hat{B} = \sqrt{\hat{\pi}_y}\hat{D}_x\sqrt{\hat{\varrho}^{(0)}} \tag{4.76}$$

and their Hermitian conjugates $\hat{A}^\dagger$ and $\hat{B}^\dagger$ and inserting the result into eq. (4.69) yields the upper bound

$$I_y \leq \frac{\zeta^2}{2}\,\mathrm{Tr}\left[\hat{\pi}_y\left(\sum_x p_x\hat{D}_x\hat{\varrho}^{(0)}\hat{D}_x\right)\right] \tag{4.77}$$

for type-$\mathcal{M}$ measurements.
Next, we turn to the contribution of type-$\mathcal{M}_\perp$ measurements. This time, however, another approach is needed. This is because for the vacuum contributions $\hat{\pi}_y\hat{\varrho}^{(0)} = \hat{\varrho}^{(0)}\hat{\pi}_y = 0$ holds due to $\mathrm{Tr}\left[\hat{\pi}_y\hat{\varrho}^{(0)}\right] = 0$ and the semidefiniteness of both operators. Consequently, $\mathrm{Tr}\left[\hat{\pi}_y\hat{\varrho}^{(1)}\right] = 0$ and the first order in the expansion of $p^{(1)}(y|x)$ vanishes. Therefore, we need to consider the second order

$$p^{(2)}(y|x) \approx \zeta^2\,\mathrm{Tr}\left[\hat{\pi}_y\tilde{\hat{\varrho}}_x^{(2)}\right]. \tag{4.78}$$

The operator $\tilde{\hat{\varrho}}_x^{(2)}$ denotes the state after projecting onto the kernel of $\hat{\varrho}^{(0)}$, i.e.

$$\tilde{\hat{\varrho}}_x^{(2)} = \hat{K}\hat{\varrho}_x^{(2)}\hat{K}, \tag{4.79}$$

where $\hat{K}$ denotes the projection onto said kernel. The projection is useful, since for type-$\mathcal{M}_\perp$ measurements $\hat{K}\hat{\pi}_y\hat{K} = \hat{\pi}_y$ holds. Therefore, mutual information generated by an outcome $y \in \mathcal{M}_\perp$ is approximately given by

$$I_y \approx \zeta^2\sum_x p_x\,\mathrm{Tr}\left[\hat{\pi}_y\tilde{\hat{\varrho}}_x^{(2)}\right]\ln\left[\frac{\mathrm{Tr}\left[\hat{\pi}_y\tilde{\hat{\varrho}}_x^{(2)}\right]}{\sum_l p_l\,\mathrm{Tr}\left[\hat{\pi}_y\tilde{\hat{\varrho}}_l^{(2)}\right]}\right]. \tag{4.80}$$

based on eq. (4.78) and eq. (4.62).
In total, combining the contributions of both types of measurements given by eqs. (4.77) and (4.80) in eq. (4.68) yields the upper bound

$$\begin{aligned} I \leq \zeta^2 \sum_{y \in \mathcal{M}_\perp} \sum_x p_x \operatorname{Tr}\left[\hat{\pi}_y \tilde{\hat{\varrho}}_x^{(2)}\right] \ln \left[\frac{\operatorname{Tr}\left[\hat{\pi}_y \tilde{\hat{\varrho}}_x^{(2)}\right]}{\sum_l p_l \operatorname{Tr}\left[\hat{\pi}_y \tilde{\hat{\varrho}}_l^{(2)}\right]} \right] \\ + \frac{\zeta^2}{2} \operatorname{Tr}\left[\left(\hat{\mathbb{1}} - \sum_{y \in \mathcal{M}_\perp} \hat{\pi}_y \right) \left(\sum_x p_x \hat{D}_x \hat{\varrho}^{(0)} \hat{D}_x \right) \right] \end{aligned} \tag{4.81}$$

on mutual information in the leading order, where for the second term the identity $\sum_{y \in \mathcal{M}} \hat{\pi}_y = \hat{\mathbb{1}} - \sum_{y \in \mathcal{M}_\perp} \hat{\pi}_y$ was used. Although eq. (4.81) certainly seems overly complicated, it will prove very useful for analyzing mutual information in the low energy regime. In the following, we will first consider mutual information of individual measurements for several types of coherent states and afterwards we will turn to collective measurements.

4.7.1 Individual Measurements

First, let us analyze mutual information for the scenario where only individual measurements are allowed. While we can not expect to find any superadditive effects, it will nevertheless illustrate the usefulness of the approach and offer some insights. For this purpose we will consider different types of coherent input states and discuss the contribution of both types of measurements, namely $\mathcal{M}$ and $\mathcal{M}_\perp$, separately.

Coherent states

Let us denote the complex amplitude of a coherent state as $\zeta\gamma_x$ where γ_x is a complex coefficient denoting the phase of the state and ζ the cost. Furthermore, the a-priori probability to send state $\hat{\varrho}_x$ is denoted by p_x. The power series of eq. (4.63) for a coherent state is given by

$$\hat{\varrho}^{(0)} = |\,0\,\rangle\langle\,0\,|\,, \tag{4.82}$$

$$\hat{\varrho}_x^{(1)} = [\gamma_x\,|\,1\,\rangle\langle\,0\,| + \gamma_x^*\,|\,0\,\rangle\langle\,1\,|]\,, \tag{4.83}$$

$$\tilde{\hat{\varrho}}_x^{(2)} = |\gamma_x|^2\,|\,1\,\rangle\langle\,1\,|\,. \tag{4.84}$$

Then, the SLD defined in eq. (4.74) can be determined and reads

$$\hat{D}_x = 2\left[(\gamma_x - \gamma_{\text{ens}})\,|\,0\,\rangle\langle\,1\,| + (\gamma_x - \gamma_{\text{ens}})^*\,|\,1\,\rangle\langle\,0\,|\right], \tag{4.85}$$

where $\gamma_{\text{ens}} = \sum_x p_x \gamma_x$. Since both, the SLD and the density matrix, depend only on the 0th and 1st photon number state, the analysis of mutual information can be carried out in the zero- and one-photon subspace only, which greatly simplifies subsequent expressions.

First, consider only type-$\mathcal{M}$ measurements. Inserting the SLD of eq. (4.85) into the appropriate bound in eq. (4.77) yields

$$I_y \leq 2\zeta^2 \langle 1 \,|\, \hat{\pi}_y \,|\, 1 \rangle \sum_x p_x \left|\gamma_x - \gamma_{\mathrm{ens}}\right|^2 \leq 2\zeta^2 \langle 1 \,|\, \hat{\pi}_y \,|\, 1 \rangle \sum_x p_x \left|\gamma_x\right|^2 . \tag{4.86}$$

Since all involved POVM elements $\hat{\pi}_y$ are of type-$\mathcal{M}$ due to our assumption, the identity $\sum_y \langle 1 \,|\, \hat{\pi}_y \,|\, 1 \rangle = 1$ holds. This fact together with the definition of the mean photon number $\sum_x p_x \zeta^2 \left|\gamma_x\right|^2 = \bar{n}$ enables us to further simplify the upper bound on mutual information of eq. (4.86) such that after summation over all outcomes y it reads

$$I \leq 2\bar{n} \tag{4.87}$$

in nats. This result agrees with the single-quadrature Shannon limit derived in eq. (3.103) and exemplifies the connection between type-$\mathcal{M}$ measurements and homodyne detection. Next, let us consider only type-$\mathcal{M}_\perp$ measurements. By inserting the second order contribution of eq. (4.84) into the bound on mutual information in eq. (4.80) we arrive at

$$I_y \approx \zeta^2 \langle 1 \,|\, \hat{\pi}_y \,|\, 1 \rangle \sum_x p_x \left|\gamma_x\right|^2 \ln \left[\frac{\left|\gamma_x\right|^2}{\sum_l p_l \left|\gamma_l\right|^2} \right] . \tag{4.88}$$

The fraction $\frac{|\gamma_x|^2}{\sum_l p_l |\gamma_l|^2}$ appearing in the logarithm corresponds to the peak-to-average power ratio. Assuming that the ratio is limited to a maximum value of

$$\mathcal{P} = \frac{\max_x \left|\gamma_x\right|^2}{\sum_l p_l \left|\gamma_l\right|^2}, \tag{4.89}$$

the upper bound on mutual information for type-$\mathcal{M}_\perp$ measurements after summation over all outcomes y reads

$$I \leq \zeta^2 \sum_x p_x \left|\gamma_x\right|^2 \ln \left[\mathcal{P}\right] = \bar{n} \ln \left[\mathcal{P}\right] . \tag{4.90}$$

The right-hand side implies that mutual information is upper bounded by the peak-to-average power ratio, if only photon counting detectors are used. It is straightforward to verify that the upper bound is achieved by a PPM encoding of order $M = \mathcal{P}$ which is consistent in the first order with the PIE derived in eq. (3.136). We see again that high peak powers are crucial for efficient communication via PPM.

Thermalized coherent states

Apart from pure coherent states also other types of states can be analyzed with the formalism presented above. Next we will turn our attention to individual measurements on thermalized coherent states (cf. section 2.8). To this end, the density matrix of eq. (2.91) needs to be expanded into a power series from which one obtains

$$\hat{\varrho}^{(0)} = \frac{1}{1+n_b} \sum_n \left(\frac{n_b}{1+n_b} \right)^n |\, n \,\rangle \langle\, n \,| \,, \tag{4.91}$$

$$\hat{\varrho}_x^{(1)} = \frac{1}{(1+n_b)^2} \sum_n \left(\frac{n_b}{1+n_b} \right)^n \sqrt{n+1} \left[\gamma_x \,|\, n+1 \,\rangle \langle\, n \,| + \gamma_x^* \,|\, n \,\rangle \langle\, n+1 \,| \right] \tag{4.92}$$

up to the first order in ζ. It is important to note that for thermalized coherent states, the zeroth order state is given by a thermal state rather than the vacuum state. This has rather drastic implications for the remaining analysis. Since a thermal state has a non-zero support over the whole Hilbert space, type-$\mathcal{M}_\perp$ do not exist in a strict sense. Consequently, we can restrict our analysis to measurements of type-$\mathcal{M}$ only.
Next, we need to determine the SLD of a thermalized coherent state. It is given by

$$\hat{D}_x = \frac{2}{2-n_b} \sum_n \sqrt{n+1} \left[(\gamma_x - \gamma_{\text{ens}}) \,|\, n+1 \,\rangle \langle\, n \,| + (\gamma_x - \gamma_{\text{ens}})^* \,|\, n \,\rangle \langle\, n+1 \,| \right]. \tag{4.93}$$

An upper bound on mutual information is obtained by inserting the above SLD into eq. (4.77). Since all measurements are of type-$\mathcal{M}$, the identity $\sum_y \hat{\pi}_y = \hat{\mathbb{1}}$ can be used and the upper bound reads

$$I = \sum_y I_y \leq \frac{2}{(1+n_b)^2 (1+2n_b)} \sum_n \left(\frac{n_b}{1+n_b} \right)^n (n+1) \sum_x \zeta^2 p_x \,|\gamma_x|^2 \,. \tag{4.94}$$

With the help of $\sum_x \zeta^2 p_x |\gamma_x|^2 = \bar{n}$ and $\sum_n \left(\frac{n_b}{1+n_b} \right)^n (n+1) = (1+n_b)^2$ this expression can be simplified further to

$$I \leq \frac{2\bar{n}}{1+2n_b} \tag{4.95}$$

in nats. A thermal background fundamentally limits the amount of information that can be transmitted via coherent states. Note that the above bound is consistent with the single-quadrature Shannon limit derived in eq. (3.106).

Phase diffused coherent states

For phase diffused coherent states, as defined in eq. (2.80), the expansion of the density matrix is given by

$$\hat{\varrho}^{(0)} = |\, 0 \,\rangle \langle\, 0 \,| \,, \tag{4.96}$$

$$\hat{\varrho}_x^{(1)} = e^{-\frac{\sigma^2}{2}} \left[\gamma_x \,|\, 1 \,\rangle \langle\, 0 \,| + \gamma_x^* \,|\, 0 \,\rangle \langle\, 1 \,| \right], \tag{4.97}$$

$$\tilde{\hat{\varrho}}_x^{(2)} = |\gamma_x|^2 \,|\, 1 \,\rangle \langle\, 1 \,| \tag{4.98}$$

and the corresponding SLD is given by

$$\hat{D}_x = 2e^{-\frac{\sigma^2}{2}} \left[(\gamma_x - \gamma_{\text{ens}}) \,|\, 0 \,\rangle \langle\, 1 \,| + (\gamma_x - \gamma_{\text{ens}})^* \,|\, 1 \,\rangle \langle\, 0 \,| \right]. \tag{4.99}$$

Analogously to coherent states, the contribution I_y to mutual information is obtained by inserting the SLD into eq. (4.77). By assuming that all measurement are type-$\mathcal{M}$ an upper bound on mutual information can be obtained which reads

$$I \leq 2\bar{n}e^{-\sigma^2} \tag{4.100}$$

in nats. Therefore, phase diffusion causes an exponential decline of mutual information. The bound can be attained with phase sensitive measurements such as homodyne detection.

Contrary, phase insensitive measurements described by type-$\mathcal{M}_\perp$ measurements, such as photon counting, are not affected at all. Their contribution is defined through eq. (4.80) and is only affected by the second order term written in eq. (4.98). Since this term is independent of the phase noise strength σ and unchanged compared to pure coherent states in eq. (4.84), the upper bound on mutual information for type-$\mathcal{M}_\perp$ is directly given by eq. (4.90).

4.7.2 Collective Measurements

Now let us analyze collective measurements, where we will restrict ourselves to consider only coherent states. Let us assume a sequence of M coherent states each drawn from a BPSK alphabet. A particular coherent state is called a *symbol* and a complete sequence of M symbols is called a *word* and denoted as

$$|\, \Psi_{\boldsymbol{x}} \,\rangle = |\, (-1)^{x_1} \zeta \,\rangle \,|\, (-1)^{x_2} \zeta \,\rangle \ldots |\, (-1)^{x_M} \zeta \,\rangle\,, \tag{4.101}$$

where the bit-string $\boldsymbol{x} = x_1 x_2 \ldots x_M$ is used to uniquely identify a particular word somewhat similar to the notation used in section 4.6. Note that the factors $(-1)^{x_i}$ take the role of the phase factors previously denoted as γ_x.

Based on the upper bound on mutual information in eq. (4.81) we want to find a probability distribution $p_{\boldsymbol{x}}$ and a measurement scheme that exceeds the mutual information compared to individual measurements. In order to make a fair comparison we will consider mutual information per symbol in the end, i.e. the fraction I/M.

For the following discussion it will be convenient to introduce a notation for normalized one-photon superposition states as

$$\begin{aligned} |\, 1_{\boldsymbol{x}} \,\rangle = \frac{1}{\sqrt{M}} \big[& (-1)^{x_1} \,|\, 1 \,\rangle \,|\, 0 \,\rangle \ldots |\, 0 \,\rangle + (-1)^{x_2} \,|\, 0 \,\rangle \,|\, 1 \,\rangle \ldots |\, 0 \,\rangle \\ & + \cdots + (-1)^{x_M} \,|\, 0 \,\rangle \,|\, 0 \,\rangle \ldots |\, 1 \,\rangle \big]. \end{aligned} \tag{4.102}$$

The M-mode vacuum state will be denoted accordingly as $|\, \boldsymbol{v} \,\rangle = |\, 0 \,\rangle \,|\, 0 \,\rangle \ldots |\, 0 \,\rangle$. We will often abbreviate the notation of several modes such that all modes are contained in

a single ket, i.e. $|\,0\,\rangle\,|\,0\,\rangle \ldots |\,0\,\rangle = |\,00\ldots 0\,\rangle$.
The M-mode input state $\hat{\varrho}_{\boldsymbol{x}} = |\,\Psi_{\boldsymbol{x}}\,\rangle\,\langle\,\Psi_{\boldsymbol{x}}\,|$ gives rise to the corresponding SLD

$$\hat{D}_{\boldsymbol{x}} = 2\sqrt{M}\,[|\,1_{\boldsymbol{x}}\,\rangle\,\langle\,\boldsymbol{v}\,| + |\,\boldsymbol{v}\,\rangle\,\langle\,1_{\boldsymbol{x}}\,|] \tag{4.103}$$

and the second-order term

$$\tilde{\hat{\varrho}}^{(2)}_{\boldsymbol{x}} = M\,|\,1_{\boldsymbol{x}}\,\rangle\,\langle\,1_{\boldsymbol{x}}\,|\,. \tag{4.104}$$

The state corresponding to the string $\boldsymbol{x}$ and its bitwise negation $\neg\boldsymbol{x}$ differ only by a global minus sign which is irrelevant within our analysis as can be easily seen in eq. (4.81). Therefore, we will choose either of them as a representative and index them with the help of the equivalence class $[\boldsymbol{x}]$ such that after inserting eqs. (4.103) and (4.104) into eq. (4.81) a two-fold reduction of the sum is achieved. The resulting upper bound reads

$$I \leq 2M\zeta^2 + M\zeta^2 \sum_{y\in\mathcal{M}_\perp}\sum_{[\boldsymbol{x}]} p_{[\boldsymbol{x}]}\,\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle \left(\ln\left[\frac{\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle}{\sum_{[\boldsymbol{l}]} p_{[\boldsymbol{l}]}\,\langle\,1_{[\boldsymbol{l}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{l}]}\,\rangle}\right] - 2\right) \tag{4.105}$$

with the overall probability $p_{[\boldsymbol{x}]} = p_{\boldsymbol{x}} + p_{\neg\boldsymbol{x}}$. The above expression can be transformed to

$$\begin{aligned} I \leq &2M\zeta^2 + M\zeta^2 \sum_{y\in\mathcal{M}_\perp}\sum_{[\boldsymbol{x}]} h\left(p_{[\boldsymbol{x}]}\right)\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle \\ &+ M\zeta^2 \sum_{y\in\mathcal{M}_\perp}\sum_{[\boldsymbol{x}]} p_{[\boldsymbol{x}]}\,\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle \ln\left[\frac{p_{[\boldsymbol{x}]}\,\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle}{\sum_{[\boldsymbol{l}]} p_{[\boldsymbol{l}]}\,\langle\,1_{[\boldsymbol{l}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{l}]}\,\rangle}\right], \end{aligned} \tag{4.106}$$

where

$$h\,(u) = -u\ln[u] - 2u. \tag{4.107}$$

The last term of eq. (4.106) is non-positive as the argument of the logarithm does never exceed one. By neglecting it one arrives at a weaker upper bound which reads

$$I \leq 2M\zeta^2 + M\zeta^2 \sum_{y\in\mathcal{M}_\perp}\sum_{[\boldsymbol{x}]} h\left(p_{[\boldsymbol{x}]}\right)\langle\,1_{[\boldsymbol{x}]}\,|\,\hat{\pi}_y\,|\,1_{[\boldsymbol{x}]}\,\rangle \tag{4.108}$$

in nats. This upper bound will play a central role in the further analysis and therefore the function $h\,(u)$ is a crucial object. Its graph is shown in fig. 4.24. It is a concave function for $0 \leq u \leq 1$ and has a maximum at $u^* = e^{-3}$ with $h\,(u^*) = u^*$. For $u > e^{-2}$ the function becomes negative. In the following, we will use eq. (4.108) to analyze the cases of 2- and 3-symbol words.

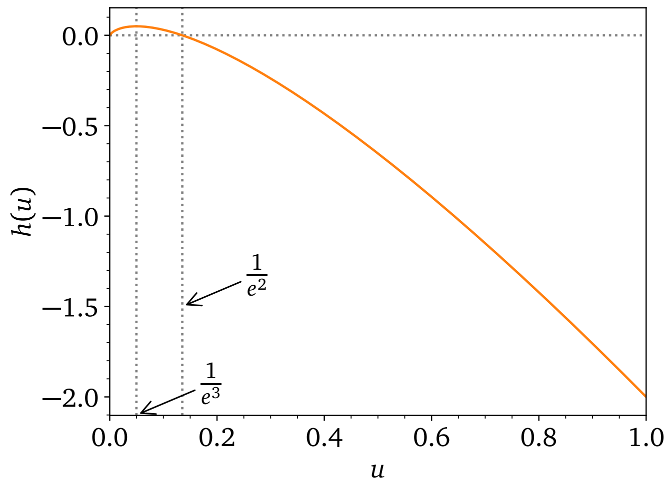


Figure 4.24: Function $h(u) = -u\ln[u] - 2u$ of eq. (4.107). The maximum value is obtained at $u^* = e^{-3}$ with $h(u^*) = u^*$ and for $u > e^{-2}$ the function becomes negative.

2-symbol words

For $M = 2$ two equivalence classes of words exist which are given by

$$|+\rangle = |1_{00}\rangle = -|1_{11}\rangle = \frac{1}{\sqrt{2}}[|10\rangle + |01\rangle], \tag{4.109}$$

$$|-\rangle = |1_{01}\rangle = -|1_{10}\rangle = \frac{1}{\sqrt{2}}[|10\rangle - |01\rangle], \tag{4.110}$$

where $+ = [00] = [11]$ and $- = [01] = [10]$ are used to denote the equivalence classes with their respective probabilities $p_+ = p_{00} + p_{11}$ and $p_- = p_{01} + p_{10}$. Inserting these expressions into eq. (4.108) yields the simplified form

$$I \leq 2\bar{n}\left[2 + h(p_+)\,p(\mathcal{M}_\perp|+) + h(p_-)\,p(\mathcal{M}_\perp|-)\right] \tag{4.111}$$

for the upper bound, where

$$p(\mathcal{M}_\perp|+) = \sum_{y\in\mathcal{M}_\perp} \langle +|\hat{\pi}_y|+\rangle, \tag{4.112}$$

$$p(\mathcal{M}_\perp|-) = \sum_{y\in\mathcal{M}_\perp} \langle -|\hat{\pi}_y|-\rangle \tag{4.113}$$

are the probabilities of obtaining any outcome $y \in \mathcal{M}_\perp$ given either state $|+\rangle$ or $|-\rangle$ respectively. In order to find the optimal probability distribution and measurement, the expression in eq. (4.111) needs to be maximized. Due to the previously discussed properties of $h(u)$ and the constraints

$$0 \leq p(\mathcal{M}_\perp|\pm) \leq 1, \tag{4.114}$$

$$p_+ + p_- = 1 \tag{4.115}$$

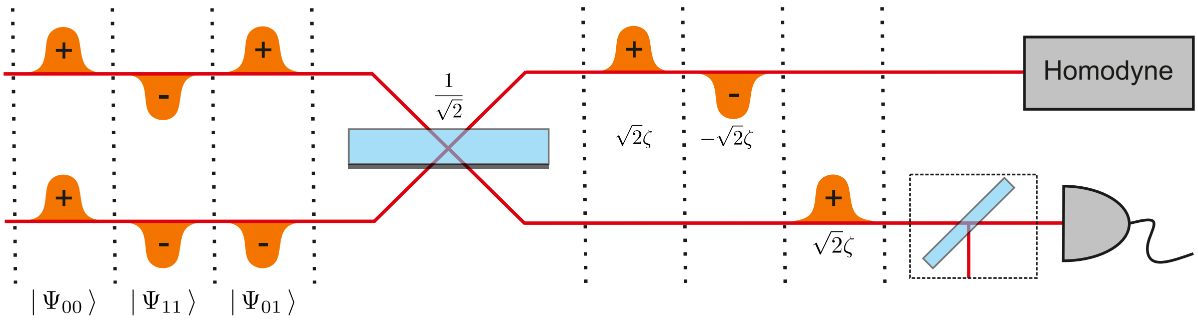


Figure 4.25: The communication scheme achieving superadditivity for 2-symbol words. They are interfered at a single 50:50 beamsplitter. The first output mode is measured by a homodyne detector whereas the second output mode is detected with a photon counter. Optionally, a displacement operation can be used before the photon counter to increase mutual information.

on the involved probabilities the maximum is achieved by choosing

$$p_- = u^*, \tag{4.116}$$

$$p\left(\mathcal{M}_\perp|-\right) = 1, \tag{4.117}$$

$$p\left(\mathcal{M}_\perp|+\right) = 0. \tag{4.118}$$

Note that swapping the labels $+$ and $-$ gives rise to an equivalent solution. Finally, the upper bound on mutual information for $M = 2$ is given by

$$\frac{I}{2} \leq 2\bar{n}\left[1 + \frac{h\left(u^*\right)}{2}\right] = 2\bar{n}\left[1 + \frac{1}{2e^3}\right] \tag{4.119}$$

in nats. Since the maximum amount of mutual information for individual measurements is given by $2\bar{n}$ in eq. (4.87), this results represents an absolute increase of $0.5e^{-3}$ nats or about 2.49% advantage relative to individual measurements.
However, one might rightfully ask whether this upper bound on mutual information is tight in the given case. Luckily, the obtained values also maximize the full expression of eq. (4.106) as the last term vanishes.
Now that we found the optimum, the next step is to translate it into an optical measurement scheme. The condition $p\left(\mathcal{M}_\perp|-\right) = 1$ implies that both states $|\,\Psi_{01}\,\rangle$ and $|\,\Psi_{10}\,\rangle$ should be detected by a type-$\mathcal{M}_\perp$ measurement which is given by a photon counting detector. Since $p_- = u^*$ specifies only the total probability of both states, it is sufficient to choose either of them. For concreteness we will choose only $|\,\Psi_{01}\,\rangle$. The other class of states, represented by $|+\rangle$, has to be measured purely with type-$\mathcal{M}$ measurements due to the condition $p\left(\mathcal{M}_\perp|+\right) = 0$. Such a measurement is given by the Dolinar or homodyne detector and in the following, we will consider the homodyne detector for our scheme. According to the Shannon Hartley limit, both states $|\,\Psi_{00}\,\rangle$ and $|\,\Psi_{11}\,\rangle$ need to be sent with equal probabilities $p_{00} = p_{11}$ in order to maximize the contribution of type-$\mathcal{M}$ measurements, which is satisfied by a BPSK alphabet in the photon starved regime.

