

# Entropy in Physics: An Overview of Definitions and Applications in Quantum Mechanics

Fabian Immanuel IJpelaar

January 22, 2021

## Abstract

In modern physics, entropy is a well-known and important quantity. At its core, the entropy is a function of the probabilities of a distribution that is meant to describe the uncertainty in outcome of a random process represented by that distribution. However, it has been used in many different fields and, as a consequence, has many interpretations. Moreover, a lot of different functions have been grouped under the name of entropy, all with their own uses and interpretations. In this work, we discuss the definitions, origins, and interpretations of many of these functions as well as how they fit together. We will also explicitly cover some of the applications that the entropies have found within physics, in particular in quantum physics. These applications include thermodynamics, measurement uncertainty in quantum mechanics, measures of mixedness in the density matrix formalism and in phase space formalisms, measures of entanglement, and parton distributions in QCD.

Master's Thesis Physics  
Van Swinderen Institute

Supervisor: Daniël Boer  
Second Examiner: Elisabetta Pallante



**rijksuniversiteit  
groningen**

# Contents

<b>Introduction</b>	<b>5</b>
<b>1 The Entropies of Statistical Mechanics and Information Theory</b>	<b>9</b>
1.1 Statistical Mechanics and the Classical Entropy . . . . .	10
1.1.1 The Liouville Equation . . . . .	10
1.1.2 The Microcanonical Ensemble . . . . .	11
1.1.3 The Microcanonical Entropy . . . . .	11
1.1.4 Extensivity of the microcanonical Entropy and Temperature . . . . .	12
1.1.5 The Canonical Ensemble . . . . .	13
1.1.6 The Canonical Entropy . . . . .	13
1.1.7 Closing Remarks . . . . .	14
1.2 The Shannon Entropy . . . . .	15
1.2.1 Comparison with the Standard deviation . . . . .	15
1.2.2 Shannon Entropy as a Measure of (Lack of) Information . . . . .	16
1.2.3 Statistical Mechanics and Information Theory: Jaynes' Maximum Entropy Principle . . . . .	18
1.2.4 Properties . . . . .	19
1.3 The Differential (Continuous Shannon) Entropy . . . . .	21
1.3.1 Properties . . . . .	22
1.3.2 The Differential Entropy and the Standard Deviation . . . . .	24
1.4 Summary and Concluding Remarks . . . . .	24
<b>2 Course Graining Classical Distributions</b>	<b>26</b>
2.1 Aggregation of Discrete Probabilities . . . . .	27
2.2 Coarse Graining Continuous Distributions . . . . .	27
2.3 The Second Law and Mixing . . . . .	29
2.4 Concluding Remarks . . . . .	31
<b>3 Generalized Entropies</b>	<b>32</b>
3.1 The Shannon-Khinchin Axioms . . . . .	33
3.1.1 Extensivity . . . . .	33
3.1.2 Experimental Robustness . . . . .	34
3.2 The Maximum Entropy Method Revisited . . . . .	35
3.3 Tsallis Entropy . . . . .	36
3.3.1 Properties . . . . .	37
3.3.2 Relation to the Shannon Entropy . . . . .	38
3.3.3 Maximization . . . . .	38
3.3.4 Extensivity . . . . .	39
3.3.5 The Zeroth Law of Thermodynamics and the Tsallis Composition Law . . . . .	39
3.4 The Rényi Entropy . . . . .	42
3.4.1 Properties . . . . .	43
3.4.2 Entropy Maximization . . . . .	44

3.5	Scaling Exponents . . . . .	45
3.6	Imposing Extensivity . . . . .	46
3.7	Summary and Concluding Remarks . . . . .	47
<b>4</b>	<b>Density Matrices, Mixed States and Quantum Operations</b>	<b>49</b>
4.1	Introduction to the Density Matrix: The Particle Source . . . . .	49
4.2	The Density Matrix . . . . .	50
4.3	Schmidt Decomposition . . . . .	51
4.4	Quantum Operations . . . . .	51
4.4.1	Kraus Operators from Unitary Time Evolution . . . . .	52
4.5	Defining Mixedness through Majorization . . . . .	53
4.5.1	Schur-Convexity (Concavity) . . . . .	53
<b>5</b>	<b>Measurement Uncertainty in Discrete Hilbert Spaces</b>	<b>54</b>
5.1	The Uncertainty Principle . . . . .	55
5.1.1	The Maassen-Uffink Uncertainty Relation in a $2D$ Hilbert Space . . . . .	57
5.2	Concluding Remarks . . . . .	58
<b>6</b>	<b>Position and Momentum Uncertainty</b>	<b>59</b>
6.1	The Uncertainty Relations . . . . .	59
6.2	Infinite Square Well . . . . .	62
6.2.1	Comparison with the Classical Infinite Square Well . . . . .	65
6.3	Harmonic Oscillator . . . . .	65
6.3.1	Comparison with the classical harmonic oscillator . . . . .	66
6.3.2	Coherent States and Squeeze States . . . . .	68
6.4	The Hydrogen Atom . . . . .	73
6.5	Maximum Entropy Wavefunctions . . . . .	75
6.6	Summary and Concluding Remarks . . . . .	78
<b>7</b>	<b>Measures of Mixedness</b>	<b>80</b>
7.1	The Von Neumann Entropy . . . . .	81
7.1.1	Von Neumann's Gedanken Experiment . . . . .	81
7.1.2	The Ensembles and Entropies of Quantum Statistical Physics . . . . .	83
7.1.3	Properties . . . . .	84
7.1.4	Entropy Maximization . . . . .	87
7.2	The Quantum Tsallis and Rényi entropies . . . . .	89
7.3	Summary and Concluding Remarks . . . . .	89
<b>8</b>	<b>Entanglement and the Entanglement Entropy</b>	<b>91</b>
8.1	Quantifying Entanglement: The LOCC Paradigm . . . . .	92
8.1.1	Maximally and Minimally Entangled States . . . . .	93
8.1.2	Locally Manipulating Two Qubits . . . . .	93
8.1.3	The Distillable Entanglement and Entanglement Cost . . . . .	94
8.1.4	Beyond Pure States . . . . .	95
8.2	The Entanglement Entropy and Area Laws . . . . .	95
8.3	The Spin XY Model and the Ground State Entropy . . . . .	96
8.4	Entanglement Entropy in Field Theory . . . . .	103
8.4.1	The Real Time Approach . . . . .	103
8.4.2	The Euclidean Method . . . . .	107
8.4.3	The Replica Trick . . . . .	109
8.4.4	The Entanglement Entropy of the Free Scalar Field . . . . .	109
8.5	Summary and Concluding Remarks . . . . .	111
<b>9</b>	<b>Phase Space Entropies</b>	<b>114</b>
9.1	The Wigner and Husimi Representations . . . . .	114

9.2	The Linear Entropy And Wigner Distributions . . . . .	116
9.2.1	Properties . . . . .	117
9.2.2	Entropy Maximization . . . . .	117
9.2.3	Entropy of Course graining . . . . .	118
9.3	Wehrl Entropy and The Husimi-Q Representation . . . . .	119
9.3.1	Quantum-Optical States . . . . .	119
9.4	Summary and Concluding Remarks . . . . .	123
<b>10</b>	<b>Entropy and Parton Distributions</b>	<b>125</b>
10.1	DIS and the Parton Model . . . . .	126
10.1.1	QCD and Corrections to Bjorken Scaling . . . . .	128
10.2	The Entropy of the Parton Distribution functions . . . . .	129
10.3	The Entropy of Ignorance . . . . .	132
10.3.1	Entropy of Ignorance for a Spin System . . . . .	133
10.3.2	The Entropy of Ignorance vs the Von Neumann entropy in the CCG model . . . . .	134
10.4	Summary and Concluding Remarks . . . . .	136
	<b>Conclusion</b>	<b>137</b>
	<b>Appendices</b>	<b>150</b>
<b>A</b>	<b>Hydrogen Atom Entropies</b>	<b>151</b>

# Introduction

Entropy, although being ubiquitous throughout physics and even many other fields, is also notorious for being hard to grasp. Von Neumann famously said to Shannon, who had developed his “own” entropy quantity but didn’t know yet what to call it [1]:

*You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.*  
(-John Von Neumann to Claude Shannon, 1949)

While this quote is a bit dated by now, entropy and entropy-like quantities are still being studied very intensively, and since then, the number of “entropies” has become even larger. Wehrl [2] has even gone as far as to say: “There is a tremendous variety of entropy-like quantities, especially in the classical case, and perhaps every month somebody invents a new one”. Moreover, as we will discuss, all these quantities are used within many different fields of research, so even for the same quantity, interpretations may vary, adding another barrier to entry for someone who wants to learn about the use of these entropies.

The aim of this work is to act as a general starting point for someone who wants to learn about these entropies and related topics, while still focusing on applications and interpretations in physics, and in particular quantum mechanics. Since we feel that a lot of confusion about the entropy comes from its past uses and interpretations, we will give a brief overview of this in the rest of the introduction. In chapter 1, we will show how the most well known entropy, now called the Boltzmann-Gibbs-Shannon (BGS) entropy, is defined in classical statistical mechanics, as well as in classical information theory. In chapter 2, we will discuss so called coarse graining of distributions, which is an important concept in information theory, and also is the precursor of modern interpretations of the second law of thermodynamics. In chapter 3 we will then discuss the shortcomings of the BGS entropy, and generalizations of it, called generalized entropies. Next, in chapter 4 we will discuss the density matrix formalism of quantum mechanics, which may be used to encode classical distributions over quantum states, as well as some related concepts. These concepts will be crucial to understand the use of entropies in quantum mechanics. Then in chapters 5 and 6 we will discuss how entropies may be used as information theoretic measures, of measurement uncertainty, similarly to how the standard deviation is normally used. We will discuss modern formulations of the uncertainty principle and applications to a few simple quantum systems. In chapter 7 we discuss the quantum counterparts of the classical entropies, and in particular the Von Neumann entropy and how it may be interpreted as quantum generalization of the thermodynamical entropy. After that, in chapter 8, we discuss how these same quantum entropies may be interpreted as measures of entanglement. Here, we will first discuss one of the common characterizations of entanglement, and how this may be viewed as a resource within certain protocols. Then we will discuss the application of the Von Neumann entropy as a measure of entanglement in a spin lattice system, as well as in the free scalar field. Then in chapter 9 we will very briefly discuss quantum phase space representations and how entropies may be defined for them. Then lastly, in chapter 10, we will discuss a specific application of the Von Neumann entropy of the parton distribution functions QCD.

Now we will give a brief overview of the uses of entropy. The first use dates back to Clausius, who coined the term in 1865 with the intent for it to be very similar to the term energy, believing the two concepts to be very similar. It was based on the observation that for a cycle of reversible processes, the ratio  $\frac{Q}{T}$  was conserved, where  $Q$  is the heat supplied to a system and  $T$  is its temperature. This led to the definition of the entropy as the function of state which satisfies

$$dS = \frac{\delta Q}{T}. \quad (1)$$

A state function is used to describe the macroscopic state of a thermodynamic system, and depends on state variables, like for example pressure, volume and temperature, which are taken as inputs<sup>1</sup> Large entropy is typically associated with a lower amount of “free” energy, or energy that can be used to do work. The second law of thermodynamics was also re-formulated in terms of this quantity, stating that the entropy of a closed system is non-decreasing for any process, only staying constant for (idealized) reversible processes.

Describing systems using state functions and variables, was based on the experimental fact that macroscopic systems have very well defined properties. Due to this there was no reason, at the time, to necessarily believe that these properties were not fundamental. However, scientists like Daniel Bernoulli had already postulated that gases consisted of separate particles beginning in the mid 1700s [3], the properties of which should be able to be used to derive the properties of the macroscopic system. However it was only when Maxwell derived the first statistical theory of gases that this idea got traction. Boltzmann followed up on Maxwell’s work and laid down the foundation of the field we now know as statistical mechanics. He formulated a differential equation which he thought at the time describes the time evolution of a gas in absence of any external force. Using this equation, he formulated his famous H-theorem, which states that the following quantity he called H, which he deemed to be related to the physical entropy by a minus sign, was non-decreasing:

$$H = k_B \int_{-\infty}^{\infty} d^3x d^3p f(\vec{x}, \vec{p}) \ln(f(\vec{x}, \vec{p})), \quad (2)$$

where  $k_B$  is Boltzmann’s constant and  $f(x, p)$  is the probability density to find a single particle at position and momentum  $\vec{x}$  and  $\vec{p}$ . For a while this seemed to be the proof for the second law of thermodynamics. However, after a while it was realized that H-theorem relied on some “unphysical” assumptions. His equation for example implicitly assumed that the velocities of the particles are initially uncorrelated and only become correlated after collision (this is most commonly referred to as molecular chaos). This essentially inserted irreversibility by hand and had as a consequence that if time was reversed, the entropy was instead non-increasing. This was in contradiction with the time-reversibility of Newtonian mechanics, and was thus deemed an unjustifiable assumption. Following this criticism he formulated the well known entropy,

$$S = k_b \ln W, \quad (3)$$

where  $W$  is the amount of ways a macroscopic thermodynamic state can be realized from a microstate. Using and improving upon the ideas of Boltzmann, Gibbs formulated the modern interpretation of statistical mechanics, in terms of distributions in phase space, called ensembles. The entropy now becomes

$$S = -k_B \sum p_i \ln p_i, \quad (4)$$

for discrete systems, where  $p_i$  is the probability of the  $i$ -th microstate, and

$$S = -k_B \int d^{3N}x d^{3N}p \rho(\vec{x}_1, \dots, \vec{x}_N; \vec{p}_1, \dots, \vec{p}_N), \quad (5)$$

---

<sup>1</sup>State functions can be used as variables, and vice versa, by inverting the thermodynamical functions. Thus, while they both describe the state of the system, the difference between the two is which whether we treat them as dependent or independent variables.

for continuous systems, where  $\rho(\cdot)$  is the probability density for the system to occupy a point in phase space and  $N$  is the number of particles. For the discrete case, Gibbs entropy just gives the Boltzmann entropy when all probabilities are equal, which is assumed to be the case for isolated<sup>2</sup> thermodynamical systems in equilibrium.

With the dawn of quantum mechanics came the realization that the microscopical dynamics of a system should indeed be described by quantum mechanics, instead of classical mechanics. However, the above entropies seem incompatible with quantum mechanical states. John Von Neumann formulated the density matrix formalism of quantum mechanics, and subsequently the Von Neumann entropy acting on these so called density matrices. Given a density matrix  $\hat{\rho}$ , the Von Neumann entropy is defined by

$$S = -k_B \text{Tr} [\hat{\rho} \ln(\hat{\rho})]. \quad (6)$$

We will explicitly define density matrices in chapter 4.

As a result of the advances made in in statistical mechanics, the idea that entropy signifies the disorder of a system has become common, even outside the scientific community [4]. The exact origin of this idea is unclear, but it seems to be related to the known examples where intuitively ordered systems have low entropies and intuitively unordered systems have higher entropies. For example, spin systems will typically align at low temperatures and have low entropy, whereas at high temperature, the spins will be mostly random and have higher entropy. However, the term disorder, although intuitively clear, is hard to define in a quantitative, rather than a qualitative, manner. The intuitive idea of disorder may sometimes even fail, for example for suspensions. For these systems, the equilibrium state of highest entropy is the state where the substance in the liquid collects in the bottom. This state thus seems ordered in some way, rather than completely disordered [5]. We will see that the interpretation of entropy as a measure of disorder is not needed, and we will instead interpret the entropy from the perspective of the field called information theory.

A central idea of information theory is that the entropy may be interpreted as a measure of the uncertainty we have about the outcome of a random process described by some distribution. The field began with Shannon, who, in parallel to much of the work done in statistical mechanics, set out to describe the statistical nature of “lost information” in phone line signals. He formulated the entropy quantity

$$- \sum p_i \ln(p_i), \quad (7)$$

where  $p_i$  are the probabilities or any probability distribution. The motivation behind formulating this entropy was not physical, and was instead motivated by studying Markov processes that produce messages. The entropy was formulated by Shannon based upon a set of axioms, which Shannon believed should reasonably be satisfied by a quantity that represents the amount of uncertainty we have about the outcome of a random process [6]. Equivalently, the quantity be interpreted as the amount of information a random process produces. Shannon’s work solidified the field of information theory and the information theoretical interpretation of entropy. Jaynes later on showed how the distributions from statistical mechanics could be derived from purely information theoretic principles, framing them purely as statistical inference<sup>3</sup> [7]. The principle he used is the now called principle of maximum entropy, and has become ubiquitous in information theory, and even shows up often in physics.

After to the tremendous success of the entropy in information theory, as well as physical theories, there have been a lot of proposed additional entropies. Some have been introduced with thermodynamical applications in mind, like most notably the Tsallis entropy. Other quantities are completely conceived from information theoretic ideas, like axiomatic characterizations. A notable example is the Rényi entropy, which also has found its way in many physical applications.

---

<sup>2</sup>Isolated means that the system cannot exchange heat, work or matter with the environment.

<sup>3</sup>Statistical inference is the process of deducing the distribution of a process from the limited amount of data that is known.

In the modern day, information theory has become more and more intertwined with physics, most notably in the field called quantum information theory. Here the entropies have found many applications such as in quantum communication, entanglement, uncertainty measures, black hole physics and more. We will discuss multiple of these applications this thesis.



# Chapter 1

## The Entropies of Statistical Mechanics and Information Theory

In this chapter we will define the conventional entropy of classical statistical mechanics, and classical information theory. In statistical mechanics, the importance of the entropy is directly related to its interpretation in the thermodynamical equations of state. Its derivation is phenomenological, in that its ultimate purpose is to satisfy those equations. However, we will see that its definition is intimately tied to the ensembles in classical mechanics, which are essentially distributions in phase space. The ensembles are meant to represent thermodynamical systems that are either isolated, or interacting with a heat bath, and are ultimately tied to the microscopic behavior of thermodynamic systems. However, they have not been derived from the actual dynamics of specific systems. Instead, in conventional statistical mechanics, we rely on physical axioms describing the behavior of the system, to derive the ensembles, which we will discuss. We will also demonstrate that the thermodynamical limit, which is the limit in which the number of constituents of a system becomes arbitrarily large, is needed for the entropy to be connected to the thermodynamical entropy. It is only in this limit that the entropy becomes extensive, meaning it scales linearly with the size of the system, when we let two systems interact, which is needed for the thermodynamical equations of state.

After discussing the entropy in the context of statistical mechanics, we will move on to the interpretation of the entropy in information theory. The function is exactly the same, apart from the absence of the Boltzmann constant, but the interpretation is different. We will show Shannon's axioms from which he derived the entropy, and which give it its meaning. We will also compare its qualitative properties against those of the standard deviation. Then we will discuss an operational interpretation of the Shannon entropy, in terms of bits produced by a random process. After that, we will present Jaynes' principle of maximum entropy, which he used to show that the distributions of statistical mechanics may be obtained merely by maximizing the entropy, given some constraints. This correspondence between statistical inference and physics has solidified the maximum entropy principle as a very powerful tool. However, we will further discuss this principle in chapter 3, and show that there is room for more general maximum entropy methods. Lastly we will discuss the generalization of the Shannon entropy to continuous distributions, which is often called the differential entropy. We discuss that this quantity is more of an analogous quantity than a proper generalization of the Shannon entropy. There is no direct limit in which the Shannon entropy becomes exactly the differential entropy, and because of this, they do not share all properties. However, in chapter 2 we will show that there still is a way to relate the two entropies through course graining.

## 1.1 Statistical Mechanics and the Classical Entropy

In this section we will give a brief review of the field of classical statistical mechanics. For a full introduction there are many good introductory books on the topics, like the one by Mandl [8] or Huang [9]. We will show the thermodynamical definition of entropy and we introduce the microcanonical ensemble by assuming the ergodic hypothesis. Following that, we define ensembles on phase space, and how the entropy is defined for specifically the so called microcanonical and canonical ensembles, which will be the main result of this section.

Experimentally we know that macroscopic systems like gases and fluids possess very stable properties in equilibrium, like temperature and pressure. In thermodynamics these properties are expressed as state functions and variables. These functions and variables, as their name implies, can be used to fully describe the macroscopic state of a system. They are thus not dependent on the history of a system. Clausius proposed the entropy as an extensive state function satisfying

$$dS = \frac{\delta Q}{T}, \quad (1.1)$$

where  $\delta Q$  is the heat supplied to the system<sup>1</sup>. Being extensive means that this property scales with the mass of the system. Using this entropy, the first law of thermodynamics has later been reformulated as

$$dU = TdS - PdV, \quad (1.2)$$

which states the conservation of energy for closed systems (meaning they cannot interchange mass).

The goal of statistical mechanics is to derive these thermodynamic properties of large systems from the microscopic laws of its constituent particles. We thus want to find expressions for energy, pressure and entropy in such a way that they reproduce the above equation, or at least in some limit. Of course, on the level of individual particles, properties like kinetic energy and pressure will fluctuate heavily, so it is clear that to reproduce the experimentally very stable properties of thermodynamics we will have to work with average quantities. A microscopic measurement will always happen over a finite time interval and thus effectively averages over time. With our knowledge of classical mechanics we might then hope to derive the thermodynamical properties by taking time averages over Newtonian configurations. The answer to this problem turns out to be more subtle and will require considerations about the behavior of large dynamical systems. To discuss such considerations we will need to use the notion of a configuration space.

In Newtonian mechanics every configuration of a multi-particle system can be parameterized by the positions and momenta of the individual particles. The space of such classical configurations is called phase-space. A single configuration will be represented by a point in all these variables, with its time evolution given by the Hamiltonian. However, in general, a phase space will be bounded. Firstly, the volume for thermodynamical considerations will be bounded, since a lot of the considered systems are gases in boxes or similar systems. Moreover, the phase space will always be bounded in momentum for an isolated system, because of the energy conservation of the Hamiltonian. States with a certain energy only evolve to states with equal energy. The momentum of a single particle may therefore not be such that its kinetic energy is more than the initial energy of the system. This bounded region of phase space, or in other words, all accessible systems, will be denoted  $\Omega(E)$  or simply  $\Omega$ . In the subsequent sections we will make a probability assignment for every such state in the case of an isolated system and in the case of a system thermally coupled to a much larger system, called a heat bath. Such a probability assignment defines what is called an ensemble. Following that, we will see how the entropies are defined for such ensembles.

### 1.1.1 The Liouville Equation

Ensembles are characterized described by probability densities in phase space. The Hamiltonian predicts reversible evolution for each state. So if we at any time assign a particular density to a

---

<sup>1</sup>Q is not a state function, which is signified by the use of  $\delta$  instead of  $d$ .

state, then at some time later, the same density must be assigned to the state it maps to through the evolution described by the Hamiltonian. As a consequence an ensemble in general is not time invariant, and moves through phase space like an incompressible fluid. This is expressed by Liouville's theorem

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \sum_{i=1}^N \left( \frac{\partial\rho}{\partial q_i} \dot{q}_i + \frac{\partial\rho}{\partial p_i} \dot{p}_i \right) = \frac{\partial\rho}{\partial t} + \{\rho, H\} = 0. \quad (1.3)$$

This essentially states that the density at a point must stay constant over time, if we view the ensemble from a frame moving along the trajectory of that initial point. We will discuss in the next sections that for the thermodynamical systems we consider, we assume that every point in phase space, given enough time, visits any accessible region in phase space. For this reason we take the equilibrium ensembles to be the ones that cover the whole accessible phase space and are static, to reflect the static properties of the thermodynamical equilibrium. For an ensemble to be time-invariant, it must satisfy

$$\frac{\partial\rho}{\partial t} = 0, \quad \text{or equivalently,} \quad \{\rho, H\} = 0. \quad (1.4)$$

### 1.1.2 The Microcanonical Ensemble

In this section we will show how we can define the ensemble that represents isolated thermodynamical systems. However, a reasonable question to ask is why we need a probabilistic description of systems that are inherently deterministic. The reason for this is that the systems we consider in statistical mechanics are considered to be “chaotic”. Such systems are characterized by an extreme sensitivity to their initial conditions [10]. For chaotic systems, any two states initially close together in phase space, will, to first order in time, diverge exponentially<sup>2</sup>. It is therefore impossible to track a given system along its trajectory in phase space for any appreciable time, without an infinitely accurate description of its initial conditions. For real systems we can therefore not hope to make accurate predictions from averages over phase-space trajectories. To then obtain a statistical description of the relevant thermodynamic systems, we assume they behave ergodically.<sup>3</sup> [2, §B.3]. An ergodic system will visit any region in  $\Omega$ , arbitrarily close, given enough time, with the amount of time spent in any area directly proportional to the size of that area. This means that any equally sized region in phase space will also be visited equally as often by such a system and that any system initially in  $\Omega$  will do the same. This means that any time averaged quantity over a time  $T$  will tend to the same value as we take  $T \rightarrow \infty$ , for any initial system. Macroscopic measurements happen over very large times compared to the timescales at which the microscopic properties change, so we will take the thermodynamic properties of an equilibrium system to be the time average over an infinite amount of time<sup>4</sup> For an isolated system with energy  $E$ , this leads us to define the probability density in state space as a constant density over all states with the Hamiltonian  $H(x_1, \dots, x_n, p_1, \dots, p_n) = E$ ,

$$\rho_{mc}(x_1, \dots, x_n, p_1, \dots, p_n) = \Omega^{-1}. \quad (1.5)$$

This defines the microcanonical ensemble. Moreover, this ensemble is invariant under the Liouville equation. This is a desirable property, since this reflects the stability of the thermodynamical properties of isolated systems in equilibrium.

### 1.1.3 The Microcanonical Entropy

The entropy of the microcanonical ensemble may be found to be given by the Boltzmann entropy

$$S_{BG} = k_B \ln W, \quad (1.6)$$

<sup>2</sup>Such behavior is typically associated with positive Lyapunov exponents everywhere in phase-space

<sup>3</sup>We do note, however that this argument has not been without scrutiny [11], and, historically, authors often instead opt for the assumption of equal a-priori probabilities, leading to the same definition of the microcanonical ensemble.

<sup>4</sup>As a side note, it is also important that the ensemble average of a property, as well as the most probable value are sufficiently close. If they are not, there may be stretches of time over which the average may actually lie closer to the most probable value, not the ensemble average

where  $W$  is the “number of states in the ensemble” that conform to the macroscopic property of having the same energy. This definition came from considerations on discrete systems, however, which is not well defined in Hamiltonian mechanics, since there is of course a continuum of states. The solution is to take  $W \propto \Omega$ , such that

$$W = \frac{1}{h^{3N}} \int d^{3N}x d^{3N}p = \frac{1}{\Omega} h^{3N} \quad (1.7)$$

where the constant  $h$  is of units  $[x][p]$ , so that  $W$  is unitless<sup>5</sup>. We will use the shorthand  $x$  and  $p$  to mean all phase space coordinates. Because we have taken the probability of every microstate to be the same, we can write the Boltzmann entropy as

$$S_B = -k_B \ln(h^{3N} \rho(x, p)) = \int_{\Omega} -k_B \rho(x, p) \ln(h^{3N} \rho(x, p)) d^{3N}x d^{3N}p, \quad (1.8)$$

because, by definition,  $\rho(x, p) = \Omega^{-1}$ . We used the short-hand notation  $x$  and  $p$  to mean all spatial coordinates and momenta. In this form it is often called the Boltzmann-Gibbs entropy.

#### 1.1.4 Extensivity of the microcanonical Entropy and Temperature

The thermodynamic definition of the entropy requires the entropy to be extensive. We can see that Boltzmann’s entropy for the microcanonical ensemble is trivially extensive for two uncorrelated systems. This is because then the total number of states  $W_{12} = W_1 W_2$  and thus the entropy  $S_{12} = k_B \ln W_1 + k_B \ln W_2$ . However, if we let the two systems make thermal contact, they are allowed to exchange energy, while keeping the total energy at  $E_T = E_1 + E_2$ . The total number of states must then be  $W_{12}(E_t) = \sum W_1(E_i) W_2(E_t - E_i)$ , which would mean that the entropy does not separate into entropies of the separate systems. However, in the thermodynamic limit (when the systems are very large), the contribution from most likely state becomes exceedingly large compared to the other states. Thus we say that  $W_{12}(E_t) \approx W_1(E'_1) W_2(E'_2)$ , which implies that  $k \ln(W_{12}(E_t)) \approx k_B \ln(W_1(E'_1)) + k_B \ln(W_2(E'_2))$ . This state is such that  $\delta \ln \Omega(E_1) \Omega_2(E_2) = 0$ , or equivalently  $\delta \ln \Omega(E_1) \Omega_2(E_2) = 0$ , since the logarithm is a monotonic increasing function. This corresponds to the systems having definite energy and being uncorrelated w.r.t. each other. Then, due to additivity of energy  $E_1 + E_2 = E$ , thus  $dE_1 = -dE_2$ , this leads to the condition

$$\frac{\partial \ln \Omega(E)}{\partial E_1} - \frac{\partial \ln \Omega(E)}{\partial E_2} = 0. \quad (1.9)$$

From this we see then

$$\frac{\partial \ln \Omega_1(E_1)}{\partial E_2} = \frac{\partial \ln \Omega_1(E_2)}{\partial E_2}, \quad (1.10)$$

or

$$\frac{\partial S_1(E_1)}{\partial E_1} = \frac{\partial S_1(E_2)}{\partial E_2} \quad (1.11)$$

which is the condition for equilibrium of the two systems in the thermodynamic limit. This defines two energies  $E'_1$  and  $E'_2$  with which the two systems stay in equilibrium. We define the temperature as

$$\frac{1}{T} \equiv \frac{\partial S_{1/2}}{\partial E_{1/2}}. \quad (1.12)$$

To illustrate extensivity of the entropy in the thermodynamic limit, we will present a simple example. Imagine  $N$  non interacting spins in a magnetic field, so that their energy is  $\epsilon$ , when they are spin up, and 0 when they are spin down. If the total energy is  $E = n\epsilon$ , then the total number of states is

<sup>5</sup>There has been some discussion on whether this definition is actually correct [12], [13]. Some claim that, instead, the correct entropy is  $k_B \ln \Sigma(E)$ , where  $\Sigma(E)$  is the phase space volume of all states with energy between zero and  $E$ , others claim they are equivalent. This discussion goes beyond the scope of this thesis however, and does not further impact the conclusions of this section.

$W = \binom{N}{n}$ . If we let two identical systems make thermal contact, then  $W_{12} = \binom{2N}{2n}$ . We now take the thermodynamic limit,  $N \rightarrow \infty$ ,  $n \rightarrow \infty$ ,  $n/N = c$ . Setting  $k_B = 1$ , it can be verified using Stirling's formula that,

$$S_{12} \approx 2N \ln N - n \ln n - 2(N - n) \ln(N - n) \approx \ln \binom{N}{n}^2 = S(W^2) = 2S(W). \quad (1.13)$$

Thus in the thermodynamic limit the entropy separates again into definite entropies of the two subsystems. Because in this case the two sub-systems are identical, the entropy divides nicely into two times the entropy of a system with  $N$  spins and energy  $n$ . As a consequence, when the systems are sufficiently large, we may think of them as being uncorrelated, even when in thermal contact, since the number of states is approximately the same regardless of being in contact or not.

### 1.1.5 The Canonical Ensemble

We will now consider a closed system (meaning no matter exchange) that is thermally coupled to a much larger system called a “heat bath”. Such a heat bath is presumed to be sufficiently large, so that any approach to equilibrium of the system doesn't change the energy of the heat bath. The system and the heat bath together are assumed to be isolated, so that the total system can be viewed as a microcanonical ensemble with a fixed total energy  $E_T$ . By the assumption of the microcanonical ensemble, the probability that the system has energy  $E$ , while the heat bath has energy  $E_T - E$ , is proportional to the number of microstates which conform to that macrostate. Thus we have that the probability of this state  $p \propto \Omega(E, E_T - E)$ . This also implies that every state with energy  $E_T$  is admissible, meaning  $\Omega(E, E_T - E) = \Omega_S(E)\Omega_B(E_T - E)$ . Assuming that the heat bath is much bigger than the system and that  $E \ll E_T - E$ , we can write

$$\begin{aligned} k \ln \Omega(E, E_T - E) &= k \ln \Omega_B(E_T - E) + k \ln \Omega_S(E) \\ &\approx k \ln \Omega_B(E_T - E) = S_B(E_T) - E \frac{dS_B(E)}{dE} + \mathcal{O}(E^2). \end{aligned} \quad (1.14)$$

Thus,

$$\Omega(E, E_T - E) \approx \exp \left[ \frac{1}{k} S_B(E_T) \right] \exp \left[ - \frac{1}{k} \frac{dS_B(E)}{dE} E \right]. \quad (1.15)$$

This means that we can write

$$\rho(E) = \frac{1}{Z} \exp \left[ - \beta H(x, p) \right], \quad (1.16)$$

where we absorbed the first exponent in the normalization  $1/Z$ .  $\rho(E)$  is the probability density at energy  $E$  and we defined  $\beta = dS_B/dE$ . The “temperature”  $\beta = 1/kT$  is the first order change in the “microcanonical entropy” of the heat bath, when we change the energy. From the normalization requirement  $\int \rho(x, p) dx dp = 1$  we see that

$$Z = \int e^{-\beta H(x, p)} dx dp. \quad (1.17)$$

### 1.1.6 The Canonical Entropy

To derive the entropy of this ensemble, we have to invoke the usual thermodynamic relations. We start out by making the identification  $U = \langle E \rangle = \int E \rho(E) dp dx$ . It can be easily seen that

$$\langle E \rangle = - \frac{\partial}{\partial \beta} \ln \left( \frac{Z}{h^{3N}} \right) \quad (1.18)$$

The quantity  $\frac{Z}{h^{3N}}$  is called the partition function. Whereas we may again think about  $Z$  as the (weighted) phase space area, the partition function may be thought of as the relation between this

area and the number of states. Using  $U = \partial(\beta F)/\partial\beta$ , this determines the Helmholtz free energy to be

$$F = -\frac{1}{\beta} \ln \left( \frac{Z}{h^{3N}} \right), \quad (1.19)$$

and through  $S = -(\partial F/\partial T)_V$ , also the entropy

$$S = k \ln \left( \frac{Z}{h^{3N}} \right) + \frac{1}{T} \langle E \rangle. \quad (1.20)$$

By using the expression for  $\rho$  from equation 1.16, it can be seen that this is equal to

$$S = - \int \rho(x, p) \ln (h^{3N} \rho(x, p)) dx dp. \quad (1.21)$$

### 1.1.7 Closing Remarks

#### The Boltzmann and the Gibbs Entropy

The entropies we have seen in this chapter were all defined on the ensembles in phase space. However, Boltzmann's original thinking was actually quite different than the interpretation we gave, which is Gibbs' interpretation. Still, Boltzmann's way of thinking about entropy seems to persist, and is sometimes conflated with Gibbs' view, which may be a source of confusion. As such we will briefly discuss the differences.