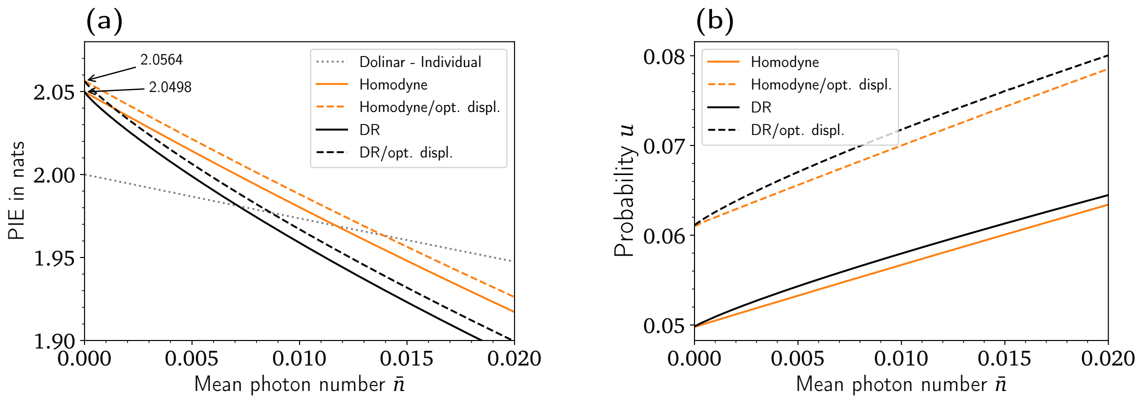


Figure 4.26: **(a)** PIE obtained by the scheme shown in fig. 4.25 with either a homodyne or displacement receiver as type-$\mathcal{M}$ measurement compared to the optimal individual measurement. The displacement receiver is limited to a PNR of 1. If an optional displacement before the photon counters for the type-$\mathcal{M}_\perp$ measurement is used, the PIE is increased which is depicted by dashed lines. **(b)** The corresponding optimal probability u for sending the state $|\,\Psi_{01}\,\rangle$.

The desired collective measurement can be achieved by interfering both input states at a 50:50 beamsplitter of the type defined in eqs. (4.56) and (4.57) resulting in the overall transformation

$$|\,\Psi_{01}\,\rangle = |\,0\,\rangle \left|\,\sqrt{2}\zeta\,\right\rangle, \tag{4.120}$$

$$|\,\Psi_{00}\,\rangle = \left|\,\sqrt{2}\zeta\,\right\rangle |\,0\,\rangle, \tag{4.121}$$

$$|\,\Psi_{11}\,\rangle = \left|\,-\sqrt{2}\zeta\,\right\rangle |\,0\,\rangle, \tag{4.122}$$

where the kets of the right-hand side refer to the two output modes of the beamsplitter. In fig. 4.25 the overall measurement scheme is depicted schematically. The homodyne detection is not the only possible choice for the type-$\mathcal{M}$ measurement, since Dolinar or displacement receivers become optimal with vanishing energy as well (cf. fig. 4.22). Note that by swapping the homodyne detector with a Dolinar receiver the measurement scheme suggested by Guha [2011] is obtained.
The final task is to determine the optimal a-priori probabilities p_{00}, p_{11} and p_{01} for a finite $\bar{n}$ as the value $p_- = u^*$ is only optimal asymptotically in the limit $\bar{n} \to 0$. To this end the actual expression for mutual information defined in eq. (3.46) has to be optimized, which was done with the parametrization $p_{01} = u$ and $p_{00} = p_{11} = \frac{1-u}{2}$. In fig. 4.26 the resulting mutual information is shown alongside the optimal probability u. In the limit $\bar{n} \to 0$ the PIE approaches the derived maximum of 2.049 and superadditivity of mutual information is achieved for $\bar{n} < 0.0117$. If instead of the homodyne receiver a displacement receiver is used, the same maximum value is achieved. However, for finite values of $\bar{n}$ the displacement receiver obtains slightly lower mutual information.

A shortcoming of the formalism presented here is that it only applies to measurement operators $\hat{\pi}$ that are independent of the energy $\bar{n}$. If the measurement is allowed to depend on $\bar{n}$, indeed higher mutual information can be attained. Physically, this can be achieved by replacing the photon counting detectors with displacement receivers where the displacement β has to be optimized depending on $\bar{n}$. In fig. 4.26 mutual information for a measurements depending on $\bar{n}$ is shown as well. It retrieves a slightly larger PIE approaching a maximum of 2.0564 for $\bar{n} \to 0$ corresponding to an improvement of 2.82% compared to individual measurements. This effect is reminiscent of a similar improvement appearing in the context of collective measurement of pairs of qubits [Buck et al. 2000].

3-symbol words

For $M = 3$ there exist four equivalence classes of words given by

$$|++\rangle = |1_{000}\rangle = -|1_{111}\rangle = \frac{1}{\sqrt{3}}\left[|100\rangle + |010\rangle + |001\rangle\right], \tag{4.123}$$

$$|+-\rangle = |1_{001}\rangle = -|1_{110}\rangle = \frac{1}{\sqrt{3}}\left[|100\rangle + |010\rangle - |001\rangle\right], \tag{4.124}$$

$$|-+\rangle = |1_{010}\rangle = -|1_{101}\rangle = \frac{1}{\sqrt{3}}\left[|100\rangle - |010\rangle + |001\rangle\right], \tag{4.125}$$

$$|--\rangle = |1_{011}\rangle = -|1_{100}\rangle = \frac{1}{\sqrt{3}}\left[|100\rangle - |010\rangle - |001\rangle\right]. \tag{4.126}$$

Unlike for 2-symbol words, the above states are mutually non-orthogonal rendering the use of eq. (4.108) problematic, since the probabilities $p\left(\mathcal{M}_\perp|1_{[\boldsymbol{x}]}\right) = \left\langle 1_{[\boldsymbol{x}]} \middle| \hat{\pi}_y \middle| 1_{[\boldsymbol{x}]} \right\rangle$ are not independent. Therefore, we are forced to base our further analysis on eq. (4.105). In order to find the optimum values for the input probabilities $p_{[\boldsymbol{x}]}$ a numerical search restricted to rank-1 type-$\mathcal{M}_\perp$ POVM elements was performed. The obtained results suggest that an optimum with a high-degree of symmetry exists. Up to a permutation of modes said optimum is characterized by

$$p_{++} = 1 - 2\nu \quad , \quad p_{+-} = p_{-+} = \nu \quad , \quad p_{--} = 0 \tag{4.127}$$

and the two type-$\mathcal{M}_\perp$ projections $\hat{\pi}_y = |\pi_y\rangle\langle\pi_y|$ with $y = 1, 2$. The projection states written in the basis $|100\rangle$,$|010\rangle$,$|001\rangle$ are given by

$$|\pi_1\rangle = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) \\ -\frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) \end{pmatrix} \quad , \quad |\pi_2\rangle = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) \\ \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) \end{pmatrix} \tag{4.128}$$

and both are orthogonal to $|++\rangle$. The POVM has an interesting geometrical interpretation which is illustrated in fig. 4.27. The first element is the projection onto $|++\rangle$ which defines a plane perpendicular to it. The remaining two elements are projections

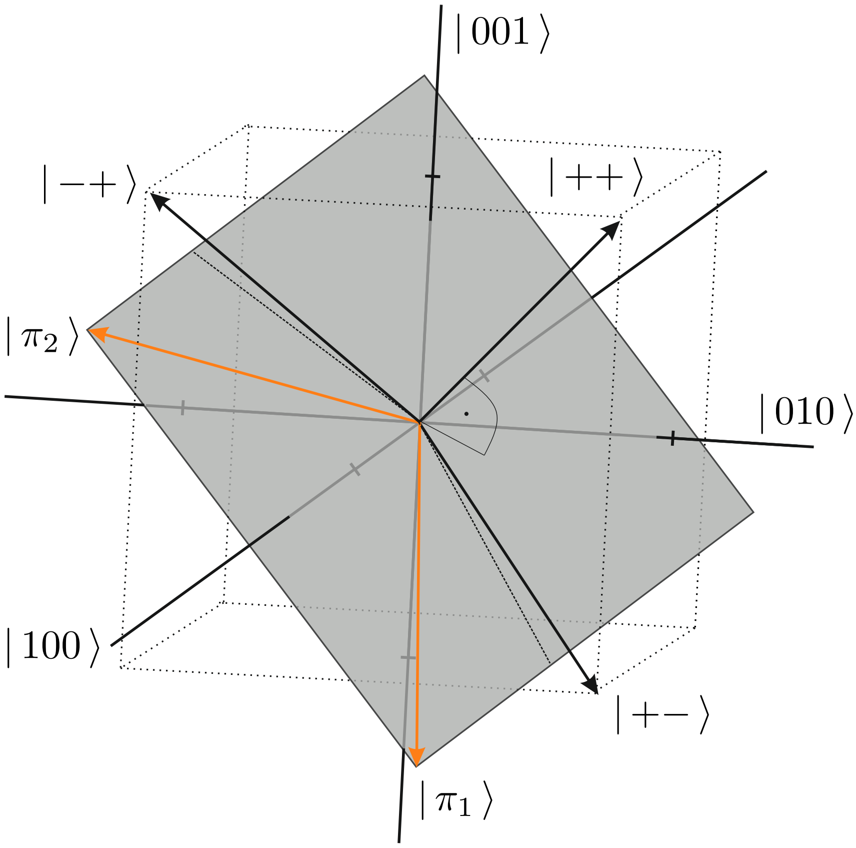


Figure 4.27: Optimal measurement for distinguishing the one-photon-states $|\,++\,\rangle, |\,+-\,\rangle$ and $|\,-+\,\rangle$ corresponding to $|\,\Psi_{000}\,\rangle, |\,\Psi_{001}\,\rangle$ and $|\,\Psi_{010}\,\rangle$. The optimal POVM consists of three projections. The first is given by a projection onto $|\,++\,\rangle$ whereas the remaining two elements are given by the minimum error measurement between the projections of $|\,+-\,\rangle$ and $|\,-+\,\rangle$ onto the plane perpendicular to $|\,++\,\rangle$. The optimal measurement vectors are $|\,\pi_1\,\rangle$ and $|\,\pi_2\,\rangle$, indicated by orange color. The figure was previously used in [Kunz et al. 2020].

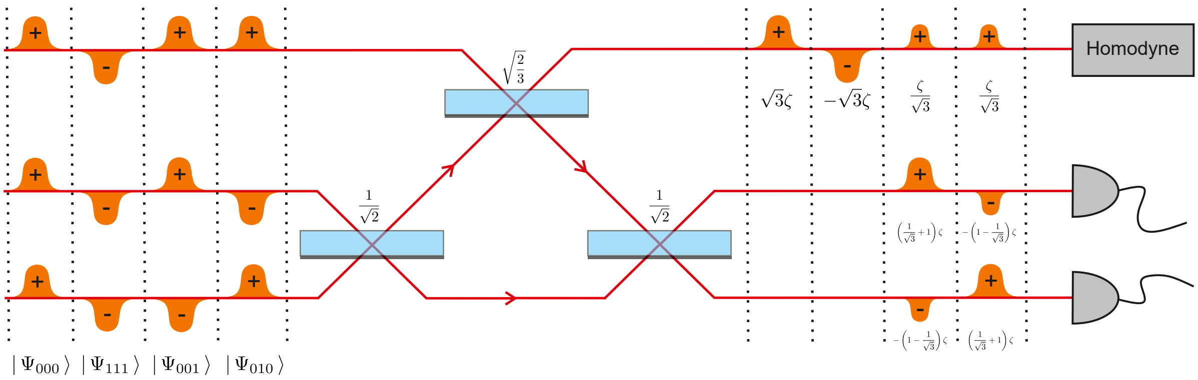


Figure 4.28: Communication scheme achieving superadditivity for 3-symbol words. The incoming states are interfered at a series of beamsplitters with varying transmittance denoted above the respective beamsplitter. The first output mode is measured by a homodyne detector whereas the other two output modes are detected by photon counting detectors.

onto the states $|\,\pi_{1,2}\,\rangle$ which are equivalent to a minimum error measurement between the states $|+-\rangle$ and $|-+\rangle$ restricted to the perpendicular plane. The bound is obtained by inserting eqs. (4.127) and (4.128) into eq. (4.105) which yields

$$\frac{I}{3} \leq 2\bar{n} + \frac{2\bar{n}}{9}\left\{4\nu\left[\sqrt{3}\ln\left[2+\sqrt{3}\right] - 4\ln\left[2\right] - \nu\right] + 8h\left(\nu\right)\right\}. \tag{4.129}$$

The maximum of the right-hand side is reached for $\nu^* \approx 0.0375$ which equals an improvement of 3.40% relative to individual measurements.
The solution presented above suggests a physical implementation similar to the 2-symbol scenario. The input states are routed through a three-port linear optical circuit consisting of three beamsplitters with varying transmissivity. All three different classes of states $|++\rangle, |+-\rangle, |-+\rangle$ are directed to a different output port. The port corresponding to $|++\rangle$ is measured by a homodyne detector and the other two ports are measured by photon counting detectors. The protocol is schematically depicted in fig. 4.28. A notable difference from the 2-symbol scenario is caused by the non-orthogonality of words. The energy cannot be collected entirely into a single mode for all words such that the obtained outcomes are not unambiguous. Nevertheless, the strategy is superior to individual or 2-symbol measurements. Maximizing mutual information obtained by the aforementioned setup over ν for finite $\bar{n}$ achieves the PIE shown in fig. 4.29. The 3-symbol scheme provides larger PIE than the 2-symbol scheme for $\bar{n} < 0.0059$ and is superior than individual measurements for $\bar{n} < 0.0093$ with an advantage of up to 3.40%.

Orthogonal words

For words consisting of more than 3-symbols numerical means are necessary to obtain the optimal solution. However, in cases where we know that different input words are

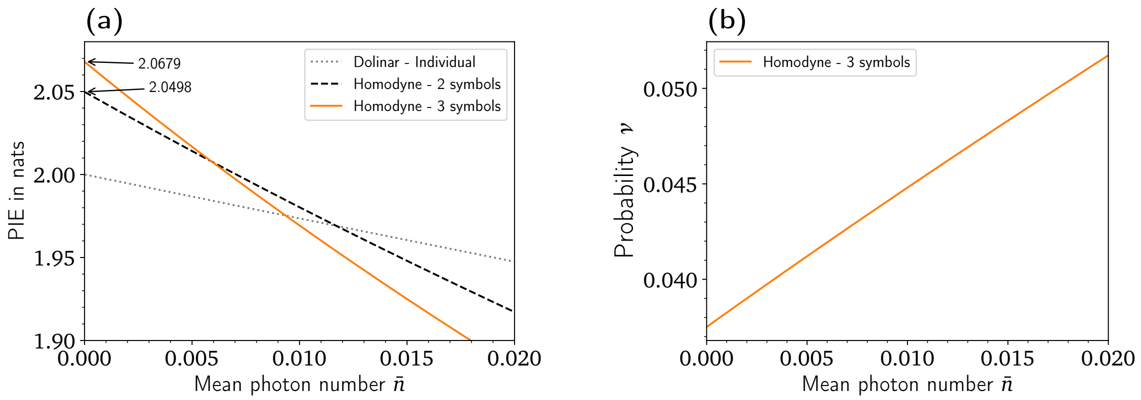


Figure 4.29: **(a)** The PIE of the 3-symbol scenario with a homodyne detector as type-$\mathcal{M}$ measurement compared to the 2-symbol scenario and the optimal individual measurement, the Dolinar receiver. For the type-$\mathcal{M}_\perp$ measurements no displacement was used. **(b)** The optimal probability ν corresponding to attained PIE.

orthogonal to each other further observation can be made based on eq. (4.108). The words are known to be orthogonal if Hadamard words can be constructed as discussed in section 4.6.

In order to maximize the right-hand side of eq. (4.108), we need to properly choose the a-priori probabilities $p_{[\boldsymbol{x}]}$ of the words and the POVM elements, i.e. the probabilities defined by $\left\langle 1_{[\boldsymbol{x}]} \middle| \hat{\pi}_y \middle| 1_{[\boldsymbol{x}]} \right\rangle$. Since the function $h\left(u\right)$ is only positive for $0 < u < e^{-2}$, we choose $\left\langle 1_{[\boldsymbol{x}]} \middle| \hat{\pi}_y \middle| 1_{[\boldsymbol{x}]} \right\rangle = 1$ if $0 < p_{[\boldsymbol{x}]} < e^{-2}$ and zero otherwise. Then, the upper bound on PIE in nats reads

$$\mathrm{PIE} = \frac{I}{M\bar{n}} \leq 2 + \sum_{y\in\mathcal{M}_\perp} \sum_{0<p_{[\boldsymbol{x}]}<e^{-2}} h\left(p_{[\boldsymbol{x}]}\right). \tag{4.130}$$

One simple strategy is to use all input words with equal probability, i.e. $p_{[\boldsymbol{x}]} = \frac{1}{M}$, and the PIE reduces to

$$\mathrm{PIE} \leq \ln\left[M\right]. \tag{4.131}$$

This strategy is entirely equivalent to PPM (cf. eq. (3.136) and section 4.6) as the input Hadamard words are converted to PPM symbols.

The second strategy is to choose the probabilities for all words but one to be at the maximum of $h\left(u\right)$, i.e. $p_{[\boldsymbol{x}]} = u^*$. The remaining word is then measured by a type-$\mathcal{M}$ measurement in accordance with setups presented above. We know that choosing the same probability for as many words as possible is optimal, since the function $h\left(u\right)$ is concave in the interval $0 < u < e^{-2}$. In said choice the bound further reduces to the expression

$$\mathrm{PIE} \leq 2 + \frac{M-1}{e^3} \tag{4.132}$$

in nats. This strategy, however, is only possible for $M \leq 20$ as the probabilities have to obey $\sum_{[\boldsymbol{x}]} p_{[\boldsymbol{x}]} = 1$.
For $M \geq 16$ the first strategy corresponding to PPM is superior, whereas for lower order words the second strategy obtains higher mutual information. These observations are consistent with [Klimek et al. 2016].

5 Noise Models in Optical Channels

During communication noise acting on the signal state is inevitable. Noise is introduced by the detector or during the propagation of a quantum system from the environment it travels through. This type of noise is commonly referred to as excess noise. In order to asses the optimal strategy and the possible information transfer of a particular communication link the type of excess noise must be known and an appropriate model developed.

In the following chapter several types of noise are considered and the corresponding models are reviewed. In chapter 6 the possible information transfer of several scenarios will be discussed based on the models developed below. All noise effects described in this chapter are aimed to represent a very simple quantum model of a glass fibre or an optical free space link. They are based on an open quantum system approach, where the propagating state interacts with an environmental system. Instead of evolution in time, the focus lies on the propagation of a single-mode quantum state in a spatial direction denoted by the variable z. In section 5.1 distributed losses and thermal background noise are discussed. In the following sections, the two types of phase noise are considered. First, in section 5.2 linear phase noise, which is independent of the signal states energy, is reviewed briefly and then, in section 5.3, nonlinear phase noise arising from the interaction of distributed losses and a Kerr nonlinearity.

5.1 Thermal Noise and Losses

The first type of noise we will consider is thermal noise which can be considered as the quantum counterpart of classical additive white Gaussian noise. It emerges, for instance, from thermal instabilities of the optical fibre, due to black body radiation in free space communication links or thermal effects in the photon detector itself [Vilnrotter and Rodemich 1984; Yuan et al. 2020]. While in atmospheric or fibre channels other noise sources, such as turbulences or Kerr nonlinearity, are overwhelming, it is a typical type of noise present in deep space communication [Haykin 2001; Essiambre et al. 2010]. Note that thermal noise usually does not come in the form of true thermal radiation as this takes place in many modes, but rather the aforementioned interactions cause a background noise with thermal statistics. Nevertheless, the below presented thermal noise model appropriately describes this type of noise as well.

While an effective description of thermal noise was already given in eq. (2.93), in terms of open quantum systems it is best described by an optical quantum state interacting with an environment at a non-zero temperature. During propagation photons in the signal state can be absorbed or created by the thermal environment such that thermal photons are incoherently added. Formally, this interaction is described by a master equation in the Kossakowski-Lindblad form where a harmonic oscillator interacts with a heated environment. It is given by [Gardiner and Zoller 2004]

$$\frac{\mathrm{d}\hat{\varrho}}{\mathrm{d}z} = \frac{\alpha}{2}\left(n_b + 1\right)\left(2\hat{a}\hat{\varrho}\hat{a}^\dagger - \hat{a}^\dagger\hat{a}\hat{\varrho} - \hat{\varrho}\hat{a}^\dagger\hat{a}\right) + \frac{\alpha}{2} n_b \left(2\hat{a}^\dagger\hat{\varrho}\hat{a} - \hat{a}\hat{a}^\dagger\hat{\varrho} - \hat{\varrho}\hat{a}\hat{a}^\dagger\right), \tag{5.1}$$

where n_b is the average number of photons contained in the environment and α denotes the interaction strength. Note that here we switched from considering the evolution in time to the propagation in z-direction following the approach of [Raymer and Mostowski 1981; Wasilewski et al. 2006]. Also, we consider the interaction picture. The stationary solution to eq. (5.1) is given by a thermal state $\hat{\varrho}_{\text{th}}$ as described in eq. (2.91).
For a coherent state $\hat{\varrho} = |\,\zeta\,\rangle\langle\,\zeta\,|$ as input state, the solution is given by a displaced thermal state that according to eq. (2.94) reads

$$\hat{\varrho}\left(z\right) = \hat{D}\left(\sqrt{\tau}\zeta\right)\hat{\varrho}_{\text{th}}\hat{D}^\dagger\left(\sqrt{\tau}\zeta\right) \tag{5.2}$$

with the power transmission of the channel defined as $\tau = e^{-\alpha z}$. The matrix elements of $\hat{\varrho}\left(z\right)$ are given by eq. (2.97) with $\zeta \to \sqrt{\tau}\zeta$.
A pure loss channel can be easily retrieved from above solution by setting $n_b = 0$. In this scenario the ground state is given by the vacuum state $|\,0\,\rangle$ and the solution for coherent states reduces to a coherent state with reduced amplitude $\sqrt{\tau}\zeta$.
A simpler model of a lossy thermal channel is given by a beamsplitter with transmission $\sqrt{\tau}$. The incoming signal state is mixed with a thermal state or the vacuum state for a pure loss channel [Cerf et al. 2007]. However, if other channel effects that do not commute with losses or thermalization are present, this model may prove to be too simplistic as it neglects further arising interactions.
Overall the interaction of coherent state with a thermal environment is entirely of Gaussian nature. The capacity is therefore given by a AWGN channel as defined in eq. (3.95).