Boltzmann original formula for the entropy was defined in his H theorem, which we discussed in the introduction. He defined the entropy to be

$$S_{B,\rho} = -k_b N \int_{\Omega} \rho(\vec{x}, \vec{p}) \ln \rho(\vec{x}, \vec{p}) d^3x d^3p, \quad (1.22)$$

where  $\rho(x, p)$  is the marginal distribution<sup>6</sup> of a single particle. Due to criticisms, he later came up with his, now famous, entropy formula  $S_B = k \ln W$  which is defined as the logarithm of the number of microstates that make up a macrostate. This entropy is now defined on the total phase space, but he thought of entropy as measure of the likelihood of a macroscopic state of a single system, not as a measure on an ensemble of systems [14]. If the entropy of a state is larger, then it is simply more likely because there are more ways to compose that same state. An approach to equilibrium is then the same as a system evolving towards the most likely state, which is the state of the maximal entropy. Gibbs' view of ensembles and entropy allowed us to directly insert assumptions on the behavior in phase space of the system, like ergodicity. In Gibbs' view, the entropy is a measure in phase space, not of a single system, which is now the conventional view. However, there is some overlap between the functions. For non-interacting particles, we have that the total probability density separates,  $\hat{\rho}(\vec{x}, \vec{p}) = \hat{\rho}_1(\vec{x}_1, \vec{p}_1) \cdots \hat{\rho}_N(\vec{x}_N, \vec{p}_N)$  into the probability densities of all separate particles. If the particles are all the same, we simply have that the Gibbs entropy is equal to the Boltzmann entropy [15]. Moreover, for the microcanonical ensemble, we define the Boltzmann macrostate as the state with energy  $E$ . Thus all microstates are all states with energy  $E$ . In Gibbs' view, we assumed all probabilities of every state to be equal, due to which the entropy becomes the same as Boltzmann's formula.

#### The Gibbs Paradox

As a simplification we have up to now ignored what is called "Gibbs' paradox". Thought experiments pertaining to the mixing of identical gases made it clear that, for the entropy to be in accordance

<sup>6</sup>If we have a distribution of multiple variables, then a marginal distribution is defined by summing or integrating over a subset of variables. E.g. from the distribution  $\rho_{12}(x_1, x_2)$  we may construct the marginal distributions  $\rho_{1/2}(x_{1/2}) = \int dx_{1/2} \rho_{12}(x_1, x_2)$ .

with experiment, the number of states  $W$  or  $Z$  should be a factor  $1/N!$  smaller for identical particles<sup>7</sup>, when calculated from the phase space volume. Additionally, this makes the entropy extensive when composing systems. Classically there was no clear fundamental reason for this, because it doesn't change the equations of state. This correct counting of states is called Boltzmann counting, and is nowadays understood through quantum mechanics, from which we know that the systems have an inherent permutation symmetry of identical particles. We have thus over estimated the number of states by integrating over every permutation of position and momentum labels individually. It can be seen that this correct counting of states causes the entropy to be smaller by an amount of  $-\ln(N!)$ .

### Discrete systems

Statistical mechanics has also found a lot of applications in systems where the number of states are inherently discrete. The most obvious example is quantum statistical mechanics. In discrete systems, the ambiguity of counting states is eliminated. This means there is no need for a  $N!$  factor nor for the constant  $h$ . From similar considerations as we used above, the entropy is then defined as

$$S = -k_B \sum p_i \ln p_i, \quad (1.23)$$

for both ensembles, where  $p_i$  is the probability of a microstate.

## 1.2 The Shannon Entropy

In this section we will introduce Shannon's entropy, which is the central quantity in the field on information theory

$$S = - \sum p_i \ln p_i. \quad (1.24)$$

where  $p_i$  are the probabilities of some distribution. This entropy of course bears a striking resemblance of the Gibbs entropy that we introduced in the previous section, the only difference being the multiplicative constant  $k_B$ . The derivation of Shannon's entropy was not from any a-priori physical considerations, however, and was developed as a measure, quantifying the uncertainty we have about the outcome of a random event. However, due to their resemblance this entropy is often also called the Boltzmann-Gibbs-Shannon, or, BGS entropy. Shannon [6, §6] formulated the following axioms as "reasonable" requirements for a quantity meant to measure the amount of uncertainty in the outcome of a random event:

1.  $S$  is a continuous function of the variables  $p_i$
2. If all  $p_i$ 's are equal,  $A(n) = S(1/n, \dots, 1/n)$  is a monotonic increasing function of  $n$
3. The entropy is invariant under how probabilities are grouped. If for example the probabilities  $p_1$  to  $p_i$  add up to  $\lambda$  and the probabilities  $p_{i+1}$  to  $p_n$  add up to  $\kappa = 1 - \lambda$ , then

$$S(p_1, \dots, p_i, p_{i+1}, \dots, p_n) = S(\lambda, \kappa) + \lambda S\left(\frac{1}{\lambda} p_1, \dots, \frac{1}{\lambda} p_i\right) + \kappa S\left(\frac{1}{\kappa} p_{i+1}, \dots, \frac{1}{\kappa} p_n\right) \quad (1.25)$$

It can then be shown that the Shannon entropy 1.24 is the only function which satisfies these requirements [6, §6].

### 1.2.1 Comparison with the Standard deviation

Anyone familiar with quantum mechanics will be used to the idea of measuring uncertainty using the standard deviation of a given operator for a given state. The entropy at first might seem like a

---

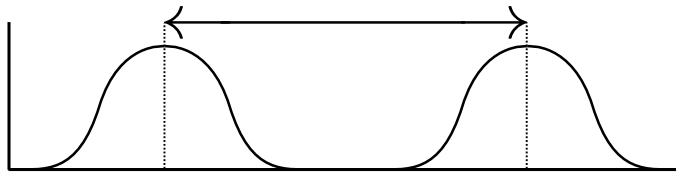
<sup>7</sup>If there are multiple types of particles, then the number of states gets a factor  $1/N_i!$  for every type of particle, where  $N_i$  is the number of particles of a specific type. This is because there is a permutation symmetry only between particles of that type. Different types are always distinguishable from each other.

very similar quantity, as it is also used as a measure of uncertainty in a distribution. We will see in this section that the two quantities are in fact very distinct, and measure uncertainty in a different way. Recall that the standard deviation of a distribution on a random variable  $X$  can be calculated from

$$\sigma_X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2}, \quad (1.26)$$

where  $\langle \cdot \rangle$  denotes the expectation value of a random variable. We see that the standard deviation is measured in units of the random variable. If we keep the distribution itself fixed (the probabilities corresponding to physical values) but change the physical values, the standard deviation changes accordingly. Of course the standard deviation, by definition, is also not invariant under a change of units. The Shannon entropy in contrast only depends on the distribution itself, not the physical values. We will further exemplify the difference of the two measures using three examples.

- First, imagine a bi-modal distribution with (sufficiently) non-overlapping peaks. If we change the distance between the peaks, the Shannon entropy remains invariant, while the standard deviation does not. The reason for this is that the Shannon entropy is invariant under the relabeling of the numerical values of the outcomes, as well as adding zero probability outcomes.



- Secondly, consider two experiments where the outcomes can have the numerical values  $-1, 0$  and  $1$ . In experiment one, the probabilities of each outcome are equal. In the second experiment  $-1$  and  $1$  occur with  $p = 1/2$  while  $0$  occurs with zero probability. Intuitively, since the first experiment had more possible outcomes, it makes sense to say it would be more uncertain. However, since the standard deviation does not only depend on the probabilities but also on the numerical difference from the mean of each quantity, we have that the standard deviation is less for the first experiment  $\sigma_1 < \sigma_2$ . On the other hand we have  $S_1 > S_2$ .
- Lastly, the standard deviation is not able to measure the uncertainty in any process with non-numerical outcomes. Consider the neutrino flavor as an observable. If there is some process that generates a distribution in outcome of neutrino flavor, there is no way to describe the uncertainty using the standard deviation. On the other hand, the Shannon entropy only is a function of the probabilities and thus is capable of describing the uncertainty for such a distribution.

We conclude that the standard deviation can not be a satisfactory measure specifically of the uncertainty in the outcome of an experiment, due to its dependence on the physical values of the outcomes. The Shannon entropy instead offers us a way to measure the uncertainty of a distribution, independent of the physical units and values.

### 1.2.2 Shannon Entropy as a Measure of (Lack of) Information

In this section we will make more clear in what way the Shannon entropy can be seen as a measure of uncertainty, by presenting an example. Let's say we have a set  $M$ , consisting of a number of  $N$  states, partitioned into non overlapping subsets,  $M = M_1 \cup \dots \cup M_k$ . Suppose we then have a system which is in one of these states. Without any prior information, for a full characterization in binary, we would need  $\log_2(N)$  bits. Say we now learn, from a measurement, that the state of our system belongs to subset  $M_i$ . This kind of knowledge could be obtained in a real experiment when all states in a partition share some macroscopic quantity that we are measuring. Denoting the amount of states in the subset  $M_i$  as  $N_i$ , the remaining amount of bits we would need to characterize the state is

$$\log_2(N_i) = \log_2\left(N \frac{N_i}{N}\right). \quad (1.27)$$



Thus the amount of states that we still have to characterize has gone down by the fraction of the size of the subset  $M_i$  compared to the size of set  $M$ . This ratio is also the probability that the actual microstate of the system is in the set  $M_i$ , thus let us denote  $N_i/N = p_i$ . We can then write 1.27 as

$$\log_2(N) + \log_2(p_i). \quad (1.28)$$

$p_i$  is smaller than one, so the right term is negative. Moreover, the left term was the information we needed without any prior knowledge. The right term can therefore be interpreted as the information we gained by knowing what subset our state was in. As such, we can express knowledge about the partitioning of our system as a quantity in bits, independent of the actual number of microstates.

By taking minus the average of this quantity, we can express the average amount of information we gain about the partitioning of a system by measurements as the positive quantity

$$-\sum p_i \log_2(p_i). \quad (1.29)$$

All our considerations have been based on representing the information in an amount of bits, from which the use of the  $\log_2$  stems. This choice, while useful for computer science, is of course arbitrary. We could have chosen ternary (3 numbers), using  $\log_3$ , quaternary (4 numbers), using  $\log_4$ , etc. Moreover, the  $\log_x$  is continuous in  $x$  so, while more abstract, we can more generally represent information in “base  $x$ ”. The choice of  $x = e$  represents the information in the number of “nats”.

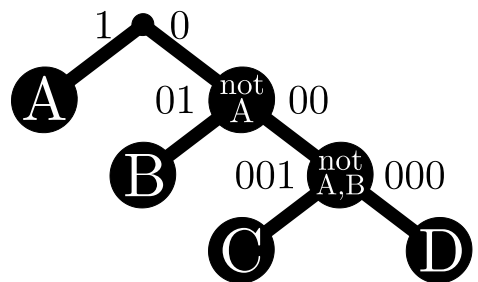
We have so far seen how it makes sense to think of the Shannon entropy as a measure of the information that a random process produces. However, we can make these claims more precise using Shannon’s source coding theorem [16, §5.2]. This states that the Shannon entropy gives the lower limit on how much we can losslessly compress the data that represents the outcomes of a random process, in the limit where the relative frequencies of the outcomes go to the probabilities. So not only does it makes sense to think of the Shannon entropy as the amount of information that a random process produces, it sets the actual limit on representations of this data. We will illustrate this using an example. In our example, our random source will produce four different letters,  $A, B, C$  and  $D$ , with probabilities  $p_A = 1/2, p_B = 1/4, p_C = 1/8$  and  $p_D = 1/8$  respectively. In such a case, where all probabilities are powers of 2, the shortest way to represent all letters is to use Huffman code. The Huffman code for general problems may be represented by so called “tree diagrams”, and is quite simple to construct for our problem. First of all we put all the letters on nodes, and connect the two nodes with the least probability to a common node and add the probabilities. We repeat this process until all nodes are connected. Assigning 1 to a left turn and 0 to a right turn, the code of our letter is then found traversing the path from the top node in the tree diagram to the node of the letter. The tree diagram for our problem is shown in figure 1.1. Thus we see that  $A, B, C$  and  $D$  are assigned 1, 2, 3 and 3 bits, or equivalently  $\log_2(2^1), \log_2(2^2), \log_2(2^3)$  and  $\log_2(2^3)$  bit, respectively. If we now send a message of  $N$  letters, we can denote the number of bits sent, by multiplying the times we sent a certain letter by the number of bits used to send that letter. By averaging the number of bit over the total number of letters sent  $N$ , we get

$$\langle bits \rangle = \frac{N_A}{N} \log_2(2^1) + \frac{N_B}{N} \log_2(2^2) + \frac{N_C}{N} \log_2(2^3) + \frac{N_D}{N} \log_2(2^3). \quad (1.30)$$

If we take the limit where  $N \rightarrow \infty$ , the relative frequencies  $N_X/N$  of the letters converge to their probabilities,  $p_X$ . In this limit we see that the average amount of bits becomes

$$\langle bits \rangle = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{4} \log_2 \frac{1}{4} - \frac{2}{8} \log_2 \frac{1}{8} = \sum_X -p_X \log_2 p_X, \quad (1.31)$$

Figure 1.1: The Huffman Compression Diagram



which is precisely the Shannon entropy in base 2.

### 1.2.3 Statistical Mechanics and Information Theory: Jaynes' Maximum Entropy Principle

Following the advances made in information theory, Jaynes published two papers viewing Statistical Mechanics in light of the ideas of this new theory [7],[17]. In his paper he re-frames Statistical Mechanics as a theory about statistical inference, as opposed to a theory based on physical arguments. What this means is that, instead of finding arguments about the physical behavior of a system to derive probabilities, we try to find the distribution that “best reflects our current state of knowledge”. Information theory tells us this should be the distribution with the largest amount of uncertainty, e.g. entropy, that still reflects this knowledge. In practice this means using maximization techniques on the entropy with as constraint normalization and some expectation value.

In this section we will briefly discuss the implications of Jaynes' work and show the power of entropy maximization. The latter is something that we will see come back, as a lot of authors show interest in finding the “most uncertain” distribution, given some measure of uncertainty. Since in these cases there is no a-priori physical theory, the meaning and usefulness of these distributions in physics, is something that can only be judged in hindsight. In any case, maximization of the Shannon entropy does nicely reproduce the distributions of statistical mechanics. Jaynes goes more in depth into the subject, so we refer the interested reader to his papers. We will, following his paper, only show maximization of discrete distributions, although it can easily be generalized to continuous distributions.

Let  $X$  be a discrete random variable, taking the values  $x_i$ . The problem we want to solve is: estimate the probabilities  $p_i$  corresponding to the values  $x_i$ , given the constraint of a known expectation value of some function dependent on  $X$

$$\langle f(X) \rangle = \sum_i p_i f(x_i) \quad (1.32)$$

with the normalization condition

$$\sum p_i = 1. \quad (1.33)$$

Whether or not a problem which such limited information seems useful to try to solve might depend on one's view on probabilities. This is where Jaynes distinguishes between the objective and subjective schools of thought regarding probabilities. The objective school of thought views probabilities as objective properties of the corresponding events. These must be dictated by some underlying theory and are in principle verifiable in every detail. The use of the ergodic hypothesis in section 1.1.2 is an example of this. Another one, which we will investigate further in section 6.3.1, is the position of a harmonic oscillator with a given energy but unspecified initial position. The chance to find the particle at any moment is dictated by the time average spend in any location, which is imposed by the underlying theory. To someone conforming to this school of thought, the above problem seems unsolvable without any additional information, perhaps from an underlying theory.

Someone adhering to the subjective school of thought, on the other hand, sees the probabilities as expressions of our ignorance. A good probability distribution should correctly represent our current state of knowledge. We can then formulate the posed problem as a problem of finding the distribution that represents the known information with the least amount of external bias. A problem like this is exactly one that information theory can guide us through. According to Shannon, the entropy is a unique and unambiguous measure for the “amount of uncertainty” we have about the outcome of an event that is described by some distribution.

To maximize the Shannon entropy w.r.t. to the constraints 1.32 and 1.33, we may equivalently maximize the following function

$$F(p_1, \dots, p_n) = - \sum_i p_i \ln p_i - \lambda \sum p_i - \mu \sum f(x_i) p_i, \quad (1.34)$$

where  $\lambda$  and  $\mu$  are Lagrange multipliers. To find the probabilities that maximize this function, we can employ the Euler-Lagrange equations because the dependence on the probabilities is continuous. We thus must solve the set of equations

$$\frac{dF(p_1, \dots, p_n)}{dp_i} = 0. \quad (1.35)$$

We find the maximum entropy probabilities

$$p_i = \exp(-\lambda - \mu f(x_i)), \quad (1.36)$$

The multipliers may be determined by substituting in the constraints 1.32 and 1.33. By imposing the normalization condition we find

$$e^\lambda = \sum_i e^{-\mu f(x_i)} \equiv Z(\mu), \quad (1.37)$$

where  $Z(\mu)$  is the familiar partition function. We see that

$$\langle f(X) \rangle = -\frac{\partial}{\partial \mu} \ln Z(\mu) \quad (1.38)$$

We may now specify a numerical value for  $\langle f(X) \rangle$ , or we may equivalently let  $\langle f(X) \rangle$  be determined by specifying a value for  $\mu$ . By making the identification  $\mu \equiv \beta$  and  $f(x_i) = E_i$ , i.e. the energy of outcome  $x_i$ , we see that the probabilities are given by

$$p_i = \frac{1}{Z} e^{-\beta E_i} \quad (1.39)$$

which is the canonical ensemble of statistical mechanics. If we were to relax constraint 1.32, we would simply find that the probabilities must be constant, in agreement with the microcanonical ensemble. The advantage of this method is that a generalization to multiple constraints, as well as to non-energy constraints (such as pressure for example) is trivial. For a general set of constraints of the form

$$\langle f_n(X) \rangle = \sum_i p_i f_n(x_i) \quad (1.40)$$

we have the maximum entropy distribution given by

$$p_i = \frac{1}{Z} \exp\left(-\sum_{n=1} \lambda_n f_n(x_i)\right) \quad (1.41)$$

where

$$Z = \sum_i \exp\left(-\sum_n \lambda_n f_n(x_i)\right) \quad (1.42)$$

and additionally

$$\langle f_n(X) \rangle = -\frac{\partial}{\partial \lambda_n} \ln Z, \quad (1.43)$$

## 1.2.4 Properties

In this section we will discuss the most important mathematical properties of the Shannon entropy. We will not prove these properties ourselves, but instead refer to the book by Cover for the proofs [16, §2]. Because the Von Neumann entropy builds upon the Shannon entropy, we will see in section 7.1.3 that a lot of the properties generalize to it. To make the analogy as close as possible, we will use the notation of probability vectors, to keep close to the mathematics of quantum mechanics. Let  $X$  and  $Y$  denote random variables being able to take the discrete values  $x_i$  and  $y_j$ , corresponding to the outcomes of measurements on systems 1 and 2. Then, let the probability vectors  $\vec{p}_1$  and  $\vec{p}_2$

define probability distributions on  $X$  and  $Y$  respectively. We define a tensor product-like operation  $(\vec{p}_1 \otimes \vec{p}_2)_{ij} \equiv (\vec{p}_1)_i (\vec{p}_2)_j$ , which defines a new probability vector  $\vec{p}_\alpha$  where  $\alpha \in \{(i, j)\}$ , describing the joint probability of events  $x_i$  and  $y_j$  happening. Similar to the tensor product in quantum mechanics, a probability distribution constructed this way is not the most general probability distribution on the random variables  $X$  and  $Y$ . A more general probability distribution can have correlations in  $X$  and  $Y$ . We will denote general probability distribution for multiple random variables as  $\vec{p}_{12\dots}$  with indices an index for every random variable. To obtain the probabilities for the outcomes of a single random variable we sum over the probabilities of the other random variable

$$(\vec{p}_1)_i = \sum_j (\vec{p}_{12})_{ij}. \quad (1.44)$$

Finally, we then denote the entropy as a “functional” of the probability vector as

$$S[\vec{p}] = - \sum_{i,j,\dots} p_{ij\dots} \ln p_{ij\dots} \quad (1.45)$$

### Simple Properties

The Shannon entropy is non-negative

$$S[\vec{p}] \geq 0. \quad (1.46)$$

### Concavity

The Shannon entropy is a concave function of all of its arguments. An illustration on a concave function of one argument can be seen in figure 1.2. Let  $\vec{p}$  and  $\vec{p}'$  be two probability distributions and let  $\lambda \in [0, 1]$ , then

$$S[\lambda\vec{p} + (1 - \lambda)\vec{p}'] \geq \lambda S[\vec{p}] + (1 - \lambda)S[\vec{p}'], \quad \forall \vec{p}, \vec{p}'. \quad (1.47)$$

The sum of probability vectors, as taken in the left side, is called a convex combination. Taking such a convex combination represents an experiment where we do not have information (except for the relative frequency) from which ensemble a particular measurement stems. As such, the entropy, as measure of lack of information, is higher than the average lack of information when we would have that information. Moreover, concavity ensures that the entropy can be maximized, and that the maximum entropy will always be the global maximum.

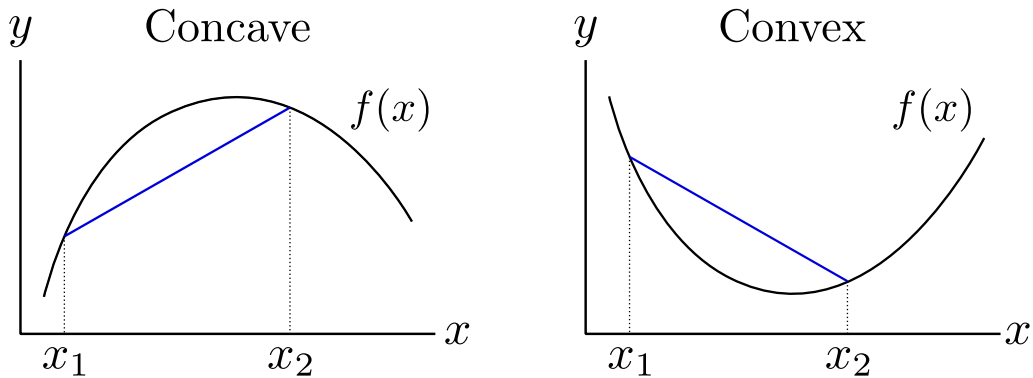


Figure 1.2: Illustrations of a concave and a convex function. The blue line represents the function  $\lambda f(x_1) + (1 - \lambda)f(x_2)$

### Additivity

The Shannon entropy is additive. If systems 1 and 2 are uncorrelated, thus  $\vec{p}_{12} = \vec{p}_1 \otimes \vec{p}_2$ , then

$$S[\vec{p}_{12}] = S[\vec{p}_1] + S[\vec{p}_2]. \quad (1.48)$$

Qualitatively this tells us that if two systems are uncorrelated, the uncertainty in outcomes of joint measurements on both systems is the same as the uncertainty of measurements on both systems separately.

### Sub-Additivity (SA)

The Shannon entropy is sub-additive

$$S[\vec{p}_{12}] \leq S[\vec{p}_1] + S[\vec{p}_2]. \quad (1.49)$$

The equality only holds when systems 1 and 2 are uncorrelated. This tells us that classical correlations always make the outcome of the total system more certain. Using this inequality the correlations can be quantified as an information quantity

$$I_{\text{correlations}} = S[\vec{p}_1] + S[\vec{p}_2] - S[\vec{p}_{12}] \geq 0 \quad (1.50)$$

### Strong-Subadditivity (SSA)

The Shannon entropy is strongly sub-additive

$$S[\vec{p}_{123}] + S[\vec{p}_2] \leq S[\vec{p}_{12}] + S[\vec{p}_{23}]. \quad (1.51)$$

This is, as the name implies, a stronger condition than SA, because SSA and additivity can be used to prove SA. This result has actually been studied much for the quantum generalization of the BGS entropy, in the context of quantum statistical mechanics and holography. We will discuss this property in a bit more detail in section 7.1.3.

## 1.3 The Differential (Continuous Shannon) Entropy

As a continuation of the Shannon entropy to continuous distributions, Shannon himself introduced the following quantity, most often called the differential entropy or sometimes just the Shannon entropy,

$$- \int_{-\infty}^{\infty} \rho(x) \ln(\rho(x)) dx, \quad (1.52)$$

where  $\rho$  is a normalized distribution. Shannon assumed this to be the correct continuation, without any formal proof. This entropy is in fact analogous to the BG entropies for continuous Hamiltonian systems. However, the quantity is not a direct analog, and it can not be directly derived from the Shannon entropy, which we will see in this section. Though the differential entropy inherits many of the properties of the Shannon entropy, it differs from it in a few regards. For one, due to the normalization condition  $\int \rho(x) dx = 1$ , probability densities get a particular scaling behavior, and, if  $x$  has units, probability densities will have units of  $[1/x]$ . If  $x$  has units, this means then that the differential entropy has units of  $[\ln 1/x]$ . Thus if we want the differential entropy to be a unitless quantity, we have to introduce an arbitrary quantity  $L$ , with units of  $[x]$ , within the logarithm, analogous to the factor of  $h^{3N}$  in the BG entropy. We write proper, unitless differential entropy as

$$S[\rho(x)] = - \int \rho(x) \ln(L\rho(x)) dx. \quad (1.53)$$

The rest of the properties, and how they differ from the Shannon entropy are discussed in section 1.3.1

In the remainder of this section, we show a method to “derive” the differential entropy from the Shannon entropy [16]. We use the quotation marks because, as we will see, the differential entropy is not just the continuous extension of the Shannon entropy. We do so by taking the Shannon entropy of a continuous distribution, averaged over intervals, called bins. We then proceed by taking the limit of the Shannon entropy where the bin size goes to zero. We divide a continuous distribution up into equally sized bins, such that the bin sizes are  $\delta x$

$$\int_{i\delta x}^{(i+1)\delta x} \rho(x) dx = p_i. \quad (1.54)$$

As we let  $\delta x$  go to zero we can approximate any  $p_i$  arbitrarily close by

$$p_i = \rho(x_i)\delta x, \quad (1.55)$$

where  $x_i \in [i\delta x, (i+1)\delta x]$ . The Shannon entropy then is arbitrarily closely approximated by

$$\begin{aligned} -\sum \rho(x_i)\delta x \ln(\rho(x_i)\delta x) &\approx -\int dx \rho(x) \ln(\rho(x)\delta x) \\ &= -\int \rho(x) \ln(L\rho(x)) - \ln\left(\frac{\delta x}{L}\right), \end{aligned} \quad (1.56)$$

where the  $L$  is some constant of units  $[x]$ . The  $L$  was introduced to keep the term in the logarithm dimensionless. We see, however, that as  $\delta x \rightarrow 0$ , the last term diverges. The differential entropy should therefore not be regarded as the actual Shannon entropy of a continuous distribution. Rather, it is a quantity that inherits some of the properties of the Shannon entropy, while differing from it by a infinite divergence. However, this term is only a constant divergence so it is usually disregarded. In fact, as we will see, the differential entropy still has a lot of properties in common with the discrete Shannon entropy, so it will behave similarly for maximization purposes. Lastly, the  $L$  term is often ignored in scientific literature because, when setting all units to 1, it only produces a constant offset.

This derivation can also be generalized to arbitrary dimension by taking course graining in rectangular boxes over a surface, volume, hyper-volume, etc. and taking the discrete Shannon entropy. In the limit we obtain a divergent term for every dimension. If we ignore every divergent term, we obtain

$$S = \int \rho(x_1, \dots, x_n) \ln\left(\rho(x_1, \dots, x_n) \prod_i L_i\right) d^n x \quad (1.57)$$

### 1.3.1 Properties

#### Simple Properties

The differential entropy is not non-negative, unlike the Shannon entropy. When a distribution becomes sufficiently peaked, such that in a particular region  $\rho(x)L > 1$ , that specific region gives a negative contribution. If the distribution is localized sufficiently within a region of size  $L$  then the entropy is negative altogether. In some situations, the constant  $L$  may be thought of as setting a smallest scale, or unit cell, the details of which we will discuss further in the next chapter. In these cases, distributions that are localized within a cell are uninteresting, or unphysical, which is then reflected by the negative entropy.

The differential entropy is not invariant under re-scaling of the integration coordinates. The origin of this, is that due to the normalization condition  $\int \rho(x) dx = 1$ , the probability density inherits the scaling of the integration measure. So, if we let  $x \rightarrow x/a$ , the density itself will transform as

$$\rho(x) \rightarrow \frac{1}{a} \rho\left(\frac{x}{a}\right), \quad (1.58)$$

The differential entropy of the transformed probability density becomes

$$\begin{aligned}
S\left[\frac{1}{a}\rho\left(\frac{x}{a}\right)\right] &= -\int \frac{1}{a}\rho\left(\frac{x}{a}\right) \ln\left(\frac{1}{a}\rho\left(\frac{x}{a}\right)\right) dx \\
&= -\int \frac{1}{a}\rho\left(\frac{x}{a}\right) \ln\rho(x) dx + \int \frac{1}{a}\rho\left(\frac{x}{a}\right) \ln(a) dx \\
&= S[\rho(x)] + \ln(a)
\end{aligned} \tag{1.59}$$

### Concavity

The differential entropy is concave [18],

$$S[\lambda\rho_1(x) + (1-\lambda)\rho_2(x)] \geq \lambda S[\rho_1(x)] + (1-\lambda)S[\rho_2(x)], \tag{1.60}$$

which is easily checked.

### Additivity

The differential is additive. If two random variables are independent, i.e. the density can be written as  $\rho(x, y) = \sigma(x)\tau(y)$ , then

$$S[\sigma(x)\tau(y)] = S[\sigma(x)] + S[\tau(y)]. \tag{1.61}$$

### Subadditivity

The differential entropy is subadditive [16, Thm. 8.6.2]

$$S[\rho(x_1, \dots, x_n)] \leq \sum S[\rho(x_i)], \tag{1.62}$$

where

$$\rho_i(x_i) = \int \rho(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \tag{1.63}$$

### Multi-modal Distributions

We discussed that the Shannon entropy is invariant under scaling of the distance between multiple peaks of a distribution. An analogous property holds for the differential entropy. Let  $f(x)$  represent some peaked distribution. We construct the bi-modal distribution  $\rho(x) = \frac{1}{2}(f(x-a) + f(x+a))$ . The differential entropy is then given by

$$\begin{aligned}
S[\rho(x)] &= \int_{-\infty}^0 \frac{1}{2}(f(x-a) + f(x+a)) \ln\left(\frac{1}{2}(f(x-a) + f(x+a))\right) dx \\
&\quad + \int_0^{\infty} \frac{1}{2}(f(x-a) + f(x+a)) \ln\left(\frac{1}{2}(f(x-a) + f(x+a))\right) dx \\
&\approx \frac{1}{2} \int_{-\infty}^0 f(x+a) \ln f(x+a) dx + \frac{1}{2} \int_0^{\infty} f(x-a) \ln f(x-a) dx.
\end{aligned} \tag{1.64}$$

The last line holds approximately if  $f(x)$  goes sufficiently quick to zero compared to  $a$ , or exactly when  $a \rightarrow \infty$  or when  $\rho(x)$  goes exactly to zero in between the peaks<sup>8</sup>. The same can obviously be shown for multiple peaks, as long as the overlap between the peaks goes to zero.

<sup>8</sup>Like is the case for the bump function  $e^{-\frac{1}{1-x^2}}$ .

### 1.3.2 The Differential Entropy and the Standard Deviation

In section 1.2.1 we discussed how the Shannon entropy and standard deviation differ from each other. Most of the discussion carries over directly to the continuous case. For bi-modal distributions (given they are made of two non-overlapping parts) the standard deviation scales with the distance of the peaks while the differential entropy does not. Moreover, the differential entropy may still be regarded as invariant under “permutation of labels” in some sense. If we discontinuously cut up a distribution into finite intervals and permute the parts of the distribution w.r.t. the  $x$  intervals, the differential entropy is still invariant. However, for specific family of functions parameterized by their standard deviation, there may now be mappings from the entropy to the standard deviation. For example for the family of normal distributions with standard deviation  $\sigma$ , the differential entropy will be, with  $L = 1$ ,

$$S[f(x, \sigma)] = \frac{1}{2} \ln(\pi e \sigma) = \frac{1}{2} + \frac{1}{2} \ln(\pi \sigma). \quad (1.65)$$

However, the differential entropy is generally not just the logarithm of the standard deviation. For example, the (unitless) Cauchy distribution  $1/(\pi[x^2 + 1])$  has an undefined standard deviation since

$$\int_{-\infty}^{\infty} \frac{x^2}{\pi(x^2 + 1)} dx, \quad (1.66)$$

diverges. However, the differential entropy is well defined

$$-\int_{-\infty}^{\infty} \frac{1}{x^2 + 1} \ln \frac{1}{x^2 + 1} dx = \pi \log(4). \quad (1.67)$$

Moreover, let us take a distribution with a well defined standard deviation

$$\frac{1}{\pi \sigma} \frac{1}{\left(\frac{x}{\sigma}\right)^4 + 1}. \quad (1.68)$$

If we calculate the differential entropy we get

$$\begin{aligned} -\int_{-\infty}^{\infty} \frac{1}{\pi \sigma} \frac{1}{\left(\frac{x}{\sigma}\right)^4 + 1} \ln \left( \frac{1}{\pi \sigma} \frac{1}{\left(\frac{x}{\sigma}\right)^4 + 1} \right) dx &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{x^4 + 1} \ln \left( \frac{1}{x^4 + 1} \right) dx + \ln(\pi \sigma) \\ &\approx 0.3597 + \ln(\pi \sigma). \end{aligned} \quad (1.69)$$

Thus we establish a different relation between the standard deviation and the entropy for the normal distribution and the above distribution. We conclude that the entropy and standard deviation are not equivalent, though there exist mappings between them, depending on the family of functions.

## 1.4 Summary and Concluding Remarks

In this chapter, we have discussed the classical entropies of thermodynamics, statistical mechanics and information theory. We have seen Clausius’ original definition,  $dS = \frac{\delta Q}{T}$ , and the continuation of that quantity in statistical mechanics, as a measure on ensembles of identical systems, which live in phase space. This continuation takes the form of the well known Boltzmann entropy  $S = k_B \ln W$  in the microcanonical ensemble, and more generally the BG entropy  $S = \int_{\omega} -\rho \ln \rho dx dp$  for either ensemble. At the end of that section we briefly discussed the Gibbs paradox, and the difference between Boltzmann’s and Gibbs’ thinking of the entropy.

After that we discussed the discrete version of the thermodynamical entropy, only from a completely different standpoint, namely that of Shannon. In this standpoint, the entropy is a measure of uncertainty of the outcome of some random process. It was axiomatically defined by Shannon based upon reasonable axioms for such a quantity. We then contrasted the standard deviation and the entropy. After that we discussed mathematically how the entropy may be thought of as describing the information produced in a random process as a number of bits. Moreover, Shannon’s coding



theorem shows that this number of bits puts an actual lower bound, in terms of data compression, on the information produced by the source.

Next, we showed where information theory and statistical mechanics meet, using Jaynes' principle of maximum entropy. From Jaynes' standpoint, statistical mechanics becomes a theory of statistical inference, and devoid of a-priori physical assumptions. In this view, we are, first and foremost, trying to find the distribution that contains the least amount of external bias, given the information we already know of the system. In practice this takes the form of maximizing the entropy given a set of constraints.

After that we discussed the properties of the Shannon entropy. We have discussed and explained most of these properties in a way that can intuitively be understood from the standpoint that the Shannon entropy is a measure of uncertainty. In specific fields of research, these properties may be very important to formulate proofs of other properties of these entropies.

Lastly we discussed the differential entropy, and showed how it may be "derived" from the Shannon entropy. We have, like for the Shannon entropy, also discussed its properties and made a comparison between it and the standard deviation. The properties made clear that the differential entropy is not as easy to interpret as an amount of bits, since it is not non-negative and not invariant under scaling, like the Shannon entropy is. In a sense this brings it closer as a measure to the standard deviation. This analogy may be taken further by the fact that for some parameter family of functions, there exists mappings between the standard deviation and the entropy. In fact, for Gaussians, for example, we have simply that  $\sigma = 1/(\pi e) \exp(2S)$ . While the direct interpretation in terms of information is lost, the differential entropy may be interpreted as a measure of the spread of a distribution and is still an important quantity in information theory [16]. There is however another correspondence between the Shannon entropy and the differential entropy, which we will discuss in the next chapter. By introducing a grid, the interpretation in terms of bits may be recovered.

A notable omission of this text are the descendant quantities of the entropy, such as the conditional entropy and relative entropy. They enjoy various applications, from statistical inference to cryptography. For a very brief review, the work by Witten is recommended [19], and for a more thorough review, we recommend the book by Cover and Thomas [16].

## Chapter 2

# Course Graining Classical Distributions

In physics we concern ourselves very often with aspects of nature that may not be directly apparent on the most fundamental scale. These emergent properties are, for example, the basis of the thermodynamical description of gases. However as we have seen, such properties are not trivially derived from the microscopic behavior of the system. Somewhere along the line, to obtain our statistical model, we had to accept that we did not have perfect information about the system we were studying. The process of going from such an accurate and fine description of a system, to a coarser, more emergent one, is broadly called course graining.

However, more often than not coarse graining is something that we don't have under control. In general, the distribution produced by some random source is not exactly the distribution that we measure. This may either be due to finite resolution of our measuring apparatus or some information scrambling inherent in the measuring process. For example in Fourier optics, a source may emit some distribution of frequencies, but perhaps due to some de-focused lens, the final distribution we measure is somewhat smeared out. As a consequence, the entropy of the final distribution may in general be different than the one theoretically produced by the source. If this is the case, then the entropies will generally be different as well.

Since coarse graining in such a context means information loss, it would seem natural that the entropy would increase. It in fact does, which we will demonstrate in this section through a few simple models. From an information theoretic perspective this allows us to define the loss of information by coarse graining

$$I_{\text{lost}} = S - S_{\text{CG}} \quad (2.1)$$

Moreover, we will see that it allows us to find a connection between the BGS entropy and the differential entropy. By dividing a distribution into “bins”, we see that the BGS entropy of the course grained distribution is related to the differential entropy of the distribution itself. This also shows why it is possible to calculate the entropy of systems with an inherently discrete spectrum of states<sup>1</sup>, using a continuous entropy.

Lastly, coarse graining has been the central ingredient in some explanations of the second law of thermodynamics, which seems violated by the fact that the entropy we have defined in classical mechanics is invariant under time evolution [20, §23]. We will briefly review this in the last section of this chapter.

---

<sup>1</sup>That is, assuming that the thermodynamical systems, which are systems in some kind of box, or atleast with some characteristic volume, behave according to quantum mechanics.

## 2.1 Aggregation of Discrete Probabilities

As a simple example we will first show a course graining process that may be performed on discrete distributions. This represents some scrambling process from the source to the final measurement. We define a new probability distribution  $\vec{p}'$  where the probabilities are obtained by averaging over two probabilities of the previous distribution  $\vec{p}$

$$p'_a = \lambda p_a + (1 - \lambda)p_b, \quad p'_b = (1 - \lambda)p_a + \lambda p_b, \quad (2.2)$$

where  $\lambda$  is some for which we have  $0 \leq \lambda \leq 1$ . This effectively brings the probabilities of the events closer together, while keeping the average unaffected,  $\vec{p}' = \vec{p}$ . Thus we can write  $p_a = \bar{p} \pm \epsilon$  and  $p_b = \bar{p} \mp \epsilon$  and also  $p'_a = \bar{p} \pm \epsilon_2$  and  $p'_b = \bar{p} \mp \epsilon_2$ . Since  $\epsilon_2 \leq \epsilon$ , this directly implies

$$-p'_a \ln p'_a - p'_b \ln p'_b \geq -p_a \ln p_a - p_b \ln p_b \quad (2.3)$$

and thus the entropy will increase as a consequence. Under such operations the hidden information will always go up

$$I_{hidden} = S' - S_0 \quad (2.4)$$

Doing the above multiple times, we may define a new probability density,

$$p''_i = \sum T_{ij} p_j, \quad (2.5)$$

where  $T$  is some bi-stochastic matrix<sup>2</sup>. We again have  $S'' \geq S$ . We will use this result in section 2.3

## 2.2 Coarse Graining Continuous Distributions

The situation is a bit more complex for the continuous case. We will identify three different methods. First of all, in the simplest case we may go from a continuous distribution to an inherently discrete one, by summing over probabilities in regions, or so called bins. This may for example, directly represent the resolution of a measurement apparatus, which aggregates all the probabilities within a certain region of space to a single event. We may construct the probabilities by integrating over a bin  $B_i$ , which represents an interval in space,

$$p_i \equiv \int_{B_i} \rho(x) d^n x. \quad (2.6)$$

The bins need not have equal size in general. The advantage of this method is that we have a proper, dimensionless distribution, to which the BGS entropy may be applied. The standard interpretation of the BGS entropy may easily be applied in this case, such that the BGS entropy describes the uncertainty we have about in which bin we will measure an event. On the other hand, this entropy should not be directly compared to the differential entropy. For example, on a finite interval, the differential entropy is bounded from above by the uniform distribution, in which case the entropy is  $\ln(|I|L)$ , where  $|I|$  is the size of the interval. However, we may make the BGS entropy of the bins arbitrarily large by dividing the interval up into any number of bins, or make the distribution  $\vec{p}$  any shape by choosing the bin sizes.

The second option is similar, in the sense that we will still divide the distributions into bins, but we keep the distribution a function of a continuous variable  $x$ . What this means is that we define a bin-wise constant distribution

$$\rho_B(x) = \frac{1}{\Omega_i} p_i, \quad x \in B_i, \quad (2.7)$$

where  $\Omega_i$  is the size of bin  $i$ . The differential entropy of this distribution becomes

$$-\int \rho_B(x) \ln L^n \rho_B(x) d^n x = \sum_i \left( - \int_{B_i} \frac{p_i}{\Omega_i} \ln \left( L^n \frac{p_i}{\Omega_i} \right) d^n x \right) = - \sum_i p_i \ln \left( L^n \frac{p_i}{\Omega_i} \right). \quad (2.8)$$

---

<sup>2</sup>This means that all its rows add to 1, and as well do all its columns.

We can prove that the box averaged differential entropy is necessarily larger. Consider the following quantity, which is completely positive [21, §4.2.1],

$$S[\rho(x)|\rho'(x)] \equiv - \int \rho(x) \ln \left( \frac{\rho(x)}{\rho'(x)} \right) d^n x \geq 0. \quad (2.9)$$

If we set  $\rho' = \rho_B$  then we see

$$- \int \rho(x) \ln \left( L^n \rho_B(x) \right) d^n x \geq - \int \rho(x) \ln \left( L^n \rho(x) \right) d^n x \quad (2.10)$$

Then, because  $\rho_B$  is constant over the boxes, we may average  $\rho(x)$  over these boxes and obtain the same integral. Thus we can see that,

$$- \int \rho(x) \ln \left( L^n \rho_B(x) \right) d^n x = - \int \rho_B(x) \ln \left( L^n \rho_B(x) \right) d^n x = S[\rho_B]. \quad (2.11)$$

Therefore

$$- \int \rho_B(x) \ln \left( L^n \rho_B(x) \right) d^n x \geq - \int \rho(x) \ln \left( L^n \rho(x) \right) d^n x. \quad (2.12)$$

The differential entropy diverges as we take the bin sizes  $\Omega_i \rightarrow \infty$ . Conversely, the above equation becomes an equality as  $\Omega_i \rightarrow 0$ . We thus may approximate  $S[\rho(x)]$  with  $S[\rho_B(x)]$ . Next assume arbitrary box sizes, i.e. the size does not have to be constant over  $i$ . We have that

$$S[\rho_B] = - \sum p_i \ln p_i + \sum p_i \ln \left( \frac{\Omega_i}{L^n} \right) \equiv S[\vec{p}] + \ln \left( \frac{\bar{\Omega}}{L^n} \right), \quad (2.13)$$

where  $\bar{\Omega} = \prod \Omega_i^{p_i}$ . Note that since we may write the Shannon entropy of the discrete probabilities as follows

$$S[\vec{p}] = - \int \rho_B(x) \ln \left( \bar{\Omega} \rho_B(x) \right) d^n x, \quad (2.14)$$

we have from equation 2.12 that for small bin sizes

$$S[\vec{p}] \approx - \int \rho(x) \ln \left( \bar{\Omega} \rho(x) \right) d^n x \quad (2.15)$$

In other words, the differential entropy with  $L^n = \bar{\Omega}$  may be approximated by the course grained BGS entropy. Conversely, this means also that the BGS entropy of the distribution  $\vec{p}$  may be approximated by the differential entropy of a continuous distribution, that gives the probabilities  $p_i$  when divided up in sufficiently small bins. The above considerations also let us define an information like quantity. Consider that

$$S_B = S[\rho_B] - \ln \left( \frac{\bar{\Omega}}{L^n} \right) \geq S[\rho] - \ln \left( \frac{\bar{\Omega}}{L^n} \right). \quad (2.16)$$

The right hand side is actually independent of  $L^n$ , which may be checked by taking the derivative. Thus we can define the non-negative quantity

$$I_{\text{hidden}} \equiv S_B - S[\rho] + \ln \left( \frac{\bar{\Omega}}{L^n} \right) \geq 0, \quad (2.17)$$

which is independent of the arbitrary parameter  $L^n$ . This motivates thinking of the  $S[\rho] - \ln \left( \frac{\bar{\Omega}}{L^n} \right)$  term as the information term representing the average amount of hidden information in the boxes.

The last sort of course graining we will mention is the smoothing or diffusion of continuous distributions, somewhat analogous to what we did in the discrete case. This may represent, for example, the de-focusing of some lens in an experiment, or some general diffusion process. We will see that this also happens in a particular phase space representation of quantum states in section 9.3. In that

case the smearing out may represent directly the resolution of a measurement apparatus. Mathematically the final distribution may be represented as a convolution of a source distribution with some smoothing kernel  $K(x - y, \lambda)$ , where  $\lambda$  is some parameter that defines the coarseness of the smoothing kernel. The smoothed distribution is given by

$$\rho_\lambda(x) = \int \rho(y)K(x - y, \lambda)dy. \quad (2.18)$$

This procedure may be followed by any of the above box averaging procedures.

## 2.3 The Second Law and Mixing

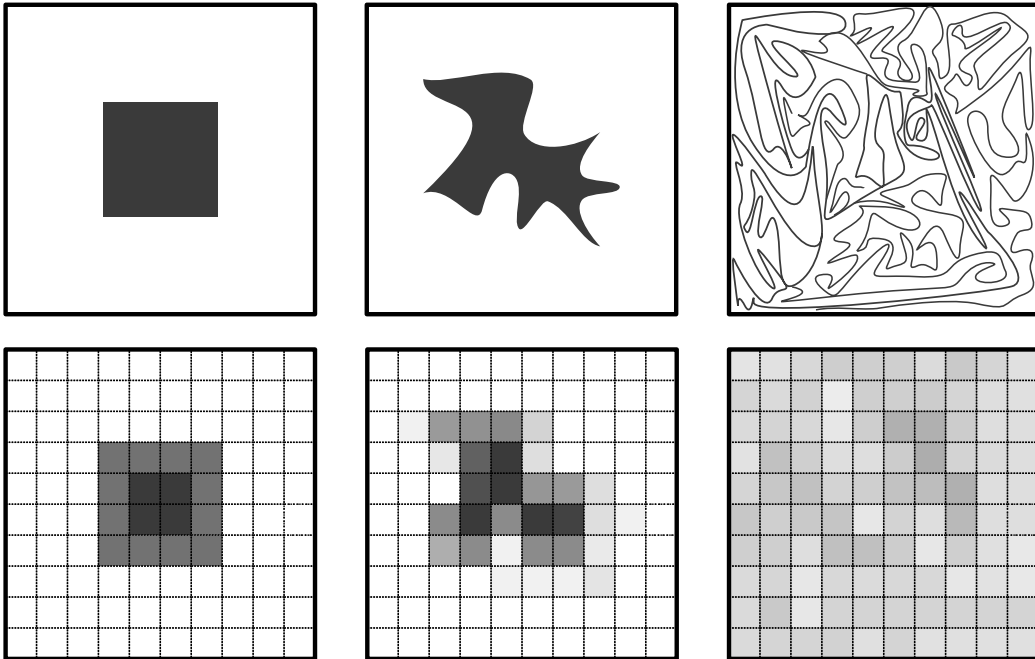


Figure 2.1: A Mixing System and its Coarse Grained Representation

In chapter 1 we have shown the entropy of classical statistical mechanics. We have mentioned that when the system is in equilibrium, the ensemble is invariant under time evolution. For the micro-canonical ensemble, which describes an isolated system, this equilibrium means, due to ergodicity, that all states with energy  $E$  are equally likely. However, what happens when we, for example, suddenly change the volume of the container of a gas, such that it ceases to be in equilibrium. Assume that the system is isolated at any point in time. Then the ensemble will spread out over all states with energy  $E$ , and position coordinates within the container. However, due to Liouville's theorem, the volume of the total ensemble is conserved, and it moves around in phase space like an incompressible fluid. In effect, its (probability) density must remain constant wherever the ensemble is nonzero. What this essentially means is that, given we knew the equilibrium state of the system at time  $t = 0$  with the precision of the size of the ensemble, we can pinpoint its state at any time  $t \neq 0$  with equal precision. As a consequence, the BG entropy of the system would seem to remain constant, in stark contradiction to the second law of thermodynamics. We will present a modern solution that has been adopted in many works [20, §23]. The key piece of this solution is the coarse graining of phase space itself. We also have to introduce the notion of mixing. Formally a mixing system is characterized by the following: Consider two subsets,  $A$  and  $B$ , of phase space  $\Omega$ . Let  $A_t$  be the time evolution of the set  $A$ . Then a system is "mixing" if for any two sets  $A$  and  $B$  we

have

$$\lim_{t \rightarrow \infty} W(A_t \cap B) = \frac{W(A)}{W(\Omega)} W(B) \quad (2.19)$$

where  $W(\cdot)$  denotes the measure of the set. This is a stronger statement than ergodicity, since that would only state

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t W(A_t \cap B) = \frac{W(A)}{W(\Omega)} W(B). \quad (2.20)$$

In other words, a system is mixing if the time evolution of any region of phase space spreads out equally over all other regions in phase space as  $t \rightarrow \infty$ . Nevertheless, the differential entropy must still remain invariant since at no point did the volume and density change of the ensemble. However, by course graining the whole phase space, e.g. by introducing a grid, the system seems to become more spread out equally over the whole space, as a consequence of its ergodic and chaotic behavior. This is illustrated in figure 2.1. By using the continuous box grained entropy we discussed, the entropy reaches its maximal value that would be expected from the equilibrium ensemble of the new volume. In general, the time evolution of the cells is not described easily by some differential equation, since in the end the dynamics of the underlying ensemble are always governed by Liouville's equation. Still, if we make some simplifying assumptions we may derive a differential equation describing the general flow of probability of the coarse grained system. Denote the probability of finding the system in cell  $i$  with  $p_i$ . Say we start with a constant ensemble, all located within the area of cell  $n$ . Its initial coarse grained distribution is  $p_n = 1$  and the rest of the probabilities are 0. Let  $p_i(t)$  denote the time evolved probabilities. If we now consider another distribution, also located within cell  $n$ , but not constant. We obtain a new time evolved distribution  $p'_i$ . However, if the system is mixing, every subset of the initial distribution should homogeneously spread out, so that the average density at time  $t$  should, in most cases, be approximately the same (in a coarse grained sense), provided that the cells are sufficiently large. We then have  $p_i \approx p'_i$ . In other words, we assert that  $p_i(t)$  may be calculated from  $p_i(0)$ . We thus have most generally

$$p_i(t) = \sum_j T_{ij}(t) p_j(0). \quad (2.21)$$

If we only require that the probability distribution remains normalized, then  $T$  being a left-stochastic matrix is sufficient<sup>3</sup>. This condition by itself does not prevent us from writing an equation that transports the whole phase-space into a single cell, thereby decreasing the entropy, which would not seem reasonable for a mixing system. However, by inserting two reasonable assumptions for such systems, we can find an equation that this does lead to an increase of entropy. First, note that for a later time  $t'$ , the probabilities may be calculated from

$$T_{ij}(t+t') = \sum_k T_{ik}(t') T_{kj}(t). \quad (2.22)$$

By inserting  $t' = dt$  we may linearize the equation to get a simple flow equation,

$$T_{ij}(t+dt) = T_{ij} - dt \sum_{k,l} \delta_{ik} W_{lk} T_{kj} + dt \sum_k W_{ik} T_{kl}. \quad (2.23)$$

which means

$$T_{ij}(dt) = \delta_{ij} (1 - dt \sum_k W_{kj}) + dt W_{ij}. \quad (2.24)$$

Next, combining the assumptions of microscopic reversibility, and that the flow out of, and into a cell must be proportional to the size of the cell (this is called detailed balancing), we find the following condition

$$W(\Omega_i) W_{ij} = W(\Omega_j) W_{ji}. \quad (2.25)$$

---

<sup>3</sup>A matrix being left-stochastic means that its columns sum to 1, but not necessarily its rows.

Then finally we may formulate the so called classical master equation

$$\dot{P}_i = \sum_k (W_{ik}p_k - W_{ki}p_i). \quad (2.26)$$

This equation is known to increase the entropy [2, B.4]. Thus, due to the reversibility and detailed balancing assumptions, the flow of probability works as a continuous version of the discrete coarse graining method we discussed in the first section of this chapter.

## 2.4 Concluding Remarks

One of the criticisms of the view of the second law that we have discussed, is that it adds subjectivity into statistical mechanics, since it essentially claims that the entropy of a system is a dependent on an observer their classical lack of knowledge about a system. However, the Clausius entropy is of course an experimental fact, and does not care about ones knowledge about a system.

On the other hand, this line of reasoning seems in line with the well known thought experiment called Maxwell's demon, which was originally conjectured to demonstrate a violation of the second law. This thought experiment essentially states that some demon with infinitely precise knowledge of the dynamics of the system, would be able to extract all energy from a thermodynamic system using very exact molecular interactions. In thermodynamics this would mean that the free energy is equal to the energy. This is because of the thermodynamic relation  $A = U - TS$ , where  $A$  is the free energy,  $U$  the internal energy and  $T$  and  $S$  the temperature and entropy. This would directly imply that the entropy must be zero. There has been a lot of debate on this subject, and generally, the statement that this demonstrates a violation of the second law is rejected. However, the general solution is that the demon may indeed extract all energy from a system, but in doing so, his own entropy must increase at least as much. Thus from the demon's perspective, ignoring his own entropy, with infinite knowledge, the entropy of the system indeed seems zero.

Perhaps this view shouldn't be interpreted as a statement that the entropy depends on ones knowledge, but more along the lines that the effective entropy, e.g. the thing that forbids us from extracting all the internal energy from a system, depends on the coarseness of operations we perform on a system. After all, if we divide a thermodynamic system in such small subsystems that the thermodynamic limit should not be considered to apply, most would agree that conventional statistical mechanics would cease to make useful predictions anyway.

## Chapter 3

# Generalized Entropies

The BGS entropy has had great successes in physical theories as well as in information theory, even being a tool which has blurred the lines between the two fields. However, with the increasing amount of research, it has become clear that there can be a need for more general information measures. For example, in the maximum entropy method, we will show that the Shannon entropy gives the most likely distribution specifically for independent stochastic processes. Many interesting processes are not independent, so this alone shows that there is room for more general information measures. Indeed a lot of alternative information measures have been introduced since the inception of Shannon's entropy, which we also feel is one of the reasons entropy as a whole has become such an challenging subject to get into. In this chapter we will give an outline of the work and ideas that have led to the introduction of all these new, so called generalized entropies. Two of the most notable ones are the Rényi and Tsallis entropies. We will discuss these entropies in their own section.

The Rényi entropy was introduced as a purely information theoretic, one-parameter generalization of the Shannon entropy. In fact, his initial reason for introducing this quantity was to find the most general additive information quantity that satisfies a set of axioms which, similar to Shannon's axioms, characterize the properties a "reasonable" information quantity should satisfy [22]. Later on Rényi recognized that the actual bit content of information for specific problems could only be quantified by the Rényi entropy, thus showing that general problems sometimes require more general information quantities [23]. Since then the Rényi entropy has been employed in many information theoretical problems, but also in physics as a measure of uncertainty.

The introduction of the Tsallis entropy, on the other hand, was purely physically motivated [24]. As we have discussed, for the BGS entropy to be useful as the thermodynamical entropy, we had to assume physical axioms like ergodicity and chaotic phase space evolution. It is also this behavior that typically causes multiplication of phase spaces, in the thermodynamic limit, when we compose two systems, which is at the root of the extensivity of the BGS entropy. For complex systems with more intricate phase space behavior, the BGS is known to lose its extensivity, which causes a breakdown of the thermodynamical equations. The Tsallis entropy, as another one parameter generalization of the BGS entropy, was introduced to be an extensive entropy where the BGS is not. For what systems it is extensive may be tweaked with the parameter. To discuss the extensivity of the entropy, we will also properly define what it means for a system to be extensive. Lastly, the Tsallis entropy also shares a few crucial properties with the BGS entropy, one of which is experimental robustness, which in short is a requirement for an entropy to be measurable. We will define that in this chapter as well.

To show how all generalized entropies fit together, we will present an axiomatic characterization of the entropy introduced by Khinchin. It is known that by relaxing only one of the axioms, all so called generalized entropies emerge. All these entropies may also be divided up into equivalence classes based on two parameters  $(c, d)$ , which characterize the behavior of the entropies in the



thermodynamic limit. We will briefly review this classification.

### 3.1 The Shannon-Khinchin Axioms

Following up on Shannon's work on information theory, Khinchin established a novel axiomatic characterization of the Shannon entropy, now called the Shannon-Khinchin axioms. He was mostly interested in ergodic and Markovian systems and thus formulated his axioms to make sense in that context [25]. Examining these axioms will give us more insight on how the generalized entropies relate to Shannon's entropy. We will adopt the same notation as in section 1.2.4 where convenient. The Shannon-Khinchin axioms can be formulated as

- SK1** Continuity:  $S[\vec{p}] = S(p_1, \dots, p_N)$ ;  
The entropy depends continuously on all probabilities  $p_i$ .
- SK2** Maximality:  $S(p_1, \dots, p_N) \leq S(\frac{1}{N}, \dots, \frac{1}{N})$ ;  
The entropy is maximal when all probabilities are equal.
- SK3** Expansibility:  $S(p_1, \dots, 0, \dots, p_N) = S(p_1, \dots, p_N)$ ;  
The entropy remains the same if a state with zero probability is added.
- SK4** Separability:  $S[\vec{p}_{12}] = S[\vec{p}_1] + \sum p_i S(\frac{p_{i1}}{p_i}, \dots, \frac{p_{iM}}{p_i}) \equiv S[\vec{p}_1] + S[\vec{p}_2|\vec{p}_1]$ ;  
The entropy of two systems is equal to the entropy of independent realizations of the first system, plus that of the second system, conditional on the outcome of the first.

Khinchin showed axioms uniquely identify the BGS entropy up to a multiplicative constant. Axiom [SK4] reduces to additivity in the case that the systems are uncorrelated (since then  $\frac{p_{ij}}{p_i} = p_j$ ). We then define generalized entropies, as all quantities that satisfy axioms [SK1]-[SK3]. In other words, generalized entropies are continuous functions of the probabilities of a probability distribution, that are maximal for the equidistribution and are invariant if we add outcomes with zero probabilities to our distribution. We may most generally write them as [26]

$$S_{[G,g]}[\vec{p}] = G\left(\sum g(p_i)\right), \quad (3.1)$$

where  $G$  and  $g$  are arbitrary functions. If  $G = id$ , e.g. the identity map, then a generalized entropy is said to be of trace form. In information theory a lot of research is done on finding axiomatic basis for a lot of these quantities [27]–[29], but we won't focus further on that. From the explicit forms of  $G$  and  $g$  we can deduce general properties of the entropies. Hanel and Thurner have proposed a characterization of the entropies, based on the asymptotic (in the thermodynamic limit) behavior for these functions [30], which we will review later on. This classification allows us to distinguish what entropy should be used as an extensive entropy for different types of phase space growth upon composition of multiple systems.

#### 3.1.1 Extensivity

In section 1.1.4 we demonstrated the extensivity of the Shannon entropy, assuming that the total number of states for two systems could be written as  $W_{12} = W_1 W_2$ . As we have mentioned in the beginning of this section, this turns out not to be universal. The answer to when this happens ultimately must lie in the behavior of specific dynamical systems. However, going directly from the microscopic dynamics to macroscopic behavior is extremely difficult, and even for the well established Boltzmann-Gibbs statistics we still rely heavily on physical axioms. While we will not try to answer what systems cause what kind of correlations, what we can try to answer is what properties are extensive, given that we assume the kind of growth the total phase space undergoes upon composition of two systems. We go deeper into this later on in this chapter, but it will be useful to properly define extensivity. We will use the following definition of extensivity [31]: A property  $I$  of  $N$  joint systems, e.g.  $I[\vec{p}_{12\dots N}]$ , is extensive when the following limit exists

$$\lim_{N \rightarrow \infty} \frac{I[\vec{p}_{12\dots N}]}{N}. \quad (3.2)$$

As we have already seen, the Shannon entropy is additive, so that for separable probabilities  $S[\vec{p}_{12}] = S[\vec{p}_1\vec{p}_2] = S[\vec{p}_1] + S[\vec{p}_2]$ . This of course still holds for  $N$  joint systems as long as the probabilities separate, thus then  $S_{12\dots N} = S_1 + S_2 + \dots + S_N$ . Given every contribution is finite and bounded from above<sup>1</sup>,  $S_{12\dots N}/N$  will be the average contribution of every system and thus finite.

### 3.1.2 Experimental Robustness

A question that comes up when studying entropy quantities that may be used for physical applications such as thermodynamics, is whether the quantity is practically measurable, e.g. whether it is experimentally robust. Lesche [32] formulated a condition he believed would ensure experimental robustness of a quantity. That condition, now called Lesche stability<sup>2</sup>, has been studied often in the context of generalized entropies. As such, we will include Lesche stability in the property sections of the Rényi and Tsallis entropy.

In essence, Lesche stability is the statement of the uniform continuity of  $I$  over the whole probability space. It may be formulated as follows: A quantity,  $I[\vec{p}]$ , that depends continuously on all the components of a probability distribution, is said to be Lesche stable if,  $\forall n, \epsilon > 0, \exists \delta > 0$ , such that

$$\|\vec{p}' - \vec{p}\|_1 \leq \delta \implies \left| \frac{I[\vec{p}'] - I[\vec{p}]}{\sup_{\{\vec{p}\}} I[\vec{p}]} \right| \leq \epsilon, \quad (3.3)$$

where

$$\|\vec{p}' - \vec{p}\|_1 \equiv \sum^n |p'_i - p_i|, \quad (3.4)$$

is the  $l_1$  norm of the probability vector. The  $l_1$  norm is used because it is considered the natural distance assignment for probability distributions [33]. For example, only for the  $l_1$  norm is the relative distance independent of the number of elements of the probability distribution [33, §4]. The factor of  $1/\sup I$  is added because of ease of calculations, since the quantity  $I$  may be unbounded in the thermodynamic limit.

We will say that a quantity is experimentally robust<sup>3</sup> if measurements on the same, or very similar systems give stable results and are repeatable. We see that Lesche stability is indeed important for a quantity to be experimentally robust by imagining the following: Say we have a system that visits a set of states with probabilities  $p_i$ . In a real experiment we measure the relative frequencies of states<sup>4</sup>, which we will interpret as probabilities of  $\vec{p}'$  in equation 3.3. It is expected that the relative frequencies  $p'_i$  only converge to the actual probabilities,  $p_i$ , as our number of samples (or equivalently the time over which the experiment happens and is averaged) goes to infinity. Due to the finiteness of real experiments, we will only observe  $\vec{p}'$  with a finite distance from  $\vec{p}$ , though the distance itself will get small as the number of samples increases. If the system is Lesche stable, we are certain that the quantity  $I[\vec{p}']$  converges to  $I[\vec{p}]$ , as the measured distribution  $\vec{p}'$  converges to the inherent distribution  $\vec{p}$ .