5.2 Linear Phase Noise

If the signal state interacts with the environment, two basic interactions of the signal and environment systems photons can appear. First, a system photon can be destroyed while an environmental photon is created and vice versa. Secondly, the signal photon can be scattered when emitting or absorbing an environmental photon. As we will see the latter scattering effect gives rise to phase noise.
The master equation in the Kossakowski-Lindblad form describing linear phase noise due to scattering is given by [Gardiner and Zoller 2004]

$$\frac{\mathrm{d}\hat{\varrho}}{\mathrm{d}z} = \frac{\Gamma}{2}\left(2\hat{n}\hat{\varrho}\hat{n} - \hat{n}^2\hat{\varrho} - \hat{\varrho}\hat{n}^2\right), \tag{5.3}$$

in the interaction picture, where Γ describes the interaction strength and $\hat{n} = \hat{a}^\dagger\hat{a}$ is the photon number operator. Note that eq. (5.3) does not incorporate losses. In terms of the matrix elements $\varrho_{m,n} = \langle m \,|\, \hat{\varrho} \,|\, n \rangle$ the master equation reads

$$\frac{\mathrm{d}\varrho_{m,n}}{\mathrm{d}z} = \frac{\Gamma}{2}\left[2mn\varrho_{m,n} - m^2\varrho_{m,n} - n^2\varrho_{m,n}\right] \tag{5.4}$$

$$= -\frac{\Gamma}{2}\left(m-n\right)^2 \varrho_{m,n}, \tag{5.5}$$

such that the solution is given by

$$\varrho_{m,n}\left(t\right) = e^{-\frac{\Gamma(m-n)^2 z}{2}} \varrho_{m,n}\left(0\right). \tag{5.6}$$

The diagonal elements, and therefore the energy of the state, remain entirely unaffected by the scattering. In contrast, off-diagonal elements decay exponentially due to the exponential factor and the decay gets stronger with rising distance to the diagonal. In short, the effect on the state $\hat{\varrho}$ is decoherence.
For a coherent state $\hat{\varrho} = |\,\zeta\,\rangle\langle\,\zeta\,|$ the output state of the evolution described by eq. (5.3) can be obtained via the transformation [Genoni et al. 2011; Olivares et al. 2013]

$$|\,\zeta\,\rangle\langle\,\zeta\,| \to \int \mathrm{d}\Phi \,\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\Phi^2}{2\sigma^2}} \,|\,\zeta e^{-i\Phi}\,\rangle\langle\,\zeta e^{-i\Phi}\,|\,, \tag{5.7}$$

which describes Gaussian phase diffusion. In the Fock basis the coherent state after phase noise is given by

$$\int \mathrm{d}\Phi \,\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\Phi^2}{2\sigma^2}} \,|\,\zeta e^{-i\Phi}\,\rangle\langle\,\zeta e^{-i\Phi}\,| \tag{5.8}$$

$$= \sum_{m,n=0}^{\infty} \langle\, m \,|\, \zeta e^{-i\Phi}\,\rangle\langle\,\zeta e^{-i\Phi}\,|\, n\,\rangle\, e^{-\frac{(m-n)^2\sigma^2}{2}} \,|\, m\,\rangle\langle\, n\,|\,. \tag{5.9}$$

By comparing eq. (5.9) and eq. (5.6) it is clearly seen that the solution to the master equation in eq. (5.3) is well described by Gaussian phase noise with a standard deviation given by $\sigma = \sqrt{\Gamma z}$. Since the noise does not depend on the energy $\bar{n} = |\zeta|^2$ of the state, it is said to be linear phase noise.

5.3 Nonlinear Phase Noise

In an optical medium with third-order nonlinearity the refractive index depends on the intensity of light. Most prominently it causes self-phase modulation, the so-called Kerr effect, which inflicts an intensity dependent phase shift and rotational shearing of the state [Milburn 1986; Milburn and Holmes 1986; Daniel and Milburn 1989; D. F. Walls 2008]. This nonlinear effect can be used for the generation of squeezed states [Rosenbluh and Shelby 1991; Schmitt et al. 1998; Werner 1998; Fiorentino et al. 2002] or Schrödinger cat states [Yurke and Stoler 1986; Paris 1999; Bužek and Knight 1995]. In optical communication, however, it is known for limiting the information transfer [Mitra and Stark 2001; Wegener et al. 2004; Temprana et al. 2015; Sorokina and Turitsyn 2014; Turitsyn

and Turitsyn 2012]. While the Kerr effect can be described by a semi-classical description, a quantum model shows additional decoherence [Milburn 1986]. The goal of the following section is to analyze the quantum model with particular focus on the additional quantum effects and their impact on optical communication and the generation of squeezed states. The developed model aims to be a very simple quantum description of a glass fibre which only considers self-phase modulation paired with linear losses distributed over the whole channel.
Self-phase modulation is described by the Hamiltonian $H_{\mathrm{K}} = \mu\hat{n}^2$ with the photon number operator $\hat{n} = \hat{a}^\dagger\hat{a}$ and a parameter μ describing the strength of the Kerr nonlinearity itself. The parameter μ is directly proportional to the third-order nonlinearity of the Kerr medium $\chi^{(3)}$ and can be linked to the parameters of an actual glass fibre link via $\mu = \gamma_{\mathrm{NL}}\hbar\omega_0/T$, with the fibre third-order nonlinearity γ_{NL}, the energy $\hbar\omega_0$ of a photon at frequency ω_0 and the pulse duration T. Distributed losses are described by a parameter α and modeled by placing the optical medium in a cold reservoir. The propagation in z-direction of a single-mode through an optical Kerr medium with linear losses is therefore given by a Master equation in the Kossakowski-Lindblad form

$$\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}z} = i\mu\left[\hat{n}^2, \hat{\rho}\left(z\right)\right] + \frac{\alpha}{2}\left(2\hat{a}\hat{\rho}\left(z\right)\hat{a}^\dagger - \hat{n}\hat{\rho}\left(z\right) - \hat{\rho}\left(z\right)\hat{n}\right). \tag{5.10}$$

Note that here we use the symbol $\hat{\rho}$ and reserve the symbol $\hat{\varrho}$ for density operators in the interaction picture. Since we only want to capture the additional non-unitary quantum effects arising in a quantum model, it is convenient to switch to the interaction picture, i.e. transform the density operator according to

$$\hat{\rho}\left(z\right) = \exp\left(i\mu z\hat{n}^2\right)\hat{\varrho}\left(z\right)\exp\left(-i\mu z\hat{n}^2\right). \tag{5.11}$$

The master equation of eq. (5.10) now reads

$$\frac{\mathrm{d}\hat{\varrho}}{\mathrm{d}z} = \frac{\alpha}{2}\left(2\hat{a}'\left(z\right)\hat{\varrho}\left(z\right)\left[\hat{a}'\left(z\right)\right]^\dagger - \hat{n}\hat{\varrho}\left(z\right) - \hat{\varrho}\left(z\right)\hat{n}\right) \tag{5.12}$$

with modified creation and annihilation operators given by

$$\hat{a}'\left(z\right) = \exp\left(-i\mu z\hat{n}^2\right)\hat{a}\exp\left(i\mu z\hat{n}^2\right), \tag{5.13}$$

which are now explicitly dependent on the position z. The photon number operator $\hat{n}$ is invariant with respect to this transformation. In order to solve eq. (5.12) we assume a coherent state $|\,\zeta_0\,\rangle$ as initial state. Later on, we will generalize the solution for coherent states to arbitrary input states.

5.3.1 Solving the Master Equation for Coherent States

An appropriate ansatz for the matrix elements $\varrho_{m,n} = \langle\, m\,|\,\hat{\varrho}\,|\,n\,\rangle$ appearing in eq. (5.12) is

$$\varrho_{m,n}\left(z\right) = \frac{\zeta_0^m\left(\zeta_0^*\right)^n}{\sqrt{m!n!}}e^{-(m+n)\alpha z/2}c_{m-n}\left(z\right), \tag{5.14}$$

following the approach of [Paris 1999]. It consists of the solution for the lossy transmission of a coherent state and an additional coefficient $c_{m-n}(z)$ that captures the propagation due to the residue interaction between losses and the nonlinearity. Note that the coefficient $c_{m-n}(z)$ only depends on the difference $m-n$ of the density matrix element indices in the Fock basis. Rewriting eq. (5.12) into a differential equation for the matrix elements results in

$$\frac{\mathrm{d}\varrho_{m,n}}{\mathrm{d}z} = \frac{\alpha}{2}\left[2\varrho_{m+1,n+1}(z)\sqrt{(m+1)(n+1)}e^{2i\mu z(m-n)} - (m+n)\varrho_{m,n}(z)\right]. \tag{5.15}$$

Inserting the ansatz of eq. (5.14) into eq. (5.15) yields a first order differential equation for the coefficient $c_{m-n}(z)$ reading

$$\frac{\mathrm{d}c_{m-n}}{\mathrm{d}z} = \alpha\left|\zeta_0\right|^2 e^{-\alpha z}e^{2i\mu z(m-n)}c_{m-n}(z). \tag{5.16}$$

This equation is solved by

$$\begin{aligned} c_{m-n}(z) &= \exp\left(-\frac{\left|\zeta_0\right|^2\alpha}{\alpha-2i\mu(m-n)}e^{-\alpha z}e^{2i\mu z(m-n)} + A_0\right) \\ &= \exp\left(-\frac{\left|\zeta_0\right|^2\tau}{1-2i\varkappa(m-n)}\tau^{-2i\varkappa(m-n)} + A_0\right), \end{aligned} \tag{5.17}$$

with the power transmission $\tau = e^{-\alpha z}$, the relative nonlinear strength $\varkappa = \mu/\alpha$ and an integration constant A_0. In the last step the identity $e^{2i\mu z(m-n)} = e^{-2i\varkappa(m-n)\log_2\tau} = \tau^{-2i\varkappa(m-n)}$ was used. Together with eq. (5.15) the solution in eq. (5.17) leads to

$$\varrho_{m,n}(z) = \frac{\left(\sqrt{\tau}\zeta_0\right)^m\left(\sqrt{\tau}\zeta_0^*\right)^n}{\sqrt{m!n!}}\exp\left(-\frac{\left|\zeta_0\right|^2\tau^{1-2i\varkappa(m-n)}}{1-2i\varkappa(m-n)} + A_0\right). \tag{5.18}$$

By using the initial condition of starting in a regular coherent state, we finally obtain the integration constant

$$A_0 = \frac{\left|\zeta_0\right|^2}{2i\varkappa(m-n)+1} - \left|\zeta_0\right|^2 \tag{5.19}$$

and the closed form solution for the density matrix elements is given by

$$\varrho_{m,n}(z) = \left\langle m \,\middle|\, \sqrt{\tau}\zeta_0 \right\rangle\left\langle \sqrt{\tau}\zeta_0 \,\middle|\, n \right\rangle \exp\left[-\left|\zeta_0\right|^2 f_\tau\left(\varkappa(m-n)\right)\right], \tag{5.20}$$

with a complex valued function

$$f_\tau(\varkappa) = 1 - \tau - \frac{1-\tau^{1-2i\varkappa}}{1-2i\varkappa}. \tag{5.21}$$

The solution consists formally of two parts. The first one, $\langle m \,|\, \sqrt{\tau}\zeta_0 \rangle\langle \sqrt{\tau}\zeta_0 \,|\, n \rangle$, is the solution for the propagation of a coherent state through a channel with linear losses. It

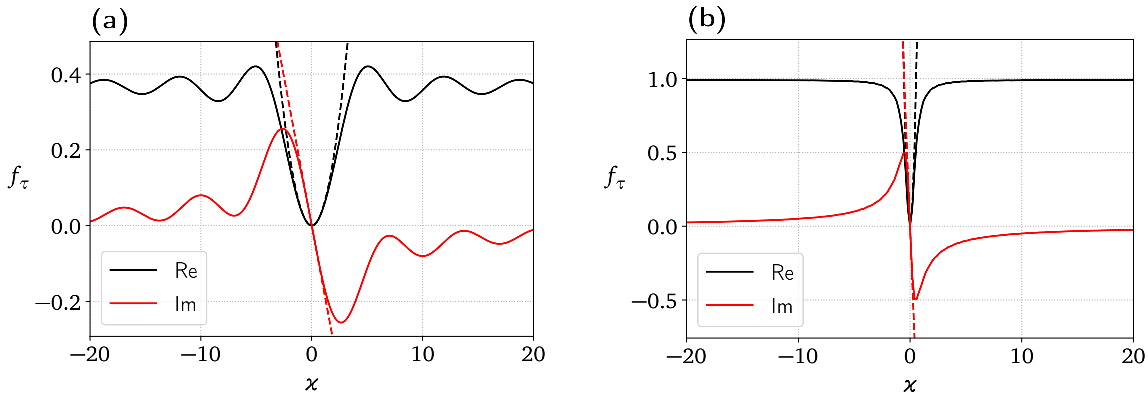


Figure 5.1: Real and imaginary part of the function f_τ of eq. (5.21) for **(a)** $\tau = 0.631$ and **(b)** $\tau = 0.100$. The dashed lines correspond to the expansion up to the first nonzero order written in eq. (5.32).

rescales the complex amplitude ζ with the transmission of the channel τ. The second part, $\exp\left[-\left|\zeta_0\right|^2 f_\tau\left(\varkappa\left(m-n\right)\right)\right]$, contains the nonlinear quantum effect. The exact nature of this quantum noise is best understood by looking at the function f_τ of eq. (5.21) depicted in fig. 5.1. The similarity of the real part of $f_\tau\left(\varkappa\right)$ to a resonance curve is explained by looking further into its structure. It can be rewritten as

$$\mathrm{Re}\left[f_\tau\left(\varkappa\right)\right] = 1 - \tau - \frac{1}{1+4\varkappa^2} - \frac{\tau\cos\left(2\varkappa\log_2\tau\right) + 2\tau\varkappa\sin\left(2\varkappa\log_2\tau\right)}{1+4\varkappa^2}. \tag{5.22}$$

Note that the middle term is a Lorentzian function, which is responsible for the resonance-like behavior. It has an amplitude and FWHM of 1 and is located at $\varkappa = 0$. The first part, $1-\tau$, defines the maximum value at which the function saturates. The last term causes a fast oscillation, whose frequency rises with decreasing τ. The imaginary part $\mathrm{Im}\left[f_\tau\left(\varkappa\left(m-n\right)\right)\right]$ only gives rise to a phase shift of the coherent state. More interestingly, the real part $\mathrm{Re}\left[f_\tau\left(\varkappa\left(m-n\right)\right)\right]$ causes an exponential damping of matrix elements $\varrho_{m,n}$ and the damping strength depends on the difference between the Fock indices $m-n$. This decoherence effect is stronger the further away the matrix element is from the 0th diagonal, whereas the 0th diagonal itself is not affected.
In fig. 5.2 the propagation of a coherent state is illustrated in phase space with the Husimi-Q-function. The coherent state gets blurred with increasing nonlinear noise strength. The same qualitative effect is obtained by increasing the propagation distance or the energy of the state. Note that the observed blurring is only due to the non-unitary effect stemming from the combination of Kerr nonlinearity and linear losses, as the unitary Kerr effect is not included in the utilized interaction picture and is therefore also not visible on the shown figures.
In a nutshell, the channel first rescales the amplitude of the coherent state by the transmission $\sqrt{\tau}$, then adds quantum excess noise in the form of decoherence and lastly applies the unitary Kerr effect.

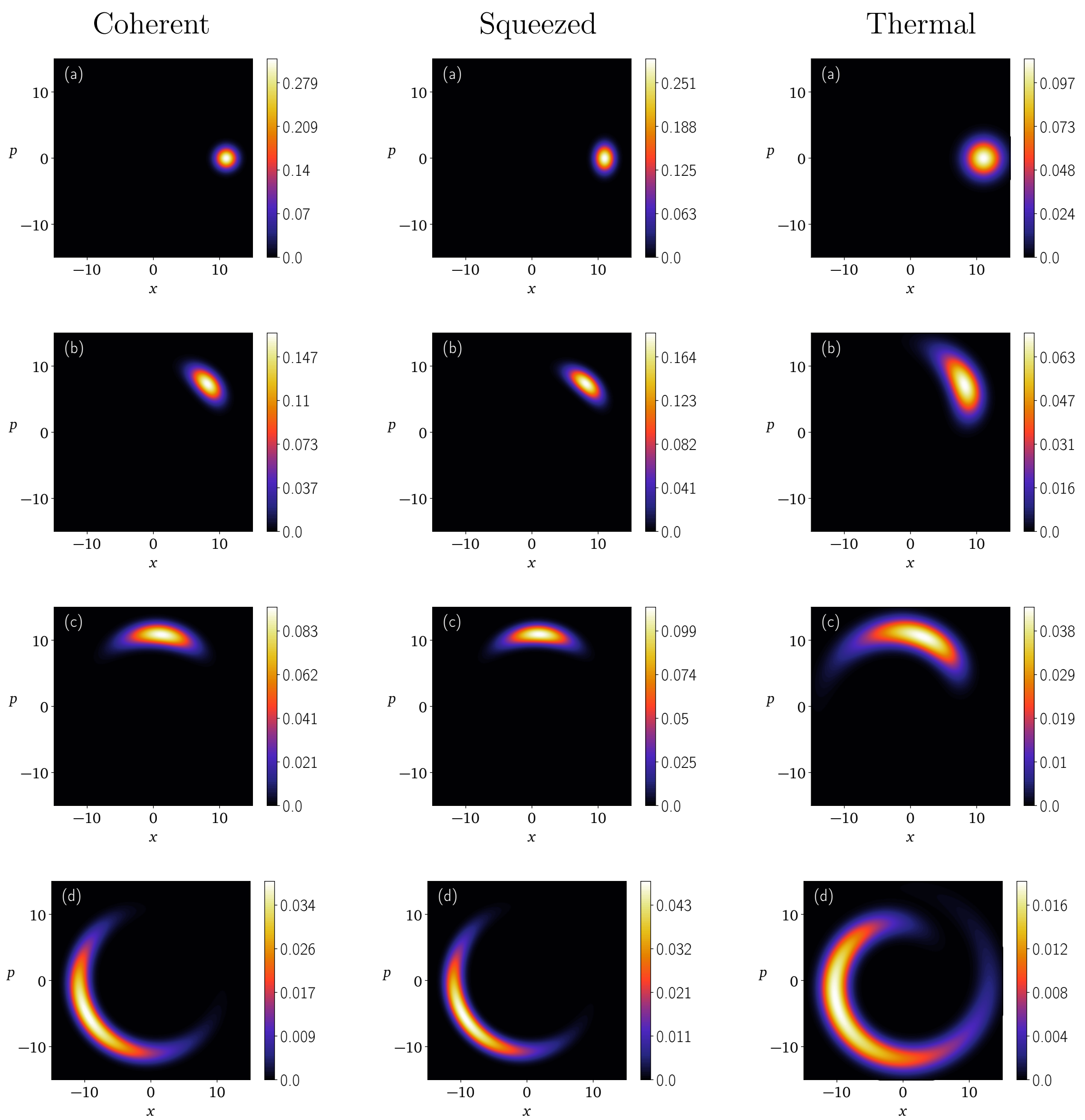


Figure 5.2: Husimi Q function of states transmitted through a nonlinear phase noise channel for power transmission $\tau = 0.631$, photon number at the output $\tau |\zeta_0|^2 = 60$ and nonlinear strength **(a)** $\varkappa = 0$, **(b)** $\varkappa = 0.05$, **(c)** $\varkappa = 0.10$ and **(d)** $\varkappa = 0.25$. The first column shows coherent states, the second column ideal squeezed states for a squeezing parameter of $r = 0.5$ and the third column shows thermal states with a thermal background of $n_b = 3$.

5.3.2 Solution for Arbitrary Input States

So far we are only able to describe the action of the channel on coherent states. Since they form an overcomplete set in the Hilbert space, it is possible to generalize the solution for coherent states to arbitrary input states. Formally, this can be achieved with the help of the Glauber-Sudarshan-P-representation as an arbitrary state can be described by the P-function. Recalling eq. (2.63), the representation is given by

$$\hat{\varrho} = \int \mathrm{d}^2\zeta \, P(\zeta) \, | \, \zeta \, \rangle \, \langle \, \zeta \, | \, . \tag{5.23}$$

A known solution for coherent states can be extended to arbitrary states by exploiting the linearity of the P-representation [Rahimi-Keshari et al. 2011]. The general effect of a channel can be written down with the help of a superoperator Φ acting on the density matrix of an arbitrary state $\hat{\varrho} = \sum_{m,n} \varrho_{m,n} \, | \, m \, \rangle \, \langle \, n \, |$ as

$$\Phi(\hat{\varrho}) = \sum_{m,n,j,k} \Phi^{m,n}_{j,k} \, \varrho_{m,n} \, | \, j \, \rangle \, \langle \, k \, | \, . \tag{5.24}$$

The elements of the tensor $\Phi^{m,n}_{j,k}$ are given by

$$\Phi^{m,n}_{j,k} = \frac{1}{\sqrt{m!n!}} \partial^m_\zeta \partial^n_{\bar{\zeta}} \left[e^{|\zeta|^2} \varrho_{j,k} \right]_{\zeta=0,\bar{\zeta}=0} , \tag{5.25}$$

where $\varrho_{j,k}$ is the matrix element of the solution for coherent states derived in eq. (5.20). The superoperator Φ for the lossy nonlinear Kerr channel can then be calculated to be

$$\Phi^{m,n}_{j,k} = \frac{F^{m-j}_{|j-k|}(\tau, \varkappa)}{(m-j)!} \tau^{\frac{j+k}{2}} \sqrt{\frac{m!n!}{j!k!}} \delta_{m-j,n-k}, \tag{5.26}$$

with the function

$$F_{|j-k|}(\tau, \varkappa) = 1 - \tau - f_\tau(\varkappa \, |j-k|) = \frac{1 - \tau^{1-2i\varkappa|j-k|}}{1 - 2i\varkappa \, |j-k|}. \tag{5.27}$$

Interestingly, the superoperator for the nonlinear channel is formally very close to the superoperator $\mathcal{L}$ of a pure loss channel given by [Rahimi-Keshari et al. 2011]

$$\mathcal{L}^{m,n}_{j,k} = \frac{(1-\tau)^{m-j}}{(m-j)!} \tau^{\frac{j+k}{2}} \sqrt{\frac{m!n!}{j!k!}} \delta_{m-j,n-k}. \tag{5.28}$$

The solution for pure losses can be obtained by setting $\varkappa = 0$ in the derivation of eq. (5.26). The sole, but defining, difference is that the real valued term $(1-\tau)^{m-j}$ of the pure loss channel is replaced by the complex valued function $F^{m-j}_{|j-k|}(\tau, \varkappa)$ of eq. (5.27). The effect of the nonlinear channel on an arbitrary input state can now be written as

$$\Phi(\hat{\varrho}) = \frac{\tau^{\frac{j+k}{2}}}{\sqrt{j!k!}} \sum_{n=0}^{\infty} \frac{F^n_{|j-k|}(\tau, \varkappa)}{n!} \sqrt{(n+j)!\,(n+k)!} \varrho_{n+j,n+k} \, | \, j \, \rangle \, \langle \, k \, | \, . \tag{5.29}$$

In fig. 5.2 the action of the nonlinear phase noise channel on ideal squeezed states and thermal states is shown. Amplitude squeezed states of the same average energy are slightly less affected than the corresponding coherent state. Thermalized coherent states, however, are strongly affected by the nonlinear channel as the added thermal photons contribute to the strength of the quantum excess noise.