Lesche stability then ensures that repeated measurements, given they happen over long enough times and the system is only slightly perturbed during or between experiments, give very similar results. That is, the results that are expected from the theoretical distribution  $\vec{p}$ . Conversely if a quantity is not Lesche stable, and therefore not uniformly continuous, there is no guarantee of either repeatability or stability of measurement. However, no guarantee does not mean that it is in principle still not observable, and this has been a point of critique of solely using Lesche stability to prove

<sup>1</sup>If every marginal entropy term is finite, but increases every term like  $N^c$ , with  $c > 1$ , then the sum still diverges in the limit  $N \rightarrow \infty$ . Indeed this construction seems very artificial and should thus not cause any relevant difficulties.

<sup>2</sup>Lesche stability should not be confused with thermodynamical stability, which has to do with concavity of the entropy w.r.t energy.

<sup>3</sup>Experimental robustness/observability and Lesche stability are often used interchangeably in literature. However, since Lesche stability has also been criticised in literature, for multiple different reasons, we make a distinction between the two.

<sup>4</sup>This may be either directly or indirectly, like through the measurements of thermodynamic variables which depend on the relative frequencies of all the states

whether a quantity is observable. For example, the support of the region where the discontinuity is, may be of zero measure, or the region where it is discontinuous may not be relevant for physical scenarios [34][35][36, B.3]. Finally, even if a quantity is discontinuous, its value may only change slightly around the discontinuity, so that it is still irrelevant for most measurements.

## 3.2 The Maximum Entropy Method Revisited

In section 1.2.3 we highlighted Jaynes' view of the maximum entropy method in terms of ideas from information theory and showed that this leads to the Boltzmann distribution. In this section we will show that the maximum entropy method naturally emerges when one tries to find the most likely distribution for independent stochastic processes. What we mean by this is that if we take multiple samples from a process with probabilities that are independent of the previous outcomes, then maximizing the Boltzmann entropy gives the most probable distribution. For physical systems we may view chaotic and ergodic systems which are in equilibrium as systems that produce sufficiently independent outcomes. What this means is that any time every region in phase space is equally likely to be measured, independent of earlier measurement outcomes. What this shows is that maximization of the Shannon entropy should not be considered the definitive information quantity when it comes to entropy maximization. Let us consider a sequence of  $N$  events,  $(x_1, \dots, x_N)$ , where  $x_i$  may be one of  $W$  outcomes, with probabilities  $q_1, \dots, q_W$ . Let us denote the histogram of outcomes as  $k = (k_1, \dots, k_W)$  where  $k_i$  denotes the number of times outcome  $i$  occurs in our sequence. The probability of any such histogram may be written as

$$P(k|q) = \frac{N!}{k_1! \dots k_W!} q_1^{k_1} \dots q_W^{k_W} \equiv M(k)G(k|q). \quad (3.5)$$

The probability thus factorizes into a multiplicity part, which stems from the amount of ways the particular histogram may be constructed and a part that contains the inherent probabilities. We now aim to find the histogram  $k^*$  such that  $P(k^*|q)$ , or equivalently  $\log P(k^*|q)$ , is maximal. We may write

$$\frac{1}{N} \log P(k|q) = \frac{1}{N} \log M(k) + \frac{1}{N} \log G(k|q). \quad (3.6)$$

For large  $N$  the second term on the right hand side becomes

$$\frac{1}{N} \log M(k) = \frac{1}{N} \log \frac{N!}{k_1! \dots k_W!} \approx \frac{1}{N} \frac{N^N}{k_1^{k_1} \dots k_W^{k_W}} \quad (3.7)$$

$$= -\frac{1}{N} \sum_i^W k_i \log \frac{k_i}{N} \equiv -\sum_i^W p_i \log p_i. \quad (3.8)$$

We defined  $p_i = k_i/N$ , e.g. the relative occurrence of outcome  $x_i$  in our sequence.

In the same way as above we may write equation 3.5 as

$$\underbrace{-\sum_i^W p_i \log \frac{p_i}{q_i}}_{\equiv -\text{Relative Entropy}} = \underbrace{-\sum_i^W p_i \log p_i}_{=\text{Entropy}} + \underbrace{\sum_i^W p_i \log q_i}_{\equiv \text{Cross Entropy}} \quad (3.9)$$

We can write the probabilities  $q_i$  in the exponential form,  $q_i = \exp(-\alpha - f_i)$  such that

$$\frac{1}{N} G(k|q) = -\alpha \sum_i^W p_i - \sum_i^W p_i f_i \quad (3.10)$$

Writing  $f_i = \beta \epsilon_i$  we obtain the familiar result from section 1.2.3,

$$\frac{1}{N} \log P(k|q) = -\sum_i^W p_i \log p_i - \alpha \sum_i^W p_i - \beta \sum_i^W p_i \epsilon_i. \quad (3.11)$$

If we wish to find the histogram that maximizes the probability we must solve the set of equations

$$\frac{\partial}{\partial k_i} P(k|q) = \frac{1}{N} \frac{\partial}{\partial p_i} P(k|q) = \frac{\partial}{\partial p_i} \left( - \sum_i^W p_i \log p_i - \alpha \sum_i^W p_i - \beta \sum_i^W p_i \epsilon_i \right), \quad (3.12)$$

which is the same set of equations as obtained from the maximum entropy method. Given that we know  $\sum q_i \epsilon_i = \langle \epsilon \rangle$ , the solution to this set of equations is of course, trivially,  $p_i = q_i$ . However, by writing  $q_i$  in the exponential form we see that there is the same knowledge contained in  $\epsilon_i$  and  $\langle \epsilon \rangle$ , given that the process is independent. In ergodic processes, even if the subsequent outcomes of a process are not independent, there exists a stationary distribution  $q_i$ , so that for any sequence of events the relative frequencies  $p_i$  will asymptotically approach the stationary frequencies  $q_i$ . Thus in such processes, may speak of an asymptotically “independent” distribution given by  $p_i = q_i$ , and we may thus expect the maximum entropy method to give the most likely distribution for such systems. However, if the system is not ergodic, we must look for generalized entropies to obtain the most likely distribution.

For more general, non-ergodic processes, the evolution of the system is history dependent. As a consequence, there is no easy way anymore to write the probability of a histogram as a multinomial multiplied by the probabilities. However it turns out that for certain problems the logarithm of the histogram probability may still be written as follows

$$\frac{1}{f(N)} \ln P(k|q) = \frac{1}{f(N)} M(k) + \frac{1}{f(N)} \ln G(k|q), \quad (3.13)$$

where  $f(N)$  is some function to remove the  $N$  dependence. For such systems it is still possible to find the most probable distribution using the method of entropy maximization. Only now, the entropy that is needed is dependent on the specifics of the system. A detailed analysis is done by Hanel and Thurner for driven systems [37]. They also show that for such more general stochastic processes, the scaling components ( $c, d$ ) may be used to classify these complex systems, by identifying them with entropy that gives the most probable distribution upon maximization [38, §6.5].

### 3.3 Tsallis Entropy

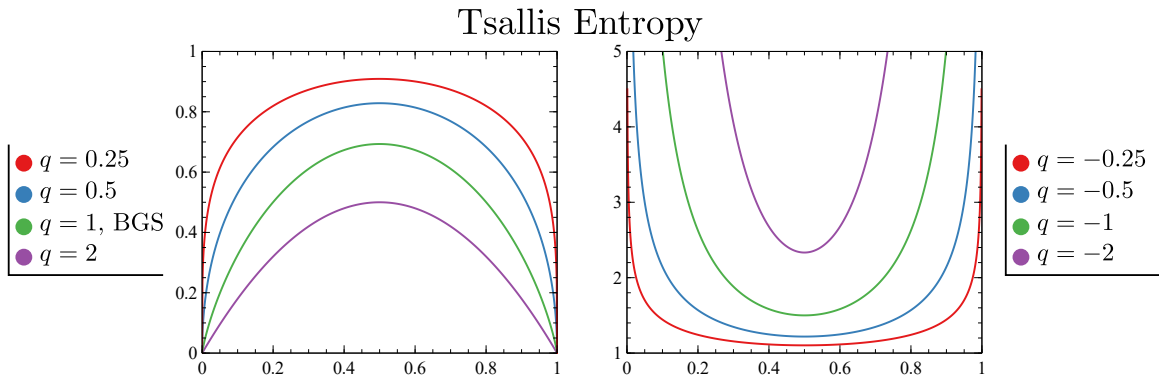


Figure 3.1: The Tsallis entropy  $T_q(x, 1-x)$ , for multiple values of  $q$ .

Tsallis first introduced his entropy in 1988 [24] as a potential generalization of the thermodynamical BGS entropy<sup>5</sup>. He defined it as follows

$$T_q = \frac{k_B}{1-q} \left( \sum_i p_i^q - 1 \right). \quad (3.14)$$

<sup>5</sup>This entropy quantity was actually already introduced by Havrda and Charvát in 1967 [39], though not for thermodynamical purposes. Because of this, it is also sometimes referred to as the Havrda-Charvát entropy.

The Tsallis entropy for a distribution of two probabilities, for different values of  $q$  is plotted in figure 3.1, which also shows its behavior relative to the BGS entropy, which occurs at  $q = 1$ . The reason for introducing this quantity was the inability of the standard BGS entropy to describe systems to which the usual arguments, the ones we used in section 1.1.2, do not apply. This means a slow rate of mixing through phase space or even non-ergodic behavior altogether. Moreover, it is clear even from simpler considerations that BGS statistics are not sufficient to describe every system. For example, in classical statistical mechanics, distributions in the form of an exponent of the Hamiltonian necessarily produce Gaussian, marginal velocity distributions [36, p.320]. However, it is known that not all systems show Gaussian behavior. Thus a thermodynamical description of such systems is not possible using the BGS entropy. In contrast, momentum distributions such as predicted by optimization of the Tsallis entropy, were shown in experiments on cold atoms in dissipative optical lattices [40]. Moreover, the statistics of the Tsallis entropy has been able to reproduce the typical power laws of cosmic rays [41], in the study of turbulent systems [42], in the study of the transverse momentum distribution of particle jets in  $e^-e^+$  annihilation [43] and in many other fields of research within, and outside of physics.

The Tsallis entropy can be written using the generalized logarithm  $\ln_q$ , which is defined as follows

$$\ln_q x = \frac{x^{1-q} - 1}{1 - q}. \quad (3.15)$$

The usual natural logarithm is recovered in the limit  $q \rightarrow 1$ . This function is called the generalized logarithm because it is the inverse function of the generalized exponent

$$e_q^x = [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (3.16)$$

which is the solution to the o.d.e.

$$\frac{dy}{dx} = y^q. \quad (3.17)$$

The usual exponential function is of course the solution for  $q = 1$ , and is recovered in the limit  $\lim_{q \rightarrow 1} e_q^x$ . Using this notation, the Tsallis entropy becomes

$$T_q = - \sum p_i \ln_q \left( \frac{1}{p_i} \right) \quad (3.18)$$

### 3.3.1 Properties

#### Non-Negativity

The Tsallis entropy is non-negative for all  $q$  by construction

$$T_q[\vec{p}] \geq 0. \quad (3.19)$$

#### Concavity

The Tsallis entropy is concave for all  $q > 0$ , and convex for all  $q < 0$  [36, §3.2.2.5],

$$T_q[\lambda \vec{p}^{(1)} + (1 - \lambda) \vec{p}^{(2)}] \geq (\leq) \lambda T_q[\vec{p}^{(1)}] + (1 - \lambda) T_q[\vec{p}^{(2)}], \quad q > 0 \quad (q < 0) \quad (3.20)$$

This is clearly seen in figure 3.1.

#### Additivity

The Tsallis entropy is not additive. It is instead “pseudo-additive”

$$T_q[\vec{p}_1 \otimes \vec{p}_2] = T_q[\vec{p}_1] + T_q[\vec{p}_2] + (1 - q) T_q[\vec{p}_1] T_q[\vec{p}_2]. \quad (3.21)$$

As we have mentioned, this does not mean the Tsallis entropy is not extensive. We will illustrate this in section 3.3.4

### Sub-Additivity

The Tsallis entropy is subadditive [44, §3]

$$T_q[\vec{p}_1 \otimes \vec{p}_2] \leq T_q[\vec{p}_1] + T_q[\vec{p}_2]. \quad (3.22)$$

When the above relation is saturated depends both on the type of correlations and the value of  $q$ . This means that the Tsallis entropy will still be maximal when it is the sum of the entropies of the subsystems. However, for  $q \neq 1$ , this maximum thus won't be attained for uncorrelated systems, in contrast to the case of the BGS entropy.

### Strong Subadditivity

The Tsallis entropy is strongly subadditive for  $q \geq 1$  [44, Thm. 3.4],

$$T_q[\vec{p}_{123}] + T_q[\vec{p}_2] \leq T_q[\vec{p}_{12}] + T_q[\vec{p}_{23}] \quad (3.23)$$

### Lesche Stability

The Tsallis entropy is Lesche stable [45].

## 3.3.2 Relation to the Shannon Entropy

The Tsallis entropy with parameter  $\alpha = 2$  is the linearization of the Shannon entropy. This is easily shown

$$-\sum_i p_i \ln p_i \approx -\sum_i p_i(p_i - 1) = 1 - \sum_i p_i^2. \quad (3.24)$$

Because of this, this entropy is often called the linear entropy. It is used more often in the context of quantum mechanics, for reasons that we will discuss later on.

## 3.3.3 Maximization

In this section we will show the functions that maximize the Tsallis entropy under the micro-canonical and canonical constraints [24].

### Micro-canonical Ensemble

Optimizing only under the constraint of normalization  $\sum^N p_i = 1$ , we recover the standard equidistribution

$$p_i = \frac{1}{N}. \quad (3.25)$$

The value of the entropy is

$$T_q = k \ln_q W \quad (3.26)$$

### The Canonical Ensemble

Optimizing under the constraints of  $\sum^N p_i = 1$  and  $\langle E \rangle = \sum^N p_i E_i$  we obtain

$$p_i = \frac{1}{Z_q} (1 - \beta(q-1)E_i)^{\frac{1}{q-1}}, \quad (3.27)$$

with

$$Z_q = \sum (1 - \beta(q-1)E_i)^{\frac{1}{q-1}}. \quad (3.28)$$

We can write this with the  $q$ -exponential as follows

$$p_i = \frac{1}{Z_q} e_{2-q}^{-\beta E_i} \quad (3.29)$$

We note that this  $\beta$  is not the actual Lagrange multiplier of the energy constraint. This is due to the factorization of the multiplier for the normalization condition,  $\alpha$ , we have  $\beta = \beta'/\alpha$ , where  $\beta'$  is the actual multiplier for the energy condition.

### 3.3.4 Extensivity

The field of generalized statistical mechanics, which is based on the Tsallis entropy, is commonly referred to as “non-extensive” statistical mechanics. This naming is arguably a bit unfortunate, and Tsallis himself hesitated a while to adopt this name. The name stems from the frequently, but wrongly referring to the non additive property of the Tsallis entropy as the non-extensive property. The Tsallis entropy is however, not always non-extensive. In fact, its appeal comes from the ability to choose  $q$  such that it *is* extensive. In the next part of this section we will show that the Tsallis can be chosen extensively in the case of strong correlations, such that some states become unavailable, but equal probabilities for the rest of the states.

#### Extensivity in Presence of Strong Correlations but Equal Probabilities

For a joint system of two subsystems, the maximal amount of states can be written as

$$W_{\max} = W_1 W_2, \quad (3.30)$$

provided there are no correlations. If there are correlations, the amount of available states does not have to change. If there are very strong correlations however, some states may become unavailable altogether (like in the case where the energy of two systems may fluctuate, but the total energy is conserved). We will denote this effective number of states that are available to the total system  $W_T$ , where

$$W_{\max} \geq W_T. \quad (3.31)$$

Let us take the case where the number of states for every subsystem is equal;  $W_1 = \dots = W_N$ . If the following relation holds

$$W_T(N) = \left( N W_1^{1-q} - (N-1) \right)^{\frac{1}{1-q}} \leq W_1^N, \quad q \leq 1. \quad (3.32)$$

This may be seen as the  $q$  analog of  $W^N$ , since that is its limiting case for  $q \rightarrow 1$ . We have

$$\ln_q W_T(N) = N \ln_q W_1. \quad (3.33)$$

In the case of  $q = 1$  then  $W_T$  is maximal and we recover the BGS entropy additivity, given we take all probabilities to be equal. If we now have  $W_T < W_1^2$ , but the probabilities of the remaining states are equally likely ( $p_i = 1/W_T$ ), then the entropy associated with the total system is

$$T_q = \frac{k_B}{1-q} \left( \sum_{i=1}^{W_T} \left( \frac{1}{W_T} \right)^q - 1 \right) = \frac{k_B}{1-q} \left( (N W_1^{1-q} - (N-1)) - 1 \right) = N k \ln_q W_1. \quad (3.34)$$

Thus, according to definition 3.2, the Tsallis entropy is extensive since

$$\lim_{N \rightarrow \infty} \frac{T_q}{N} = \ln_q W_1 \quad (3.35)$$

### 3.3.5 The Zeroth Law of Thermodynamics and the Tsallis Composition Law

We have seen that the Tsallis entropy is an extensive entropy for systems where the Shannon entropy is not. Along with this there are several other properties of the Tsallis entropy, some of which it shares with the Shannon entropy, that make it specifically interesting for thermodynamical applications. In this section we will show that requiring composability, which states that the entropy of a joint system may be determined from the entropies of the subsystems, together with the zeroth law of thermodynamics, we uniquely find the Tsallis pseudo-additivity property 3.22. Of course, this includes the Shannon entropy composition law for  $q \rightarrow 1$

The Tsallis entropy is the only entropy of trace form that is composable, e.g. it satisfies the following condition

$$S[\vec{p}_{12}] \equiv S_{12} = \Phi(S[\vec{p}_1], S[\vec{p}_2]), \quad (3.36)$$

where  $\Phi$  is a symmetric function. Composability is a natural condition for thermodynamics because it states that the entropy of a joint system is fully determined by the entropies of the subsystems. If an entropy were not composable, additional knowledge of the microscopic dynamics would be necessary to determine the total entropy. Moreover, we will now show that assuming composability along with the zeroth law of thermodynamics uniquely identifies the Tsallis pseudo-additivity 3.3.1 [28].

Let  $X_{12} = X_1 + X_2$  be some fixed, additive quantity of our system like, for example, the internal energy or the volume. The stationary state of the system is given by the maximum entropy state

$$dS_{12} = d\Phi(S_1, S_2) = \frac{dS_{12}}{dX_1}dX_1 + \frac{dS_{12}}{dX_2}dX_2 = 0. \quad (3.37)$$

Since  $X_{12}$  is fixed and additive, we additionally have

$$dX_1 = -dX_2. \quad (3.38)$$

From these both conditions we have

$$\frac{\partial\Phi(S_1, S_2)}{\partial S_1} \frac{dS_1}{dX_1} = \frac{\partial\Phi(S_1, S_2)}{\partial S_2} \frac{dS_2}{dX_2}. \quad (3.39)$$

Now we will invoke the zeroth law. For our purposes, it suffices to interpret the zeroth law as the statement that the two systems can be at equilibrium with each other, and thus share an equilibrium property. The entropy should be additive as well, at least at equilibrium, so that  $dS_{12}/dX_1 = dS_1/dX_1$ . We see immediately from the above equation that the usual definition for the temperature  $\beta = dS/dE$ , for example, can only hold if  $\Phi(S_1, S_2) = S_1 + S_2$ . Thus for more general functions, to be able to formulate an equilibrium property, we need that the above equation separates as follows

$$J_1 = J_2, \quad (3.40)$$

where  $J_1$  only depends on variables of the first system and likewise for  $J_2$ .  $J$  may be identified with the temperature, pressure or chemical potential if  $X$  is the internal energy, volume or particle number, respectively. The above equation separates if

$$\frac{d\Phi(S_1, S_2)}{dS_1} = k(S_1, S_2)g(S_1)h(S_2) \quad (3.41)$$

$$\frac{d\Phi(S_1, S_2)}{dS_2} = k(S_1, S_2)g(S_2)h(S_1), \quad (3.42)$$

where  $k$  is a symmetric function, and  $g$  and  $h$  are arbitrary functions. This form is required because of the symmetry of  $\Phi$ . We will however assume, for ease of calculation, that  $k = id$ . This assumption is not strictly necessary and assuming an arbitrary form for  $k$  leads to the same conclusions [46]. Imposing integrability, we get

$$\frac{\partial^2\Phi(S_1, S_2)}{\partial S_1\partial S_2} = \frac{\partial^2\Phi(S_1, S_2)}{\partial S_2\partial S_1}, \quad (3.43)$$

together with equations 3.41 and 3.42 this leads to

$$\frac{1}{g(S_1)} \frac{dh(S_1)}{dS_1} = \frac{1}{g(S_2)} \frac{dh(S_2)}{dS_2} \equiv r. \quad (3.44)$$

Substituting this into equations 3.41 and 3.42 give us

$$\frac{d\Phi(S_1, S_2)}{dS_1} = \frac{1}{r} \frac{dh(S_1)}{dS_1} \quad (3.45)$$

$$\frac{d\Phi(S_1, S_2)}{dS_2} = \frac{1}{r} \frac{dh(S_2)}{dS_2}. \quad (3.46)$$



These equations can simply be integrated so that we see that

$$\Phi(S_1, S_2) = \frac{1}{r}h(S_1)h(S_2) + c. \quad (3.47)$$

To determine  $c$ , we impose that when system 1 and 2 are completely ordered, we have  $S_1 = S_2 = S_{12} = 0$ . This means

$$c = -\frac{1}{r}h^2(0), \quad (3.48)$$

and then we may set

$$h(0) = 1. \quad (3.49)$$

From the boundary condition that if  $S_2$  is completely ordered, then  $S_{12} = S_1$  we get

$$h(S) = rS + 1. \quad (3.50)$$

Finally, we can substitute this into 3.47 to get

$$S_{12} = S_1 + S_2 + rS_1S_2, \quad (3.51)$$

which is exactly the Tsallis composition law if we set  $r = 1 - q$ .

### 3.4 The Rényi Entropy

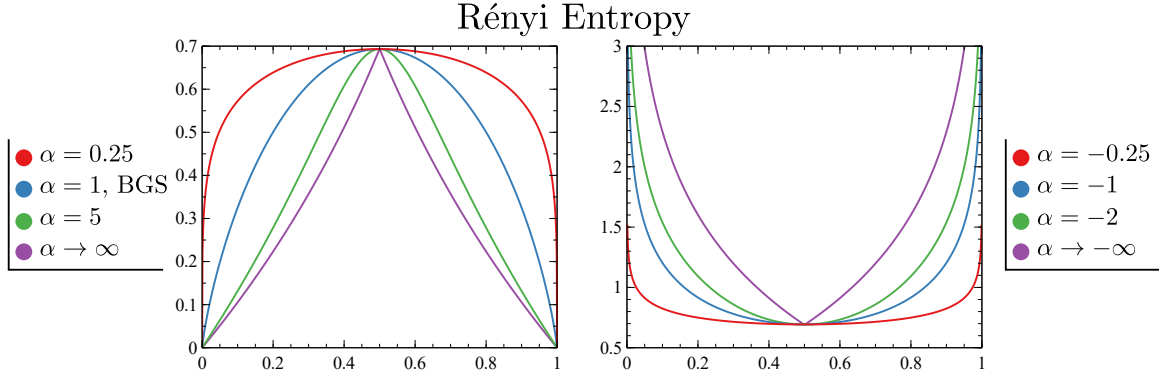


Figure 3.2: The Rényi entropy  $R_q(x, 1 - x)$ , for multiple values of  $q$ .

In search of the most general, yet additive entropy, Rényi found a new entropy like quantity in 1961 [22], generalizing the notion of the information theoretic BGS entropy. Its definition is

$$R_\alpha = \frac{1}{1 - \alpha} \ln \left( \sum p_i^\alpha \right), \quad (3.52)$$

and it can easily be checked that the Rényi entropy goes to the BGS entropy in the limit  $\alpha \rightarrow 1$ . The Rényi entropy for a distribution of two probabilities, for different values of  $\alpha$  is plotted in figure 3.1, which also shows its behavior relative to the BGS entropy, which occurs at  $\alpha = 1$ .

The Rényi entropy may be considered a generalized entropy in the sense that it obeys the Shannon-Khinchin axioms SK1-SK3. Moreover, it may be thought of as generalizing the information function  $I(p_i) = -\ln p_i$ . Whereas the Shannon entropy is the linear mean of this function, in the theory of generalized expectation values, we consider expectation values of the following form

$$\langle x \rangle_f = f^{-1} \left( \sum_i p_i f(x_i) \right), \quad (3.53)$$

where  $f$  is some function. Whereas the Shannon entropy gives the linear average of the information function  $\ln 1/p_i$ ,  $f(x) = id$ , the Rényi entropy has  $f(x) = e^{(1-\alpha)x}$ . Taking  $x = \ln 1/p_i$  and plugging in the expression for  $f(x)$ , we see that

$$\langle \ln 1/p_i \rangle_f = \frac{1}{1 - \alpha} \ln \left( \sum_i p_i e^{-(1-\alpha) \ln p_i} \right) = R_\alpha \quad (3.54)$$

Later on, the Rényi entropy was recognized by Rényi himself as the sensible measure of information in random search problems [23]. As such, we may say that the Rényi entropy was not introduced with physics applications in mind, but rather information theoretic ones. This reflects in the fields it has found applications in, which include: cryptography [47][48], the study of multifractal- systems and dimension [49], among other applications.

Though thermodynamic applications were not borne in mind when the Rényi entropy was initially conceived, there has been research on the role of this entropy within thermodynamics. For example, Rényi statistics, obtained from either entropy maximization or from the fundamental equation of thermodynamics, was studied by Parvan and Biró [50]. It was found by them, however, that in the thermodynamical limit, Rényi statistics reproduces BGS statistics. We will see later that the Rényi entropy indeed falls into the same equivalence class as the BGS entropy, based on their asymptotic behavior in the thermodynamic limit.

### 3.4.1 Properties

#### Concavity

The Rényi entropy is concave for  $0 < \alpha \leq 1$  but neither concave nor convex for  $q > 1$  [26]. On the other hand, it is convex for  $\alpha < 0$ . This may be seen in figure 3.2.

#### Additivity

The Rényi entropy is additive by construction. If  $\vec{p}_{ij} = \vec{p}_i \vec{p}_j$ , then

$$R_\alpha[\vec{p}_{ij}] = \frac{1}{1-\alpha} \ln \left( \sum \vec{p}_i^\alpha \vec{p}_j^\alpha \right) = R_\alpha[\vec{p}_i] + R_\alpha[\vec{p}_j] \quad (3.55)$$

#### Sub-Additivity

The Rényi entropy does not satisfy any sub-additivity nor strong sub-additivity properties, except for  $\alpha = 1$ , which is the case of the Shannon entropy. As such, correlations between subsystems do not necessarily cause the Rényi entropy to decrease.

#### Connection to the Tsallis Entropy

Using that  $\ln(x) \approx x - 1$  we see that the first order of the Rényi entropy around  $\sum \vec{p}_i^\alpha = 1$  is given by

$$\frac{1}{1-\alpha} \ln \left( \sum \vec{p}_i^\alpha \right) \approx \frac{1}{1-\alpha} \left( \sum \vec{p}_i^\alpha - 1 \right) = T_\alpha[\vec{p}_i]. \quad (3.56)$$

Thus we see that the Tsallis entropy is the linearization of the Rényi entropy.

#### Min-, Max-Entropy

Two special cases worth noting, are the min-entropy ( $\alpha \rightarrow \infty$ ) and max-entropy ( $\alpha = 0$ ). As we take  $\alpha \rightarrow \infty$ , the Rényi entropy converges to

$$R_\infty = \max_i(\ln p_i). \quad (3.57)$$

This can be seen by considering that  $\sum p_i^\alpha \approx \max_i p_i^\alpha$ , as  $\alpha \rightarrow \infty$  and then using l'Hôpital's rule. This limiting behavior causes the cusp as  $\alpha \rightarrow \infty$ , which may be seen in figure 3.2.

The min-entropy, on the other hand, is always smaller than any other order of  $\alpha$  of the Rényi entropy. When we take  $\alpha = 0$ , we have

$$R_0 = \ln n, \quad (3.58)$$

where  $n$  is the maximum index of  $p_i$ . This entropy is sometimes also called the Hartley entropy.

#### Collision Entropy

The Rényi entropy of order 2 is typically called the collision entropy.

$$R_2[\vec{p}] = -\ln \left( \sum_i p_i^2 \right) \quad (3.59)$$

This name stems from its relation to the probability  $P(X = Y)$  that two values i.i.d. (independent and identically distributed) drawn, are the same (collide). We have that

$$P(X = Y) = e^{-R_2}. \quad (3.60)$$

### 3.4.2 Entropy Maximization

#### Microcanonical Ensemble

Imposing only the normalization constraint we solve for the Euler-Lagrange equations

$$\frac{d}{dp_i} \left[ \frac{\ln(\sum_i p_i^\alpha)}{1-\alpha} - \lambda \sum_i p_i \right] = 0 \quad (3.61)$$

$$\implies \frac{\alpha}{\alpha-1} \frac{p_i^{\alpha-1}}{\sum_i p_i^\alpha} = \lambda. \quad (3.62)$$

We see that  $p_i$  must be constant over  $i$ . Therefore, due to the normalization condition, the Rényi entropy is maximized by

$$p_i = \frac{1}{N} \quad (3.63)$$

#### Canonical Ensemble

We now solve for the following Euler-Lagrange Equations

$$\frac{d}{dp_i} \left[ \frac{\ln(\sum_i p_i^\alpha)}{1-\alpha} - \lambda \sum_i p_i - \beta \sum_i E_i p_i \right] = 0. \quad (3.64)$$

$$\implies \frac{\alpha}{1-\alpha} \frac{p_i^{\alpha-1}}{\sum_j p_j^\alpha} - \lambda - \beta E_i. \quad (3.65)$$

We can multiply by  $p_i$  and sum over  $i$ , using the normalization condition  $\sum_i p_i = 1$  to find

$$\lambda = \frac{\alpha}{1-\alpha} - \beta \langle E \rangle \quad (3.66)$$

We may then write the probabilities as

$$p_i = \left( \sum_j p_j^\alpha \right)^{\frac{1}{\alpha-1}} \left( 1 - \frac{\alpha}{1-\alpha} \beta (E_i - \langle E \rangle) \right)^{\frac{1}{\alpha-1}}, \quad (3.67)$$

Imposing the normalization once again we find

$$\frac{\sum_i p_i}{\left( \sum_j p_j^\alpha \right)^{\frac{1}{\alpha-1}} \sum_i \left( 1 - \frac{\alpha}{1-\alpha} \beta (E_i - \langle E \rangle) \right)^{\frac{1}{\alpha-1}}} = 1. \quad (3.68)$$

We then have as the maximum entropy distribution

$$p_i = \frac{1}{Z_\alpha} \left( 1 - \frac{\alpha}{1-\alpha} \beta (E_i - \langle E \rangle) \right)^{\frac{1}{\alpha-1}} \quad (3.69)$$

with

$$Z_\alpha = \sum_i \left( 1 - \frac{\alpha}{1-\alpha} \beta (E_i - \langle E \rangle) \right)^{\frac{1}{\alpha-1}} \quad (3.70)$$

	Entropy	c	d
BGS	$S = -\sum_i p_i \ln p_i$	1	1
Rényi	$R_\alpha = \frac{1}{1-\alpha} \ln \sum_i p_i^\alpha$	1	1
Tsallis 1	$T_q = \frac{1-\sum_i p_i^q}{q-1}, (q < 1)$	q	0
Tsallis 2	$T_q = \frac{1-\sum_i p_i^q}{q-1}, (q > 1)$	1	0
Kaniadakis	$S_\kappa = -\sum_i p_i (p_i^\kappa - p_i^{-\kappa})/2\kappa, (0 < \kappa \leq 1)$	1 - $\kappa$	0
Abe	$A_q = -\sum_i (p_i^q - p_i^{1/q})/(q-1/q), (0 < q \leq 1)$	q	0
Sharma-Mittal	$SM_{r,\kappa} = -\sum_i p_i^r (p_i^\kappa - p_i^{-\kappa})/2\kappa$	r - $\kappa$	0
Landsberg-Vedral	$LV_q = \left( (\sum_i p_i^q)^{-1} - 1 \right) / (q-1), (q > 1)$	2 - q	0
Curado	$C_b = \sum_i (1 - e^{-bp_i}) + e^{-b} - 1, (b > 0)$	1	0
Exponential	$S_E = \sum_i p_i (1 - e^{-(p_i-1)})$	1	0
Anteonodo-Plastino	$AP_\eta = \sum_i \left( \Gamma\left(\frac{\eta+1}{1}, -\ln p_i\right) - p_i \Gamma\left(\frac{\eta+1}{\eta}\right) \right)$	1	$\frac{1}{\eta}$
Ubricco	$U_\gamma = \sum_i p_i \ln^{\frac{1}{\gamma}}\left(\frac{1}{p_i}\right)$	1	$\frac{1}{\gamma}$
Shafee	$S_\beta = -\sum_i p_i^\beta \ln\left(\frac{1}{p_i}\right)$	$\beta$	1
Hanel-Thurner 1	$HT_{c,d} = \sum_i \left( \frac{e}{c} \Gamma(d+1, 1 - c \ln p_i) - p_i \right) / (1 - c + cd)$	c	d
Hanel-Thurner 2	$HT_{c,d}^{(2)} = \sum_i p_i^c \ln^d\left(\frac{1}{p_i}\right), d \neq 0$	c	d

Table 3.1: The equivalence classes of multiple entropies. The data of this table is taken from Hanel *et al.*, [38]. The function  $\Gamma(s, x)$  is the incomplete gamma function, given by  $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t}$

### 3.5 Scaling Exponents

In this section we will discuss a classification of generalized entropies (entropies that satisfy SK1-SK3), proposed by Hanel and Thurner [30]. This classification is based on scaling behavior of the entropy functions for values of  $p_i$  near zero, and thus on the behavior of such functions in the thermodynamic limit. We will first consider entropies of trace form. For the equidistribution  $p_i = 1/N$ , such entropies take the form

$$S_g(N) = \sum_i^N g\left(\frac{1}{N}\right) = Ng\left(\frac{1}{N}\right) \quad (3.71)$$

The first scaling exponent will be obtained from the relative change of the entropy, when going from  $N$  states to  $\lambda N$  states, when  $N$  is large. If all probabilities are equal, we get

$$\lim_{N \rightarrow \infty} \frac{S_g(\lambda N)}{S_g(N)} = \lim_{N \rightarrow \infty} \lambda \frac{g\left(\frac{1}{\lambda N}\right)}{g\left(\frac{1}{N}\right)} = \lambda^{1-c}. \quad (3.72)$$

The last equality is a consequence of the fact that, for a function  $g$  satisfying SK1-SK3, the following property will be satisfied [30, Thm 1]

$$f(z) \equiv \lim_{x \rightarrow 0} \frac{g(xz)}{g(x)} = z^c. \quad (3.73)$$

From this we see that

$$\lim_{N \rightarrow \infty} \frac{S_g(\lambda N)}{S_g(N)} \lambda^{c-1} = 1 \quad (3.74)$$

We obtain the second property by taking  $\lambda \rightarrow N^a$ . We obtain [30, Thm 2]

$$\lim_{N \rightarrow \infty} \frac{S_g(N^{1+a})}{S_g(N) N^{a(1-c)}} = \lim_{N \rightarrow \infty} \frac{g\left(\frac{1}{N^{1+a}}\right)}{g\left(\frac{1}{N}\right)} N^{ac} = (1+a)^d \quad (3.75)$$

In the following section we will see how we can use these scaling exponents to find the extensive entropy for a given type of growth of phase space, upon composition of multiple systems.