5.3.3 Effective Description - Gaussian Phase Noise

The propagation of coherent states according to eq. (5.20) can be described and understood by deriving an effective description. The behavior of the function $f_\tau(\varkappa)$ in fig. 5.1 suggests an expansion around $\varkappa = 0$. The expansion becomes relevant when the matrix elements $\langle m \,|\, \sqrt{\tau}\zeta_0 \rangle \langle \sqrt{\tau}\zeta_0 \,|\, n \rangle$ are substantially different from zero. The range of m, n in which this occurs can be approximated considering the Poissonian distribution defined by $|\langle \sqrt{\tau}\zeta_0 \,|\, n \rangle|^2$ with mean value and standard deviation $\tau\,|\zeta_0|^2$. This distribution is non-vanishing roughly for

$$\tau\,|\zeta_0|^2 - \sqrt{\tau}\,|\zeta_0| \lesssim n \lesssim \tau\,|\zeta_0|^2 + \sqrt{\tau}\,|\zeta_0|\,. \tag{5.30}$$

Therefore, the argument of the function f in eq. (5.20) should be $\varkappa\,(m-n) \lesssim 2\varkappa\sqrt{\tau}\,|\zeta_0|$ and an expansion of $f_\tau(\varkappa)$ around zero is valid if

$$2\varkappa\sqrt{\tau}\,|\zeta_0| \ll 1. \tag{5.31}$$

Up to the second order in $\varkappa$ eq. (5.21) is approximated by

$$f_\tau(\varkappa) \approx f_\tau^G(\varkappa) = -2i\,(1-\tau+\tau\log_2\tau)\,\varkappa + 2\left[2-\tau-\tau\,(1-\log_2\tau)^2\right]\varkappa^2. \tag{5.32}$$

A density matrix element of eq. (5.20) itself can therefore be estimated by

$$\varrho_{m,n}^G(z) = \langle m \,|\, \sqrt{\tau}\zeta_0 \rangle \langle \sqrt{\tau}\zeta_0 \,|\, n \rangle\, e^{-|\zeta_0|^2 \mathrm{Re}\left[f_\tau^G(\varkappa(m-n))\right]} e^{-i|\zeta_0|^2 \mathrm{Im}\left[f_\tau^G(\varkappa(m-n))\right]}. \tag{5.33}$$

This approximation is quite accurate for a wide range of parameters, even beyond the range implied by the assumptions used to derive eq. (5.33). The approximation becomes inaccurate, if $|\zeta_0|$ is too large. However, since the real part of the approximation gives rise to an exponential damping of the matrix element that grows with $|\zeta_0|$ the inaccuracy of the approximation in this regime becomes irrelevant. A good measure to verify this effect is the infidelity [Uhlmann 1976]

$$1 - F = 1 - \mathrm{Tr}\left[\sqrt{\sqrt{\hat{\varrho}}\hat{\varrho}^G\sqrt{\hat{\varrho}}}\right] \tag{5.34}$$

between the approximated density matrix $\hat{\varrho}^G$ and the actual matrix $\hat{\varrho}$, where F denotes the fidelity defined in eq. (3.26). A small infidelity ensures that the separation in Hilbert space between the states is small and they may therefore be seen as states whose physical properties are close [Benedetti et al. 2013; Bina et al. 2014; Mandarino et al. 2016]. In fig. 5.3 the infidelity is shown for a transmission of $\tau = 0.631$. It is very close to zero,

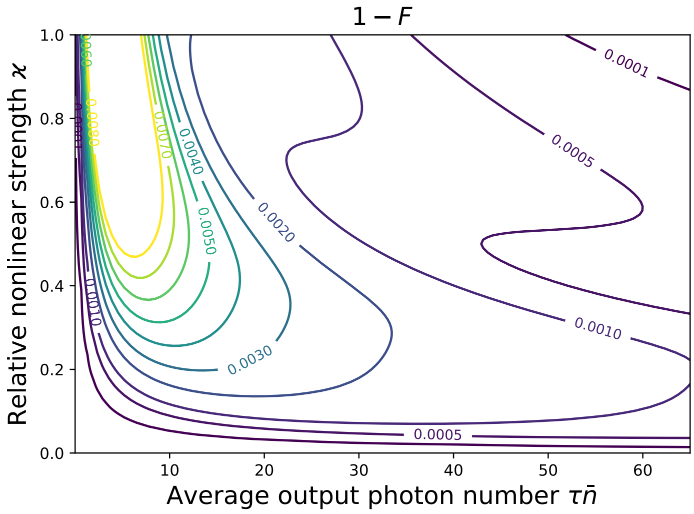


Figure 5.3: Infidelity $1 - F$ between the approximate output of eq. (5.33) and the exact state of eq. (5.20) as a function of the output photon number $\tau\bar{n}$ and the nonlinear strength $\varkappa$ for $\tau = 0.631$.

except for a region around $\tau\bar{n} \approx 6$ and large values of the relative nonlinear strength $\varkappa$. However, the infidelity in this regime is of the order of 1% and large values of $\varkappa$ are rarely reached in realistic scenarios. A coherent state with amplitude $\sqrt{\tau}\zeta_0$ that is affected by Gaussian phase noise with standard deviation σ and a phase shift Φ_0 is given by

$$\begin{aligned}
&\int_{-\infty}^{\infty} \mathrm{d}\Phi \, \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\Phi-\Phi_0)^2}{2\sigma^2}} \left| e^{i\Phi}\sqrt{\tau}\zeta_0 \right\rangle \left\langle e^{i\Phi}\sqrt{\tau}\zeta_0 \right| \\
&= \sum_{m,n=0}^{\infty} \left\langle m \middle| \sqrt{\tau}\zeta_0 \right\rangle \left\langle \sqrt{\tau}\zeta_0 \middle| n \right\rangle e^{i(m-n)\Phi_0} e^{-(m-n)^2 \frac{\sigma^2}{2}} \left| m \right\rangle \left\langle n \right| . \qquad (5.35)
\end{aligned}$$

A direct comparison of eq. (5.33) and eq. (5.35) yields

$$\Phi_0 = 2\varkappa \left|\zeta_0\right|^2 \left(1 - \tau + \tau \log_2 \tau\right), \qquad (5.36)$$

$$\sigma^2 = 4\varkappa^2 \left|\zeta_0\right|^2 \left[2 - \tau - \tau \left(1 - \log_2 \tau\right)^2\right]. \qquad (5.37)$$

Most notably the standard deviation σ depends on the amplitude $|\zeta_0|$ of the propagating coherent state. The excess noise can therefore be effectively described as nonlinear Gaussian phase noise.

5.3.4 Nonlinear Phase Noise and Squeezing

Within an optical Kerr medium squeezed states can be generated with the Kerr effect [Rosenbluh and Shelby 1991; Schmitt et al. 1998; Werner 1998; Fiorentino et al. 2002]. In this section we analyze the impact of the previously described quantum excess noise

in form of nonlinear Gaussian phase noise on the amount of squeezing that can be generated. Note that squeezed states generated by the Master equation of eq. (5.10) are not Gaussian states like the squeezed states described in section 2.7. Gaussian squeezing is generated by a Hamiltonian depending on $\hat{a}^2$ and $\left(\hat{a}^\dagger\right)^2$, while here the Hamiltonian is proportional to $\hat{n}^2$.
We start with a phase diffused state according to eq. (5.35)

$$\hat{\varrho}^G = \int_{-\infty}^{\infty} \mathrm{d}\phi \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\phi^2}{2\sigma^2}} \left| e^{i\phi}\zeta \right\rangle \left\langle e^{i\phi}\zeta \right| \tag{5.38}$$

with the standard deviation of eq. (5.37). To capture the full evolution under lossy self-phase modulation it is necessary to include the unitary transformation done in eq. (5.11), such that

$$\hat{\rho}^G = \exp\left(i\mu z\hat{n}^2\right) \hat{\varrho}^G \exp\left(-i\mu z\hat{n}^2\right). \tag{5.39}$$

This is best solved in the interaction picture in which the propagation for annihilation operator is given by

$$\frac{\mathrm{d}\hat{a}}{\mathrm{d}z} = -i\mu\left[\hat{n}^2, \hat{a}(z)\right] = i\mu\left[2\hat{a}^\dagger(z)\,\hat{a}(z) + 1\right]\hat{a}(z). \tag{5.40}$$

Since we will be interested in the fluctuations around the average amplitude of a coherent state, it is convenient to use the substitution

$$\hat{a}(z) = e^{i\mu z\left(2\zeta^2+1\right)} e^{i\phi}\left[\zeta + \hat{b}(z)\right]. \tag{5.41}$$

The operator $\hat{b}(z)$ describes the fluctuations in a reference frame, where the average amplitude is zero, i.e. where a coherent state $\left| e^{i\phi}\zeta \right\rangle$ is mapped onto the vacuum state $\left| 0 \right\rangle$. In order to obtain quadrature variances it will be sufficient to calculate expectation values based on the operators $\hat{b}(z)$ and $\hat{b}^\dagger(z)$. Applying the substitution of eq. (5.41) to eq. (5.40) we arrive at

$$\frac{\mathrm{d}\hat{b}}{\mathrm{d}z} = 2i\mu\zeta^2\left[\hat{b}(z) + \hat{b}^\dagger(z)\right]. \tag{5.42}$$

The equation contains only linear terms in $\hat{b}(z)$ and $\hat{b}^\dagger(z)$, since they describe only small fluctuations around zero. The solution is given by

$$\hat{b}(z) = \left(1 + 2i\mu z\zeta^2\right)\hat{b}(0) + 2i\mu z\zeta^2\hat{b}^\dagger(0). \tag{5.43}$$

In order to calculate the expectation values of the variance $\Delta\hat{x}_\theta = \hat{x}_\theta - \langle \hat{x}_\theta \rangle$ for a general quadrature $\hat{x}_\theta = \hat{q}\cos\theta + \hat{p}\sin\theta$, we first need to define the quadrature operators which are

$$\hat{q} = e^{-i\mu z\left(2\zeta^2+1\right)}\hat{a}(z) + e^{i\mu z\left(2\zeta^2+1\right)}\hat{a}^\dagger(z) = e^{i\phi}\left[\zeta + \hat{b}(z)\right] + e^{-i\phi}\left[\zeta^* + \hat{b}^\dagger(z)\right] \tag{5.44}$$

and

$$\hat{p} = -ie^{-i\mu z\left(2\zeta^2+1\right)}\hat{a}\left(z\right) + ie^{i\mu z\left(2\zeta^2+1\right)}\hat{a}^{\dagger}\left(z\right) = -ie^{i\phi}\left[\zeta + \hat{b}\left(z\right)\right] + ie^{-i\phi}\left[\zeta^* + \hat{b}^{\dagger}\left(z\right)\right]. \tag{5.45}$$

Then the quantum mechanical expectation values of the quadratures are

$$\langle\, \hat{q}\,\rangle = 2\zeta\,\langle\, \cos\left(\phi\right)\rangle\,, \tag{5.46}$$

$$\langle\, \hat{p}\,\rangle = 2\zeta\,\langle\, \sin\left(\phi\right)\rangle \tag{5.47}$$

and their second moments

$$\left\langle\, \hat{q}^2\,\right\rangle = 1 + 4\zeta^2\left\langle\, \cos^2\left(\phi\right)\right\rangle - 4\mu z\zeta^2\,\langle\, \sin\left(2\phi\right)\rangle + \left(4\mu z\zeta^2\right)^2\left\langle\, \sin^2\left(\phi\right)\right\rangle, \tag{5.48}$$

$$\left\langle\, \hat{p}^2\,\right\rangle = 1 + 4\zeta^2\left\langle\, \sin^2\left(\phi\right)\right\rangle + 4\mu z\zeta^2\,\langle\, \sin\left(2\phi\right)\rangle + \left(4\mu z\zeta^2\right)^2\left\langle\, \cos^2\left(\phi\right)\right\rangle. \tag{5.49}$$

Note that the fluctuation operators $\hat{b}\left(0\right)$ and $\hat{b}^{\dagger}\left(0\right)$ are acting on the vacuum state. The expectation value of the symmetrically ordered product of the quadrature operators is equal to

$$\langle\, \hat{q}\hat{p} + \hat{p}\hat{q}\,\rangle = 4\zeta^2\,\langle\, \sin\left(2\phi\right)\rangle + 8\mu z\zeta^2\,\langle\, \cos\left(2\phi\right)\rangle - \left(4\mu z\zeta^2\right)^2\langle\, \sin\left(2\phi\right)\rangle\,. \tag{5.50}$$

The second round of expectation values over ϕ are to be taken for a Gaussian distribution $p\left(\phi\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{\phi^2}{2\sigma^2}}$ which gives

$$\begin{aligned} &\langle\, \sin\left(\phi\right)\rangle = 0, && \langle\, \sin\left(2\phi\right)\rangle = 0, \\ &\langle\, \cos\left(\phi\right)\rangle = e^{-\sigma^2/2}, && \langle\, \cos\left(2\phi\right)\rangle = e^{-2\sigma^2}, \\ &\left\langle\, \sin^2\left(\phi\right)\right\rangle = \frac{1}{2}\left(1 - e^{-2\sigma^2}\right), && \left\langle\, \cos^2\left(\phi\right)\right\rangle = \frac{1}{2}\left(1 + e^{-2\sigma^2}\right). \end{aligned} \tag{5.51}$$

The expectation values of the quadratures then read

$$\langle\, \hat{q}\,\rangle = 2\zeta e^{-\sigma^2/2}, \tag{5.52}$$

$$\langle\, \hat{p}\,\rangle = 0, \tag{5.53}$$

$$\left\langle\, \hat{q}^2\,\right\rangle = 1 + 2\zeta^2 + 2\sinh^2\left(r\right) - 2e^{-2\sigma^2}\left(\sinh^2\left(r\right) - \zeta^2\right), \tag{5.54}$$

$$\left\langle\, \hat{p}^2\,\right\rangle = 1 + 2\zeta^2 + 2\sinh^2\left(r\right) + 2e^{-2\sigma^2}\left(\sinh^2\left(r\right) - \zeta^2\right), \tag{5.55}$$

$$\langle\, \hat{q}\hat{p} + \hat{p}\hat{q}\,\rangle = 4\sinh\left(r\right)e^{-2\sigma^2} \tag{5.56}$$

with the squeezing parameter r defined by $\sinh\left(r\right) = 2\mu z\zeta^2$. Now we are finally ready to calculate the expectation values of a general quadrature $\hat{x}_\theta = \hat{q}\cos\left(\theta\right) + \hat{p}\sin\left(\theta\right)$ and its variance $\left\langle\, \left(\Delta\hat{x}_\theta\right)^2\,\right\rangle$. The first and second quadrature moments are equal to

$$\langle\, \hat{x}_\theta\,\rangle = 2\zeta e^{-\sigma^2/2}\cos\left(\theta\right), \tag{5.57}$$

$$\begin{aligned} \left\langle\, \hat{x}_\theta^2\,\right\rangle = {} & 1 + 2\zeta^2 + 2\sinh^2\left(r\right) - 2e^{-2\sigma^2}\left(\sinh^2\left(r\right) - \zeta^2\right)\cos\left(2\theta\right) \\ & + 2\sinh\left(r\right)e^{-2\sigma^2}\sin\left(2\theta\right) \end{aligned} \tag{5.58}$$

and the variance follows directly from $\left\langle \left(\Delta\hat{x}_\theta\right)^2 \right\rangle = \left\langle \hat{x}_\theta^2 \right\rangle - \left\langle \hat{x}_\theta \right\rangle^2$ as

$$\begin{aligned}\left\langle \left(\Delta\hat{x}_\theta\right)^2 \right\rangle = 1 + 2\zeta^2\left(1 - e^{-\sigma^2}\right) + 2\sinh^2(r) \\ - 2e^{-2\sigma^2}\left\{\cos(2\theta)\left[\sinh^2(r) + \zeta^2\left(e^{\sigma^2} - 1\right)\right] - \sin(2\theta)\sinh(r)\right\}.\end{aligned} \tag{5.59}$$

In the case without phase noise $\sigma = 0$ the quadrature with minimal variance is given by $\tan(\theta_0) = -e^{-r}$ with the corresponding variance $\left\langle \left(\Delta\hat{x}_{\theta_0}\right)^2 \right\rangle = e^{2r}$. In the presence of phase noise the minimal variance is reached for the angle θ_{opt} defined through

$$\tan(2\theta_{\text{opt}}) = -\frac{\sinh(r)}{\sinh^2(r) + \zeta^2\left(e^{\sigma^2 - 1}\right)} \tag{5.60}$$

and reads

$$\begin{aligned}\left\langle \left(\Delta\hat{x}_{\theta_{\text{opt}}}\right)^2 \right\rangle = 1 + 2\zeta^2\left(1 - e^{-\sigma^2}\right) + 2\sinh^2(r) \\ - 2e^{-2\sigma^2}\sqrt{\sinh^2(r) + \left[\sinh^2(r) + \zeta^2\left(e^{\sigma^2} - 1\right)\right]^2}.\end{aligned} \tag{5.61}$$

The standard deviation σ is given by eq. (5.37) and the amplitude by $\zeta = \sqrt{\tau}\zeta_0$. To derive a limit on the attainable squeezing we expand eq. (5.59) up to the first order in σ^2. The main contribution is then given by

$$\left\langle \left(\Delta\hat{x}_\theta\right)^2 \right\rangle = \left\langle \left(\Delta\hat{x}_{\theta_0}\right)^2 \right\rangle + 4\zeta^2\sigma^2\sin^2(\theta_0) \tag{5.62}$$

$$= \left\langle \left(\Delta\hat{x}_{\theta_0}\right)^2 \right\rangle + 4g(\tau)\sinh^2(r)\sin^2(\theta_0) \tag{5.63}$$

$$= \left\langle \left(\Delta\hat{x}_{\theta_0}\right)^2 \right\rangle + 4g(\tau)\left(1 - e^{-2r}\right)\tanh(r), \tag{5.64}$$

where in the second line we used eq. (5.37) and $\sinh(r) = 2\mu z\zeta^2$ with

$$g(\tau) = \frac{2 - \tau - \tau\left(1 - \log_2(\tau)\right)^2}{\tau^2\log_2^2(\tau)} \approx \frac{1 - \tau}{3}. \tag{5.65}$$

The approximation of $g(\tau)$ is valid for low attenuation $\tau \approx 1$. Since for strong squeezing $r \ll 1$, we have $\left(1 - e^{-2r}\right)\tanh(r) \approx 1$ and therefore $\frac{1-\tau}{3}$ defines the maximal attainable amount of squeezing possible. Compared to the case without nonlinear phase noise the difference is the factor $1/3$. In fig. 5.4 the variance of eq. (5.64) and the approximation of eq. (5.65) are shown for several values of τ.

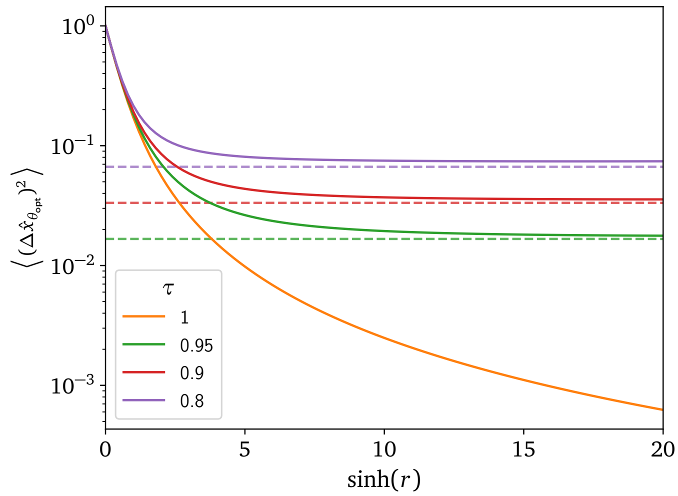


Figure 5.4: Variance of the optimal quadrature $\left\langle \left(\Delta\hat{x}_{\theta_{\rm opt}}\right)^2 \right\rangle$ according to eq. (5.61) as a function of the nonlinear strength $\varkappa$ for different values of the power transmission τ. For convenience the x-axis is parametrized with the squeezing parameter $\sinh(r) = 2\mu z\tau\zeta_0^2$. The input photon number is $\zeta_0^2 = 1 \times 10^8$. The dashed grey lines show the corresponding approximate values $\frac{1-\tau}{3}$ of the variance for large squeezing.

6 Noisy Communication via Optical Channels

In this chapter several communication scenarios will be discussed based on the noise models developed in chapter 5. We consider only alphabets that consist of coherent states. The obtained output distributions depend on the particular type of channel as well as the type of utilized receiver. The goal is to analyze the impact of the channel on the information transfer and to identify the optimal strategy for each particular scenario. First, a thermal noise channel will be investigated in section 6.1. Section 6.2 is dedicated to the effect of transmitting information encoded in the phase variable of coherent state via a channel that is subject to losses and a Kerr nonlinearity. Finally, in section 6.3 strategies for communication under constraints on average and maximum energy in the presence of phase noise, both linear and nonlinear, are discussed.

6.1 Thermal Noise

The impact of a thermal channel on the transmission of coherent states is a thermalization of the states. In the following, we will consider the influence of thermal noise on the performance of various receivers based on the model discussed in section 5.1 with a focus on the displacement receiver and deduce strategies to improve their performance. In particular, we will consider communication scenarios with binary coherent alphabets as well as PIE in the low energy limit.

Binary coherent alphabets

Let us assume we have on average $\bar{n}$ photons available to distribute between two equiprobable coherent states placed on the real axis x in phase space. After transmission through a thermal channel an average number of n_b thermal photons is added and the state at the channel output is given by eq. (2.97).
If a homodyne receiver is used for detection the output distribution is given by

$$p\left(x|\zeta\right) = \frac{1}{\sqrt{\pi\left(2n_b+1\right)}} e^{-\frac{\left(x-\sqrt{2}\zeta\right)^2}{2n_b+1}}. \tag{6.1}$$

The photon count statistics which can be obtained by a photon counting detector are directly defined by eq. (2.97) with $m = n$ and read

$$p\left(n|\zeta, n_b\right) = \frac{n_b^n}{\left(1+n_b\right)^{n+1}} e^{-\frac{|\zeta|^2}{1+n_b}} L_n\left(-\frac{|\zeta|^2}{n_b\left(1+n_b\right)}\right). \tag{6.2}$$

The statistics obtained by a displacement receiver are also given by eq. (6.2) with the substitution $\zeta \to \zeta + \beta$.
Based on these probability distributions the error probability and mutual information can be calculated. For all receivers the complex amplitudes of both coherent states are optimized in order to determine the optimal alphabet. Additionally, for the displacement receiver the displacement β has to be optimized. In fig. 6.1 the results for both error probability p_{err} and mutual information I are shown as a function of the average number of thermal photons n_b for the example of $\bar{n} = 2$.
The optimal alphabet when using a simple SPD remains OOK as was expected. Also BPSK remains the optimal binary alphabet for both, homodyne and displacement receiver. Since thermal noise affects all states equally strong, regardless of their position in phase space, there is no benefit in changing the alphabet.
While a homodyne receiver is fairly robust to thermal noise, the displacement receiver is highly susceptible. Homodyne detection only starts to deteriorate around $n_b \approx 0.1$, whereas the displacement receiver already degrades for very small values of n_b. The PNR necessary to accommodate decent performance rises with the number of thermal background photons. Without PNR, i.e. PNR $= 1$, homodyne detection surpasses the error probability of the displacement receiver already for $n_b \approx 0.004$, while for higher PNR the displacement receiver remains superior until the thermal background reaches $n_b \gtrsim 0.1$. This results is in agreement with [Yuan et al. 2020]. For instance, the displacement receiver remains the preferable choice until $n_b \approx 0.2$ with PNR $= 5$. Interestingly, for a hard decision rule the point where homodyne detection becomes superior does not depend on the mean photon number $\bar{n}$ as can be seen in fig. 6.1(e).
However, for soft decision decoding homodyne detection surpasses the mutual information of a displacement receiver earlier. In this case, the turning point does depend on the mean photon number $\bar{n}$.
In summary, the displacement receiver is highly susceptible to thermal noise. With PNR detection the receiver becomes much more robust and remains able to perform sub-shot-noise detection in a larger regime depending on whether a hard or soft decision rule is employed. While a soft decision provides a better information transfer in general, it is also more susceptible to thermal noise.

PPM in the low energy limit

In the regime where the available energy is very limited $\bar{n} \ll 1$, it becomes beneficial to use PPM to get closer to the Holevo limit as shown in fig. 3.14. Its success is based on the idea to concentrate the available energy into a single time bin and leave all other slots empty. This way the information can be easily retrieved by an SPD. If thermal noise is present, however, also the empty slots contain photons and can trigger a detector click. These false clicks severely deteriorate the attainable PIE with rising number of thermal photons n_b .
In principle, also receivers other than SPDs can be used to detect PPM. While in the noiseless scenario they are able to achieve the same PIE as an SPD at best, they become beneficial when thermal noise is present. The PIE for PPM of order M can be

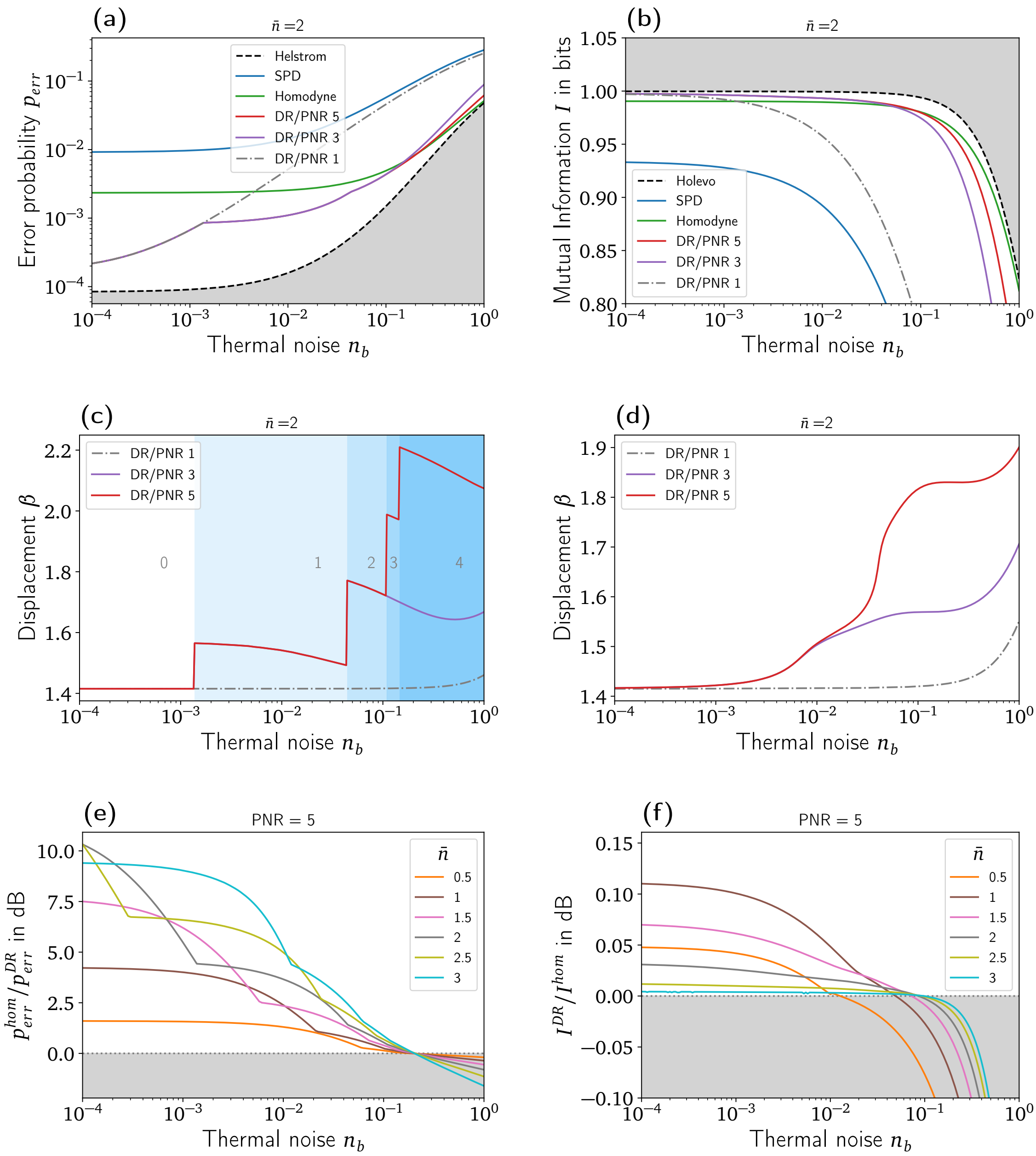


Figure 6.1: **(a)** Error probability and **(b)** soft decision mutual information for different receivers discriminating a binary coherent alphabet as a function of the average number of thermal photons n_b per signal. The optimal displacement for the displacement receiver is shown in **(c)** and **(d)** for error probability and mutual information respectively. The grey numbers denote the optimal decision threshold for the displacement receiver in the marked area. The ratio between **(e)** the error probability or **(f)** mutual information of homodyne and displacement receivers are shown for PNR = 5. Positive values of the ratio in dB signify a superior performance of the displacement receiver.