### 3.6 Imposing Extensivity

When composing two systems that are uncorrelated, the phase spaces are simply multiplied. For any of the  $W_1$  states of system A, system B may, with independent probabilities, be in any state of the  $W_2$  states. The phase space for the composition of  $N$  identical systems thus grows like  $W^N$ . In general, for a joint system, the amount of states available to the total system must satisfy

$$W_T < W_1 \cdots W_N. \quad (3.76)$$

We will see that the way the space space grows will determine the entropy equivalence class. Imposing extensivity will mean that we require that, according to property 3.2, as  $N \rightarrow \infty$ ,  $S(W_T(N)) \approx \phi N$  where  $\phi$  is some constant. We thus impose that

$$\left. \frac{d}{dN} S_g(W_T(N)) \right|_{N \rightarrow \infty} = W'_T(N) \left( g \left( \frac{1}{W_T(N)} \right) - \frac{1}{W_T(N)} g' \left( \frac{1}{W_T(N)} \right) \right) = \phi \quad (3.77)$$

Taking the derivative of equation 3.72 w.r.t.  $\lambda$  we see that

$$\lim_{N \rightarrow \infty} g' \left( \frac{\lambda}{N} \right) = \lim_{N \rightarrow \infty} cN \lambda^{c-1} g(x) \quad (3.78)$$

Using this and setting  $\lambda = 1$ , we can write equation 3.77 as

$$W'_T g \left( \frac{1}{W_T} \right) (1 - c) = \phi \quad (3.79)$$

$$\implies \lim_{N \rightarrow \infty} N \frac{W'_T}{W_T} = \frac{1}{1 - c}. \quad (3.80)$$

Equation 3.80 was obtained by imposing the extensivity condition  $W_T g(1/W) \approx \phi N$ . To find  $d$ , we use

$$\lim_{N \rightarrow \infty} \frac{d}{da} \frac{g \left( \frac{1}{W_T^{1+a}} \right)}{\frac{1}{W_T^c} g \left( \frac{1}{W_T} \right)} = d(1+a)^{d-1} \quad (3.81)$$

$$\implies_{a \rightarrow 0} d = \lim_{N \rightarrow \infty} \ln W_T \frac{g' \left( \frac{1}{W_T} \right) \frac{1}{W_T} - c g \left( \frac{1}{W_T} \right)}{g \left( \frac{1}{W_T} \right)} \quad (3.82)$$

Then using,

$$g' \left( \frac{1}{W_T} \right) = N\phi - \frac{W_T}{W'_T} \phi, \quad (3.83)$$

which is obtained from the extensivity condition we find that

$$d = \lim_{N \rightarrow \infty} \ln W_T \left( \frac{1}{N} \frac{W}{W'} + c - 1 \right) \quad (3.84)$$

#### Examples of Phase Space Growth and Their Extensive Entropy

- **Exponential Phase Space Growth:**  $W(N) \propto e^N$ . As we have mentioned, this is the standard type of growth that happens when composing independent systems. As such we expect to find the BGS equivalence class. From equations 3.80 and 3.82 we find

$$c = \lim_{N \rightarrow \infty} 1 - \frac{1}{N} = 1 \quad (3.85)$$

$$d = \lim_{N \rightarrow \infty} \ln N \frac{1}{N} = 1. \quad (3.86)$$

Thus, the equivalence class of the extensive entropy is indeed  $(c, d) = (1, 1)$

- **Power Law Phase Space Growth:**  $W(N) \propto N^b$ . We have already seen this type of growth as well for the Tsallis entropy 3.3.4. We find

$$c = \lim_{N \rightarrow \infty} 1 - \frac{1}{bNN^{-1}} = 1 - \frac{1}{b} \quad (3.87)$$

$$d = \lim_{N \rightarrow \infty} b \log N \left( \frac{1}{N} \frac{N}{b} - \frac{1}{b} \right) = 0 \quad (3.88)$$

The equivalence class is thus  $(1 - \frac{1}{b}, 0)$  which is Tsallis like.

- **Stretched Exponential Phase Space Growth:**  $W(N) \propto \exp \lambda N^\gamma$ . We have

$$c = \lim_{N \rightarrow \infty} 1 - \frac{1}{\lambda \gamma N^{\gamma-1}} = 1 \quad (3.89)$$

$$d = \lim_{N \rightarrow \infty} \lambda N^\gamma \left( \frac{1}{N} \frac{1}{\lambda \gamma N^{\gamma-1}} \right) = \frac{1}{\gamma} \quad (3.90)$$

Thus the equivalence class is  $(1, \frac{1}{\gamma})$ , which is associated with the so called Anteonodo-Plastino entropy.

### 3.7 Summary and Concluding Remarks

In this chapter we have discussed the generalized entropies. All these generalized entropies follow from relaxing one of Khinchin's axioms that define the Shannon entropy. The axiom relaxed is the separability axiom. In general, the generalized entropies that follow are continuous functions of all the probabilities of a probability distribution and are of the form

$$G\left(\sum_i g(p_i)\right), \quad (3.91)$$

Moreover, adding a zero probability should leave these entropies invariant, and they should all be maximal for the equidistribution. If  $G = id$ , then the entropies are said to be of trace form.

Next, we properly defined two qualities of the entropies. First, we defined extensivity of an entropy to hold when we can define a mean entropy when composing  $N$  systems, in the limit where we take  $N \rightarrow \infty$ . After that, we discussed what it means for an entropy to be experimentally robust, and how Lesche stability may be used to show robustness.

The first generalized entropy that we discussed separately was the Tsallis entropy. Although the Tsallis entropy has seen some interest in information theory, the largest part of its interest comes from the fact that it may be an extensive entropy for systems where the BGS entropy is not. Whereas the Shannon entropy is extensive when the composition of two systems causes a multiplication of phase space, the Tsallis entropy is extensive when there is a smaller amount of states available upon increasing the size of the system. In this light, the parameter  $q$  becomes a tweakable parameter, to make the entropy extensive for different types of systems. More than that, the Tsallis entropy also possesses a few characteristics that make it interesting for thermodynamical applications. Among other things, it is the only composable entropy of trace form, that allows for some kind of definition of equilibrium properties, in accordance with the first law. Of the properties we discussed, the only thing that it doesn't share with the BGS entropy is the additivity, but it is exactly due to this that the Tsallis entropy is extensive for different systems than the BGS entropy.

One important aspect that we have not discussed due to the scope of this work, is that the current Tsallis framework, is based upon the use of escort probabilities,  $P_i \equiv p_i^q / (\sum_j p_j^q)$  and so called  $q$  expectation values  $\langle A \rangle_q \equiv \sum P_i A_i$ . This again opens up discussion about some of the established properties such as Lesche stability, as well as about the meaning of these escort probabilities [51][35]. For a comprehensive guide on the modern interpretation of the Tsallis statistics, the book by Tsallis himself is recommended [36].

Next, we have discussed the Rényi entropy. This entropy is based mostly information on theoretic concepts, such as an axiomatic characterization and the idea of generalized means. There has been some research into thermodynamical applications, but, overall, due to the additivity of the Rényi entropy, it is extensive for the same systems as the BGS entropy. As a consequence, it governs the same statistics as the BGS entropy in the thermodynamic limit. Moreover, it does not generally possess the concavity and sub-additivity properties.

We have also revisited Jaynes' principle of maximum entropy. Mathematically, the exact same principle may be found when trying to find the most likely distribution of a random process where we have history-independent probabilities for the events. This shows the limitations of entropy maximization, based solely on the BGS entropy as the fundamental measure of uncertainty.

The failure of the BGS entropy to be extensive for all but uncorrelated systems, and the narrowness of Jaynes' maximum entropy method, show clearly where there is a need for generalized entropies. These generalized entropies come in many different forms, so we have also discussed a classification for them, based on their behavior in the thermodynamic limit. These classes are based on two scaling exponents,  $c$  and  $d$ , and are proposed by Hanel and Thurner [52]. We have discussed how these scaling exponents may be used to characterize the entropies based on their type of growth in phase space, when all probabilities of the states are equal. The assumption that all probabilities of the available states are equal is may be seen as a bit limiting, though, since this obviously does not hold for all systems. At the same time, when an entropy is extensive for a particular system, then we may, by definition, assume that there is some mean entropy per particle. For a detailed work in the case where not all probabilities are equal, see the work by Balogh *et al.* [53]. The generalized entropies still apply, but a bit more care may be needed in specific cases.

Moreover, we have remarked that, when formulating equations to find the most probable distribution for more general processes, the scaling exponents may also be used to characterize certain processes. In information theory there is also some interest in the entropy production rate of a process. This is defined as  $\lim_{N \rightarrow \infty} 1/N \log(p_{i_1 \dots i_N})$ , where  $p_{i_1 \dots i_N}$  is joint probability of finding the sequence of outcomes  $i_1 \dots i_N$ . The specifics of this topics are beyond the scope of this text, but we will make a brief remark about it. In Markovian systems, which are ergodic, the entropy rate goes to the BGS entropy [38]. However, for non-Markovian systems, the scaling exponents may also be used to characterize the entropy that gives the entropy rate of the system. Hanel *et al.* [38] conclude that for ergodic and Markovian systems, we have that

$$S_{\text{ext}} = S_{\text{max}} = S_{\text{prod}}, \quad (3.92)$$

or in other words, the extensive entropy, the entropy that gives the most likely distribution, and the entropy that gives the entropy production rate are equal. However, for complex systems, like non-ergodic and non-Markovian systems, we have in general that

$$S_{\text{ext}} \neq S_{\text{max}} \neq S_{\text{prod}}. \quad (3.93)$$



## Chapter 4

# Density Matrices, Mixed States and Quantum Operations

This chapter will serve as a short preliminary chapter to the following chapters. We will introduce the density matrix formalism of quantum mechanics. This formalism will be crucial to better understand quantum correlations, marginal quantum states, analogous to the marginal distributions in the classical case, as well as the related topic of encoding classic probabilities or a classical lack of knowledge into states of quantum systems. States that have such information encoded in them are called mixed states. We will use this formalism in most chapters after this one. Then in the final section of this chapter, we will discuss a way of defining the mixedness of a density matrix, through so called majorization. This will allow us to define measures of mixedness in chapter 7 and also connect them to entanglement measures in chapter 8

### 4.1 Introduction to the Density Matrix: The Particle Source

Consider a particle source that emits a particle, while at the same time entangling itself with that particle. We will consider the simple case of an emitted particle characterized by a spin. Every time the source emits a particle, the quantum state of particle together with the source may be written as follows

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow_z\rangle |S_\uparrow\rangle + |\downarrow_z\rangle |S_\downarrow\rangle). \quad (4.1)$$

This may be done by, for example, entangling the spin of the emitted particle with the spin of another particle at the source, but it may also be entangled with something else, as long as the states are orthogonal. A question now arises when we ask how to describe the state of only the emitted particle. On first thought we might think that  $\frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle)$  would suffice, since it describes the correct spin measurements in the  $z$  direction. However, we must consider that the state is entangled with the source. If a third party would collapse the source by measurement, the emitted particle would only arrive as an up or down spin, not the quantum mechanical superposition. This matters when decomposing our state into a different representation. If the source collapsed beforehand, then the probability to measure up or down in the  $x$  direction would always be  $1/2$ . In contrast, the state  $\frac{1}{\sqrt{2}}(|\uparrow_z\rangle + |\downarrow_z\rangle)$  has a probability of 1 to measure spin up in the  $x$  direction.

However, on further inspection we may see that it shouldn't, and doesn't matter if the source is collapsed beforehand or not. The entanglement between the particle and the source causes the particle to behave like  $|\uparrow_z\rangle$  or  $|\downarrow_z\rangle$  half the time. To describe such systems, we have to introduce the density matrix formalism of quantum mechanics.

## 4.2 The Density Matrix

The density matrix of a general state  $|\psi\rangle$  is given by  $\hat{\rho} = |\psi\rangle\langle\psi|$ . For the state of equation 4.1 we then get

$$\hat{\rho} = |\psi\rangle\langle\psi| = \frac{1}{2} (|\uparrow_z S_\uparrow\rangle\langle\uparrow_z S_\uparrow| + |\downarrow_z S_\downarrow\rangle\langle\downarrow_z S_\downarrow| + |\uparrow_z S_\uparrow\rangle\langle\downarrow_z S_\downarrow| + |\downarrow_z S_\downarrow\rangle\langle\uparrow_z S_\uparrow|). \quad (4.2)$$

We will denote the indices of the densities as follows: for a two system wavefunction we write  $\hat{\rho}_{ij,kl}$  with pairwise indices  $ij$  and  $kl$  for the kets and the bras. In other words, the indices  $ik$  denote the states in the kets and bras respectively for the first system, and the same holds for the indices  $jl$  for the second system. We may find the regular expectation values of some operator  $\hat{A}$  through the trace operation as follows

$$\langle\hat{A}\rangle = \text{Tr} [\hat{\rho}\hat{A}] \quad (4.3)$$

We see this by noting that any Hermitian operator  $\hat{A}$  may be written as a sum of projectors multiplied by the corresponding eigenvalues,  $\hat{A} = \sum_i \lambda_i |i\rangle\langle i|$ . Equation 4.3 may then be written as

$$\text{Tr} \left[ \sum_i \lambda_i \hat{\rho} |i\rangle\langle i| \right] = \sum_{i,j} \lambda_i \langle j|\psi\rangle\langle\psi|i\rangle\langle i|j\rangle = \sum_{i,j} \delta_{ij} \langle j|\psi\rangle\langle\psi|i\rangle = \sum_i \lambda_i |\langle\psi|i\rangle|^2. \quad (4.4)$$

We now wish to measure the spin of only the first system, which may be represented by the following operator  $\sigma_z \otimes \mathbb{1}$ . We thus calculate

$$\langle\sigma_z \otimes \mathbb{1}\rangle = \text{Tr} [\hat{\rho}\sigma_z \otimes \mathbb{1}] = \sum_{i,j,l,k} \hat{\rho}_{ijkl} (\sigma_z)_{ki} \mathbb{1}_{lj} = \sum_{i,j,k} \hat{\rho}_{ijkj} (\sigma_z)_{ki}. \quad (4.5)$$

The sum over  $j$  in the above equation is independent of the operator  $\sigma_z$ . This motivates us to define a new density matrix for the first system, which reproduces the same expectation value when only contracted with  $\sigma_z$ ,

$$\hat{\rho}_{ij}^A = \text{Tr}_B [\hat{\rho}_{ikjl}] \equiv \sum_k \hat{\rho}_{ikjk}. \quad (4.6)$$

We call this operation the partial trace. Computing the partial trace over the source states of equation 4.2, we obtain

$$\hat{\rho}^A = \frac{1}{2} (|\uparrow_z\rangle\langle\uparrow_z| + |\downarrow_z\rangle\langle\downarrow_z|), \quad (4.7)$$

which we call the reduced density matrix. It may be regarded as the quantum analog to the marginal probability distributions in statistics. According to this state we have a chance of  $\frac{1}{2}$  to measure the spin of the particle in either spin up or down. The standard interpretation of this density matrix is then that it describes the results of measurements on the emitted particle over an infinite amount of independent measurements.

At the same time, the density matrix may be used to encode a classical lack of knowledge about the state of the system. If we wish to describe an ensemble of systems where the quantum state is always either  $|\uparrow_z S_\uparrow\rangle$  or  $|\downarrow_z S_\downarrow\rangle$  half of the time, we may describe the outcomes over all ensembles with the so-called mixed state

$$\hat{\rho} = \frac{1}{2} (|\uparrow_z S_\uparrow\rangle\langle\uparrow_z S_\uparrow| + |\downarrow_z S_\downarrow\rangle\langle\downarrow_z S_\downarrow|). \quad (4.8)$$

The reduced density matrix of the emitted particle in this case is the same as in equation 4.7. Such states in general are called mixed states. When the mixing comes from tracing away another system, the mixture is called an improper mixture, while a mixture that represents a classical lack of knowledge is called a proper mixture. Mathematically, these mixtures are actually completely equivalent.

Lastly we note that, since the density matrix  $\hat{\rho}$  is Hermitian, it is always diagonalizable. The eigenvalues of the density matrix may be interpreted as probability that the system turns out to be in the pure state that corresponds to that eigenvalue. In this basis, the density matrix may be written as a sum of density matrices corresponding to orthonormal pure states,  $\hat{\rho} = \sum_i \lambda_i |i\rangle \langle i|$  with  $\sum_i \lambda_i = 1$ . Such a sum is called an incoherent sum, in the sense that no quantum coherence between the states  $|i\rangle \langle i|$  is added.

### 4.3 Schmidt Decomposition

We will now discuss the Schmidt decomposition, which is an important tool when using density matrices. Let  $\mathcal{H}$  be a Hilbert space which is a tensor product of smaller Hilbert spaces. Let us now introduce some arbitrary partitioning on this Hilbert space such that  $\mathcal{H} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$ . Due to Schmidt's theorem, we know that we can always decompose any state as follows

$$|\Psi\rangle = \sum_i c_i |i_A\rangle |i_B\rangle, \quad (4.9)$$

where  $|i_A\rangle$  are states of  $\mathcal{H}_A$ , and  $|i_B\rangle$  of  $\mathcal{H}_B$ . The coefficients  $c_i$  are called the Schmidt coefficients. If we take the partial trace of the density matrix, in this basis this means that we get for the first system

$$\hat{\rho}_1 = \sum_i |c_i|^2 |i\rangle \langle i|, \quad (4.10)$$

and for the second

$$\hat{\rho}_2 = \sum_i |c_i|^2 |i'\rangle \langle i'|. \quad (4.11)$$

If the states are separable then this implies is that one of the Schmidt coefficients is 1 while the others are 0. In this way the Schmidt rank, which is the number of non-zero Schmidt coefficients, is often considered as a measure for entanglement. If the Schmidt rank is 1, then the state is separable, while any other Schmidt rank indicates entanglement. Thus this measures entanglement in a binary way. In chapter 8 we discuss continuous measures of entanglement.

### 4.4 Quantum Operations

Following the density matrix formulation of quantum mechanics, the quantum operation formalism has been introduced to describe general quantum evolution a system can undergo, that cannot be directly written as unitary evolution. Quantum operations may be derived from a set physically motivated axioms<sup>1</sup>, or may be directly implemented using a special set of operators, which induce non-unitary behaviour on a system. However, quantum operations may always be represented in an equivalent manner as the effect of a unitary operation on the system in question coupled with an environment [54, §8.2]. Quantum operations may also always be represented as the action of so called Kraus operators, which we will show in the next section. Following this, we will show as an example how the Kraus operators naturally emerge from unitary evolution after the partial trace. The quantum operation formalism has found many uses in the description of open quantum systems, quantum computation, and quantum information theory, in which case they are usually referred to as quantum channels. In chapter 8 we will define a way of characterizing entanglement which is based on quantum operations. Moreover, quantum operations are used a lot within quantum information and communication, in which entropy quantities are also used extensively. As such, we feel it is important to briefly discuss quantum operations, though we won't rely too heavily on them ourselves.

Quantum operations are most generally defined as linear maps on the density matrices

$$\hat{\rho} \rightarrow \hat{\rho}' = \Phi[\hat{\rho}], \quad (4.12)$$

---

<sup>1</sup>Master equations like the Lindblad equation are an example of this.

which are completely positive and trace preserving, CPTP in short. However, more generally, sometimes in literature a different type of map is also considered. Namely, completely positive, non-trace-increasing maps. We will treat these separately and explain their relevance.

All such quantum maps, which are linear, completely positive and non-trace increasing, may be written in the operator sum representation [54, Thm 8.1], such that

$$\hat{\rho}' = \Phi[\hat{\rho}] = \sum_n \hat{K}_n \hat{\rho} \hat{K}_n^\dagger. \quad (4.13)$$

The operators  $\hat{K}_n$  and  $\hat{K}_n^\dagger$  are called the Kraus operators and obey the relation<sup>2</sup>

$$\sum_n K_n^\dagger K_n = \mathbb{1} \quad (4.14)$$

Note that quantum maps need not to map on Hilbert spaces of equal dimension<sup>3</sup>. Let  $\Phi[\rho]$  is a quantum operation that maps Hilbert space  $\mathcal{H}$  to  $\mathcal{G}$ , of dimensions  $n$  and  $m$ , respectively. Then, the Kraus operators representing this map are  $m \times n$  matrices.

#### 4.4.1 Kraus Operators from Unitary Time Evolution

Let the time evolution of a state be generated by the unitary operator  $\hat{U}_t$ . We write  $|\psi(t)\rangle = \hat{U}_t |\psi(0)\rangle$ . In the same way, the time evolved density matrix is given by

$$|\psi(t)\rangle \langle \psi(t)| = \hat{U}_t |\psi(0)\rangle \langle \psi(0)| \hat{U}_t^\dagger. \quad (4.15)$$

For our example we will consider a product state of two systems, where the second system is in some pure state  $\hat{\sigma} = |e_m\rangle \langle e_m|$ , such that  $|e_m\rangle$  forms an orthonormal basis. If we take the partial trace of the time evolved state, we obtain

$$\text{Tr}_2 \left[ \hat{U}_t |\psi_1\rangle \langle \psi_1| \otimes |\psi_2\rangle \langle \psi_2| \hat{U}_t^\dagger \right] = \sum_n \langle n | \hat{U}_t |e_m\rangle |\psi_1\rangle \langle \psi_1| \langle e_m | \hat{U}_t^\dagger |n\rangle, \quad (4.16)$$

where the states  $|n\rangle$  form some orthonormal basis for the second system. We see we have obtained a new operator, only acting on the first state, namely

$$\hat{K}_n \equiv \langle n | \hat{U}_t |e_m\rangle, \quad (4.17)$$

which are in fact a set of Kraus operators. We can see they satisfy the completeness relation

$$\sum_n \hat{K}_n^\dagger \hat{K}_n = \sum_n \langle e_m | \hat{U}_t^\dagger |n\rangle \langle n | \hat{U}_t |e_m\rangle = \langle e_m | \hat{U}_t^\dagger \mathbb{1} \hat{U}_t |e_m\rangle = \mathbb{1}, \quad (4.18)$$

More generally we may consider the case where the two systems are initially uncorrelated, but the second system is in a general state  $\hat{\sigma} = \sum_m p_m |e_m\rangle \langle e_m|$ . Much in the same way we find the Kraus operators

$$\hat{K}_{mn} = \sqrt{p_m} \langle n | \hat{U}_t |e_m\rangle, \quad (4.19)$$

obeying the completeness relation

$$\sum_{m,n} \hat{K}_{mn}^\dagger \hat{K}_{mn} = \mathbb{1} \quad (4.20)$$

<sup>2</sup>Quantum operations are also sometimes classified as all operations  $\sum_n K_n^\dagger K_n \leq c \cdot \mathbb{1}$ , where  $|c| \leq 1$ . In this case the quantum operation projects onto the post measurement state, but doesn't produce a normalized density matrix.

<sup>3</sup>A quantum operation may, for example, map a two spin state on the partial traced one spin state, or couple additional particles to our original state.

## 4.5 Defining Mixedness through Majorization

A way of defining mixedness, which was formulated by Uhlmann [2] is through the majorization. Majorization is a preorder<sup>4</sup>, which we can define on the set of density matrices through their spectrum. In the next section, we will see that this definition of mixedness will give us a way to link entropic measures with entanglement.

Let  $\hat{\rho}_1$  and  $\hat{\rho}_2$  be two density matrices with eigenvalues  $\lambda_i[\hat{\rho}_1]$  and  $\lambda_i[\hat{\rho}_2]$  all in descending order w.r.t.  $i$ . We say that  $\hat{\rho}_1$  majorizes  $\hat{\rho}_2$  if the following relation holds

$$\sum_i^n \lambda_i[\hat{\rho}_1] \geq \sum_i^n \lambda_i[\hat{\rho}_2], \quad (4.21)$$

for all  $n$ . We denote the majorization relation as

$$\hat{\rho}_1 \succ \hat{\rho}_2. \quad (4.22)$$

This relation comes from the notion of mixing enhancing, which states that if a state  $\hat{\rho}_2$  may be written as a convex sum in the form of  $\sum p_i \hat{U}_i \hat{\rho}_1 \hat{U}_i^\dagger$ , where  $\hat{U}_i$  are unitary matrices, then  $\hat{\rho}_2$  must be more mixed than  $\hat{\rho}_1$ . Using information theoretic concepts, the underlying assumptions here are that  $\hat{\rho}_2$  and  $\hat{U}_i \hat{\rho}_1 \hat{U}_i^\dagger$  contain the same amount of information about the state, while the convex sum of density matrices decreases the amount of information, just as in classical information theory. To show that mixing enhancement implies majorization, we may use Ky Fan's [2, Eq. 2.6] inequality, which states that for any orthonormal basis  $\{|\phi_i\rangle\}$ ,  $1 \leq i \leq N$ ,

$$\sum_i^n \lambda_i[\hat{\rho}] \geq \sum_i^n \langle \phi_i | \hat{\rho} | \phi_i \rangle. \quad (4.23)$$

Of course, equality must be obtained when  $n = N$ . Next, let  $|\phi_i\rangle$  be the eigenvector belonging to  $\lambda_i[\hat{\rho}_2]$ , we then have

$$\sum_i^n \lambda_i[\hat{\rho}_2] = \sum_i^n \sum_j^n p_j \langle \phi_i | \hat{U}_j \hat{\rho}_1 \hat{U}_j^\dagger | \phi_i \rangle \leq \sum_i^n \sum_j^n p_j \lambda_i[\hat{\rho}_1] = \sum_i^n \lambda_i[\hat{\rho}_1]. \quad (4.24)$$

In fact, we can extend the notion of mixing enhancing to general quantum operations. If the Kraus operators representing the CPTP operation satisfy the following condition [55, Thm. 2],

$$\sum_n \hat{K}_n \hat{K}_n^\dagger = \mathbb{1} \quad (4.25)$$

then the map is always mixing enhancing. In other words, the map is mixing enhancing if the Kraus operators are normal operators. We see that Uhlmann's notion of mixing enhancing is only a special case of a larger group of quantum operations.

### 4.5.1 Schur-Convexity (Concavity)

Of course, the question now arises what measures exist that are monotone in the majorization relation. With monotonicity of a measure  $\text{Tr}[f(\hat{\rho})]$ , where  $f$  is some function of the density matrix, we mean that  $\hat{\rho}_1 \succ \hat{\rho}_2$  implies that we have that either  $\text{Tr}[f(\hat{\rho}_1)] \geq \text{Tr}[f(\hat{\rho}_2)]$  or  $\text{Tr}[f(\hat{\rho}_1)] \leq \text{Tr}[f(\hat{\rho}_2)]$ . For majorization, these properties are called Schur convex and concave, respectively. Schur convexity (concavity) may be checked by the following criterion [56, Thm. 12.25]: If the function  $f$  is symmetric, then it is Schur convex (concave) if and only if

$$(\lambda_i - \lambda_j) \left( \frac{\partial f(\lambda_1, \dots, \lambda_n)}{\partial \lambda_i} - \frac{\partial f(\lambda_1, \dots, \lambda_n)}{\partial \lambda_j} \right) \underset{(\leq)}{\geq} 0, \forall i \neq j. \quad (4.26)$$

In chapter 7 we define entropies on the spectrum of the density matrix and we will discuss whether they are Schur-convex (concave).

<sup>4</sup>A preorder is a reflexive and transitive relation.

## Chapter 5

# Measurement Uncertainty in Discrete Hilbert Spaces

When we perform a measurement on a quantum system, there is an inherent uncertainty in outcome, if the state is not already in an eigenstate of the measurement. This uncertainty is always present if we wish to measure two non-commuting operators on a single state, where we can never be certain about the outcomes of both measurements. Historically this fact is most commonly expressed using standard deviations, in Heisenberg's uncertainty relation, and its discrete counterpart, the Robertson relation. The standard deviation of the state, for the measurement of operator  $\hat{A}$  is given by

$$\sigma_{\hat{A}}(|\psi\rangle) = \sqrt{\langle\psi|\hat{A}^2|\psi\rangle - (\langle\psi|\hat{A}|\psi\rangle)^2} \quad (5.1)$$

In more recent years, entropic measures have been considered as measures of uncertainty for quantum measurements, giving rise to, among other things, new formulations of the uncertainty principle. The entropies may be applied directly to the probabilities of a certain outcome obtained by the Born rule. Let  $\hat{A}$  be a measurement operator, with eigenbasis  $|a_i\rangle$ . The probability the measured state  $|\psi\rangle$  is collapsed to state  $|a_i\rangle$  is given by the Born rule

$$p_a \equiv |c_a|^2 = \langle a_i|\hat{\rho}|a_i\rangle. \quad (5.2)$$

We denote the measurement BGS entropy in the eigenbasis of  $\hat{A}$  as

$$S[\hat{A}, \hat{\rho}] \equiv \sum_a -p_a \ln p_a. \quad (5.3)$$

If the state is pure, we will use  $\psi$  instead of  $\hat{\rho}$ . Part of the interest in the formulation of novel uncertainty relations in terms of the entropies comes from quantum cryptography. An excellent review of such applications is given by Coles *et al.* [57].

The Rényi and Tsallis entropies are often considered as well and may be applied in the same way to the probabilities. Using Kraus operators, we may also express this entropy as a functional of the post measurement state. Such a state represents the a-priori classical lack of information we have about a quantum state after a measurement has been performed. If we take as the Kraus operators the set of orthogonal projectors in the eigenbasis of  $\hat{A}$ ,  $\hat{K}_i = |a_i\rangle\langle a_i|$ , then we obtain as post measurement state

$$\sum_i \hat{K}_i^\dagger \hat{\rho} \hat{K}_i = \sum_i p_i |a_i\rangle\langle a_i| \equiv \hat{\rho}_{\text{pm}}. \quad (5.4)$$

The measurement uncertainty is now simply given by

$$S[\hat{A}, \hat{\rho}] = \text{Tr} [-\hat{\rho}_{\text{pm}} \ln \hat{\rho}_{\text{pm}}]. \quad (5.5)$$

This is actually the well known Von Neumann entropy, which we will discuss in further detail in chapter 7.

## 5.1 The Uncertainty Principle

The standard textbook description of the uncertainty of the simultaneous measurement of two non-commuting observables involves the well known Robertson relation,

$$\sigma_{\hat{A}}(|\psi\rangle)\sigma_{\hat{B}}(|\psi\rangle) \geq \frac{1}{2} |\langle \psi | [\hat{A}, \hat{B}] | \psi \rangle|^2. \quad (5.6)$$

This relation intuitively portrays the uncertainty principle quite nicely, showing that there must indeed be a lower bound on the amount of certainty we can have about the outcomes of two different measurements on a single state. Heisenberg's famous uncertainty principle for the position and momentum representations accomplishes this by expressing a global minimum over all states  $\phi$ , which is only attained by a few special states, called coherent states. It states the absolute limit on the knowledge that we may have about two different measurement outcomes of a single state. The above relation does not accomplish the same, and we may broadly identify two problems with it. First, both sides are state dependent. As such, it does not express a global minimum over the space of states. To solve this, one might think that we could take the infimum over all  $|\phi\rangle$ , but this actually leads us to the second problem. This is that the infimum over  $|\phi\rangle$  for given  $\hat{A}$  and  $\hat{B}$  is always zero. This happens when  $|\psi\rangle$  is an eigenvector of either  $\hat{A}$  or  $\hat{B}$ . We may easily see this from the fact that in this case, either  $\sigma_{\hat{A}}$  or  $\sigma_{\hat{B}}$  must be zero. Since the right hand side is positive, it must be zero as well. It may also be checked from the right hand side directly. Since the state  $|\psi\rangle$  may always be chosen to be an eigenvector of either operator, any formulation in terms of products of the standard deviations in the outcomes is doomed to have the same trivial global bound<sup>1</sup>. Schrödinger has proposed a stronger version of the relation, but it suffers from the same problems [58, Eq. 9]. Alternatively, relations in terms of sums of the variances have been proposed by Maccone and Pati [59]. One of them takes the following form

$$\sigma_{\hat{A}}^2(|\phi\rangle) + \sigma_{\hat{B}}^2(|\phi\rangle) \geq \pm i \langle \psi | [\hat{A}, \hat{B}] | \psi \rangle + |\langle \psi | \hat{A} \pm i\hat{B} | \psi^\perp \rangle|^2, \quad (5.7)$$

where  $|\psi^\perp\rangle$  is a state, orthogonal to  $|\psi\rangle$  and the choice of sign should be such that  $\pm i \langle \psi | [\hat{A}, \hat{B}] | \psi \rangle$  is positive. Of course since the uncertainty is now stated in terms of a sum of a unitful quantity, we must introduce a conversion factor if  $\hat{A}$  and  $\hat{B}$  are not measured in the same units. If we want to find the minimal uncertainty state however, we still need to find the minimum over all states. Though, it is unclear whether the infimum over all states of this expression would express the exact lowest bound, and it is hard to say whether there is any computational gain w.r.t. directly computing the infimum over all states of the standard deviations directly.

A proposed alternative to variance based uncertainty relations, are those that are based on entropic measures. Deutsch, in search of a relation with the same quantities as Heisenberg's one proposed such a relation [60], which was later improved by Maassen and Uffink [61]. Applying the definition of the entropy on pure states we established at the beginning of this chapter, the relation by Maassen and Uffink may be stated as follows

$$S[\hat{A}, \psi] + S[\hat{B}, \psi] \geq 2 \ln \frac{1}{c} \equiv u_{\text{MU}} \quad (5.8)$$

where  $c$  is the largest overlap between eigenvectors of  $\hat{A}$  and  $\hat{B}$

$$c \equiv \sup_{a,b} |\langle a|b \rangle|. \quad (5.9)$$

As a result, when  $\hat{A}$  and  $\hat{B}$  have at least one common eigenvector,  $c = 1$  and thus the right hand side is equal to 0. This is not surprising, since if we take this eigenvector as our state, then the outcomes

<sup>1</sup>Although it could be argued that formulations of the uncertainty principle in terms of products of any measure of uncertainty should realistically have the same problem. For if  $|\psi\rangle$  is an eigenvector of either operator, we know what the outcome of that measurement will be, so it may be reasonable for a proper measure of uncertainty to give 0 in that case, which is certainly the case for entropic measures.

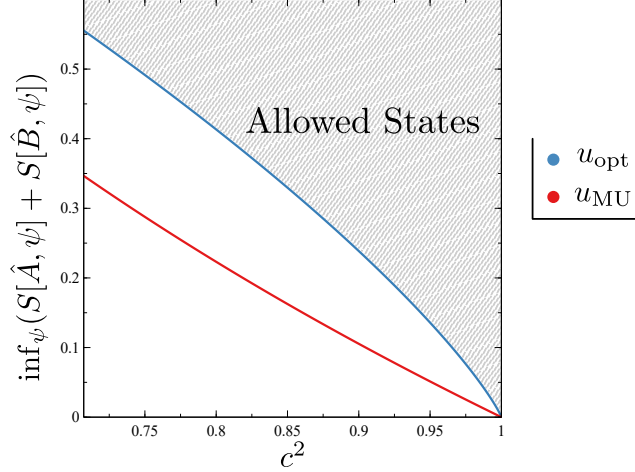


Figure 5.1: A comparison of the Maassen-Uffink bound and the bound found by Ghirardi *et al.*, in the interval  $c^2 \in [\frac{1}{\sqrt{2}}, 1]$

of both measurements are known and contain no uncertainty. We will demonstrate the application of this bound to a  $2D$  Hilbert space in the next section. Moreover, Maassen and Uffink proved [61] that their uncertainty relation is actually part of a family of uncertainty relations that may be defined with the Rényi entropy, where we define the measurement Rényi entropy analogously to the BGS entropy in equation 5.3. Define the parameters  $\frac{1}{\alpha} = \frac{1}{\beta} = 2$ , then

$$R_{\alpha}[\hat{A}, \hat{\rho}] + R_{\beta}[\hat{B}, \hat{\rho}] \geq -2 \ln c. \quad (5.10)$$

The uncertainty relation in terms of the BGS entropy is recovered in the limit  $\alpha = \beta = 1$ .

However, while this fixes the problem of the state dependent bounds, trivial bounds and unital quantities, this relation is known to be not sharp, something that we will also demonstrate in the next section. This means that it does not give the global minimum uncertainty. Since then, tighter bounds have proposed. In  $2D$  Hilbert spaces, exact results were found by Ghirardi *et al.* [62] in the region where  $c^2 \leq \frac{1}{\sqrt{2}}$  that the optimal bound is given by the minimization problem

$$u_{\text{opt}} = \inf_{\theta} \left( s_{\text{bin}} \left( \frac{1 + \cos \theta}{2} \right) + s_{\text{bin}} \left( \frac{1 + \cos(\gamma_c - \theta)}{2} \right) \right), \quad (5.11)$$

where  $s_{\text{bin}}$  is the binary entropy

$$s_{\text{bin}}(x) = -x \ln x - (1 - x) \ln(1 - x), \quad (5.12)$$

and

$$\gamma_c = 2 \arccos \sqrt{c}. \quad (5.13)$$

However, they could only find an analytic solution for  $c^2 \geq \frac{1}{\sqrt{2}}$ , which is given by

$$u_{\text{opt}} = 2s_{\text{bin}} \left( \frac{1 + c}{2} \right). \quad (5.14)$$

This bound is plotted in figure 5.1 alongside the Maassen-Uffink bound. Other numerical solutions in the region  $c^2 < \frac{1}{\sqrt{2}}$  were found as well by De Vicente *et al.* [63]. The Maassen-Uffink bound still holds, but becomes more loose for systems in a Hilbert space with  $D > 2$  as well as for describing the uncertainty in mixed states. In these situations finding tighter bounds becomes quite complex, though several have been proposed [64][65].



### 5.1.1 The Maassen-Uffink Uncertainty Relation in a 2D Hilbert Space

In the following we will apply the measurement entropies in the simple case of a 2D Hilbert space. We will compare results to the bounds that we have found above. We will also demonstrate that the Maassen-Uffink bound is indeed not tight for all pairs of operators  $\hat{A}, \hat{B}$  by showing a particular example. For the first example we will use the eigenvectors Pauli matrices  $\hat{\sigma}_z$  and  $\hat{\sigma}_x$  as our basis. Then  $c = \frac{1}{\sqrt{2}}$ . For this state we can easily see that the uncertainty relation can be minimized. If we take

$$|\psi\rangle = |\uparrow_z\rangle = \frac{1}{\sqrt{2}}(|\uparrow_x\rangle + |\downarrow_x\rangle), \quad (5.15)$$

then

$$S[\hat{\sigma}_z, \psi] + S[\hat{\sigma}_x, \psi] = 0 + \frac{1}{2} \ln 2 + \frac{1}{2} \ln 2, \quad (5.16)$$

which indeed the Maassen-Uffink bound. Moreover, the Rényi entropies are given by

$$R_\alpha[\hat{\sigma}_z, \psi] + S_\beta[\hat{\sigma}_x, \psi] = \frac{1}{1-\alpha} \ln \left( \frac{1}{2^{\alpha-1}} \right) = \ln 2, \quad (5.17)$$

which thus also saturates the Maassen-Uffink bound.

Let us now use the eigenvectors of  $\sigma_z$  and  $\sigma_{\frac{\pi}{3}} \equiv (\cos(\frac{\pi}{3})\sigma_z + \sin(\frac{\pi}{3})\sigma_x)$  as basis. The relation between the eigenvectors is

$$|\uparrow_z\rangle = \frac{\sqrt{3}}{2} |\uparrow_{\frac{\pi}{3}}\rangle - \frac{1}{2} |\downarrow_{\frac{\pi}{3}}\rangle, \quad (5.18)$$

$$|\downarrow_z\rangle = \frac{1}{2} |\uparrow_{\frac{\pi}{3}}\rangle + \frac{\sqrt{3}}{2} |\downarrow_{\frac{\pi}{3}}\rangle, \quad (5.19)$$

meaning  $c = \frac{\sqrt{3}}{2}$ . The most general state can be written as

$$|\psi\rangle = \cos \theta |\uparrow_z\rangle + \sin \theta e^{i\phi} |\downarrow_z\rangle = \left( \frac{\sqrt{3}}{2} \cos \theta + \frac{1}{2} \sin \theta e^{i\phi} \right) |\uparrow_{\frac{\pi}{3}}\rangle + \left( \frac{\sqrt{3}}{2} \sin \theta e^{i\phi} - \frac{1}{2} \cos \theta \right) |\downarrow_{\frac{\pi}{3}}\rangle \quad (5.20)$$

The minimum for the sum of the entropies may be found to be at  $\theta = \frac{\pi}{12}$  and  $\phi = 0$ . In this case

$$|\psi\rangle = \cos \frac{\pi}{12} |\uparrow_z\rangle + \sin \frac{\pi}{12} |\downarrow_z\rangle = \cos \frac{\pi}{12} |\uparrow_{\frac{\pi}{3}}\rangle - \sin \frac{\pi}{12} |\downarrow_{\frac{\pi}{3}}\rangle \quad (5.21)$$

with

$$S[\hat{\sigma}_z, \psi] + S[\hat{\sigma}_{\frac{\pi}{3}}, \psi] \approx 0.4916 > q_{\text{MU}} = \ln \frac{4}{3} \approx 0.2877 \quad (5.22)$$

Thus we have demonstrated that the Maassen-Uffink bound is loose for these two operators. On the other hand,  $q_{\text{opt}} = 2s_{\text{bin}} \left( \frac{2-\sqrt{3}}{2} \right) \approx 0.4916$  gives the exact value for the minimum uncertainty. Letting  $\beta = \frac{\alpha}{2\alpha-1}$ , such that  $\alpha$  and  $\beta$  satisfy constraint  $\frac{1}{\alpha} + \frac{1}{\beta} = 2$ , the Rényi entropy sum as a function of  $\alpha$  is plotted in figure 5.2

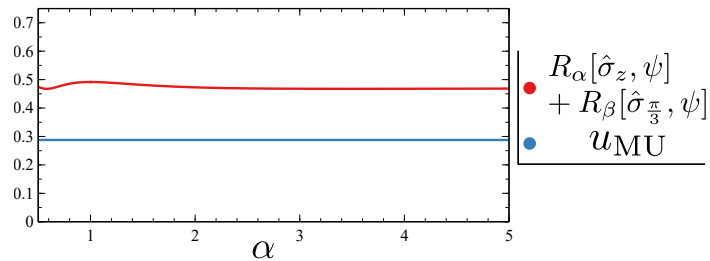


Figure 5.2: The Rényi entropy sum for the operators  $\hat{\sigma}_z$  and  $\hat{\sigma}_{\frac{\pi}{3}}$ , as a function of  $\alpha$  and the Maassen-Uffink bound.

## 5.2 Concluding Remarks

We have seen that the entropy may effectively be used as a measure of uncertainty in discrete quantum mechanical applications. The proposed lower bound by Maassen and Uffink is a lower bound, independent of the state, and is easily computable, although it is not tight. At the same time, we have seen that it is not tight, and that there is still research in tighter bounds. These state independent bounds have seen interest in different fields, most notably cryptography. Most often these applications require the use of the conditional entropies, to which the Maassen-Uffink bound may easily be generalized. For an overview of such applications, and some others, see the work by Coles *et al.* [57].

## Chapter 6

# Position and Momentum Uncertainty

In this chapter we will review uncertainty measures for position and momentum distributions that arise in quantum mechanics. By constructing a probability density from the wavefunction  $\rho(x) = |\psi(x)|^2$  we may directly apply the differential entropy or continuous forms of the Rényi and Tsallis entropies. These entropies are measures of the broadness of a distribution and have the upside of being relatively simple to implement. On the other hand, they are less easily interpreted as the uncertainty of a specific outcome happening. Instead, we may also apply coarse graining techniques, reflecting the fact that in actual experiment infinitely fine information about the distribution of a particle is impossible to obtain. Coarse graining gives us a method to directly represent the experimental resolution. Moreover, if we apply the bin averaging technique to obtain a discrete distribution, which we discussed in section 2.2, we may directly interpret the Shannon entropy as a measure of uncertainty w.r.t. the bin that a particle may be found in. Interest in the application of information theoretic measures to quantum mechanical distributions comes from the possibility that they may reveal some fundamental properties of the wavefunction, or give a comprehensive quantitative description of such properties. A tangible example of such a property are again the uncertainty relations, which have found multiple different forms in the continuous case as well. Moreover, they may be properties that are inherently tied to the spread of the wavefunction, like the electron correlation in atoms [66].

In this chapter, we will discuss forms of the uncertainty relations using entropic measures. Moreover, we will investigate the application of entropic measures for simple Hamiltonians; the infinite square well, the harmonic oscillator and finally the Hydrogen atom.

### 6.1 The Uncertainty Relations

In the continuous case, there also has been a lot of interest in formulating novel uncertainty relations using entropic measures, though, in this case that is not because any supposed shortcomings of Heisenberg's uncertainty principle. In terms of the differential entropy with  $L = 1$ , Białynicki-Birula and Mycielski proved the following uncertainty principle: Defining the Fourier transform

$$\mathcal{F}[f](k) = \frac{1}{\sqrt{2\pi\hbar}} \int f(q) e^{-\frac{iqk}{\hbar}} dq \quad (6.1)$$

we have for any function and its Fourier conjugate

$$S[|\psi(x)|^2] + S[|\mathcal{F}[\psi](p)|^2] \equiv S_x + S_p \geq \ln(\hbar\pi e). \quad (6.2)$$

Comparing it to Heisenberg's uncertainty relation for  $d = 1$ , it has been proven that this relation is stronger, in the sense that it provides a lower bound for the products of standard deviations [67, p. 282]

$$\ln(2\pi e \Delta x \Delta p) \geq S_x + S_p \geq \ln(\hbar \pi e), \quad (6.3)$$

where we have simply  $S[\rho(x, y, z)] = -\int \rho \ln \rho dx dy dz$ . Since it is known that for coherent states the standard deviations are minimal, which we show in the section investigating the entropies for the Harmonic oscillator, we that  $S_x + S_p = \ln(\hbar \pi e)$ . Thus the equation is tight. For  $d$ -dimensions the uncertainty principle for the entropies becomes [68]

$$S[\rho(x, y, z)] + S[\rho(p_x, p_y, p_z)] \geq d \ln(\hbar \pi e) \quad (6.4)$$

Interestingly, this entropy sum actually possesses scale invariance, unlike the entropies in themselves. For a family of wavefunctions  $\sqrt{\lambda} \psi(\lambda x)$ , their Fourier conjugate is given by

$$\sqrt{\lambda} \frac{1}{\sqrt{2\pi\hbar}} \int dx \psi(\lambda x) e^{\frac{ixp}{\hbar}} = \frac{1}{\sqrt{2\pi\hbar}} \int dx \frac{1}{\sqrt{\lambda}} \psi(x) e^{\frac{ix}{\hbar} \frac{p}{\lambda} x} = \frac{1}{\sqrt{\lambda}} \mathcal{F}[\psi] \left( \frac{p}{\lambda} \right). \quad (6.5)$$

Thus we see that the Fourier transformed wavefunction scales inversely to the wavefunction itself. Then, denoting the scaled position entropy  $S_{x,\lambda}$  and the momentum entropy  $S_{p,\lambda}$ , we invoke the scaling property 1.65 to see that

$$S_{x,\lambda} = S_{x,\lambda=0} + \lambda \quad (6.6)$$

$$S_{p,\lambda} = S_{p,\lambda=0} - \lambda. \quad (6.7)$$

Thus we have for the entropy sum

$$S_{x,\lambda} + S_{p,\lambda} = S_{x,\lambda=0} + S_{p,\lambda=0}, \quad (6.8)$$

due to which, the entropy sum considered an attractive quantity to study in literature. On the other hand, taking the harmonic oscillator as an example, the units ratio of the constants  $m$  and  $\omega$  to  $\hbar$  physically determine the extent of the spread of the position and momentum distributions. If a particle in the ground state is very massive, it *should* be more localized around the minimum of the potential than a less massive particle. However, the decrease in uncertainty for a position measurement is exactly counterbalanced by the increase in uncertainty for the momentum of the particle, so this doesn't show in the entropy sum. As such, the separate entropies and the entropy sum may be useful for different applications.

For mixed states we immediately may see that this bound is actually not tight. The density matrix for continuous states  $\psi(x)$  may be constructed from  $\hat{\rho}(x, y) = \sum_i p_i \psi_i(x) \psi_i^\dagger(y)$ . In position representation the post measurement state is given by  $\int \hat{\rho}(x, y) \delta(x - y) dy = \hat{\rho}(x, x) = \sum_i p_i \rho_i(x)$ . Thus, the positional entropy only depends on  $\sum_i p_i \rho_i(x)$ , and similarly only on  $\sum_i p_i \rho_i(p)$  for the momentum representation. Hence, the entropy of a mixed state depends on the convex sum of the probability densities, whereas for a pure state it could only depend on the square of the sum of amplitudes. As a consequence we see that its entropy must be larger than for a pure state as a direct consequence of concavity of the differential entropy 1.60, since a mixed state is always a convex sum of probability densities. Therefore it cannot saturate the uncertainty relation 6.2. To improve this relation, Frank and Lieb [69] proposed the following relation for mixed states, setting  $\hbar = 1$ ,

$$S_x + S_p \geq \ln 2\pi - \sum_i \lambda_i \ln \lambda_i \quad (6.9)$$

where  $\lambda_i$  are the eigenvalues of  $\hat{\rho}(x, y)$ , such that  $\int \hat{\rho}(x, y) f_i(y) dy = \lambda_i f_i(x)$ . This entropy term on the right hand side is actually the Von Neumann entropy again. The Von Neumann entropy may be seen as a measure of mixedness which we discuss in chapter 7, thus this relation says that the minimum uncertainty grows as the state becomes more mixed.

The uncertainty relation 6.2 has been generalized to the Rényi entropy as well, where it takes the form, for  $d = 1$  [70]

$$\frac{1}{1-\alpha} \ln \left( \int dx (\hat{\rho}(x))^\alpha \right) + \frac{1}{1-\alpha} \ln \left( \int dx (\hat{\rho}(x))^\alpha \right) \quad (6.10)$$

$$\equiv R_\alpha^x + R_\alpha^p \geq \frac{1}{2(1-\alpha)} \ln \left( \frac{\pi}{\alpha} \right) + \frac{1}{2(1-\beta)} \ln \left( \frac{\pi}{\beta} \right) + \ln \hbar \quad (6.11)$$

Next we will investigate uncertainty relations that may be defined on the coarse grained entropies. We define the bin sizes for the position and momentum coordinates as  $\delta x$  and  $\delta p$  respectively. These bins may be thought of as representing the experimental accuracy of a measurement device. The probabilities to measure a particle in bin  $l$  for position and  $k$  for momentum are given by

$$q_l = \int_{l\delta x}^{(l+1)\delta x} \rho(x) dx, \quad p_k = \int_{k\delta p}^{(k+1)\delta p} \rho(p) dp. \quad (6.12)$$

It is shown by Białynicki-Birula [70] that the sum of position and momentum entropies is bounded from below by

$$S_x + S_p \geq \ln \left( \frac{e\pi\hbar}{\delta x \delta p} \right) \equiv u_{CG}. \quad (6.13)$$

Qualitatively, this uncertainty relations shows that our uncertainty grows as we try to pinpoint the position and momentum of a particle into smaller regions of phase space. Interestingly, the relation becomes negative if  $\delta x \delta p > e\pi\hbar$ , so this immediately shows that the relation is not sharp. This is mentioned by Białinicky-Birula himself, though he also mentions that bound becomes sharp as  $\delta x, \delta p \rightarrow 0$ . To demonstrate how this bound behaves, we have plotted it in figure 6.1 with  $\delta x = \delta p \equiv \delta$ , for the first three harmonic oscillator energy eigenstates, for which the formula is shown in equation 6.37. In this figure it can be seen that for the ground state, which is a Gaussian, the bound does indeed seem to become tight as  $\delta \ll 1$ . This relation turns out to be part of a family of relations for the Rényi entropies as well. Using the definition of the Rényi entropy from 3.4, with  $\frac{1}{\alpha}, \frac{1}{\beta} = 2$ , we have [70, §5],

$$R_\alpha[\rho(x)] + R_\beta[\rho(p)] \geq -\frac{1}{2(1-\alpha)} \ln(\alpha) - \frac{1}{2(1-\beta)} \ln(\beta) + \ln \left( \frac{\pi\hbar}{\delta x \delta p} \right). \quad (6.14)$$

It is easily checked that we recover relation 6.13 in the limit where  $\alpha, \beta \rightarrow 1$ . These relations also hold for mixed states, which has been proven by Białinicky-Birula [70].

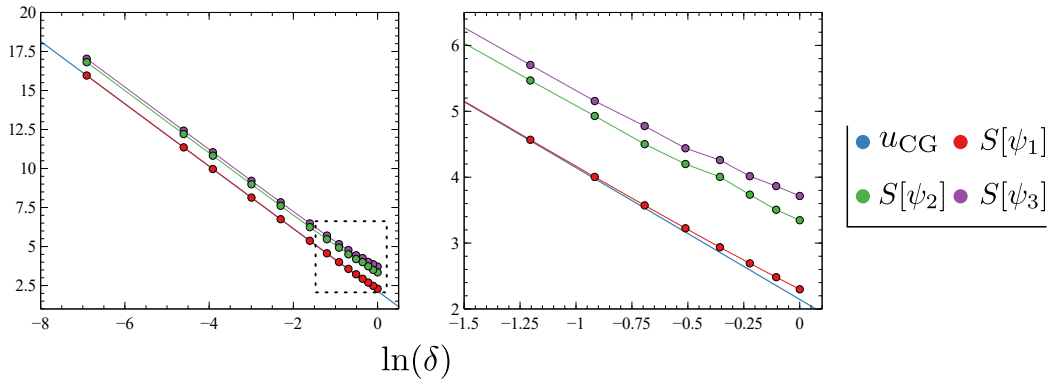


Figure 6.1: The coarse grained entropy for the first three energy eigenstates for the harmonic oscillator. The  $\delta$  axis is plotted in logarithmic scale. In the legend we have used the abbreviation  $S[\psi]$  to mean the position and momentum entropy sum.

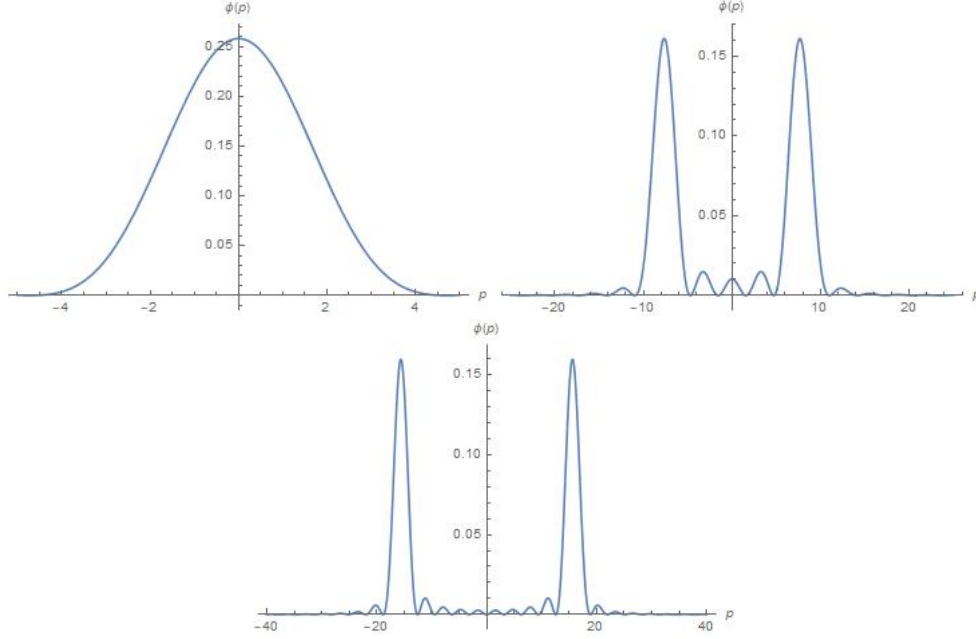


Figure 6.2: Momentum distributions for the infinite square well with  $a = 1$  and  $n = 1, n = 5$  and  $n = 10$  respectively.

## 6.2 Infinite Square Well

In the following sections, we will investigate the differential entropies of the eigenstates of simple quantum systems, and specifically their dependence on the quantum number  $n$ . We will investigate their entropies in the position representation as well as the momentum representation.

We start with infinite square well, with it's potential defined as

$$V(x) = \begin{cases} 0 & |x| \leq a, \\ \infty & \text{everywhere else,} \end{cases} \quad (6.15)$$

meaning that the width of the well is equal to  $2a$ . The energy eigenstates  $\psi_n$  represented in position and momentum space respectively are given by

$$\psi_n(x) = \sqrt{\frac{1}{a}} \sin\left(\frac{\pi n}{2a}(x - a)\right), \quad -a \leq x \leq a \quad (6.16)$$

and

$$\mathcal{F}[\psi_n](p) = \sqrt{\frac{\pi n^2 \hbar^3}{2a^3}} \frac{\sin\left(\frac{a}{\hbar}p - \frac{\pi n}{2}\right)}{p^2 - \frac{\pi^2 n^2 \hbar^2}{4a^2}} e^{i\alpha_n}, \quad (6.17)$$

where  $\alpha_n$  is some phase dependent on  $n$ . The standard deviations for  $x$  and  $p$  are

$$\Delta x(a, n) = \frac{a}{\sqrt{3}} \sqrt{1 - \frac{6}{\pi^2 n^2}} \quad (6.18)$$

$$\Delta p(a, n) = \frac{\hbar \pi n}{2a} \quad (6.19)$$

Meaning that their product is

$$\Delta x \Delta p = \frac{\hbar}{2\sqrt{3}} \sqrt{\pi^2 n^2 - 6}, \quad (6.20)$$

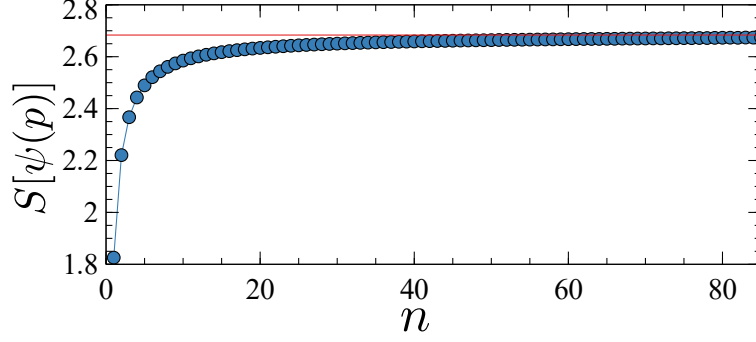


Figure 6.3: The momentum entropy of the infinite square well with  $a = 1$  as a function of  $n$ , and its asymptotic value.

which goes as  $n$  when  $n \rightarrow \infty$ . The differential entropy in the position representation stays constant over  $n$

$$S_x(n, a) = \ln(4a) - 1. \quad (6.21)$$

The origin of this is the periodicity of the wavefunction, as well as that  $n$  in this problem effectively only scales the wavefunction. Because  $n$  may be seen as a scaling factor, we have that

$$\int_{-a}^a S[\rho(nx)]dx = \frac{1}{n} \int_{-n \cdot a}^{n \cdot a} S[\rho(x)]dx = n \cdot \frac{1}{n} \int_{-a}^a S[\rho(x)]dx, \quad (6.22)$$

which shows why the entropy stays constant. The entropy of the momentum representations is harder to compute and, to our knowledge, no analytic expression has been found that links the Shannon entropy and the quantum number  $n$ . Numerical results for the momentum entropy may be seen in figure 6.3. The asymptotic value, for  $n \rightarrow \infty$ , of the momentum entropy is known, however. Majernik *et al.* found that the asymptotic value is given by [71],

$$\lim_{n \rightarrow \infty} S_p(a, n) = \ln(8\pi) - \ln\left(4\frac{a}{h}\right) + 2(1 - \gamma), \quad (6.23)$$

where  $\gamma$  is the Euler's constant. Thus as  $n \rightarrow \infty$  the sum of the position and momentum entropies is independent of  $a$  and goes to

$$\lim_{n \rightarrow \infty} [S_x + S_p] = \ln(8\pi) + 2(1 - \gamma) - 1 \approx 3.0697 \quad (6.24)$$

The fact that the entropy saturates is also not too surprising. The momentum distribution is a bi-modal distribution, and the  $n$  acts only as a shift parameter for the peaks. If  $|x - a|$  is sufficiently large in between the peaks (which is only possible for large  $n$ ) then the distribution becomes two distinct peaks with vanishing overlap. Like we discussed in section 1.3.1, the differential entropy becomes asymptotically invariant of the shift induced by changing  $n$ .

### The Rényi and Tsallis Entropies

In contrast to the case of the BGS entropy, where we have not been able to find an analytic expression for the momentum representation as a function of  $n$ , Aptekarev *et al.* [72], have found that the Rényi and Tsallis entropies, for both representations, can be written as a function of the quantum number  $n$ . In the position representation the Rényi and Tsallis entropies are actually independent of  $n$  as well, and are given by

$$R_\alpha^x = \ln(a) + \frac{1}{1 - \alpha} \ln\left(\frac{2\Gamma(\alpha + \frac{1}{2})}{\sqrt{\pi}\Gamma(\alpha + 1)}\right) \quad (6.25)$$

and

$$T_q^x = \frac{1}{q - 1} \left[ 1 - \frac{2}{a^{q-1}\sqrt{\pi}} \frac{\Gamma(q + \frac{1}{2})}{\Gamma(q + 1)} \right] \quad (6.26)$$

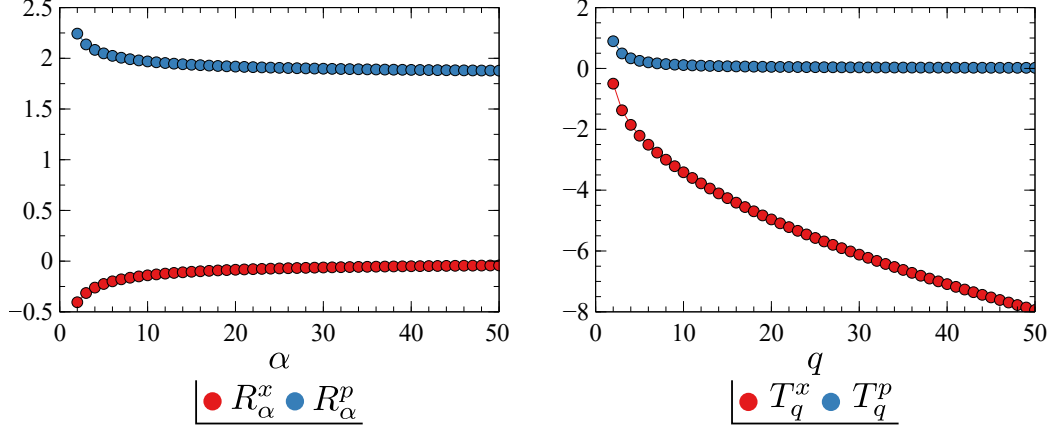


Figure 6.4: The large  $n$  limits of the Rényi and Tsallis entropies of the position and momentum representations of the infinite square well with  $a = 1$ , as a function of  $\alpha$  and  $q$ .

respectively. For the momentum entropies the authors have found an expression for the entropic moment  $W_\alpha[\rho_n]$ ,

$$W_\alpha[\rho_n(p)] = \left(\frac{\pi n^2}{2}\right)^\alpha a^{\alpha-1} I_{n,\alpha} \quad (6.27)$$

where

$$I_{n,\alpha} = (\pi n)^{-2\alpha} \sum_{j=0}^{\alpha-1} \binom{2j+2\alpha-1}{2\alpha-1} \frac{2\pi(\pi n)^{-2j}(-\frac{1}{4})^j}{(2\alpha-2j-1)!} \sum_{i=0}^{\alpha-1} (-1)^i \binom{2\alpha}{i} (\alpha-i)^{2\alpha-2j-1}. \quad (6.28)$$

The entropies are then found to be

$$R_\alpha^p = \frac{1}{1-\alpha} \ln(W_\alpha[\rho_n(p)]) = -\ln(a) + \frac{1}{1-\alpha} \ln\left(\left(\frac{\pi n^2}{2}\right)^\alpha I_{n,\alpha}\right) \quad (6.29)$$

and

$$T_q^p = \frac{1}{q-1} (1 - W_q[\rho_n(p)]) = \frac{1}{q-1} \left[1 - \left(\frac{\pi n^2}{2}\right)^q a^{q-1} I_{n,q}\right] \quad (6.30)$$

To obtain the BGS entropy in the limit  $\alpha, q \rightarrow 1$  we would need to perform a derivative w.r.t.  $\alpha$  or  $q$ , by L'Hôpital's rule. However, as we can see, these expressions are only defined for integer  $\alpha$  and  $q$ , and it is not obvious how we could analytically continue them. As such it is not possible to derive an analytical expression for the BGS entropy, directly from these expressions. Lastly, we will show the limiting behavior as  $n \rightarrow \infty$  for both entropies. For  $\alpha, q > 1$ , we have as  $n \rightarrow \infty$

$$W_\alpha[\rho_n(p)] \simeq \frac{b_\alpha}{2^\alpha \pi^{\alpha-1}} a^{\alpha-1}, \quad (6.31)$$

with

$$b_\alpha = 4\alpha \sum_{i=0}^{\alpha-1} \frac{(-1)^i (\alpha-i)^{2\alpha-1}}{i!(2\alpha-i)!} \quad (6.32)$$

This can be plugged into 6.29 and 6.30 to obtain the final result, which are plotted as a function of  $\alpha$  and  $q$  in figure 6.4. Do note that using the above expressions to construct the Rényi entropy sum as discussed in the previous section is not possible using these analytical expressions. This is because we needed to satisfy the constraint  $\frac{1}{\alpha} + \frac{1}{\beta} = 2$ . Thus if  $\alpha$  is integer, then  $\beta$  is necessarily fractional. However, the above results only work for integer parameters. As such, we have plotted the position and momentum entropies separately but not their sum.