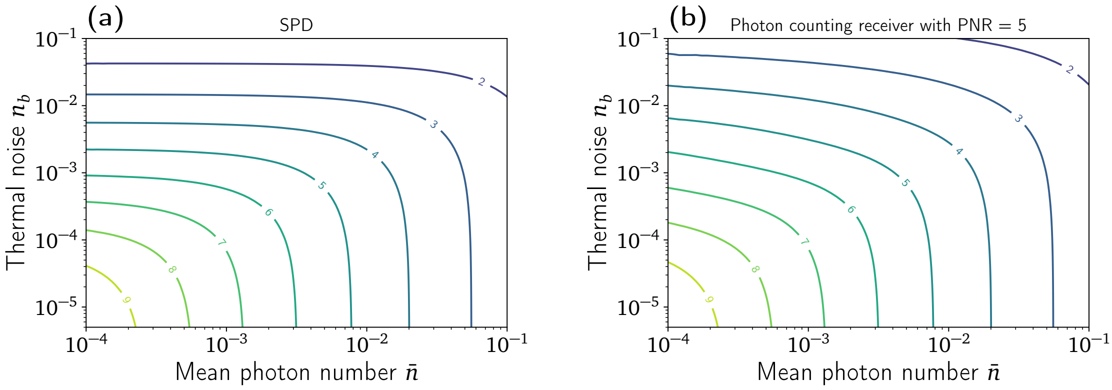


Figure 6.2: PIE for PPM detected by an **(a)** SPD and **(b)** photon counting detector with PNR = 5 as a function of the average number of thermal photons n_b and the mean photon number $\bar{n}$.

determined based on eq. (3.46) with the probabilities given by eqs. (6.1) and (6.2) for an OOK alphabet. The a-priori probabilities are given by $p_0 = \frac{M-1}{M}$ for the thermal state in the empty slots and $p_1 = \frac{1}{M}$ for the coherent signal state with amplitude $\zeta = \sqrt{M\bar{n}}$. In order to find the optimal strategy for each type of detector the PPM order M, and additionally for the displacement receiver the displacement β, have to be optimized.

Let us first investigate whether a detector with photon counting capability can yield an improvement over a simple SPD. In fig. 6.2(a) the PIE of an SPD is shown. There are two regions with different behaviors. In the upper left region where $\bar{n}/n_b \lesssim 10$, the PIE is limited by the amount of thermal noise. A change in the mean photon number $\bar{n}$ does not change the PIE. Contrary, in the lower right region with $\bar{n}/n_b \gtrsim 10$, the PIE is only limited by the energy $\bar{n}$. Here, reducing the thermal noise does not yield an improvement.

If a photon counter is used instead of an SPD the situation changes. In fig. 6.2(b) the PIE of a photon counter with PNR = 5 is depicted. Now the PIE does scale with the mean photon number, even though in the region $\bar{n}/n_b \lesssim 10$ this effect is much less pronounced.

Apart from photon counting detector also a homodyne or displacement receiver can be used for detection. In fig. 6.3 the results are shown as a function of n_b for the example of $\bar{n} = 10^{-3}$. For not too large thermal noise photon counting detectors constitute the optimal measurement. Displacement receivers, which differ only by the displacement operation from photon counters, offer little to no improvement. However, with rising n_b the displacement receiver offers a significant advantage over a simple photon counting detector. For overwhelming thermal noise a large PNR becomes necessary to retain a decent PIE. In this regime the homodyne detector starts to offer a similar PIE to a displacement receiver and both approach the Holevo limit.

Overall, photon counting offers an increased robustness against thermal noise and improves the scaling with the mean photon number $\bar{n}$. An added displacement operation

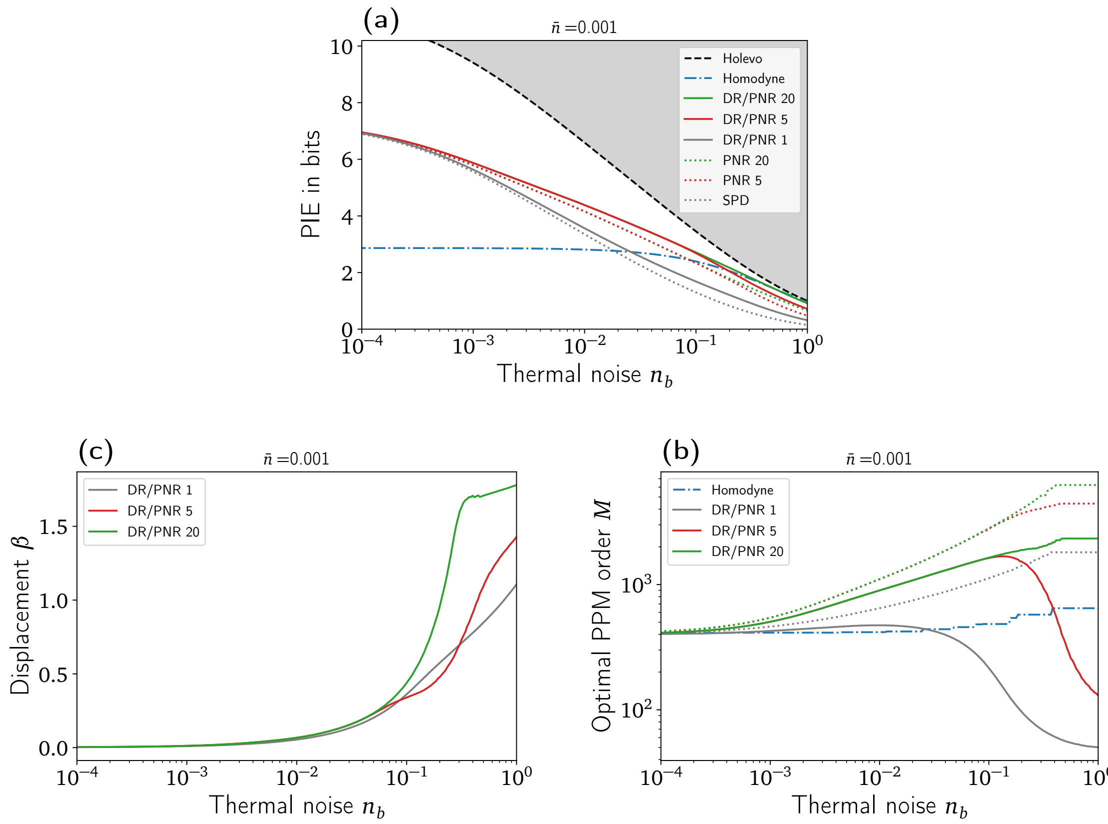


Figure 6.3: **(a)** Mutual information for different receivers detecting PPM for $\bar{n} = 10^{-3}$ as a function of the average number of thermal photons n_b. The Holevo limit is given by eq. (3.95). Dotted lines show a photon counting detector without displacement with a PNR of the respective color. The optimal PPM order M and optimal displacement for the displacement receivers are shown in **(b)** and **(c)** respectively.

increases this effect. For all types of detectors the optimal PPM order M increases with the strength of thermal noise. By increasing M the signal amplitude increases as well, since it becomes more important to distinguish the signal from the underlying thermal background. Note that in the limit $\bar{n} \to 0$ the PIE for PPM detected by a photon counter approaches the Holevo limit [Jarzyna 2019].

6.2 Kerr Nonlinearity

In this section we want to analyze the limits that a nonlinearity imposes on encoding information in the phase variable. As a test alphabet we consider coherent states with equal energy, which defines a ring in phase space on which the states are positioned. The location on this ring is controlled solely by the phase. Such an alphabet consists of coherent states $\left| \zeta_0 e^{-i\phi_i} \right\rangle$ with the same absolute value of the amplitude $|\zeta_i| = \zeta_0$ but

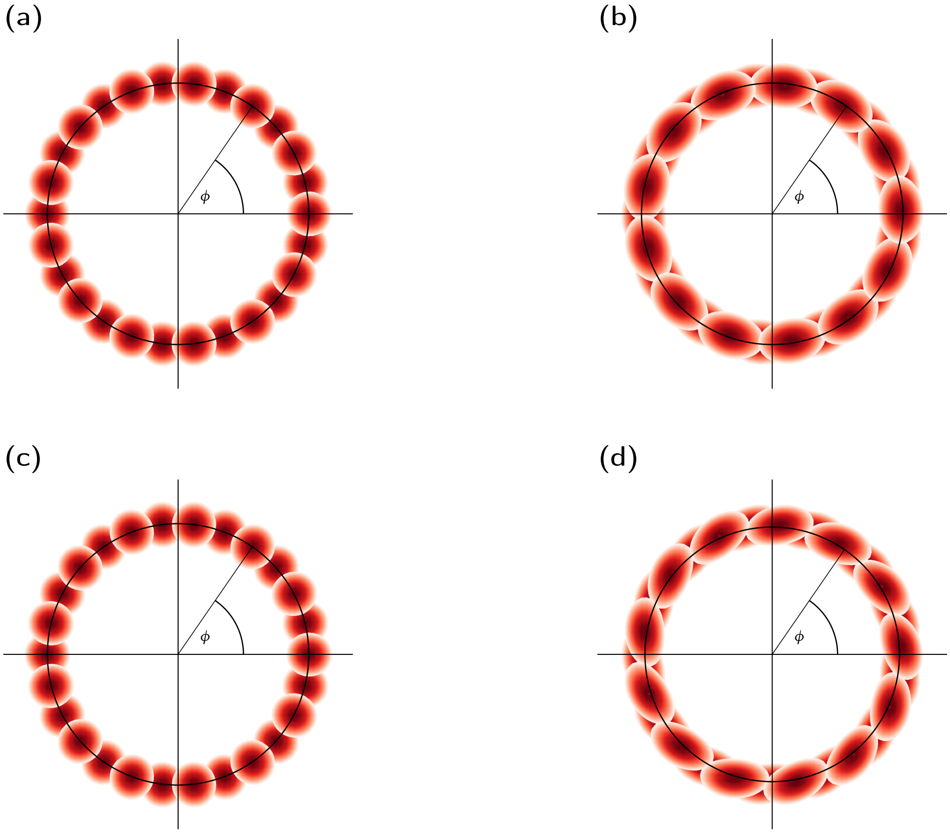


Figure 6.4: A sketch of the Husimi-Q-function of an evenly distributed alphabet with $p_\phi = \frac{1}{2\pi}$ for $\tau\bar{n} = 9$ after propagation via a channel with transmission $\tau = 0.631$ and a nonlinear strength **(a,c)** $\varkappa = 0$ and **(b,d)** $\varkappa = 0.25$. The output of initially coherent states is depicted in the top row **(a,b)**, whereas the output of squeezed states is shown in the bottom row **(c,d)**. The chosen squeezing parameter is **(c)** $z = 0.06e^{\pi}$ and **(d)** $z = 0.64e^{0.2\pi}$.

different phases ϕ_i. Let us assume they are distributed continuously over the range of 2π such that their a-priori probability is uniform $p_\phi = \frac{1}{2\pi}$ [Kahn and Ho 2004]. Furthermore, the alphabet satisfies a constraint on the mean photon number $\bar{n} = |\zeta_0|^2$. The continuous spacing ensures that changes to the shape of a state due to phase noise immediately affect the distinguishability. A sketch of this alphabet in phase space is depicted in fig. 6.4. Note that this alphabet is not necessarily optimal and does not provide us with an ultimate limit on the transmittable information as one can always add amplitude encoding.

The more states on the ring are distinguishable the more information we can retrieve in principal. If the available energy $\bar{n}$ increases the ring widens and more states become distinguishable, since each state still occupies the same amount of space in phase space. A nonlinear lossy channel as discussed in section 5.3, however, changes this picture.

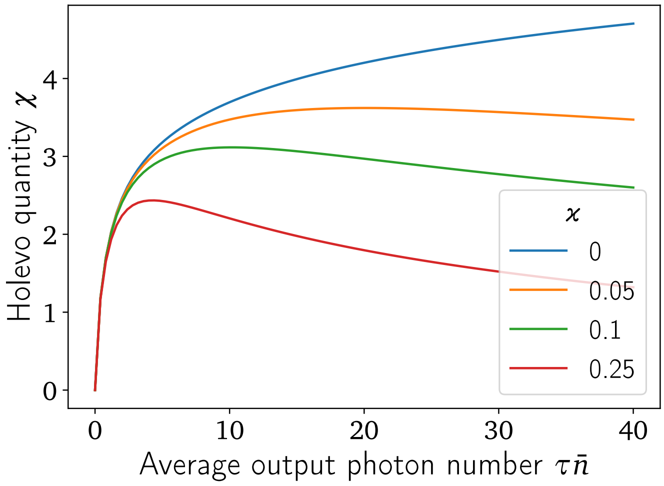


Figure 6.5: The Holevo quantity χ as a function of the mean photon number at the output $\tau\bar{n}$. It is assumed that the alphabet consists of coherent states where the information is continuously encoded in the phase variable with a uniform a-priori probability distribution $p_\phi = \frac{1}{2\pi}$. The states are transmitted via a channel with power transmission $\tau = 0.631$ and several values of the nonlinear strength $\varkappa$.

States become distorted and less information is transmittable. We use the Holevo quantity χ defined in eq. (3.89) as a figure of merit, since it assumes the best possible (but with unknown practical realization) measurement for the task. Therefore, a reduction in the Holevo quantity translates into a reduced amount of information which is transmitted via the phase variable ϕ. Since we are considering the best possible measurement an improvement can only be achieved by increasing the distinguishability of states. However, because of the considered continuous encoding the states densely fill the ring in phase space and no improvement is to be expected by altering the type of phase encoding such that we can consider the loss in Holevo information as fundamental.
In fig. 6.5 the Holevo quantity as a function of the average mean photon number after losses $\tau\bar{n}$ is shown. Without nonlinearity the amount of information one can transmit rises continuously with the available energy. Contrary, if a nonlinearity is present the Holevo information exhibits a distinct peak at low or moderate mean photon numbers. Using more energy for communication now actually lowers the amount of transmitted information. An intuitive understanding of this fact can be obtained by looking at the phase space representation of the output states. When increasing the energy the ring in phase space widens and more states become distinguishable. However, the larger the energy of the state the more it becomes distorted by the nonlinearity. At some value of the mean photon number $\bar{n}$, depending on $\varkappa$, the latter effect overtakes the advantage of widening the ring and the Holevo information χ starts to decrease. This effect is not reversible as it is caused by a non-unitary propagation and is of fundamental nature.

Using squeezed states instead of coherent ones can increase the Holevo information even for $\varkappa = 0$, since more states can be distinguished. Furthermore, squeezing can help to mitigate the effects of the nonlinearity. The required amount of squeezing depends on the particular parameters of the channel and finding the optimal value is non-trivial. Moreover, squeezing requires energy which shrinks the ring in phase space when compared to the case of coherent states and lowers the distinguishability of the states.
Let us assume we want to communicate over a channel with transmission $\tau = 0.631$ and a mean photon number at the output of $\tau\bar{n} = 9$. In fig. 6.6 the Holevo quantity is shown as a function of the squeezing direction ϕ_z and strength r for several values of the nonlinear strength $\varkappa$. Due to numerical limitations we consider an alphabet of 26 coherent states with equal amplitudes ζ_i equally distributed in phase with the prior probability distribution $p_\phi = \frac{1}{26}$. In the case that no nonlinearity is present with $\varkappa = 0$, the optimum choice is to squeeze with $z \approx 0.06e^{i\pi}$. In the presence of a nonlinearity the optimal squeezing parameter changes and becomes $z \approx 0.46e^{i0.36\pi}$ for $\varkappa = 0.1$, while for $\varkappa = 0.25$ it is given by $z \approx 0.64e^{i0.2\pi}$. These values were obtained from a numerical analysis and might contain a small inaccuracy due to the precision of the calculations. The corresponding alphabets for $\varkappa = 0$ and $\varkappa = 0.25$ after transmission are depicted in fig. 6.4. As can be seen the squeezing is chosen such that the overlap between neighboring states is minimized.
While squeezed states improve the information transmission rate, the task to find the optimal alphabet for a nonlinear phase noise channel remains unclear. However, the presented and previously known results give some hints. In the regime of large output powers $\tau\bar{n} \gg 1$ the optimal alphabet for a loss only ($\varkappa = 0$) channel is given by a Gaussian ensemble of coherent states [Giovannetti et al. 2004] and in this limit the capacity can be approximated by $C \approx \log_2(\tau\bar{n})$. Here, information is encoded equally into the phase and amplitude. With increasing phase noise strength the amount of information that can be encoded into the phase variable decreases, until the phase can no longer carry information. In this regime the states arrive completely dephased at the channel output and are diagonal in the Fock basis. Of course, it is still possible to encode information in the amplitude, which amounts to exactly half of the capacity of a pure loss channel, i.e. it is given by $C \approx \frac{1}{2}\log_2(\tau\bar{n})$ [Martinez 2007].
It can be expected that in the regime between no phase noise and complete dephasing the optimal encoding consists of two parts. On the one hand, encoding information into the amplitude remains unchanged as it is unaffected by phase noise. Phase encoding, on the other hand, is severely impeded and utilizes squeezed states as depicted in fig. 6.4. For a given energy the squeezed states are placed on a "ring" in phase space in such a way that minimizes the impact of the noise. The optimal amount and direction of the squeezing can be determined depending on the strength of the phase noise $\varkappa$ and the particular energy as shown in fig. 6.6.

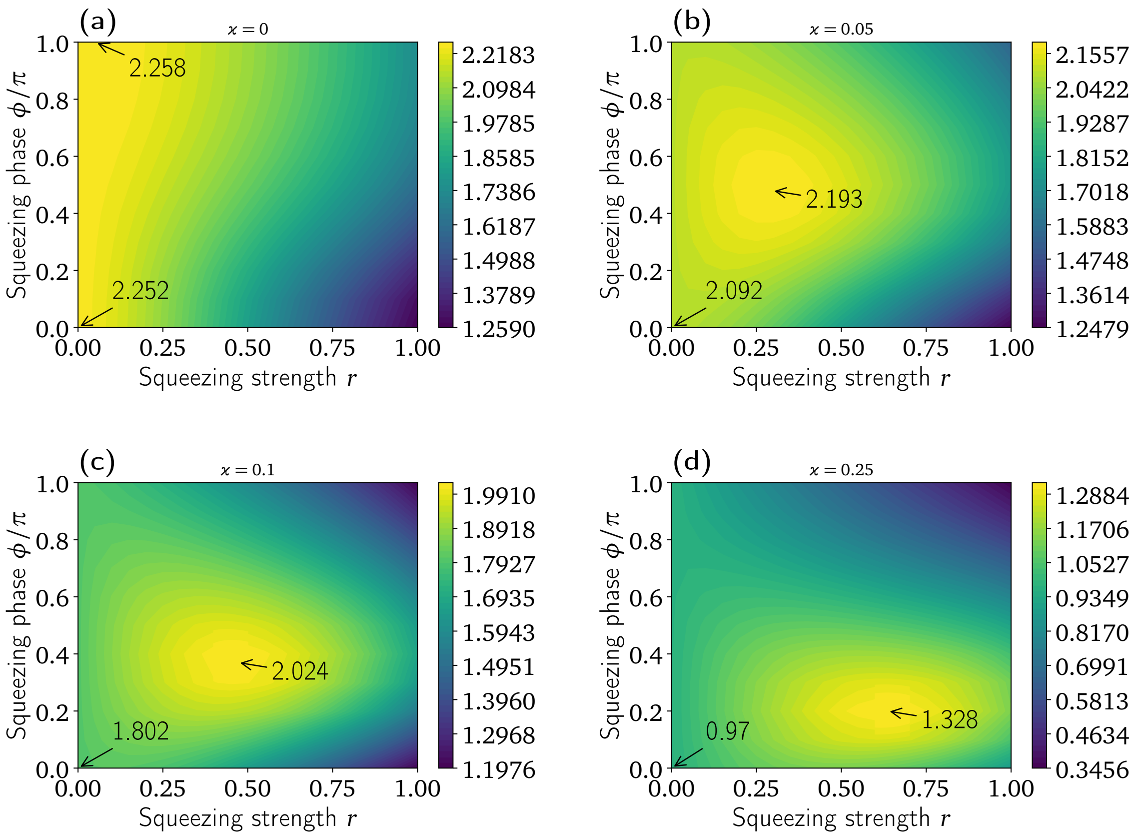


Figure 6.6: The Holevo quantity χ as a function of the squeezing direction ϕ_z and strength r for a mean photon number at the output of $\tau\bar{n} = 9$. It is assumed that the alphabet consists of 26 coherent states where the information is encoded in the phase variable with a uniform a-priori probability distribution $p_\phi = \frac{1}{26}$. The states are transmitted via a channel with power transmission $\tau = 0.631$ and a nonlinear strength of **(a)** $\varkappa = 0$, **(b)** $\varkappa = 0.05$,**(c)** $\varkappa = 0.1$ and **(c)** $\varkappa = 0.25$. The arrows show the Holevo quantity for no squeezing and the optimal squeezing respectively.

6.3 Phase Noise

Phase noise, whether it is linear or nonlinear, has a striking effect on communication. So far we discussed its influence on the information carrier itself assuming the optimal measurement. But carrying out the optimal measurement is a highly non-trivial task as in most cases for optical communication a physical realization is not even known. In cases where it is known, such as the Dolinar receiver for noiseless BPSK, it is a technically challenging task. Therefore, we will investigate how commonly used, but suboptimal, receivers perform under phase noise. Instead of fixing the alphabet to the optimal choice of the noiseless scenario, any binary coherent alphabet is allowed and the state amplitudes are optimized under the given constraint on the available energy. We will show that this optimization can lead to a significant improvement and robustness against phase noise for the displacement receiver and extends its sub-shot-noise detection capability.

In the following sections, we consider two different constraints on the available energy. First, a constraint on the mean photon number $\bar{n}$ is assumed and in the second part only the maximum energy to prepare a single state is limited.

Constraint on mean energy

In this first part we consider a constraint on the mean energy of the alphabet, where we are able to distribute the available energy within a limited time frame. Let the alphabet consist of two coherent states with complex amplitudes ζ_0 and ζ_1 and a-priori probabilities $p_0 = p_1 = \frac{1}{2}$. For simplicity we assume that the states are located on the real axis x in phase space. Choosing any other quadrature will lead to the same results as one can always rotate phase space accordingly. The mean energy used in the protocol is given by

$$\bar{n} = \frac{1}{2}\left(|\zeta_0|^2 + |\zeta_1|^2\right). \tag{6.3}$$

The states $|\,\zeta_i\,\rangle$ are subject to either linear or nonlinear phase noise with standard deviation σ. Nonlinear phase noise, as discussed in section 5.3, is characterized by a standard deviation σ that depends on the energy of the affected state $|\zeta_i|^2$, while for linear phase noise it does not. Note that the energy taken into account for nonlinear noise is not the mean energy $\bar{n}$ of the alphabet, but the energy $|\zeta_i|^2$ of the particular state.

Not every binary alphabet is affected by phase noise in the same way. For instance, the vacuum state is not affected rendering OOK more robust to noise. However, the overlap between states for OOK is inherently larger than for other binary alphabets. Consequently, simply using the most robust alphabet is not the best choice. For each receiver we will optimize the state amplitudes according to the used figure of merit. This ensures that the obtained performance is indeed the best possible.