## 6.2.1 Comparison with the Classical Infinite Square Well

To make a comparison to a classical system, we need to establish how to obtain a probabilistic description of a classical system. Like in quantum mechanics, we will only assume to know the energy of the particle, but not its actual position, or its position at any time before. To assign a likelihood to find a particle in a certain region, we coarse grain the space in the well into bins, and find the time the particle spends in each bin on average. We then assign the probability of finding the particle in a bin to the ratio of this time to the total average time spent in any bin is. The time spent in a bin  $\delta t$ , in the limit where the bin size  $\delta x \ll 1$ , is approximately the bin size divided by the average velocity in the bin. Thus infinitesimally we have that the probability is proportional to  $dt$ , with

$$dt = \frac{dx}{\dot{x}}. \quad (6.33)$$

The probability density is then

$$\rho(x) = \frac{dt(x)}{\int dt(x)}. \quad (6.34)$$

We will apply this result to the harmonic oscillator, but in the case of the infinite square well the solution is very simple. Since the velocity is always the same at any point in the well, the distribution is simply the equidistribution, which holds for any energy. In this sense, the entropy actually behaves similarly to the classical case, in the sense that it is not dependent on the energy. As for the momentum distribution, we may obtain the distribution in the same way

$$dt = \frac{d\dot{x}}{\ddot{x}} = \frac{dp}{m\ddot{x}}. \quad (6.35)$$

However, the entropy diverges in the classical case. This is because, in the limit where particle the instantaneously changes the direction of its momentum when it collides with the wall, the momentum distribution becomes two infinitely localized peaks at  $p = \pm\sqrt{2mE}$ .

## 6.3 Harmonic Oscillator

In this section we will calculate and discuss aspects of the differential entropy applied to the position and momentum distributions of the quantum harmonic oscillator

$$V(x) = \frac{1}{2}m\omega^2x^2. \quad (6.36)$$

The eigenstates of the Hamiltonian in the position representation, are

$$\psi_n(x) = \frac{1}{\sqrt{\pi^{\frac{1}{2}}2^n n! \sigma_x}} H_n\left(\frac{x}{\sigma_x}\right) e^{-\frac{x^2}{2\sigma_x^2}}. \quad (6.37)$$

The Functions  $H_n$  are the Hermite polynomials and  $\sigma_x^2 = \frac{\hbar}{m\omega}$ . In the momentum representation the states take a similar form

$$\mathcal{F}[\psi_n](p) = \frac{1}{\sqrt{\pi^{\frac{1}{2}}2^n n! \sigma_p}} H_n\left(\frac{p}{\sigma_p}\right) e^{-\frac{p^2}{2\sigma_p^2}}, \quad (6.38)$$

with  $\sigma_p^2 = \omega\hbar m$ . The wavefunctions therefore have the same shape in both representations, but are just scaled compared to each other. The standard deviations for the n-th eigenstate for position and momentum are given by

$$\Delta x = \sqrt{\frac{\hbar}{2m\omega}}(2n+1), \quad (6.39)$$

$$\Delta p = \sqrt{\frac{\hbar m\omega}{2}}(2n+1), \quad (6.40)$$

and are easily obtained from the ladder operator representations of  $x$  and  $p$ , which we will discuss later on. The product is thus

$$\Delta x \Delta p = \frac{\hbar}{2} (2n + 1)^2 \quad (6.41)$$

For large  $n$  this diverges as  $n^2$ .

Since the momentum wavefunction is actually the same function, but with  $x/\sigma_x \rightarrow p/\sigma_p$ , we can easily see from the scaling property 1.65 that

$$S[\mathcal{F}[\psi_n](p)] = S[\psi_n(x)] + \ln\left(\frac{\sigma_p}{\sigma_x}\right). \quad (6.42)$$

Thus we only need to specify the entropies in the position representation, which can be found analytically, relatively easily for the ground state and first excited state

$$S[\psi_0(x)] = \frac{1}{2}(1 + \ln(\pi)) + \ln(\sigma_x) = \ln(\sqrt{e\pi}\sigma_x) \quad (6.43)$$

$$S[\psi_1(x)] = \ln(2\sqrt{\pi}\sigma_x) + \gamma - \frac{1}{2}, \quad (6.44)$$

where  $\gamma$  is again Euler's constant. It can be checked that the sum of the position and momentum entropies saturates the inequality 6.2. The large  $n$  behavior of the energy eigenstates was found by van Assche and Dehesa *et al.* [73] [74]. Unlike the square well, both the position and momentum entropies diverge. Their asymptotical behavior, with all physical constants set to 1, is given by

$$S_x = S_p \approx \ln\left(\frac{\sqrt{2n\pi}}{e}\right). \quad (6.45)$$

The entropy sum thus is given by

$$S_x + S_p \approx \ln\left(\frac{2n\pi^2}{e^2}\right) \quad (6.46)$$

### 6.3.1 Comparison with the classical harmonic oscillator

Using the procedure we discussed in the previous section, we will compare the distributions and the entropy of the classical harmonic oscillator (CHO) and the quantum harmonic oscillator (QHO). To get a good comparison we will fill the quantum energies into the classical equations of motion. We have

$$\frac{1}{2}m\dot{x}^2 + kx^2 = E \quad (6.47)$$

Inserting  $E = \hbar\omega(n + \frac{1}{2})$  and using  $k = m\omega^2$ , we get

$$\dot{x}^2 + \omega^2 x^2 = \frac{\hbar\omega}{m}(2n + 1) \quad (6.48)$$

Setting  $\dot{x}$  to zero, we find the classical amplitude

$$A_n = \sqrt{\frac{\hbar}{m\omega}(2n + 1)} \quad (6.49)$$

Filling this in, into the equation of motion

$$dt = \pm \frac{1}{\omega\sqrt{A_n^2 - x^2}} dx. \quad (6.50)$$

We are only interested in magnitude of the velocity, so we may take the absolute value. Then finally we may normalize over the region  $-A_n$  to  $A_n$ , to find the probability density

$$\rho_{cl}(x) = \frac{1}{\pi\sqrt{A_n^2 - x^2}} \quad (6.51)$$

Calculating the entropy we may find [75]

$$S_{x,cl} = \frac{1}{2} \ln \left( \frac{\pi^2 A_n^2}{4m\omega} \right) = \frac{1}{2} \ln \left( \frac{\pi^2 \hbar (n + \frac{1}{2})}{2m\omega} \right). \quad (6.52)$$

Since we have that  $x = \pm \frac{1}{\omega} \sqrt{\omega A_n^2 - \dot{x}^2}$  and  $\dot{x} = -\omega^2 x$  we find for the momentum distribution

$$\rho_{cl}(p) = \frac{1}{\pi \sqrt{m^2 \omega^2 A_n^2 - p^2}} \quad (6.53)$$

and thus the entropy is

$$S_{p,cl} = \frac{1}{2} \ln \left( \frac{\pi^2 m \omega A_n^2}{4} \right). \quad (6.54)$$

As such, the classical momentum entropy behaves, as should be expected, the same as the position entropy, except for the dependence on the physical constants. In figure 6.5 we have plotted the behavior of the entropy of the CHO and the QHO. We see that for low energies the uncertainty of the classical system is lower, while for high energies,  $n \geq 15$ , the quantum system is less uncertain. On the other hand, the classical standard deviation is given by  $\Delta x_{cl} = n + \frac{1}{2}$ , compared to the  $\Delta x = \frac{1}{\sqrt{2}}(2n + 1)$  for the quantum distribution. Thus, the standard deviation estimates the classical uncertainty as always being lower than the quantum mechanical case. The reason for this may be that the classical distribution is mostly constant in the middle, while the quantum mechanical distribution fluctuates. Thus in the middle, for the quantum mechanical distribution, there are regions where we are relatively certain we can find the particle, and regions where we are relatively certain we won't find it.

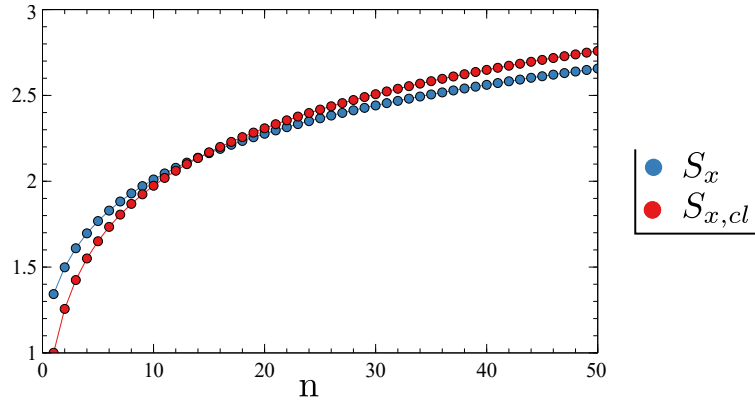


Figure 6.5: The entropy of the classical and quantum harmonic oscillator with  $m = \omega = \hbar = 1$

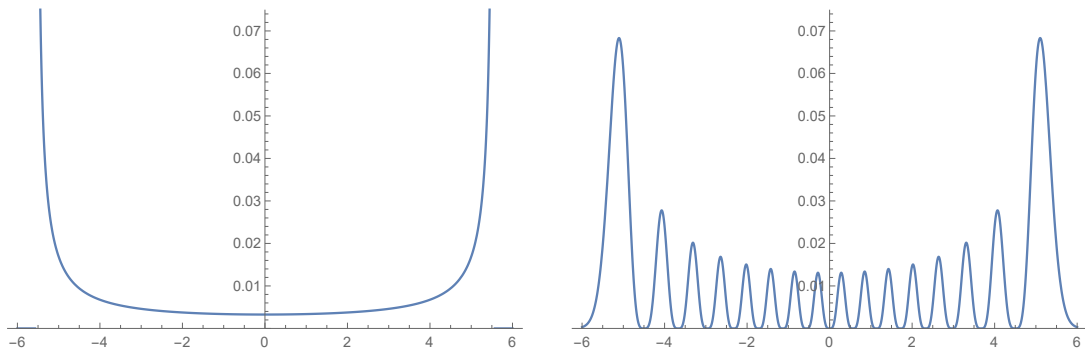


Figure 6.6: The classical (left) and quantum (right) position distributions of the harmonic oscillator for  $n = 15$ .

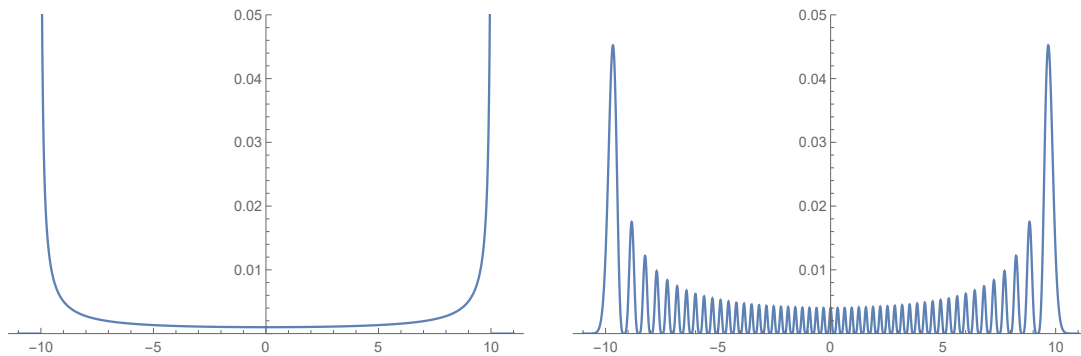


Figure 6.7: The classical (left) and quantum (right) position distributions of the harmonic oscillator for  $n = 50$ .

### 6.3.2 Coherent States and Squeeze States

In this section we will discuss coherent and squeeze states, which in the context of the harmonic oscillator are sometimes called optical states. Coherent states are minimal uncertainty states. Practically speaking, their distributions are of the Gaussian form but displaced with respect to the origin of the harmonic oscillator potential. Though the distributions of these states have the general shape of the harmonic oscillator, they are not energy eigenstates. These states are very special in the sense they behave very similar to states of the CHO. They are also very important in quantum optics, since they are the closest analog to classical waves of pure frequency. Squeezed states are generalizations of coherent states. They behave similarly to coherent states but are typically characterized by having scaled standard deviations, though they still minimize Heisenberg's uncertainty principle. For example the position standard deviation may be  $1/2$  of the standard deviation of the ground state of the harmonic oscillator at some point in time. This means, due to the uncertainty relation, that the standard deviation of the momentum distribution must be scaled by at least 2. As such, squeezed states may, but do not have to, minimize the uncertainty principle. In this chapter we will give an overview of coherent and squeeze states. We feel it is important to give a little bit of background information on these quantum optical states, because that will also explain why the entropies behave the way they do. Moreover, we will return to quantum optical states in chapter 9. So, in this section we will briefly discuss their properties, interpretation in phase space, and, of course, their position and momentum standard deviations and entropies.

Coherent states of the harmonic oscillator are minimal uncertainty states (e.g. they minimize the Heisenberg's uncertainty principle), meaning that their distributions have the same general shape as the ground state, but are displaced in position, momentum or both. Their time evolution is non-trivial, but at any time the state remains a coherent state. Hence, all the time evolution of the state is contained in the displacements. A general coherent state is actually an eigenstate of the annihilation operator  $a = \frac{1}{\sqrt{2}}(x + ip)$ . These eigenstates have as eigenvalue  $\alpha$  and may easily be checked to be given by

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{(\alpha a^\dagger)^n}{n!} |0\rangle, \quad (6.55)$$

where  $|0\rangle$  is the ground state of the harmonic oscillator. It can be shown that the wavefunction is given by [76, Ch. 2]

$$\psi_c(x) = \langle x|\alpha\rangle = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} \exp\left(-\frac{m\omega}{2\hbar} \left(x - \sqrt{\frac{2\hbar}{m\omega}} \operatorname{Re}[\alpha]\right)^2 + i\sqrt{\frac{2m\omega}{\hbar}} \operatorname{Im}[\alpha]x\right) \quad (6.56)$$

The position and momentum expectation values are

$$\langle x \rangle_\alpha = \sqrt{\frac{2\hbar}{m\omega}} \operatorname{Re}[\alpha], \text{ and } \langle p \rangle_\alpha = \sqrt{2m\omega\hbar} \operatorname{Im}[\alpha]. \quad (6.57)$$

In other words, the distributions are the shifted Gaussians

$$\rho(x) = \sqrt{\frac{m\omega}{\pi\hbar}} e^{-\frac{m\omega}{\hbar} \left(x - \sqrt{\frac{2\hbar}{m\omega}} \operatorname{Re}[\alpha]\right)^2}, \text{ and } \rho(p) = \sqrt{\frac{m\omega\hbar}{\pi}} e^{-\frac{1}{m\omega\hbar} \left(p - \frac{1}{\sqrt{2m\omega\hbar}} \operatorname{Im}[\alpha]\right)^2} \quad (6.58)$$

Lastly, the time evolved coherent state is given by

$$|\alpha(t)\rangle = e^{-\frac{i\omega t}{2}} |e^{-i\omega t} \alpha\rangle. \quad (6.59)$$

In other words, we could plot the means  $\langle x \rangle_\alpha(t)$  and  $\langle p \rangle_\alpha(t)$  and would see that they would rotate on a circle over time, just like the position of momentum of the CHO in phase space.

We can take this analogy a bit further by using a phase space representation of the coherent state, using the Wigner distribution function (WDF),  $W(x, p)$ . We will go more in depth into the subject of phase space representations and their entropies in chapter 9. For now we only need to know that WDF's are real functions and that we can obtain both the position and momentum distribution of the state  $\psi$  by integrating over  $p$  and  $x$  respectively. Thus the distributions may be seen as a sort of projection onto the  $x$  and  $p$  axes. The WDF itself should not be considered a probability distribution itself. In this picture, setting the physical constants to 1, coherent states become 2D Gaussian

$$W(x, p) = \frac{1}{\pi} e^{-\frac{1}{2s_x^2}(p - A \sin t)^2 - \frac{1}{2s_p^2}(x - A \cos t)^2}, \quad (6.60)$$

which rotates around the origin over time. It is symmetric around its mean, so we see that the distributions only shift over time. This representation also gives the displacement operator a very clear interpretation, generating a displacement in the  $x, p$  plane of the mean of the ground state.

Since the distributions of the position and momentum operators are simply shifted ground state distributions at any time, the entropy is constant with respect to time and equal to that of the ground state of the harmonic oscillator 6.43. Thus any coherent state minimizes the entropic uncertainty relation 6.2.

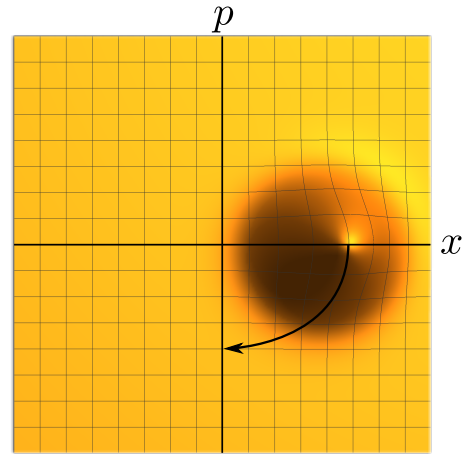
However, the structure of the Hilbert space of the QHO is generally much more easy to describe in terms of the ladder operators. In fact, in most literature the structure of coherent states and the squeeze states is described in terms of operators working on the vacuum state. The action of the sum of operators on the ground state can also be written as a single operator, using the definition of the exponential, the Baker-Campbell-Hausdorff formula and the fact that the lowering operator annihilates the ground state, so that we get

$$|\alpha\rangle = D(\alpha) |0\rangle \equiv e^{\alpha a^\dagger - \alpha^* a} |0\rangle. \quad (6.61)$$

Here,  $D(\alpha)$ , is an Hermitian operator called the displacement operator. In phase space its interpretation is that it displaces a whole state in phase space. Using that we can write the position and momentum operators in terms of the ladder operators

$$x = \sqrt{\frac{\hbar}{2m\omega}} (a + a^\dagger), \quad p = i\sqrt{\frac{\hbar m\omega}{2}} (a - a^\dagger), \quad (6.62)$$

Figure 6.8: The WDF of a coherent state.



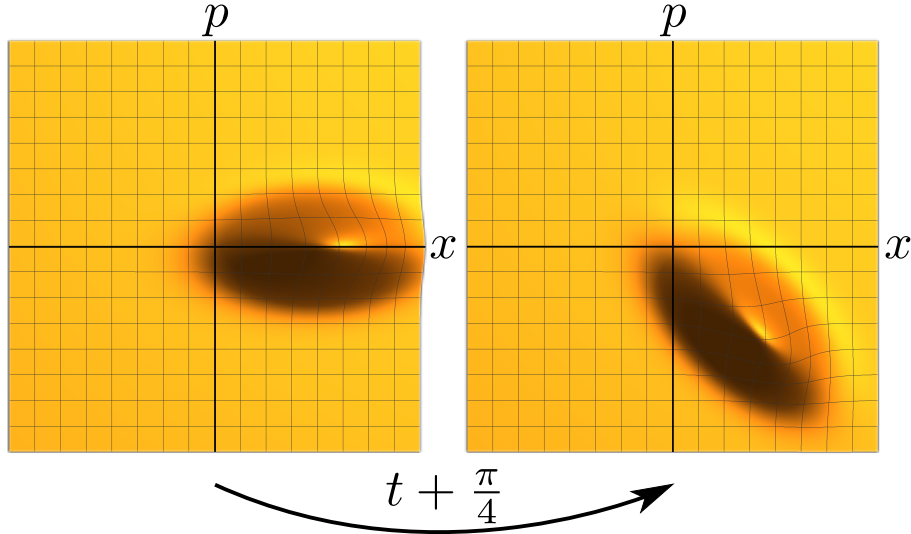


Figure 6.9: The Displaced Squeeze State at a time  $t$  and  $t + \pi/4$ .

and that<sup>1</sup>

$$D^\dagger(\alpha)aD(\alpha) = a + \alpha, \quad D^\dagger(\alpha)a^\dagger D(\alpha) = a^\dagger + \alpha^*. \quad (6.63)$$

From the above we can easily calculate the standard deviations, which are the same as the ground state

$$\Delta x = \sqrt{\frac{\hbar}{2m\omega}}, \quad \text{and} \quad \Delta p = \sqrt{\frac{\hbar m\omega}{2}} \quad (6.64)$$

$$\Delta x \Delta p = \frac{\hbar}{2}. \quad (6.65)$$

For the entropy we must use the explicit distributions, in which case we find

$$S_x = \ln(\sqrt{e\pi 2\Delta x}), \quad \text{and} \quad S_p = \ln(\sqrt{e\pi 2\Delta p}), \quad (6.66)$$

which makes the sum

$$S_x + S_p = \ln(e\pi 2\Delta x \Delta p) = \ln(e\pi\hbar). \quad (6.67)$$

Thus the entropy is minimal, which was expected from relation 6.3

Squeezed coherent states<sup>2</sup> are a generalization of coherent states and in general are eigenstates of the operator  $b = \mu a + \nu a^\dagger$  with eigenvalue  $\beta$  and  $|\mu|^2 - |\nu|^2 = 1$ . Due to this constraint, this transformation preserves the commutation relation  $[b, b^\dagger] = 1$  [77]. When  $\mu$  and  $\nu$  are real, e.g. we can write  $\mu = \cosh \lambda$  and  $\nu = \sinh \lambda$ , then the eigenstate is a minimal uncertainty state [78]. In terms of the Wigner distribution, they are still given by a double Gaussian, but are squeezed in one axis, and elongated in the perpendicular axis, which can be seen in figure 6.9. Most generally their Wigner distribution may be parameterized, setting all constants to unity, as

$$W(x, p) = \frac{1}{\pi} e^{-\frac{1}{s^2} - (q_+(x,p) - a)^2(x,p) - s^2 (q_-(x,p) - b)^2(x,p)} \quad (6.68)$$

<sup>1</sup>These relations may be derived from the following equation

$$e^{\hat{A}} \hat{B} e^{-\hat{A}} = \hat{B} + [\hat{A}, \hat{B}] + \frac{1}{2!} [\hat{A}, [\hat{A}, \hat{B}]] + \dots$$

<sup>2</sup>The squeezed states we mention in this section are also sometimes called two photon coherent states. This has to do with the optical process which is used in the lab to create squeeze states. This process generally may be thought of as converting one photon with higher frequency into two photons with half the frequency. However, adding a photon to a state in quantum optics is represented by the creation operator  $a^\dagger$ . Since these squeezed states are still described by an infinite sum of energy eigenstates, one should not think of these states as actually consisting of two photons.

where  $q_+(x, p) = x \cos \theta + y \sin \theta$ ,  $q_- = y \cos \theta - x \sin \theta$  and  $a$  and  $b$  are general parameters that parametrize the shift in the  $x, p$  plane. If we let the state evolve, then we may see that actually all time evolution is governed by the parameter  $\theta$ . As such, the means of squeeze states still follow the classical trajectory, but now the squeezing axis also rotates. This behavior is also what makes these states very interesting in experiment, since the means behave like classical harmonic oscillator states, but they may have much less uncertainty in amplitude at specific phases.

We may generate squeeze states with the help of the squeeze operator

$$S(\zeta) = e^{\frac{1}{2}(\zeta^* a^2 - \zeta (a^\dagger)^2)}. \quad (6.69)$$

Assuming that  $\zeta = r e^{2i\theta}$  we have that

$$S^\dagger(\zeta) a S(\zeta) = a \cosh(r) - a^\dagger e^{i2\theta} \sinh(r), \quad (6.70)$$

and

$$S^\dagger(\zeta) a^\dagger S(\zeta) = a^\dagger \cosh(r) - a e^{-i2\theta} \sinh(r). \quad (6.71)$$

If we let  $b = S^\dagger(\zeta) a S(\zeta)$ , then the eigenstate of  $b$  is then generated from the vacuum as follows<sup>3</sup>

$$|\beta, \zeta\rangle = S(\zeta) D(\beta) |0\rangle. \quad (6.72)$$

In the literature, squeezed states may also be constructed by first acting on the vacuum with the squeeze operator, followed by the displacement operator, which is the method we will use for the rest of this section. These squeezed states are related to the above ones by the following relation

$$|\beta, \zeta\rangle = D(\beta \cosh r - \beta^* e^{2i\theta} \sinh r) S(\zeta) |0\rangle \quad (6.73)$$

Using the above identities, we may find the expectation of powers of the position and momentum operators directly using the ladder operators. We find,

$$\langle 0 | S^\dagger(\zeta) D^\dagger(\alpha) x D(\alpha) S(\zeta) | 0 \rangle = \sqrt{\frac{\hbar}{2m\omega}} (2\text{Re}[\alpha]) \quad (6.74)$$

and

$$\begin{aligned} \langle 0 | S^\dagger(\zeta) D^\dagger(\alpha) x^2 D(\alpha) S(\zeta) | 0 \rangle = \\ \frac{\hbar}{2m\omega} (4\text{Re}[\alpha] + \cosh^2 r + \sinh^2 r - 2 \cosh r \sinh r \cos 2\theta). \end{aligned} \quad (6.75)$$

Moreover, for the momentum expectations we have

$$\langle 0 | S^\dagger(\zeta) D^\dagger(\alpha) p D(\alpha) S(\zeta) | 0 \rangle = \sqrt{\frac{\hbar m \omega}{2}} (2\text{Im}[\alpha]) \quad (6.76)$$

and

$$\begin{aligned} \langle 0 | S^\dagger(\zeta) D^\dagger(\alpha) p D(\alpha) S(\zeta) | 0 \rangle = \\ \frac{\hbar m \omega}{2} (4\text{Im}[\alpha] + \cosh^2 r + \sinh^2 r + 2 \cosh r \sinh r \cos 2\theta) \end{aligned} \quad (6.77)$$

Focussing for now on the minimum uncertainty states, which have  $\theta = 0$  we have for the standard deviations

$$\Delta x = \sqrt{\frac{\hbar}{2m\omega}} e^{-r}, \text{ and } \Delta p = \sqrt{\frac{\hbar m \omega}{2}} e^r \quad (6.78)$$

---

<sup>3</sup>This is possible if the transformation  $b = \mu a + \nu a^\dagger$  preserves the commutation relations. In that case the transformation is called canonical and may always be written as a unitary transformation.

Thus we have trivially

$$\Delta x \Delta p = \frac{\hbar}{2} \quad (6.79)$$

As such, the states still minimize the uncertainty relation, though they can have a smaller uncertainty in one of the components. The reason for this is that the phase of  $\zeta$  controls the rotation of the coherent state in the  $x, p$  plane. If we take  $\theta \neq 0$  then we have for the standard deviations

$$\Delta x = \sqrt{\frac{\hbar}{2m\omega}} \sqrt{\cosh(2r) + \sinh(2r) \cos(2\theta)} \quad (6.80)$$

and

$$\Delta p = \sqrt{\frac{\hbar}{2m\omega}} \sqrt{\cosh(2r) - \sinh(2r) \cos(2\theta)}. \quad (6.81)$$

Thus for the product we have

$$\Delta x \Delta p = \frac{\hbar}{2} \sqrt{\cosh^2(2r) - \sinh^2(2r) \cos^2(2\theta)}. \quad (6.82)$$

The time evolution is given from the following relation

$$|\beta(t), \zeta(t)\rangle = e^{i\phi(t)} |\alpha e^{-i\omega t}, \zeta e^{-i2\omega t}\rangle \quad (6.83)$$

where  $\phi(t)$  is some function of time. Hence, the means will rotate on a circle with radius  $|\alpha|$ , and the squeezing axis will rotate as well. This means that if we start with a squeezed state with  $\theta = 0$  initially, we can take  $\theta \rightarrow \omega t$  in the above relations to obtain the time dependent uncertainties.

For calculation of the entropy, we need explicit expressions for the wavefunctions. In the case of the minimal uncertainty squeeze states the most general wavefunction is

$$\psi_{sq}(x) = C \exp\left(-\frac{m\omega}{\hbar} \frac{(x - x_0)^2}{2s^2} + \frac{i}{\hbar} p_0 x\right). \quad (6.84)$$

This simply has as standard deviations

$$\Delta x = s \sqrt{\frac{\hbar}{2m\omega}}, \text{ and } \Delta p = \frac{1}{s} \sqrt{\frac{\hbar m\omega}{2}}. \quad (6.85)$$

where  $s$  is the squeezing parameter, and  $x_0$  and  $p_0$  give the expectation values of position and momentum. We see that this is simply the square root of a Gaussian, with an position dependent phase. Because of this, the Fourier transform has exactly the same form

$$\tilde{\psi}_{sq}(p) = C \exp\left(-\hbar m\omega \frac{s^2(p - p_0)^2}{2} + \frac{i}{\hbar} x_0 p\right). \quad (6.86)$$

Due to the scaling property of the differential entropy, we have that the entropy of scaled states is equal to the entropy of the state with no scaling, plus the logarithm of the scaling factor. Thus we have

$$S_x + S_p = \ln(\hbar\pi e) + \ln(s) - \ln(s). \quad (6.87)$$

However, squeezed states, as we have mentioned, do not minimize their uncertainty at every point in time. As an example we will use the wavefunction obtained by Majernik *et al.* [75]. This wavefunction is for the squeezed vacuum, however, the entropy is invariant under displacements, so for our purposes this will cover all cases. The wavefunction is given by

$$\psi(x, t) = \frac{e^{i(\frac{\omega t}{2} + \frac{\pi}{4})}}{\sqrt{\sqrt{\pi} \left(\frac{1}{s_0} \sin \omega t - i s_0 \cos \omega t\right)}} \exp\left(\frac{m\omega}{\hbar} \left(-\frac{x^2}{2s^2(\omega t)} - i \frac{\left(s_0^2 - \frac{1}{s_0^2}\right) \sin 2\omega t}{4s^2(\omega t)} x^2\right)\right), \quad (6.88)$$



where  $s_0$  is the initial squeeze parameter at  $t = 0$  and  $s^2(t) = \frac{1}{s_0^2} \sin^2 \omega t + s_0^2 \cos^2 \omega t$ . From the phase space representation above, we see that we can obtain the momentum representation by taking  $t \rightarrow t + \frac{\pi}{2}$ . Thus the distributions are simply

$$\rho(x, t) = \frac{1}{s(t)} \sqrt{\frac{\pi m \omega}{\hbar}} e^{-\frac{m \omega}{\hbar} \frac{x^2}{s^2(t)}}, \text{ and } \rho(p, t) = \frac{1}{s(t + \frac{\pi}{2})} \sqrt{\frac{1}{\pi \hbar m \omega}} e^{-\frac{1}{\hbar m \omega} \frac{p^2}{s^2(t + \frac{\pi}{2})}}. \quad (6.89)$$

Thus we have for the entropies

$$S_x = \ln \left( s(t) \sqrt{\frac{e \pi \hbar}{m \omega}} \right), \text{ and } S_p = \ln \left( s(t + \frac{\pi}{2}) \sqrt{e \pi \hbar m \omega} \right), \quad (6.90)$$

which makes the sum

$$S_x + S_p = \ln \left( e \pi \hbar s(t) s(t + \frac{\pi}{2}) \right). \quad (6.91)$$

If we take  $s_0 = e^r$  in the definition of  $s(t)$ , then we can see that this is just equal to

$$\ln \left( e \pi \hbar \sqrt{\cosh^2(2r) - \sinh^2(2r) \cos^2(2\theta)} \right) = \ln(2e\pi \Delta x \Delta p). \quad (6.92)$$

From this we see again that the entropy sum is just a monotone of the product of standard deviations.