Conceptually, the simplest way to detect a binary alphabet is to use an SPD. As it is only able to detect amplitude, but not phase, OOK is the best distinguishable alphabet and

its performance is naturally unaffected by phase noise. Consequently, error probability and mutual information are simply given by eqs. (3.74) and (3.112) for all types and values of σ. The OOK/SPD strategy provides an upper bound on the error probability and a lower bound on the mutual information as it is in general suboptimal due to the smaller state separation. It can be seen as the safe way of communicating under phase noise. Only when states are completely diffused it becomes optimal, since in this regime information can only be encoded in the intensity of light.
The homodyne receiver does not suffer from the same limitations, since it can resolve any binary alphabet. However, it is susceptible to phase noise. The probability distribution provided by the homodyne receiver under phase noise is given by

$$p\left(x|\zeta,\sigma\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{d}\phi\; e^{-\frac{\phi^2}{2\sigma^2}}p\left(x|\zeta e^{i\phi}\right) \tag{6.4}$$

with the homodyne probability distribution $p\left(x,\zeta\right)$ defined in eq. (4.9). Note that for nonlinear phase noise the standard deviation σ depends on $\left|\zeta\right|^2$. The error probability and mutual information are determined by eqs. (4.13) and (4.16) together with the phase diffused probability distribution of eq. (6.4).
The last type of receiver we will discuss is the displacement receiver. By adjusting the displacement to the utilized alphabet, the displacement receiver can distinguish the states of every binary alphabet. If phase noise is present the probability distribution is given by

$$p\left(n|\zeta+\beta,\sigma\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\mathrm{d}\phi\; e^{-\frac{\phi^2}{2\sigma^2}}e^{-\left|\zeta e^{i\phi}+\beta\right|^2}\frac{\left|\zeta e^{i\phi}+\beta\right|^{2n}}{n!}. \tag{6.5}$$

Again note that the standard deviation σ depends on $\left|\zeta\right|^2$ for nonlinear phase noise. Based on eq. (6.5) and eqs. (3.44) and (4.22) we can calculate the error probability p_{err} and mutual information I.
In order to maximize the receivers performance we distribute the available energy $\bar{n}$ between the two states and, if possible and needed, adjust the receivers parameters to match the used alphabet. This optimization is done numerically based on either error probability or mutual information as a figure of merit.
In fig. 6.7 the results of the optimization for the error probability are shown for the mean photon number $\bar{n}=2$. A SPD has a constant error probability and the optimal alphabet is always OOK. The performance of a homodyne receiver constantly declines with rising noise strength σ and BPSK remains the optimal alphabet even under phase noise. In contrast, the displacement receiver utilizes non-standard alphabets and both, the robustness against phase noise as well as the performance, are increased drastically by adjusting the amplitudes together with the displacement. Only for an intermediate regime of the standard deviation a displacement receiver performs worse than the more conventional receivers. If the PNR is increased, however, the displacement receiver approaches the performance of the homodyne receiver. For large standard deviations the displacement receiver becomes the optimal Helstrom measurement. Note that the state amplitudes have been also optimized for calculating the Helstrom minimum error bound.

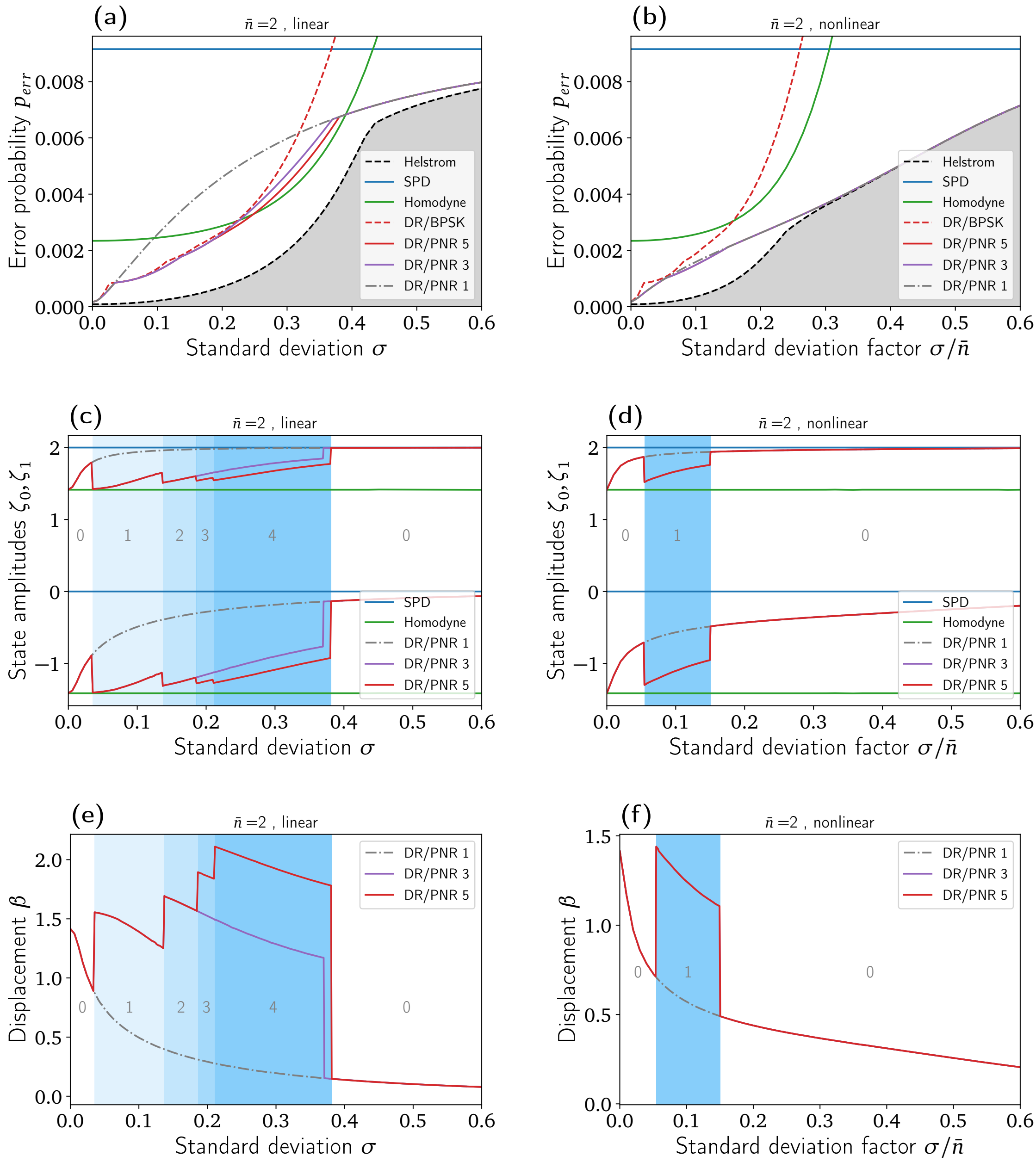


Figure 6.7: On **top (a,b)** the error probability p_{err}, in the **middle (c,d)** the state amplitudes and on the **bottom (e,f)** the displacement introduced by the detector for a mean photon number $\bar{n} = 2$ as a function of the phase noise strength is shown. The phase noise is assumed to be linear on the **left (a,c,e)** and non-linear on the **right (b,d,f)**. The grey numbers denote the optimal decision threshold k_n in the marked area for the displacement receiver. If the PNR of the receiver is not high enough the next possible lower decision threshold should be chosen.

Optimizing both, the state amplitudes and the displacement, is crucial. If BPSK is used and only the displacement is optimized, the performance of the displacement receiver declines fast, even with PNR capability. The optimal alphabet for the displacement receiver starts at BPSK for $\sigma = 0$ and gets closer to OOK with rising phase noise strength. The corresponding optimal displacement β depends also on the phase noise and the given PNR of the receiver. The sudden jumps in the displacement and state amplitudes are not an artifact of the optimization, but occur due to the discreteness of the number of photons. At these particular values of σ it becomes beneficial to change the decision threshold of the displacement receiver.

The behavior of the displacement receiver can be understood intuitively by looking at the advantages of different possible alphabets. In general, the largest possible state separation is desired as it provides the best distinguishability. Counteracting that, phase noise distorts the shape of the state in phase space and renders the states less distinguishable. OOK, the alphabet with the least state separation, is immune to phase noise. The displacement receiver is able to simultaneously utilize the superior state separation of BPSK and the robustness of OOK by using a non-standard alphabet with intermediate amplitudes. The stronger the phase noise the closer the alphabet is to OOK as more robustness is wanted. This flexible utilization of different alphabets is only possible, since the displacement receiver can be adjusted to fit the particular amplitudes by changing the displacement. Other receivers do not possess this flexibility.

The results for mutual information as a figure of merit with $\bar{n} = 2$ are shown in fig. 6.8. While qualitatively the results stay the same, there are some quantitative differences. The homodyne receiver performs considerably better when paired with a soft decision than a hard decision strategy. In an intermediate regime of phase noise it outperforms the displacement receiver. However, the latter not only approaches the performance of the homodyne receiver with increasing PNR, but surpasses it for weak or strong phase noise. For very large values of the standard deviation the displacement receiver approaches the Holevo bound and can be considered the optimal measurement.

While we now understand the behavior and advantage of the displacement receiver, the question remains in which regime this advantage is present and whether PNR can improve the performance. In fig. 6.9 the error probability ratio $p_{\text{err}}^{\text{conv}}/p_{\text{err}}^{DR}$ between conventional receivers and the displacement receiver is depicted in dB. The conversion to dB is given by $10\log_{10}(x)$. The conventional receiver is chosen to be either the SPD or homodyne detector, depending on which performs better. For linear phase noise there exists a regime for intermediate phase noise strength σ and not too small mean energies $\bar{n}$ where conventional detection performs better. With PNR this region can be made considerably smaller and the performance of the displacement receiver improved. In the case of nonlinear phase noise the displacement receiver always performs better than conventional methods. PNR can still be used for larger mean photon numbers to reduce the errors.

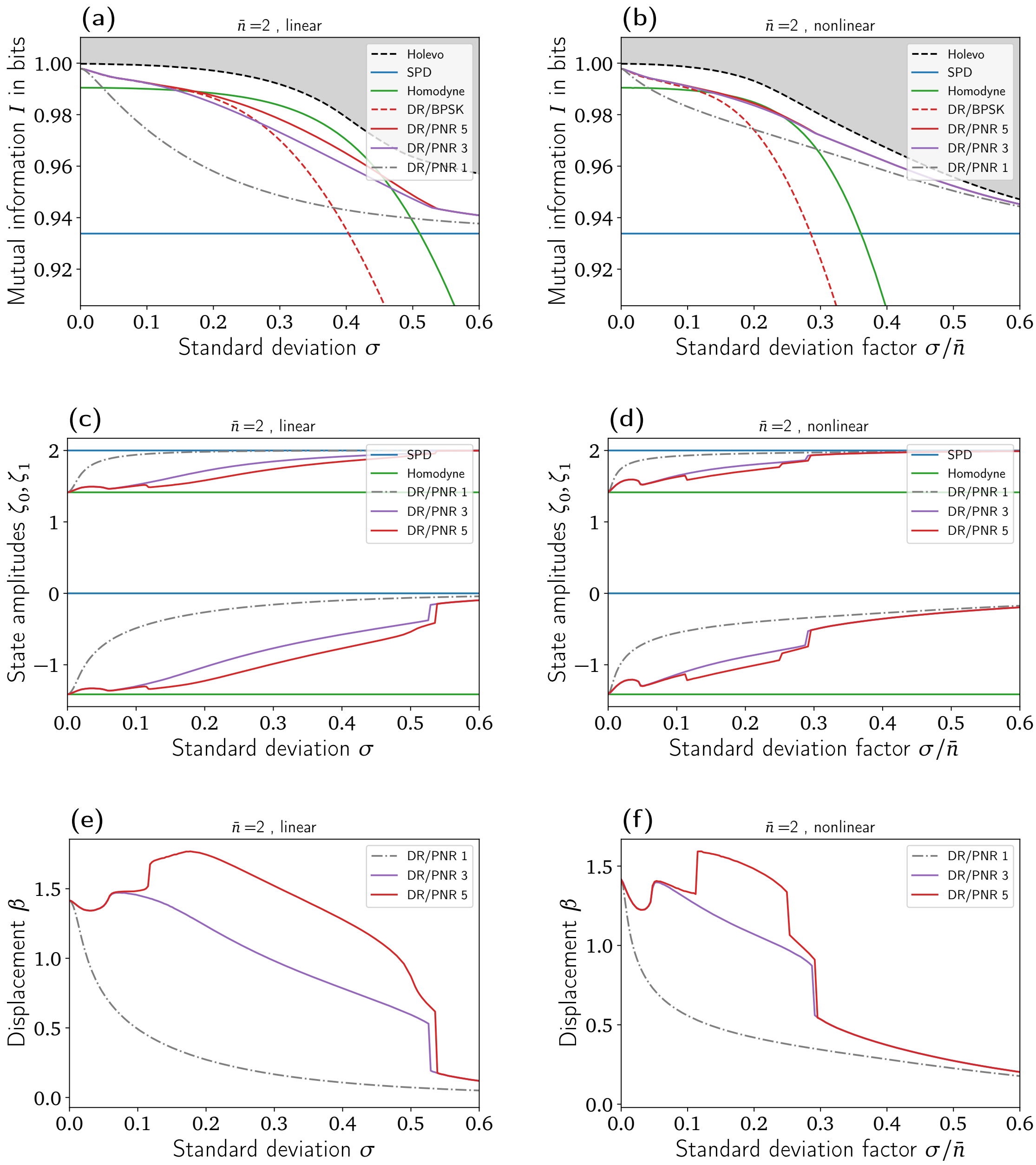


Figure 6.8: On **top (a,b)** mutual information I, in the **middle (c,d)** the state amplitudes and at the **bottom (e,f)** the displacement introduced by the detector for a mean photon number $\bar{n} = 2$ as a function of the phase noise strength is shown. The phase noise is assumed to be linear on the **left (a,c,e)** and nonlinear on the **right (b,d,f)**.

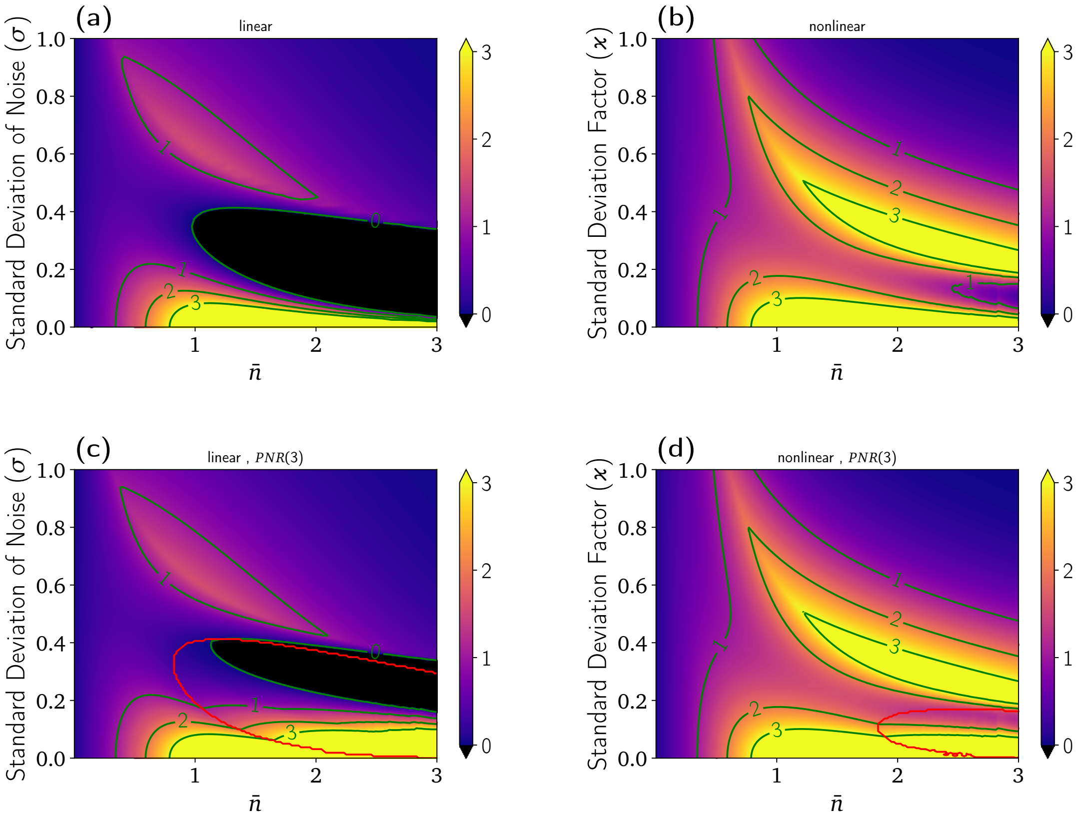


Figure 6.9: The ratio $p_{\text{err}}^{\text{conv}}/p_{\text{err}}^{DR}$ in dB as a function of the mean photon number $\bar{n}$ and the phase noise strength for linear on the **left (a,c)** and nonlinear phase noise on the **right (b,d)**. The displacement receiver is assumed to have a PNR of 1 on **top (a,b)** and of 3 at the **bottom (c,d)**. Larger values of the ratio correspond to a larger advantage of the displacement receiver over conventional methods and 0 marks equal performance. The error probability $p_{\text{err}}^{\text{conv}}$ for conventional detection is taken from either the SPD or homodyne receiver, depending on which performs better. Black areas mark the region where conventional methods outperform the displacement receiver. Red contour lines show the area, where $PNR > 1$ is necessary. Discrete jumps in the contour lines occur due to a finite resolution. They are expected to be smooth for an unlimited resolution.

Constraint on maximum energy

In this section we discuss optical communication with binary coherent alphabets under a limit on the maximum available energy to prepare a signal state. This directly translates into a constraint on the maximum amplitude of either of the states, such that $|\zeta_i|^2 < n_{\max}$. Unlike for a constraint on the mean energy, we are not able to store unused energy and transfer it to the other state.

The error probability and mutual information offered by the receivers can still be calculated based on the probability distributions presented in previous sections, i.e. eqs. (4.1), (6.4) and (6.5). However, the optimization now has to be done with the new constraint and an additional parameter needs to be optimized as well, since the amplitude of the second state is not directly dictated by the constraint.

In fig. 6.10 the error probability for $n_{\max} = 2$ is shown alongside the corresponding state amplitudes ζ_i and displacements β. Homodyne detection performs very close to the Helstrom bound for not too strong linear and nonlinear phase noise. For strong nonlinear phase noise, it is beneficial to lower the energy of the states while retaining the BPSK encoding.

The displacement receiver performs well for low and very strong phase noise. Again it reaches the Helstrom bound in the latter regime. However, the performance of the displacement receiver is much more reliant on high PNR than for a mean photon number constraint. The optimal alphabet shifts towards OOK with rising phase noise strength. This effect is more pronounced for nonlinear noise.

In fig. 6.11 the results for mutual information are presented. Unlike for the mean photon number constraint, soft decision homodyne detection performs similar to a hard decision strategy. In fact for not too strong phase noise it comes close to the Helstrom bound. Similar to previous results the optimal alphabet shifts away from BPSK towards OOK for the displacement receiver.

An overview of the performance of the displacement receiver compared to conventional detection is given in fig. 6.12. The depicted value is the ratio between the error probability for conventional detection and the displacement receiver in dB. Only for very strong or weak phase noise the latter is able to outperform conventional detection. The regime where PNR is needed spans a much larger area in the parameter space than for the mean photon number constraint. The advantages of the displacement receiver over conventional detection are more pronounced with nonlinear phase noise.

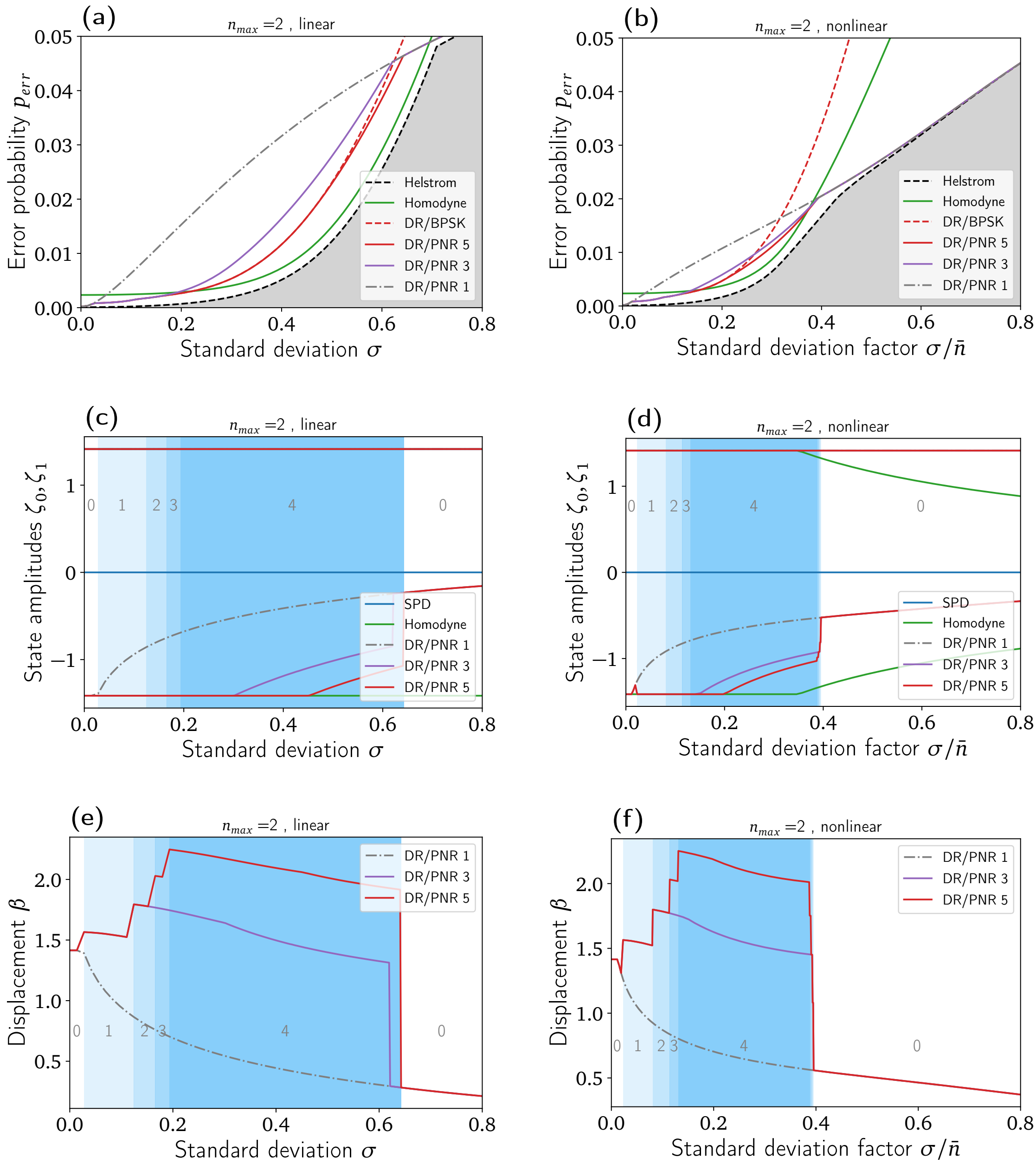


Figure 6.10: On **top (a,b)** the error probability p_{err}, in the **middle (c,d)** the state amplitudes and on the **bottom (e,f)** the displacement introduced by the detector are shown as a function of the phase noise strength for a maximum photon number of $n_{\max} = 2$. The phase noise is assumed to be linear on the **left (a,c,e)** and nonlinear on the **right (b,d,f)**. The grey numbers denote the optimal decision threshold k_n in the marked area for the displacement receiver. If the PNR of the receiver is not high enough the next possible lower decision threshold should be chosen.

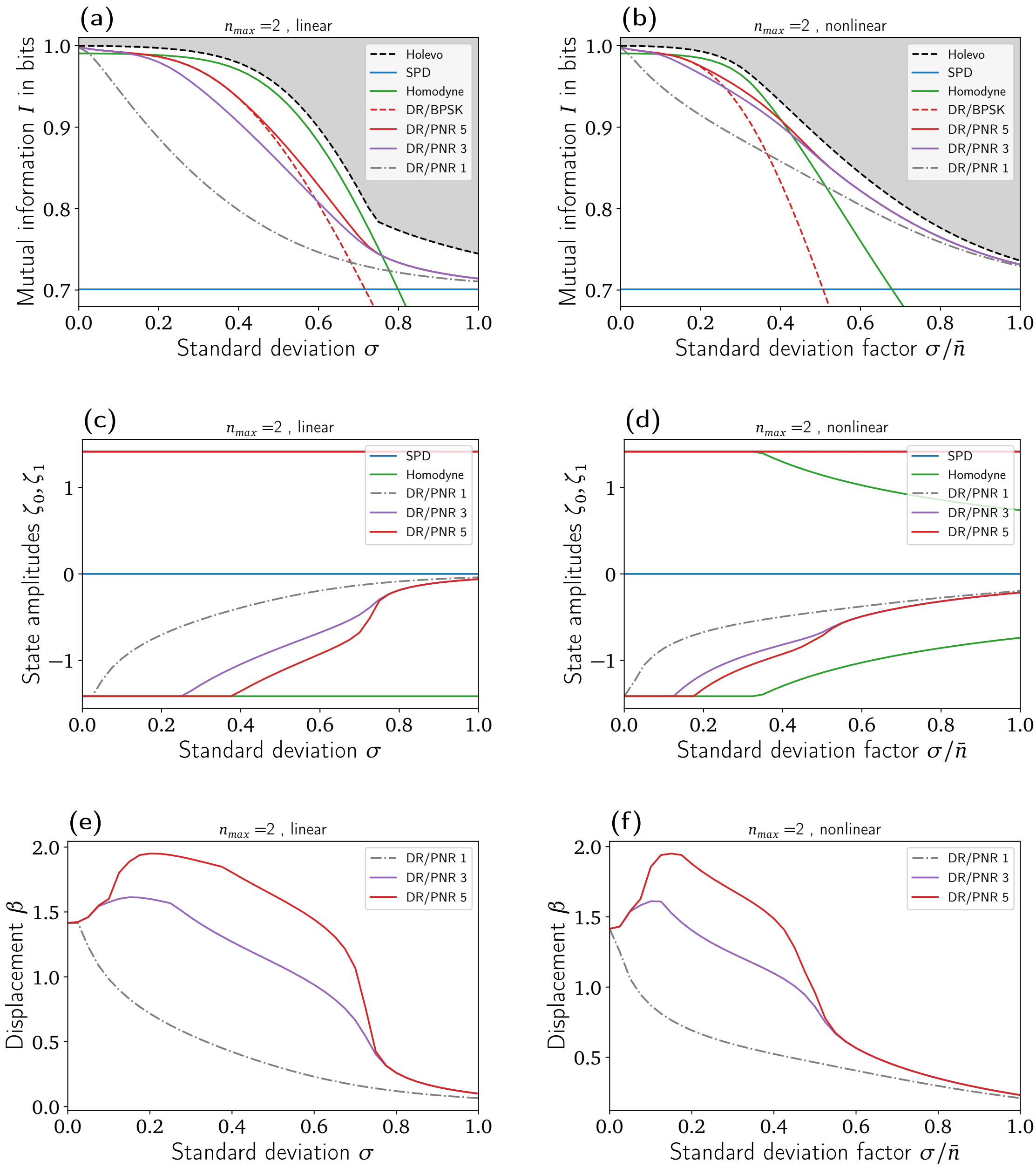


Figure 6.11: On **top (a,b)** mutual information I, in the **middle (c,d)** the state amplitudes and at the **bottom (e,f)** the displacement introduced by the detector are shown as a function of the phase noise strength for a maximum photon number of $n_{\max} = 2$. The phase noise is assumed to be linear on the **left (a,c,e)** and nonlinear on the **right (b,d,f)**.

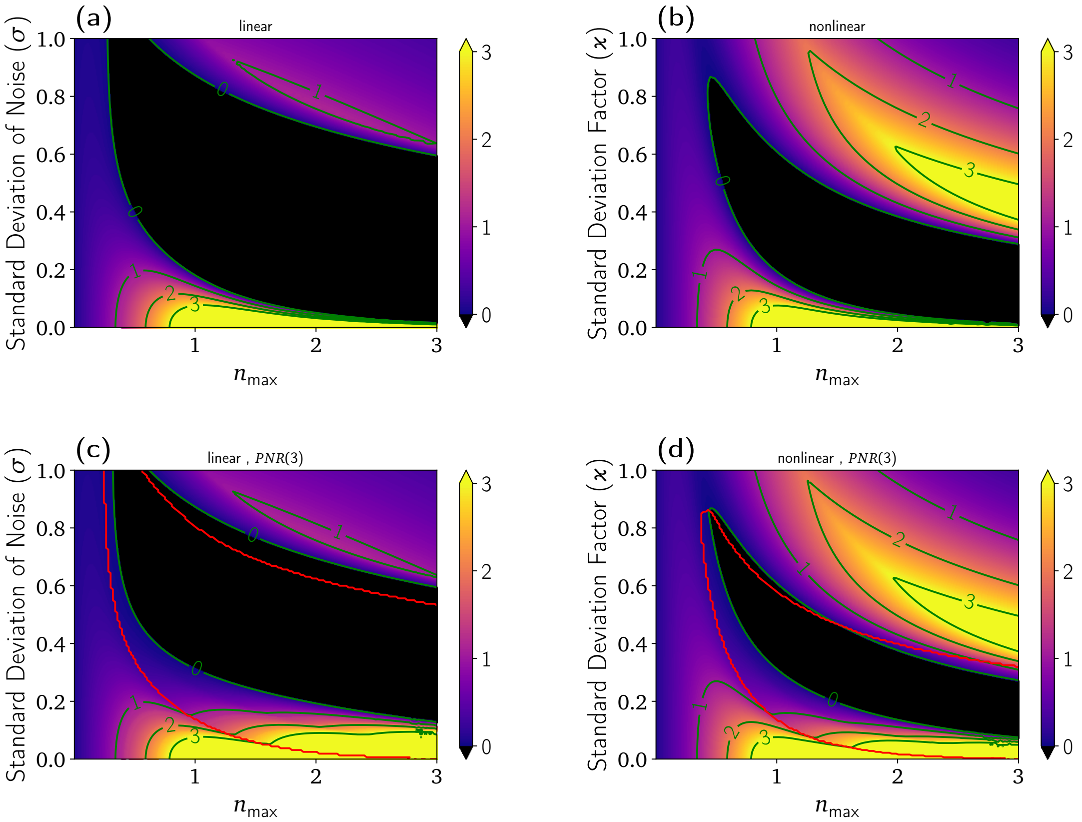


Figure 6.12: The ratio $p_{\text{err}}^{\text{conv}}/p_{\text{err}}^{DR}$ in dB as a function of the maximum photon number $n_{\max}$ and the phase noise strength for linear on the **left (a,c)** and nonlinear phase noise on the **right (b,d)**. The displacement receiver is assumed to have a PNR of 1 on **top (a,b)** and 3 at the **bottom (c,d)**. Larger values of the ratio correspond to a larger advantage of the displacement receiver over conventional methods and 0 marks equal performance. The error probability $p_{\text{err}}^{\text{conv}}$ for conventional detection is taken from either the SPD or homodyne receiver, depending on which performs better. Black areas mark the region where conventional methods outperform the displacement receiver. Red contour lines show the area, where $PNR > 1$ is necessary. Discrete jumps in the contour lines occur due to a finite resolution. They are expected to be smooth for an unlimited resolution.

7 Conclusions

In this thesis, we have outlined the features and the capability of a photon number resolving displacement receiver for the detection of coherent states. The displacement receiver is intimately connected to homodyne detection and mainly differs by the strength of the local oscillator. By employing only a weak local oscillator to act as a displacement operation the amount of noise injected into the system is drastically limited which allows for sub-shot-noise detection. Assuming an infinite photon-number resolution (PNR) and arbitrary strong local oscillator the displacement receiver essentially operates as a homodyne detector. Also, it is possible to interpolate between operating as a photon counting detector and a homodyne detector by controlling the displacement amplitude.

The displacement receiver offers a near-optimum performance for discriminating binary coherent alphabets and in the photon-starved regime approaches the optimal measurement. We derived a more realistic model of a displacement receiver by considering various imperfections, most of which can be readily combated by PNR and an optimization of the displacement amplitude. Out of all discussed imperfections the displacement receiver is most prone to a non-unit visibility.
In the presence of channel noise the performance of a displacement receiver deteriorates quickly without PNR and conventional detection techniques become superior. We showed, however, that by increasing the PNR capability and optimizing the displacement, the receiver becomes much more robust to noise and it remains able to perform at a sub-shot-noise level. Only for a very large number of thermal background photons the noise is overwhelming. For thermal noise it is sufficient to numerically optimize the displacement operation based on the noise strength. The optimal displacement crucially depends on the chosen figure of merit.
Phase noise acting on the signal requires a more involved strategy. In this scenario the state amplitudes have to be optimized additionally. The flexibility of being able to choose the displacement amplitude allows the displacement receiver to combine the superior state separation of BPSK and the robustness to phase noise of OOK such that the detection performance remains below the shot-noise level. For very strong phase noise the displacement receiver approaches the Helstrom limit and can be seen as a minimum error measurement. The feasibility of this strategy was shown in a proof-of-principle type of experiment which was carried out in collaboration with M. DiMario and F. E. Becerra [DiMario et al. 2019].

We also developed a simple quantum model of a glass fibre which takes into account self-phase modulation and losses based on a master equation approach. The interplay between the Kerr nonlinearity and losses gives rise to quantum excess noise. The solu-

tion to master equation for this type of noise was obtained for arbitrary input states. For coherent input states the noise behaves as nonlinear Gaussian phase noise whose standard deviation depends on the states energy. Furthermore, the nonlinear phase noise limits the amount of squeezing that can be obtained with a Kerr nonlinear medium.
For communication scenarios the nonlinear phase noise imposes new fundamental limitations. In a loss only scenario, the amount of information that can be encoded into the phase variable rises continuously with the mean photon number. If nonlinear phase noise is present, there exists an optimum for the Holevo information at a lower mean photon number instead of a continuous growth with the available energy.

In the photon starved regime, where the available energy for communication is very limited, the displacement receiver offers an improved PIE for PPM if thermal background noise is present. The photon-number resolving detection allows to better distinguish the signal from the background and therefore also limits the required PPM order.
We also presented a formalism which allows to analyze mutual information in the limit of vanishing energy and derived several measurement schemes that can attain superadditivity. Their physical realizations are based on a series of beamsplitters combined with SPDs, displacement receiver or homodyne detection.

Further steps towards the practicality of displacement receivers require the consideration of more advanced noise models. In free space links, for instance, atmospheric turbulences can cause major distortions of the signal. Also for fibre based communication many nonlinear effects have been neglected in this work. A fully quantum model that also includes phenomena such as cross-phase modulation or four-wave mixing could yield interesting insights. Furthermore, the formalism derived in [Kunz et al. 2020] is not only applicable to classical communication, but could also be used to investigate other scenarios such as private communication. Another direction for future work is to apply the concept of optimizing the alphabet to non-binary alphabets with more states or considering more complex measurements, for instance, displacement receiver that utilize feedback.

Bibliography

A. Acín, E. Bagan, M. Baig, L. Masanes, and R. Muñoz-Tapia. Multiple-copy two-state discrimination with individual measurements. *Physical Review A*, 71(3), mar 2005. doi: 10.1103/physreva.71.032338.

U. L. Andersen, T. Gehring, C. Marquardt, and G. Leuchs. 30 years of squeezed light generation. *Physica Scripta*, 91(5):053001, apr 2016. doi: 10.1088/0031-8949/91/5/053001.

J. M. Arrazola and N. Lütkenhaus. Quantum communication with coherent states and linear optics. *Physical Review A*, 90(4):042335, oct 2014. doi: 10.1103/physreva.90.042335.

A. Assalini, N. D. Pozza, and G. Pierobon. Revisiting the Dolinar receiver through multiple-copy state discrimination theory. *Physical Review A*, 84(2), aug 2011. doi: 10.1103/physreva.84.022342.

K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete. Discriminating states: the quantum chernoff bound. *Physical Review Letters*, 98(16):160501, apr 2007. doi: 10.1103/physrevlett.98.160501.

K. M. R. Audenaert, M. Mosonyi, and F. Verstraete. Quantum state discrimination bounds for finite sample size. *Journal of Mathematical Physics*, 53(12):122205, dec 2012. doi: 10.1063/1.4768252.

K. Banaszek and M. Jachura. Structured optical receivers for efficient deep-space communication. In *2017 IEEE International Conference on Space Optical Systems and Applications (ICSOS)*, pages 34–37, Naha, Okinawa, Japan, nov 2017. IEEE. doi: 10.1109/icsos.2017.8357208.

K. Banaszek, C. Radzewicz, K. Wódkiewicz, and J. S. Krasinski. Direct measurement of the Wigner function by photon counting. *Physical Review A*, 60(1):674–677, jul 1999. doi: 10.1103/physreva.60.674.

K. Banaszek, L. Kunz, M. Jarzyna, and M. Jachura. Approaching the ultimate capacity limit in deep-space optical communication. In H. Hemmati and D. M. Boroson, editors, *Free-Space Laser Communications XXXI*, page 109100A, San Francisco, CA, mar 2019a. SPIE. doi: 10.1117/12.2506963.

K. Banaszek, W. Zwoliński, L. Kunz, and M. Jarzyna. Photon efficiency limits in the presence of background noise. In *Proceedings of the 2019 IEEE International Conference on Space Optical Systems and Applications (ICSOS), Portland, OR, USA*, 2019b. URL https://arxiv.org/abs/1911.06772.

K. Banaszek, L. Kunz, M. Jachura, and M. Jarzyna. Quantum limits in optical communications. *Journal of Lightwave Technology*, 38(10):2741–2754, may 2020. doi: 10.1109/jlt.2020.2973890.

A. Bandilla. Correlations and excess noise in one-port and balanced two-port homodyne detection. *Journal of Modern Optics*, 37(12):1947–1954, dec 1990. doi: 10.1080/09500349014552151.

S. M. Barnett and S. Croke. Quantum state discrimination. *Adv Opt Photonics*, 1(2): 238–278, feb 2009. doi: 10.1364/AOP.1.000238.

F. E. Becerra, J. Fan, G. Baumgartner, S. V. Polyakov, J. Goldhar, J. T. Kosloski, and A. Migdall. M-ary-state phase-shift-keying discrimination below the homodyne limit. *Physical Review A*, 84(6):062324, dec 2011. doi: 10.1103/PhysRevA.84.062324.

F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdall. Experimental demonstration of a receiver beating the standard quantum limit for multiple nonorthogonal state discrimination. *Nature Photonics*, 7(2):147–152, jan 2013. doi: 10.1038/nphoton.2012.316.

F. E. Becerra, J. Fan, and A. Migdall. Photon number resolution enables quantum receiver for realistic coherent optical communications. *Nature Photonics*, 9(1):48–53, nov 2015. doi: 10.1038/nphoton.2014.280.

C. Benedetti, A. P. Shurupov, M. G. A. Paris, G. Brida, and M. Genovese. Experimental estimation of quantum discord for a polarization qubit and the use of fidelity to assess quantum correlations. *Physical Review A*, 87(5):052136, may 2013. doi: 10/gf3wmq.

E. Bimbard, N. Jain, A. MacRae, and A. I. Lvovsky. Quantum-optical state engineering up to the two-photon level. *Nature Photonics*, 4(4):243–247, feb 2010. doi: 10.1038/nphoton.2010.6.

M. Bina, A. Mandarino, S. Olivares, and M. G. A. Paris. Drawbacks of the use of fidelity to assess quantum resources. *Physical Review A*, 89(1):012305, jan 2014. doi: 10.1103/PhysRevA.89.012305.

A. Biswas, D. Boroson, and B. Edwards. Mars laser communication demonstration: what it would have been. In G. S. Mecherle, editor, *Free-Space Laser Communication Technologies XVIII*. SPIE, feb 2006. doi: 10.1117/12.669551.

A. Biswas, M. Srinivasan, R. Rogalin, S. Piazzolla, J. Liu, B. Schratz, A. Wong, E. Alerstam, M. Wright, W. T. Roberts, J. Kovalik, G. Ortiz, A. Na-Nakornpanom, M. Shaw,

C. Okino, K. Andrews, M. Peng, D. Orozco, and W. Klipstein. Status of NASA's deep space optical communication technology demonstration. In *2017 IEEE International Conference on Space Optical Systems and Applications (ICSOS)*. IEEE, nov 2017. doi: 10.1109/icsos.2017.8357206.

R. S. Bondurant. Near-quantum optimum receivers for the phase-quadrature coherent-state channel. *Optics Letters*, 18(22):1896, nov 1993. doi: 10.1364/ol.18.001896.

M. Born. Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 37(12):863–867, dec 1926. doi: 10.1007/bf01397477.

D. M. Boroson. On achieving high performance optical communications from very deep space. In H. Hemmati and D. M. Boroson, editors, *Free-Space Laser Communication and Atmospheric Propagation XXX*. SPIE, feb 2018. doi: 10.1117/12.2291542.

S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439–3443, may 1994. doi: 10.1103/PhysRevLett.72.3439.

H.-P. Breuer and F. Petruccione. *The theory of open quantum systems*. Oxford University Press, jan 2007. ISBN 9780198520634. doi: 10.1093/acprof:oso/9780199213900.001.0001.

J. R. Buck, S. J. van Enk, and C. A. Fuchs. Experimental proposal for achieving superadditive communication capacities with a binary quantum alphabet. *Physical Review A*, 61(3):032309, feb 2000. doi: 10.1103/physreva.61.032309.

I. A. Burenkov, O. V. Tikhonova, and S. V. Polyakov. Quantum receiver for large alphabet communication. *Optica*, 5(3):227, feb 2018. doi: 10.1364/optica.5.000227.

I. A. Burenkov, M. V. Jabir, A. Battou, and S. V. Polyakov. Time-resolving quantum measurement enables energy-efficient, large-alphabet communication. *PRX Quantum*, 1(1), sep 2020. doi: 10.1103/prxquantum.1.010308.

I. A. Burenkov, M. V. Jabir, and S. V. Polyakov. Practical quantum-enhanced receivers for classical communication. *AVS Quantum Science*, 3(2):025301, jun 2021. doi: 10.1116/5.0036959.

V. Bužek and P. L. Knight. Quantum interference, superposition states of light, and nonclassical effects. In *Progress in Optics*, volume 34, pages 1–158. Elsevier, 1995. doi: 10.1016/s0079-6638(08)70324-x.

J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan. Quantum Chernoff bound as a measure of distinguishability between density matrices: application to qubit and Gaussian states. *Physical Review A*, 77(3):032311, mar 2008. doi: 10.1103/physreva.77.032311.

H. B. G. Casimir and D. Polder. The influence of retardation on the london-van der waals forces. *Physical Review*, 73(4):360–372, feb 1948. doi: 10.1103/physrev.73.360.

C. M. Caves. Quantum-mechanical noise in an interferometer. *Physical Review D*, 23 (8):1693–1708, apr 1981. doi: 10.1103/PhysRevD.23.1693.

C. M. Caves and P. D. Drummond. Quantum limits on bosonic communication rates. *Reviews of Modern Physics*, 66(2):481–537, apr 1994. doi: 10.1103/RevModPhys.66.481.

N. J. Cerf, G. Leuchs, and E. S. Polzik. *Quantum information with continuous variables of atoms and light*. Imperial college press, feb 2007. doi: 10.1142/p489.

A. Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, nov 2000. doi: 10.1080/00107510010002599.

V. Chorvalli, L. L. Hors, L. Vaillon, and G. Planche. Optical communications between an aircraft and a GEO relay satellite: design and flight results of the LOLA demonstrator. In J. Costeraste, E. Armandillo, and N. Karafolas, editors, *International Conference on Space Optics — ICSO 2008*. SPIE, nov 2017. doi: 10.1117/12.2308195.

H. W. Chung, S. Guha, and L. Zheng. Superadditivity of quantum channel coding rate with finite blocklength joint measurements. *IEEE Transactions on Information Theory*, 62(10):5938–5959, oct 2016. ISSN 2157-8095. doi: 10.1109/TIT.2016.2597285.

R. L. Cook, P. J. Martin, and J. M. Geremia. Optical coherent state discrimination using a closed-loop quantum measurement. *Nature*, 446(7137):774–777, apr 2007. doi: 10.1038/nature05655.

M. Cooper, L. J. Wright, C. Söller, and B. J. Smith. Experimental generation of multi-photon Fock states. *Optics Express*, 21(5):5309, feb 2013. doi: 10.1364/oe.21.005309.

T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley, apr 2005. doi: 10.1002/047174882x.

J. Czajkowski, M. Jarzyna, and R. Demkowicz-Dobrzanski. Super-additivity in communication of classical information through quantum channels from a quantum parameter estimation perspective. *New Journal of Physics*, 19(7):073034, jul 2017. doi: 10.1088/1367-2630/aa7a7b.

G. J. M. D. F. Walls. *Quantum optics*. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-28574-8. doi: 10.1007/978-3-540-28574-8.

D. J. Daniel and G. J. Milburn. Destruction of quantum coherence in a nonlinear oscillator via attenuation and amplification. *Physical Review A*, 39(9):4628–4640, May 1989. doi: 10.1103/PhysRevA.39.4628.

M. T. DiMario and F. E. Becerra. Robust measurement for the discrimination of binary coherent states. *Physical Review Letters*, 121(2):023603, jul 2018. doi: 10.1103/PhysRevLett.121.023603.

M. T. DiMario, E. Carrasco, R. A. Jackson, and F. E. Becerra. Implementation of a single-shot receiver for quaternary phase-shift keyed coherent states. *Journal of the Optical Society of America B*, 35(3):568–574, Mar. 2018. doi: 10.1364/JOSAB.35.000568.

M. T. DiMario, L. Kunz, K. Banaszek, and F. E. Becerra. Optimized communication strategies with binary coherent states over phase noise channels. *npj Quantum Information*, 5(1):65, jul 2019. doi: 10.1038/s41534-019-0177-4.

D. Ding, D. S. Pavlichin, and M. M. Wilde. Quantum channel capacities per unit cost. *IEEE Transactions on Information Theory*, 65(1):418–435, jan 2019. doi: 10.1109/tit.2018.2854747.

P. A. M. Dirac. *The principles of quantum mechanics*. Oxford University Press, 1930. ISBN 0198520115.

P. A. M. Dirac. A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):416–418, jul 1939. doi: 10.1017/s0305004100021162.

S. J. Dolinar. An optimum receiver for the binary coherent state quantum channel. *Quarterly Progress Report*, 111, 1973.

M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov. Single-photon sources and detectors. *Review of Scientific Instruments*, 82(7):071101, jul 2011. doi: 10.1063/1.3610677.

A. D. Ellis, J. Zhao, and D. Cotter. Approaching the non-linear Shannon limit. *Journal of Lightwave Technology*, 28(4):423–433, feb 2010. doi: 10.1109/jlt.2009.2030693.

R.-J. Essiambre and R. W. Tkach. Capacity trends and limits of optical communication networks. *Proceedings of the IEEE*, 100(5):1035–1055, may 2012. doi: 10.1109/jproc.2012.2182970.

R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel. Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4):662–701, feb 2010. doi: 10.1109/jlt.2009.2039464.

A. Feinstein. A new basic theorem of information theory. *Transactions of the IRE Professional Group on Information Theory*, 4(4):2–22, sep 1954. doi: 10.1109/tit.1954.1057459.

A. R. Ferdinand, M. T. DiMario, and F. E. Becerra. Multi-state discrimination below the quantum noise limit at the single-photon level. *npj Quantum Information*, 3(1): 43, oct 2017. doi: 10.1038/s41534-017-0042-2.

M. Fiorentino, J. E. Sharping, P. Kumar, A. Porzio, and R. S. Windeler. Soliton squeezing in microstructure fiber. *Optics Letters*, 27(8):649, apr 2002. doi: 10.1364/ol.27.000649.

C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, may 1999. doi: 10.1109/18.761271.

D. Fukuda, G. Fujii, T. Numata, K. Amemiya, A. Yoshizawa, H. Tsuchida, H. Fujino, H. Ishii, T. Itatani, S. Inoue, and T. Zama. Titanium-based transition-edge photon number resolving detector with 98% detection efficiency with index-matched small-gap fiber coupling. *Optics Express*, 19(2):870, jan 2011. doi: 10.1364/oe.19.000870.

C. Gardiner and P. Zoller. *Quantum noise*. Springer Berlin Heidelberg, 2004. ISBN 3540223010.

M. G. Genoni, S. Olivares, and M. G. A. Paris. Optical phase estimation in the presence of phase diffusion. *Physical Review Letters*, 106(15):153603, apr 2011. doi: 10.1103/physrevlett.106.153603.

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. *Physical Review Letters*, 92 (2), jan 2004. doi: 10.1103/PhysRevLett.92.027902.

V. Giovannetti, R. Garcia-Patrón, N. J. Cerf, and A. S. Holevo. Ultimate classical communication rates of quantum optical channels. *Nature Photonics*, 8(10):796–800, sep 2014. doi: 10.1038/nphoton.2014.216.

N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Reviews of Modern Physics*, 74(1):145–195, mar 2002. doi: 10.1103/revmodphys.74.145.

R. J. Glauber. Coherent and incoherent states of the radiation field. *Physical Review*, 131(6):2766–2788, sep 1963. doi: 10.1103/PhysRev.131.2766.

J. P. Gordon. Quantum effects in communications systems. *Proceedings of the IRE*, 50 (9):1898–1908, sep 1962. doi: 10.1109/JRPROC.1962.288169.

V. Gorini, A. Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of N-level systems. *Journal of Mathematical Physics*, 17(5):821, 1976. doi: 10.1063/1.522979.

D. J. Griffiths and D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge University Press, aug 2018. ISBN 9781316995433. doi: 10.1017/9781316995433.

F. Grosshans, G. V. Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier. Quantum key distribution using gaussian-modulated coherent states. *Nature*, 421(6920):238–241, jan 2003. doi: 10.1038/nature01289.

S. Guha. Structured optical receivers to attain superadditive capacity and the Holevo limit. *Physical Review Letters*, 106(24):240502, jun 2011. doi: 10.1103/PhysRevLett.106.240502.

K. Günthner, I. Khan, D. Elser, B. Stiller, Ömer Bayraktar, C. R. Müller, K. Saucke, D. Tröndle, F. Heine, S. Seel, P. Greulich, H. Zech, B. Gütlich, S. Philipp-May, C. Marquardt, and G. Leuchs. Quantum-limited measurements of optical signals from a geostationary satellite. *Optica*, 4(6):611–616, jun 2017. doi: 10.1364/optica.4.000611.

R. H. Hadfield. Single-photon detectors for optical quantum information applications. *Nature Photonics*, 3(12):696–705, dec 2009. doi: 10.1038/nphoton.2009.230.

R. Han, J. A. Bergou, and G. Leuchs. Near optimal discrimination of binary coherent signals via atom–light interaction. *New Journal of Physics*, 20(4):043005, apr 2018. doi: 10.1088/1367-2630/aab2c5.

M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, mar 2009. doi: 10.1038/nphys1224.

P. Hausladen and W. K. Wootters. A "pretty good" measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, dec 1994. doi: 10.1080/09500349414552221.

P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54(3):1869–1876, sep 1996. doi: 10.1103/physreva.54.1869.

M. Hayashi. *Quantum information*. Springer, 2006. ISBN 9783540302650. doi: 10.1007/3-540-30266-2.

S. Haykin. *Communication systems*. Wiley, New York, 2001. ISBN 0471178691.

W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4):172–198, mar 1927. doi: 10.1007/bf01397280.

C. W. Helstrom. *Quantum detection and estimation theory*. Academic Press, 1976. ISBN 9780080956329.

H. Hemmati. *Deep space optical communications*. John Wiley & Sons, Inc., mar 2006. doi: 10.1002/0470042419.

H. Hemmati, A. Biswas, and I. B. Djordjevic. Deep-space optical communications: future perspectives and applications. *Proceedings of the IEEE*, 99(11):2020–2039, nov 2011. ISSN 0018-9219. doi: 10.1109/JPROC.2011.2160609.

M. Hofheinz, E. M. Weig, M. Ansmann, R. C. Bialczak, E. Lucero, M. Neeley, A. D. O'Connell, H. Wang, J. M. Martinis, and A. N. Cleland. Generation of Fock states in a superconducting quantum circuit. *Nature*, 454(7202):310–314, jul 2008. doi: 10.1038/nature07136.

A. Holevo and R. Werner. Evaluating capacities of bosonic Gaussian channels. *Physical Review A*, 63(3), feb 2001. doi: 10.1103/physreva.63.032312.

A. S. Holevo. Some estimates of information transmitted through quantum communication channel. *Probl. Peredachi Inf.*, 9(3):3–11, 1973. ISSN 0555-2923.

A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9(3):177–183, 1973. URL http://mi.mathnet.ru/eng/ppi903.

A. S. Holevo. On asymptotically optimal hypothesis testing in quantum statistics. *Theory of Probability & Its Applications*, 23(2):411–415, mar 1979. doi: 10.1137/1123048.

A. S. Holevo. *Probabilistic and statistical aspects of quantum theory*. Edizioni della Normale, Amsterdam, 1982. ISBN 978-88-7642-375-8. doi: 10.1007/978-88-7642-378-9.

A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, Jan 1998. ISSN 0018-9448. doi: 10.1109/18.651037.

L. P. Hughston, R. Jozsa, and W. K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, nov 1993. doi: 10.1016/0375-9601(93)90880-9.

K. Husimi. Some formal properties of the density matrix. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 22(4):264–314, 1940. doi: 10.11429/ppmsj1919.22.4$_2$64.

E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn. Coherent detection in optical fiber systems. *Optics Express*, 16(2):753–791, Jan 2008. doi: 10.1364/OE.16.000753.

S. Izumi, M. Takeoka, M. Fujiwara, N. D. Pozza, A. Assalini, K. Ema, and M. Sasaki. Displacement receiver for phase-shift-keyed coherent states. *Physical Review A*, 86(4):042328, oct 2012. doi: 10.1103/PhysRevA.86.042328.

S. Izumi, M. Takeoka, K. Ema, and M. Sasaki. Quantum receivers with squeezing and photon-number-resolving detectors for M-ary coherent state discrimination. *Physical Review A*, 87(4), apr 2013. doi: 10.1103/physreva.87.042328.

S. Izumi, J. S. Neergaard-Nielsen, S. Miki, H. Terai, and U. L. Andersen. Experimental demonstration of a quantum receiver beating the standard quantum limit at telecom wavelength. *Physical Review Applied*, 13(5), may 2020. doi: 10.1103/physrevapplied.13.054015.

S. Izumi, J. S. Neergaard-Nielsen, and U. L. Andersen. Adaptive generalized measurement for unambiguous state discrimination of quaternary phase-shift-keying coherent states. *PRX Quantum*, 2(2):020305, apr 2021. doi: 10.1103/PRXQuantum.2.020305.

M. V. Jabir, I. A. Burenkov, N. F. R. Annafianto, A. Battou, and S. V. Polyakov. Experimental demonstration of the near-quantum optimal receiver. *OSA Continuum*, 3(12):3324, nov 2020. doi: 10.1364/osac.409200.

M. Jarzyna. Attaining classical capacity per unit cost of noisy bosonic Gaussian channels. 2019. URL `https://arxiv.org/abs/1912.09931`.

M. Jarzyna and K. Banaszek. Efficiency of optimized pulse position modulation with noisy direct detection. In *2017 IEEE International Conference on Space Optical Systems and Applications (ICSOS)*, pages 176–181. IEEE, nov 2017. doi: 10.1109/icsos.2017.8357229.

M. Jarzyna and J. Kolodynski. Geometric approach to quantum statistical inference. *IEEE Journal on Selected Areas in Information Theory*, 1(2):367–386, aug 2020. doi: 10.1109/jsait.2020.3017469.

M. Jarzyna, P. Kuszaj, and K. Banaszek. Incoherent on-off keying with classical and non-classical light. *Optics Express*, 23(3):3170–3175, feb 2015. doi: 10.1364/OE.23.003170.

M. Jarzyna, V. Lipinska, A. Klimek, K. Banaszek, and M. G. A. Paris. Phase noise in collective binary phase shift keying with Hadamard words. *Optics Express*, 24(2): 1693–1698, jan 2016. doi: 10.1364/OE.24.001693.

J. M. Kahn and K.-P. Ho. Spectral efficiency limits and modulation/detection techniques for DWDM systems. *IEEE Journal of Selected Topics in Quantum Electronics*, 10(2): 259–272, mar 2004. ISSN 1077-260X. doi: 10.1109/JSTQE.2004.826575.

K. Kato, M. Osaki, M. Sasaki, and O. Hirota. Quantum detection and mutual information for QAM and PSK signals. *IEEE Transactions on Communications*, 47(2): 248–254, 1999. doi: 10.1109/26.752130.

R. S. Kennedy. A near-optimum receiver for the binary coherent state quantum channel. *Quarterly Progress Report*, 108, 1973.

K. Kikuchi. Fundamentals of coherent optical fiber communications. *Journal of Lightwave Technology*, 34(1):157–179, jan 2016. ISSN 0733-8724. doi: 10.1109/JLT.2015.2463719.

M. S. Kim, F. A. M. de Oliveira, and P. L. Knight. Properties of squeezed number states and squeezed thermal states. *Physical Review A*, 40(5):2494–2503, sep 1989. doi: 10.1103/physreva.40.2494.

A. Klimek, M. Jachura, W. Wasilewski, and K. Banaszek. Quantum memory receiver for superadditive communication using binary coherent states. *Journal of Modern Optics*, 63(20):2074–2080, apr 2016. doi: 10.1080/09500340.2016.1173731.

L. Kunz, M. G. A. Paris, and K. Banaszek. Noisy propagation of coherent states in a lossy Kerr medium. *Journal of the Optical Society of America B*, 35(2):214, jan 2018. doi: 10.1364/josab.35.000214.

L. Kunz, M. Jarzyna, W. Zwolinski, and K. Banaszek. Low-cost limit of classical communication with restricted quantum measurements. *New Journal of Physics*, 22(4): 043010, apr 2020. doi: 10.1088/1367-2630/ab7bd3.

W. E. Lamb and R. C. Retherford. Fine structure of the hydrogen atom by a microwave method. *Physical Review*, 72(3):241–243, aug 1947. doi: 10.1103/physrev.72.241.

W. R. Leeb. Degradation of signal to noise ratio in optical free space data links due to background illumination. *Applied Optics*, 28(16):3443, aug 1989. doi: 10.1364/ao.28.003443.

G. Li. Recent advances in coherent optical communication. *Advances in Optics and Photonics*, 1(2):279, feb 2009. doi: 10.1364/aop.1.000279.

K. Li. Second-order asymptotics for quantum hypothesis testing. *The Annals of Statistics*, 42(1), feb 2014. doi: 10.1214/13-aos1185.

K. Li, Y. Zuo, and B. Zhu. Suppressing the errors due to mode mismatch for M-ary PSK quantum receivers using photon-number-resolving detector. *IEEE Photonics Technology Letters*, 25(22):2182–2184, nov 2013. doi: 10.1109/lpt.2013.2282155.

G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, jun 1976. doi: 10.1007/bf01608499.

A. E. Lita, A. J. Miller, and S. W. Nam. Counting near-infrared single-photons with 95% efficiency. *Optics Express*, 16(5):3032, 2008. doi: 10.1364/oe.16.003032.

J. Liu, J. Chen, X.-X. Jing, and X. Wang. Quantum Fisher information and symmetric logarithmic derivative via anti-commutators. *Journal of Physics A: Mathematical and Theoretical*, 49(27):275302, may 2016. doi: 10.1088/1751-8113/49/27/275302.

K. Lukanowski and M. Jarzyna. Capacity of a lossy photon channel with direct detection. *IEEE Transactions on Communications*, pages 1–1, 2021. doi: 10.1109/tcomm.2021.3075714.

D. J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, Sept. 2003. ISBN 0521642981.

A. Mahalov and M. Moustaoui. Characterization of atmospheric optical turbulence for laser propagation. *Laser & Photonics Reviews*, 4(1):144–159, jan 2010. doi: 10.1002/lpor.200910002.

A. Mandarino, M. Bina, C. Porto, S. Cialdi, S. Olivares, and M. G. A. Paris. Assessing the significance of fidelity as a figure of merit in quantum state reconstruction of discrete and continuous-variable systems. *Physical Review A*, 93(6):062118, jun 2016. doi: 10.1103/PhysRevA.93.062118.

L. Mandel and E. Wolf. *Optical coherence and quantum optics*. Cambridge University Press, sep 1995. doi: 10.1017/cbo9781139644105.

F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam. Detecting single infrared photons with 93% system efficiency. *Nature Photonics*, 7(3):210–214, feb 2013. doi: 10.1038/nphoton.2013.13.

A. Martinez. Spectral efficiency of optical direct detection. *Journal of the Optical Society of America B*, 24(4):739–749, mar 2007. doi: 10.1364/JOSAB.24.000739.

J. C. Maxwell. VIII. a dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, 155:459–512, dec 1865. doi: 10.1098/rstl.1865.0008.

G. J. Milburn. Quantum and classical Liouville dynamics of the anharmonic oscillator. *Physical Review A*, 33(1):674–685, jan 1986. doi: 10.1103/physreva.33.674.

G. J. Milburn and C. A. Holmes. Dissipative quantum and classical Liouville mechanics of the anharmonic oscillator. *Physical Review Letters*, 56(21):2237–2240, may 1986. doi: 10.1103/PhysRevLett.56.2237.

P. P. Mitra and J. B. Stark. Nonlinear limits to the information capacity of optical fibre communications. *Nature*, 411(6841):1027–1030, jun 2001. doi: 10.1038/35082518.

T. Moon. *Error correction coding : mathematical methods and algorithms*. Wiley-Interscience, Hoboken, N.J, may 2005. ISBN 0471648000. doi: 10.1002/0471739219.

J. Neyman and E. S. Pearson. IX. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, feb 1933. doi: 10.1098/rsta.1933.0009.

M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2009. doi: 10.1017/cbo9780511976667.

S. Olivares, S. Cialdi, F. Castelli, and M. G. A. Paris. Homodyne detection as a near-optimum receiver for phase-shift-keyed binary communication in the presence of phase diffusion. *Physical Review A*, 87(5):050303, may 2013. doi: 10.1103/physreva.87.050303.

J. Oz-Vogt, A. Mann, and M. Revzen. Thermal coherent states and thermal squeezed states. *Journal of Modern Optics*, 38(12):2339–2347, dec 1991. doi: 10.1080/09500349114552501.

M. G. A. Paris. Generation of mesoscopic quantum superpositions through Kerr-stimulated degenerate downconversion. *Journal of Optics B: Quantum and Semiclassical Optics*, 1(6):662–667, nov 1999. doi: 10.1088/1464-4266/1/6/308.

S. Rahimi-Keshari, A. Scherer, A. Mann, A. T. Rezakhani, A. I. Lvovsky, and B. C. Sanders. Quantum process tomography with coherent states. *New Journal of Physics*, 13(1):013006, jan 2011. doi: 10.1088/1367-2630/13/1/013006.

M. G. Raymer and J. Mostowski. Stimulated Raman scattering: unified treatment of spontaneous initiation and spatial propagation. *Physical Review A*, 24(4):1980–1993, oct 1981. doi: 10.1103/PhysRevA.24.1980.

M. Rosenbluh and R. M. Shelby. Squeezed optical solitons. *Physical Review Letters*, 66 (2):153–156, jan 1991. doi: 10.1103/PhysRevLett.66.153.

D. Šafránek. Discontinuities of the quantum Fisher information and the Bures metric. *Physical Review A*, 95(5):052320, may 2017. doi: 10.1103/physreva.95.052320.

J. J. Sakurai and J. Napolitano. *Modern quantum mechanics*. Cambridge University Press, sep 2017. doi: 10.1017/9781108499996.

M. Sasaki, K. Kato, M. Izutsu, and O. Hirota. Quantum channels showing superadditivity in classical capacity. *Physical Review A*, 58(1):146–158, jul 1998. doi: 10.1103/PhysRevA.58.146.

C. Sayrin, I. Dotsenko, X. Zhou, B. Peaudecerf, T. Rybarczyk, S. Gleyzes, P. Rouchon, M. Mirrahimi, H. Amini, M. Brune, J.-M. Raimond, and S. Haroche. Real-time quantum feedback prepares and stabilizes photon number states. *Nature*, 477(7362): 73–77, aug 2011. doi: 10.1038/nature10376.

W. P. Schleich. *Quantum optics in phase space*. Wiley, feb 2001. doi: 10.1002/3527602976.

S. Schmitt, J. Ficker, M. Wolff, F. König, A. Sizmann, and G. Leuchs. Photon-Number Squeezed Solitons from an Asymmetric Fiber-Optic Sagnac Interferometer. *Physical Review Letters*, 81(12):2446–2449, sep 1998. doi: 10.1103/physrevlett.81.2446.

E. Schrödinger. Quantisierung als Eigenwertproblem. *Annalen der Physik*, 384(4):361–376, 1926. doi: 10.1002/andp.19263840404.

B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, jul 1997. doi: 10.1103/PhysRevA.56.131.

M. O. Scully and M. S. Zubairy. *Quantum optics*. Cambridge University Press, Sept. 1997. doi: 10.1017/cbo9780511813993.

J. Senior. *Optical fiber communications: principles and practice*. Financial Times/Prentice Hall, Harlow, England New York, 2009. ISBN 9780130326812.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(4):623–656, oct 1948. doi: 10.1002/j.1538-7305.1948.tb00917.x.

C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1): 10–21, jan 1949. doi: 10.1109/JRPROC.1949.232969.

J. Shapiro. Quantum noise and excess noise in optical homodyne and heterodyne receivers. *IEEE Journal of Quantum Electronics*, 21(3):237–250, mar 1985. doi: 10.1109/jqe.1985.1072640.

J. Shapiro. The quantum theory of optical communications. *IEEE Journal of Selected Topics in Quantum Electronics*, 15(6):1547–1569, 2009. doi: 10.1109/JSTQE.2009.2024959.

M. L. Shcherbatenko, M. S. Elezov, G. N. Goltsman, and D. V. Sych. Sub-shot-noise-limited fiber-optic quantum receiver. *Physical Review A*, 101(3), mar 2020. doi: 10.1103/physreva.101.032306.

J. S. Sidhu, S. Izumi, J. S. Neergaard-Nielsen, C. Lupo, and U. L. Andersen. Quantum receiver for phase-shift keying at the single photon level. *PRX Quantum*, 2(1):010332, feb 2021. doi: 10.1103/PRXQuantum.2.010332.

C. Silberhorn, T. C. Ralph, N. Lütkenhaus, and G. Leuchs. Continuous variable quantum cryptography: beating the 3 dB loss limit. *Physical Review Letters*, 89(16), sep 2002. doi: 10.1103/physrevlett.89.167901.

D. E. Smith. Two-way laser link over interplanetary distance. *Science*, 311(5757):53–53, jan 2006. doi: 10.1126/science.1120091.

B. Smutny, R. Lange, H. Kämpfner, D. Dallmann, G. Mühlnikel, M. Reinhardt, K. Saucke, U. Sterr, B. Wandernoth, and R. Czichy. In-orbit verification of optical inter-satellite communication links based on homodyne BPSK. In S. Mecherle, editor, *Free-Space Laser Communication Technologies XX*. SPIE, feb 2008. doi: 10.1117/12.774673.

Z. Sodnik, C. Heese, P.-D. Arapoglou, K.-J. Schulz, I. Zayer, R. Daddato, and S. Kraft. Deep-space optical communication system (DOCS) for ESA's space weather mission to Lagrange orbit L5. In *2017 IEEE International Conference on Space Optical Systems and Applications (ICSOS)*. IEEE, nov 2017. doi: 10.1109/icsos.2017.8357207.

M. A. Sorokina and S. K. Turitsyn. Regeneration limit of classical Shannon capacity. *Nature Communications*, 5(1), may 2014. doi: 10.1038/ncomms4861.

E. C. G. Sudarshan. Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams. *Physical Review Letters*, 10(7):277–279, apr 1963. doi: 10.1103/PhysRevLett.10.277.

D. Sych and G. Leuchs. Practical receiver for optimal discrimination of binary coherent signals. *Physical Review Letters*, 117(20), nov 2016. doi: 10.1103/physrevlett.117.200501.

M. Takeoka and M. Sasaki. Discrimination of the binary coherent signal: Gaussian-operation limit and simple non-Gaussian near-optimal receivers. *Physical Review A*, 78(2):022320, aug 2008. doi: 10.1103/PhysRevA.78.022320.

J. Tang. The multispan effects of Kerr nonlinearity and amplifier noises on Shannon channel capacity of a dispersion-free nonlinear optical fiber. *Journal of Lightwave Technology*, 19(8):1110–1115, Aug 2001. ISSN 0733-8724. doi: 10.1109/50.939790.

B. Teklu, J. Trapani, S. Olivares, and M. G. A. Paris. Noisy quantum phase communication channels. *Physica Scripta*, 90(7):074027, jun 2015. doi: 10.1088/0031-8949/90/7/074027.

E. Temprana, E. Myslivets, B. P. P. Kuo, L. Liu, V. Ataie, N. Alic, and S. Radic. Overcoming Kerr-induced capacity limit in optical fiber transmission. *Science*, 348 (6242):1445–1448, jun 2015. ISSN 0036-8075. doi: 10.1126/science.aab1781.

G. S. Thekkadath, S. Sempere-Llagostera, B. A. Bell, R. B. Patel, M. S. Kim, and I. A. Walmsley. Single-shot discrimination of coherent states beyond the standard quantum limit. *Optics Letters*, 46(11):2565, may 2021. doi: 10.1364/ol.421646.

N. Tomassoni and M. G. A. Paris. Quantum binary channels with mixed states. *Physics Letters A*, 373(1):61–64, dec 2008. ISSN 0375-9601. doi: 10.1016/j.physleta.2008.11.002.

J. Trapani, B. Teklu, S. Olivares, and M. G. A. Paris. Quantum phase communication channels in the presence of static and dynamical phase diffusion. *Physical Review A*, 92(1), jul 2015. doi: 10.1103/physreva.92.012317.

K. Tsujino, D. Fukuda, G. Fujii, S. Inoue, M. Fujiwara, M. Takeoka, and M. Sasaki. Sub-shot-noise-limit discrimination of on-off keyed coherent signals via a quantum receiver with a superconducting transition edge sensor. *Optics Express*, 18(8):8107–8114, apr 2010. doi: 10.1364/OE.18.008107.

K. Tsujino, D. Fukuda, G. Fujii, S. Inoue, M. Fujiwara, M. Takeoka, and M. Sasaki. Quantum receiver beyond the standard quantum limit of coherent optical communication. *Physical Review Letters*, 106(25):250503, jun 2011. doi: 10.1103/PhysRevLett.106.250503.

K. S. Turitsyn and S. K. Turitsyn. Nonlinear communication channels with capacity above the linear Shannon limit. *Optics Letters*, 37(17):3600, aug 2012. doi: 10.1364/ol.37.003600.

A. Uhlmann. The “transition probability” in the state space of a ∗-algebra. *Reports on Mathematical Physics*, 9(2):273–279, apr 1976. doi: 10.1016/0034-4877(76)90060-4.

V. Vilnrotter and E. Rodemich. A generalization of the near-optimum binary coherent state receiver concept (corresp.). *IEEE Transactions on Information Theory*, 30(2): 446–450, mar 1984. doi: 10.1109/tit.1984.1056853.

J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer Berlin Heidelberg, 1932. ISBN 9783642614095. doi: 10.1007/978-3-642-61409-5.

W. Wasilewski, A. I. Lvovsky, K. Banaszek, and C. Radzewicz. doi:10.1098/rsta.1933.0009pulsed squeezed light: simultaneous squeezing of multiple modes. *Physical Review A*, 73(6), jun 2006. doi: 10.1103/physreva.73.063819.

C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84(2): 621–669, may 2012. doi: 10.1103/revmodphys.84.621.

L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood. The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: cross-phase modulation and four-wave mixing. *Physica D: Nonlinear Phenomena*, 189(1-2):81–99, feb 2004. doi: 10.1016/j.physd.2003.09.038.

M. J. Werner. Quantum soliton generation using an interferometer. *Physical Review Letters*, 81(19):4132–4135, nov 1998. doi: 10.1103/physrevlett.81.4132.

E. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5):749–759, jun 1932. doi: 10.1103/physrev.40.749.

M. M. Wilde, S. Guha, S.-H. Tan, and S. Lloyd. Explicit capacity-achieving receivers for optical communication and quantum reading. In *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, jul 2012. doi: 10.1109/isit.2012.6284251.

C. Wittmann, M. Takeoka, K. N. Cassemiro, M. Sasaki, G. Leuchs, and U. L. Andersen. Demonstration of near-optimal discrimination of optical coherent states. *Physical Review Letters*, 101(21):210501, nov 2008. doi: 10.1103/PhysRevLett.101.210501.

C. Wittmann, U. L. Andersen, M. Takeoka, D. Sych, and G. Leuchs. Demonstration of coherent-state discrimination using a displacement-controlled photon-number-resolving detector. *Physical Review Letters*, 104(10):100505, mar 2010. doi: 10.1103/PhysRevLett.104.100505.

T. Yamashita, S. Miki, H. Terai, and Z. Wang. Low-filling-factor superconducting single photon detector with high system detection efficiency. *Optics Express*, 21(22):27177, nov 2013. doi: 10.1364/oe.21.027177.

R. Yuan, M. Zhao, S. Han, and J. Cheng. Kennedy receiver using threshold detection and optimized displacement under thermal noise. *IEEE Communications Letters*, 24 (6):1313–1317, jun 2020. doi: 10.1109/lcomm.2020.2980537.

H. Yuen and J. Shapiro. Optical communication with two-photon coherent states – Part I: quantum-state propagation and quantum-noise. *IEEE Transactions on Information Theory*, 24(6):657–668, nov 1978. doi: 10.1109/tit.1978.1055958.

H. P. Yuen. Two-photon coherent states of the radiation field. *Physical Review A*, 13 (6):2226–2243, jun 1976. doi: 10.1103/PhysRevA.13.2226.

H. P. Yuen and M. Ozawa. Ultimate information carrying limit of quantum systems. *Physical Review Letters*, 70(4):363–366, jan 1993. doi: 10.1103/physrevlett.70.363.

M. Yukawa, K. Miyata, T. Mizuta, H. Yonezawa, P. Marek, R. Filip, and A. Furusawa. Generating superposition of up-to three photons for continuous variable quantum information processing. *Optics Express*, 21(5):5529, feb 2013. doi: 10.1364/oe.21.005529.

B. Yurke and D. Stoler. Generating quantum mechanical superpositions of macroscopically distinguishable states via amplitude dispersion. *Physical Review Letters*, 57(1): 13–16, jul 1986. doi: 10.1103/PhysRevLett.57.13.