## 6.4 The Hydrogen Atom

As is well known, the wavefunction of the hydrogen separates as follows

$$\psi_{nlm} = R_{nl}(r) Y_{lm}(\theta, \phi). \quad (6.93)$$

We will express the radial part  $R$  and angular part  $Y$  in Hartree atomic units, such that

$$R_{nl} = 2 \left( \frac{2x}{n} \right)^l e^{-\frac{x}{n}} \sqrt{\frac{k!}{n^4(n+l)!}} L_k^\alpha \left( \frac{2r}{n} \right) \quad (6.94)$$

with  $k = n - l - 1$ ,  $\alpha = 2l + 1$  and  $L$  the Laguerre polynomials. The momentum representation of the radial part of the wavefunction is given by [79, Eq. 8.8]

$$P_{nl}(p) = \sqrt{\frac{2k!}{\pi(n+l)!}} 2^{2(l+1)} n^{l+2} l! \frac{p^l}{(1+n^2 p^2)^{l+2}} C_k^{(l+1)} \left( \frac{1-n^2 p^2}{1+n^2 p^2} \right), \quad (6.95)$$

where  $C$  are the Gegenbauer polynomials. The spherical part is given by [79, Eq. 1.7]

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos(\theta)) e^{im\phi}, \quad (6.96)$$

and they have the same form in momentum space.

The question now is how we have to apply the entropy. In principle we have the normalized densities  $r^2 R_{nl}(r)$  and  $\sin \theta Y_{lm}(\theta, \phi)$ , so we could naively apply the entropy. We would get

$$\begin{aligned} S[r^2 \sin \theta R_{nl}(r) Y_{lm}(\theta, \phi)] &= S[r^2 R_{nl}(r)] + S[\sin \theta Y_{lm}(\theta, \phi)] \\ &= - \int r^2 R_{nl}(r) \ln(r^2 R_{nl}(r)) dr - \int \sin \theta Y_{lm}(\theta, \phi) \ln(\sin \theta Y_{lm}(\theta, \phi)) d\theta \end{aligned} \quad (6.97)$$

However, the uncertainty principle 6.2 is only defined for a variable and its Fourier conjugate. The momentum representation of the Hydrogen atom is not simply the Fourier transform of  $\psi_{nlm}$ , w.r.t.

$r, \theta$  and  $\phi$ . Thus, to define the entropy in a way that has a connection with the minimal uncertainty bounds, we need to define it as an integral over Euclidean space<sup>4</sup>

$$\begin{aligned} S[|\psi_{nlm}|^2] &= - \int |R_{nl}(r)|^2 |Y_{lm}(\theta, \phi)|^2 \ln (|R_{nl}(r)|^2 |Y_{lm}(\theta, \phi)|^2) r^2 dr d\Omega \\ &= - \int |R_{nl}(r)|^2 \ln (|R_{nl}(r)|^2) r^2 \sin \theta dr - \int |Y_{lm}(\theta, \phi)|^2 \ln (|Y_{lm}(\theta, \phi)|^2) d\Omega \end{aligned} \quad (6.98)$$

due to the additivity of the differential entropy. We can separate the radial and the spherical part,

$$S_{r;nl} \equiv - \int_0^\infty |R_{nl}(r)|^2 \ln (|R_{nl}(r)|^2) r^2 dr, \quad (6.99)$$

and

$$S_{\theta,\phi;lm} \equiv \int_0^{2\pi} \int_0^\pi |Y_{lm}(\theta, \phi)|^2 \ln (|Y_{lm}(\theta, \phi)|^2) \sin \theta d\theta d\phi. \quad (6.100)$$

Note that the factors from the Jacobian do not enter the logarithm, in contrast to expression 6.97. Similarly, we can separate the integrals in momentum space as

$$S_{p;nl} \equiv - \int_0^\infty |R_{nl}(r)|^2 \ln (|R_{nl}(r)|^2) p^2 dr, \quad (6.101)$$

and

$$S_{p_\theta,p_\phi;lm} \equiv \int_0^{2\pi} \int_0^\pi |Y_{lm}(p_\theta, p_\phi)|^2 \ln (|Y_{lm}(p_\theta, \phi)|^2) \sin(p_\theta) dp_\theta dp_\phi. \quad (6.102)$$

To our knowledge no completely analytical solution exists of these integrals. As such, we will present the results of numerical integration of the above functions using Mathematica. Our target accuracy was five decimals. However, we were not able to obtain result with good accuracy for every wavefunction, using the integration package of Mathematica, and letting it pick the upper bound of integration. For these cases the upper limit had to be picked manually. This was needed because at too large of an upper limit, the integration starts to become unstable. We are confident in our end results, however, since our results reproduce those of the work done by Jiao *et al.* [80] up to our target accuracy. They did not have any results for  $n = 11$  however. For the position integrals convergence to a precision of under three decimals was easily obtained, but for the momentum integrals this was not the case. As such, we expect an error larger than  $10^{-5}$  for these results. The results of the numerical integration are presented in figures 6.10-6.15, and the actual numerical results themselves can be found in the appendix A.

Comparing the entropy with the standard deviation turns out to be a bit less straightforward this time around. Ideally, we would like to compare uncertainty measures that are subject to the same lower bound. This way, we can clearly discuss the measures w.r.t. the minimal uncertainty. In one dimension, the entropy and standard deviation had a clear relation, which is seen in equation 6.3. In three dimensions, we have the uncertainty principle for the entropy given by 6.4. The analogous uncertainty measure in terms of standard deviations, would be the logarithm of the products of all the Euclidean standard deviations,  $\ln (\pi^3 e^3 \Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z)$ , since this is subject to the same lower bound. However, consider as an example the standard deviation in  $x$ . This is obtained by “projecting” the density onto the  $x$ -axis, through integration over the  $y$  and  $z$  axes. As a consequence, the standard deviation product is not invariant under rotations. Of course, the uncertainty principle must still be satisfied whatever rotation we choose. The aforementioned adds a factor of arbitrariness to the product of standard deviations as a measure of uncertainty. The standard deviation still has a clear interpretation in one of the dimensions, but we must conclude that it does not satisfactorily

---

<sup>4</sup>For clarification, note that now, we don't have the factors from the Jacobian within the logarithm. Thus we are just calculating the entropy of the density defined on the Cartesian coordinates  $x, y$  and  $z$ , even though we do not have an explicit expression for it.

express the overall uncertainty over all of space, like the entropy does<sup>5</sup>. In figure 6.15 we have added the results of the logarithm of the standard deviation product,  $\ln(\pi^3 e^3 \Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z)$ . Here we have defined the the spherical coordinates of the Hydrogen atom in the canonical way, w.r.t. the Cartesian coordinate coordinates,  $(x, y, z) = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)$ .

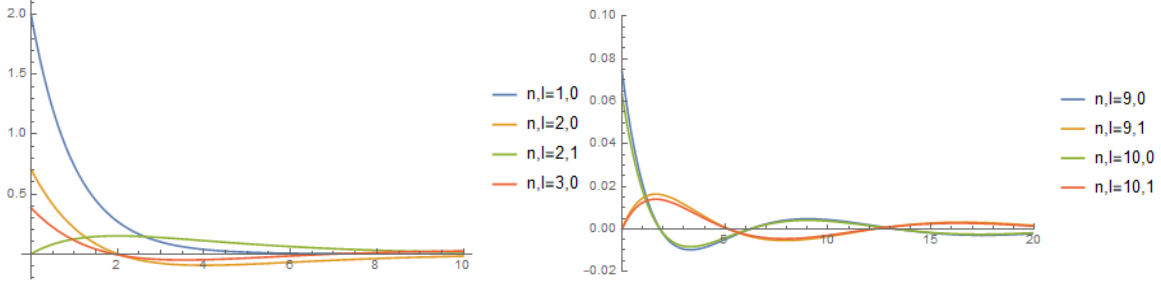


Figure 6.10: Several radial wavefunctions in the position representation for different  $n, l$ .

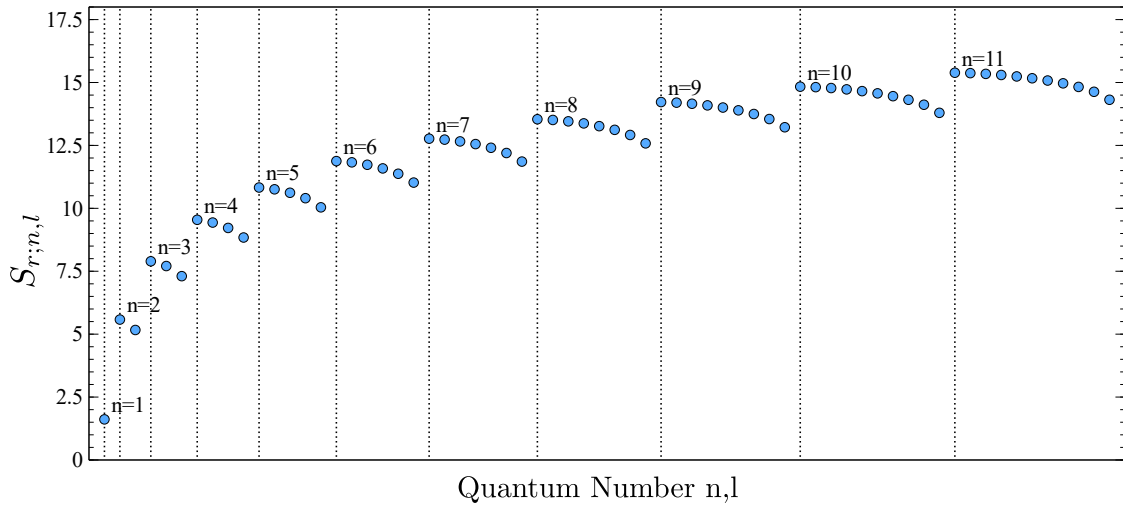


Figure 6.11: The differential entropy of the radial part of the hydrogen atom wavefunction in position representation. Every vertical dotted line indicates a new  $n$  value, the subsequent data points have a different value for  $l$ , in order of their magnitude.

## 6.5 Maximum Entropy Wavefunctions

We have discussed maximum entropy inference, and how it may be used to obtain the ensembles of statistical mechanics. If we maximize the BGS entropy we find the Boltzmann distribution, which bears no significance for pure state quantum mechanics, given that  $\psi^*(x)\psi(x) = \rho(x)$ . However, if we do not merely require the distribution to maximize the entropy, but instead the wavefunctions themselves, we may find an interesting result [81].

<sup>5</sup>There are different ways to define the standard deviations in multiple dimensions. One might consider, for example, the radial standard deviation  $\Delta r = \sqrt{\langle x^2 + y^2 + z^2 \rangle}$  and its momentum counterpart for the comparison. However, note that  $\Delta r$  is not the analogous measure compared to the radial entropy that we have calculated. Instead it would be analogous to the entropy of equation 6.97. The product  $\Delta r \Delta p$  is also not subject to the same bound lower bound. Moreover, comparisons of the uncertainty of states of systems with different symmetries would again have the same problems. Since the entropy may always be defined unambiguously in Euclidean space, it does not have these problems.

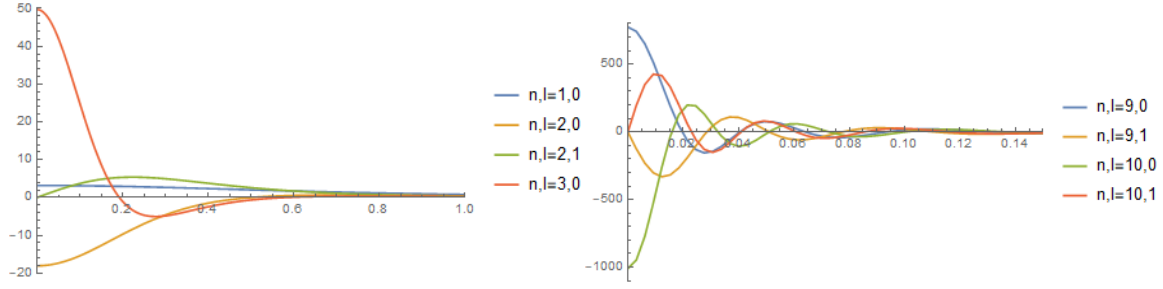


Figure 6.12: Several radial wavefunctions in the momentum representation for different  $n, l$ .

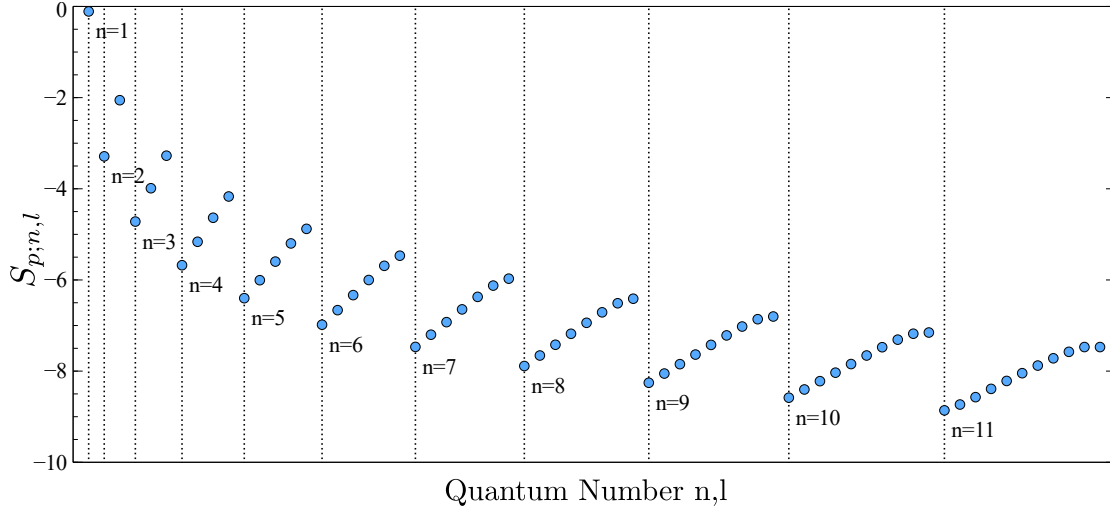


Figure 6.13: The differential entropy of the radial part of the hydrogen atom wavefunction in momentum representation. Every vertical dotted line indicates a new  $n$  value, the subsequent data points have a different value for  $l$ , in order of their magnitude.

We will thus maximize

$$S(\psi^*, \psi) = - \int \psi^* \psi \ln(\psi^* \psi) dx, \quad (6.103)$$

with the constraints of, analogously to Jaynes' procedure, normalization of the probability density, and the expectation value  $E$  of some operator  $\hat{H}$ ;

$$\langle \psi | \psi \rangle = 1 \quad (6.104)$$

$$\langle \psi | \hat{H} | \psi \rangle = E. \quad (6.105)$$

We can solve this by adding these constraints with Lagrange multipliers to the entropy and solving the Euler-Lagrange equations, so that we obtain the equations

$$\beta \hat{H} \psi + \psi \ln \left( \frac{\psi^* \psi}{\rho} \right) + (\lambda + 1) \psi = 0 \quad (6.106)$$

$$\beta \hat{H}^* \psi + \psi \ln \left( \frac{\psi^* \psi}{\rho} \right) + (\lambda + 1) \psi = 0. \quad (6.107)$$

These equations can only simultaneously hold when  $\hat{H}$  is a self adjoint operator. Thus the only equation we are left with is then

$$\hat{H} \psi + T \ln \left( \frac{\psi^* \psi}{\rho} \right) \psi = \gamma \psi, \quad (6.108)$$

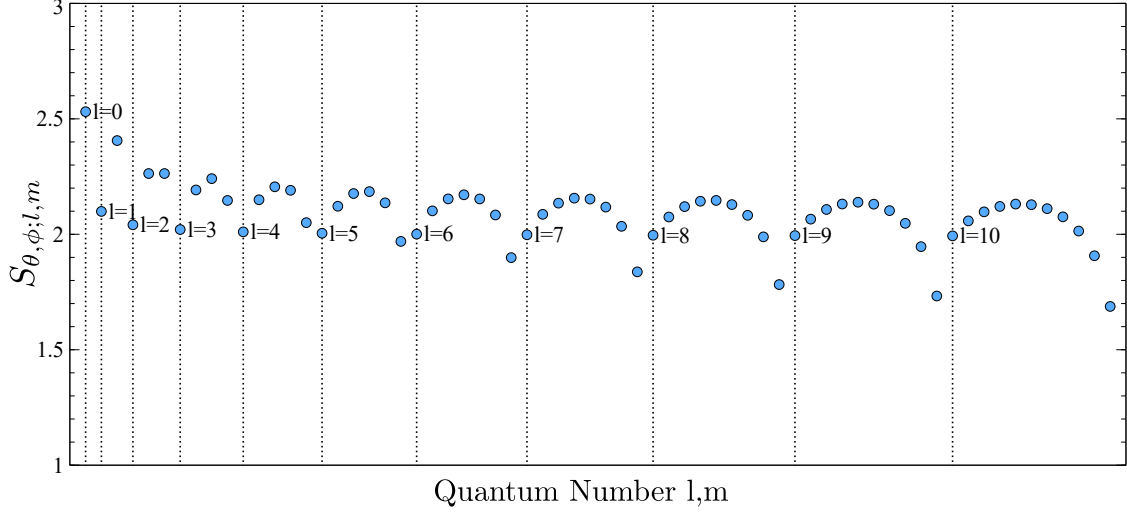


Figure 6.14: The differential entropy of the spherical part of the hydrogen atom. Every vertical dotted line indicates a new  $l$  value, the subsequent data points have a different value for  $m$ , in order of their magnitude.

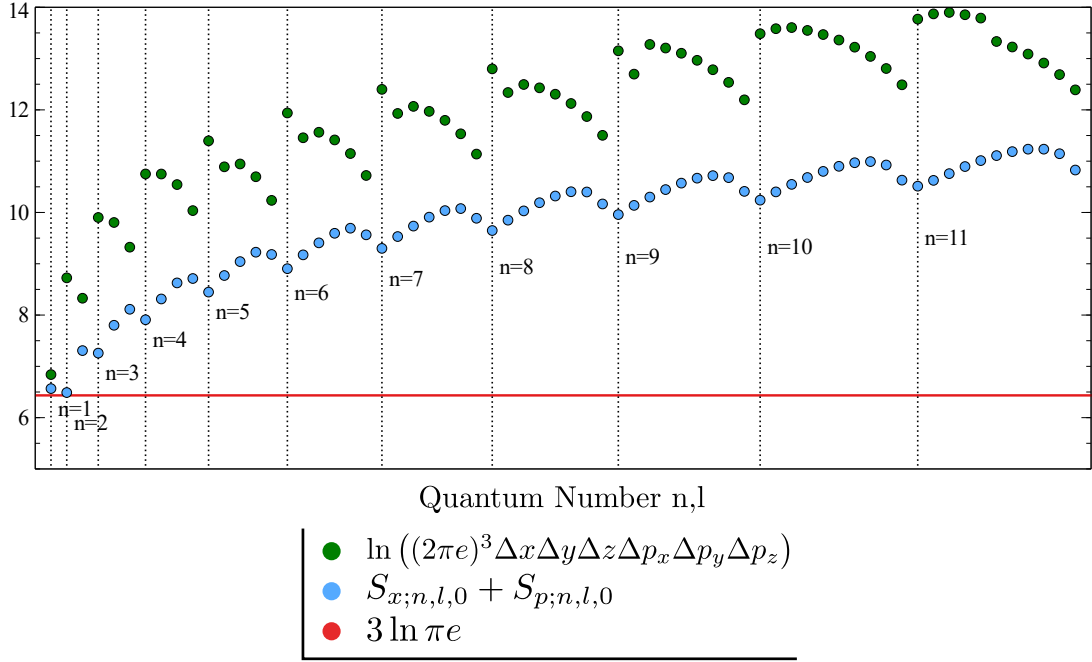


Figure 6.15: The total differential entropy of the position and momentum representations, together with the logarithm of the standard deviation product  $\ln(\pi^3 e^3 \Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z)$  and the lower bound 6.4. All values for  $m$  were chosen to be 0.

where we have defined  $T = 1/\beta$  and  $\gamma = -(\lambda + 1)/\beta$ . The Lagrange multipliers are determined by the constraints and vice versa. If we set  $T = 0$  then we see that  $\psi$  must be an eigenfunction of  $\hat{H}$ , with eigenvalue  $\gamma$ . Thus the zero temperature limit corresponds to eigenfunctions of the operator  $\hat{H}$ . Note that this “temperature” should not be interpreted as physical temperature, but merely a parameter that controls the average energy of the system.

If  $\hat{H}$  is the Hamiltonian and  $E$  is an eigenvalue, we find that for zero temperature we recover the

time-independent Schrödinger equation.

$$H\psi = E\psi. \quad (6.109)$$

The further properties and systems with non-zero  $T$  are investigated by Jakobsen *et al.* [81].

## 6.6 Summary and Concluding Remarks

In this section, we have discussed the application of the differential entropy, as well as the continuous Tsallis and Rényi entropies, to the continuous position and momentum distributions of quantum states. Most importantly, these entropies are used as alternative uncertainty measures to the standard deviation. Importantly, many forms of Heisenberg's uncertainty relation have been proven for these entropies. We have seen that some obtain a similar form, while others are quite different due to the use of course graining. Most notable is Białynick-Birula and Mycielski's relation

$$S_x + S_p \geq \ln(\hbar\pi e). \quad (6.110)$$

After that, we have discussed the application of these entropies to simple quantum systems and their dependence on the quantum number  $n$ . First we discussed the infinite square well. An interesting difference between the standard deviation and the differential entropy that we found, is that the differential entropy sum has a limiting value as  $n \rightarrow \infty$ . The reason for this difference in behavior, is due to two things. First is that the quantum number  $n$  works as a scaling parameter for  $x$ , while at the same time effectively scaling the width of the well  $a$  compared to  $x$ . Due to the behavior of the entropy, w.r.t. scaling, it is invariant under a change of  $n$ . The standard deviation, on the other hand, does change but has a limit when  $n \rightarrow \infty$ . The momentum distribution on the other hand separates into two peaks, where asymptotically  $n$  only controls the shift between the two peaks. We discussed how the differential entropy behaves for such distributions in section 1.3.1, which is that it goes to two times the entropy of one of those separate peaks. Notably, the expression of the standard deviation is quite easily obtained through the operator representation of  $p$ . On the other hand, to our knowledge, no analytic expression is known for the entropy as a function of  $n$ . Thus in this case, the standard deviation seems to have an advantage when it comes to calculation. There are known analytic expressions for the Rényi and Tsallis entropies when  $q$  and  $\alpha$  are not equal to 1. However, these are quite difficult expressions to work with, and due to the nature of the uncertainty principle in terms of the Rényi entropy, there is no clear connection between these expression and the minimal uncertainty.

The second system we have looked at, is the harmonic oscillator. The behavior of the standard deviation and entropy this time is more similar. They both diverge as  $n \rightarrow \infty$ , and  $\ln(\Delta x \Delta p)$  diverges like  $\ln(n)$ , just like the entropy. Moreover, while the standard deviations are very easily obtained using the ladder operator representations of  $x$  and  $p$ , there is again, to our knowledge, no analytical expression of the entropy sum as a function of  $n$  known. Moreover, we used the entropy to compare the quantum and classical distributions of the harmonic oscillator. In contrast to the standard deviation, the entropy measures the quantum states with  $n \geq 15$  as being less uncertain than its classical counterpart. The reason for the difference in behavior in the standard deviation and the entropy is unclear, though it could stem from the sensitivity of the standard deviation for values far from the origin [82], and the fact that the quantum density has infinite support. Next, we discussed the uncertainty measures for coherent states and squeezed states. Uncertainty measures are particularly interesting in this setting, since there are non-classical states of the harmonic oscillator, that can have smaller uncertainties for either position or momentum, than the "classical" coherent states. These states are the squeezed states that we discussed. For these states the standard deviations are a good classification of the uncertainty, and allow us to discuss what states are better than others for particular applications. The expressions for the standard deviations, are reasonably easy to obtain, once the commutation relations between the ladder operators and the displacement and squeeze operators are known. In fact, quantum optical states are actually most often discussed in terms of these operators. The entropies, on the other

hand, require analytic expressions of the wavefunctions. The wavefunction is needed, because the Fourier transform acts on it, instead of on the density. Although, the probability densities for these states are always Gaussian, when the squeeze states are not of minimal uncertainty, we cannot just assume that the momentum distribution has the inverse squeeze factor in its standard deviation compared to the position standard deviation. Due to this, using the Fourier transform is a necessity. Once the explicit densities are known however, the entropies are easy to obtain, due to the simple expression for the entropy of a Gaussian.

The last system that we looked at, is the hydrogen atom. Here, we have discussed results that we have obtained for the entropies of the orbital wavefunctions of the hydrogen atom. As a result of the symmetries of the Hamiltonian, the wavefunction only nicely separates in spherical coordinates. However, the uncertainty principles that we have discussed earlier, are expressed in Cartesian coordinates. Still, the Euclidean entropy in spherical coordinates separates into a radial part and an angular part as well, though those separate parts separately cannot be connected with an uncertainty principle, as far as we know. We have also discussed the problems with the use of the product of standard deviations, due to the arbitrariness in choosing a coordinate system. Although, the separate standard deviations in Cartesian coordinates have a clear interpretation, the product does not. Thus from our reasoning, the differential entropy seems like a better measure for overall uncertainty in multiple dimensions. Moreover, the entropy was generally less difficult to calculate using the integration package of Mathematica than the standard deviations. Since the orthogonal sets of quantum states need sets of orthogonal functions to be expressed in continuous forms, there has been some research into the entropy of these entropies of these sets of orthogonal functions. Though we have not used them here, we can then often express the entropic integrals into somewhat simpler equations involving the entropies of these orthogonal functions [83][80, §2.3-2.4], which makes the entropy even simpler to work with for numerical calculations.

In the last section we briefly discussed the effect of entropy maximization on wavefunctions. By only assuming the normalization condition of the probability density, and a mean energy given as the expectation value of some operator  $H$ , we retrieved the Schrödinger equation if  $H$  is interpreted as the Hamiltonian.

Finally, we wish to mention a few applications of the differential entropy in quantum theories that we have not discussed in the text. The differential entropy sum has been studied as a measure of the quality of atomic trial wave functions [84], it has been studied as a measure for electron correlation [66][85], and it has seen applications in quantum chemistry [86], among other things.

## Chapter 7

# Measures of Mixedness

In chapters 5 and 6 we have concerned ourselves with the information theoretic aspects of the entropy quantities in quantum mechanics. In a similar manner to the standard deviation, the BGS, Tsallis and Rényi entropies were regarded as a measure of the broadness of a distribution or uncertainty in outcome of a distribution. We also have discussed in chapter 1, how the BGS entropy emerges in statistical physics, from the point of view of classical physics. In this point of view, the ergodic principle allowed us to make a link between the averages over disjoint systems (an ensemble of systems) in different states and the time average of a single system. This link crucially allowed us to establish the thermodynamical properties of a single system. Of course, we know now that the microscopic world is fundamentally described by quantum mechanics, and not by classical mechanics. One might raise the question whether Shannon entropy applied to pure states may have a connection to the thermodynamical entropy. After all, it measures the uncertainty in outcome of a quantum measurement exactly in the same way as it does for the uncertainty in outcome in a classical experiment. However, the averages that we calculate in quantum mechanics are measures over an ensemble of identical systems, not over a single system. For a single system, the state is collapsed upon measurement and the information theoretical entropy may then only be considered zero. Moreover, we have seen that the entropies of pure states are actually representation dependent, which would be unacceptable for any entropy that has to be connected to a thermodynamical entropy. Thus, it should be clear that for a connection to the thermodynamical entropy, we must consider other quantities.

We will show in this section that the actual thermodynamical quantum entropy must be defined on mixed states, be it proper or improper. This entropy is known as the Von Neumann entropy. We will actually show two different interpretations of the Von Neumann entropy as a thermodynamical entropy. The first interpretation is shown through a thought experiment that Von Neumann himself used to derive his entropy. In this thought experiment we consider an ensemble of separate particles all described by the same density matrix. This density matrix is meant to encode the classical ignorance we have about the actual pure state that particular particle is in, while we know the distribution of states within our ensemble.

The second interpretation is more typical for quantum statistical mechanics (QSM), and uses the time averaged density matrix. Similarly as in classical statistical mechanics, we rely on a few unproven postulates that tell us how we get from the typical description of an isolated pure state to the statistical description. Thus, just like in the classical case, these postulates tell us what the behavior of a system should be to obtain a QSM description, but not what systems show this behavior. Moreover, we will also discuss why in quantum mechanics the notion of a completely isolated state is not sufficient to obtain QSM.

Since these concepts rely on the idea of mixed states, we will first set up a definition of mixedness that is independent of entropy. This definition is based on a reflexive and transitive relation on



the set of density matrices which is called majorization. Then using this definition, we will show a general criterion, called Schur concavity, for a quantity to be monotone in this relation. In other words, we can show when a quantity may be considered a measure of mixedness.

Lastly, we will briefly discuss the properties of the Von Neumann entropy and of the quantum generalizations of the Tsallis and Rényi entropies, which also turn out to be Schur concave, and thus measures of mixedness.

## 7.1 The Von Neumann Entropy

In this chapter we will introduce the Von Neumann entropy, which was introduced by Von Neumann as a continuation of the classical BGS entropy on quantum systems. His entropy is defined as follows

$$S[\hat{\rho}] = -\text{Tr} [\hat{\rho} \ln \hat{\rho}] \quad (7.1)$$

where  $\hat{\rho}$  is the density matrix representation of some state. We may also write the Von Neumann entropy as

$$S[\hat{\rho}] = -\sum_i \lambda_i \ln \lambda_i, \quad (7.2)$$

where  $\lambda_i$  are the eigenvalues of the density matrix  $\hat{\rho}$ . We see that this is equivalent to equation 7.1 by considering that the trace is preserved under unitary transformations, and thus stays the same if we diagonalize the matrix which is traced over. Moreover, functions of matrices can be defined through the Taylor series, and all powers of a matrix may be simultaneously diagonalized. Thus the trace of some function  $f$  of a matrix may be computed from the diagonal form. In this basis, the function of the diagonal matrix,  $f(\hat{\rho}_D)$ , becomes a diagonal matrix where the diagonal entries are given by  $f(\lambda_i)$ , i.e. the function evaluated at the eigenvalues. This may again be seen from the Taylor expansion. If  $|\phi_i\rangle$  are orthogonal states, we then have

$$\text{Tr} [f(\hat{\rho})] = \text{Tr} \left[ f \left( \sum_i \lambda_i |\phi_i\rangle \langle \phi_i| \right) \right] = \text{Tr} \left[ \sum_i f(\lambda_i) |\phi_i\rangle \langle \phi_i| \right] = \sum_i f(\lambda_i). \quad (7.3)$$

Taking  $f(\hat{\rho}) = \hat{\rho} \ln \hat{\rho}$  we see that equation 7.8 is exactly the Von Neumann entropy of equation 7.1.

In this section we will discuss this form of the entropy. We will now first show a simple argument based on thermodynamic considerations to derive the Von Neumann entropy. After that we will discuss the properties of the Von Neumann entropy.

### 7.1.1 Von Neumann's Gedanken Experiment

In this section we present a gedanken experiment along the lines of what Von Neumann used to derive his entropy [87, Ch. V][88] [89]. A schematic representation of the experiment is given in figure 7.1.

Consider a gas in a box that behaves according to quantum mechanics, and is described by the mixed state  $\hat{\rho} = \lambda |\phi_1\rangle \langle \phi_1| + (1 - \lambda) |\phi_2\rangle \langle \phi_2|$ . We mean this in the sense that there is a classical lack of knowledge about the state of each particle. If there are then a total of  $N$  particles,  $\lambda N$  are in the pure state  $\phi_1$  and  $(1 - \lambda)N$  in  $\phi_2$ . Initially, the box has volume  $V$ . Next we add an empty box to the original box, and replace the shared wall by two new walls. Wall  $A$  is completely impenetrable, while wall  $B$  is permeable only by the  $\phi_2$  particles. Note that for this to be possible,  $\phi_1$  and  $\phi_2$  must be orthogonal. If they were not,  $|\phi_2\rangle$  may be written as  $|\phi_1\rangle + |\phi_3\rangle$  where  $\phi_1 \perp \phi_3$ . The wall has an interaction with  $\phi_1$ , by definition, so it cannot be completely permeable to  $\phi_2$  in this case. We now also add a wall,  $C$ , to the right of the box, left to the original wall. We assume this wall is permeable to  $\phi_1$  but not  $\phi_2$ . Next, we push walls  $A$  and  $C$  very slowly (reversibly) to the left, without changing the relative distance of the walls. When this process is done, we are left with a

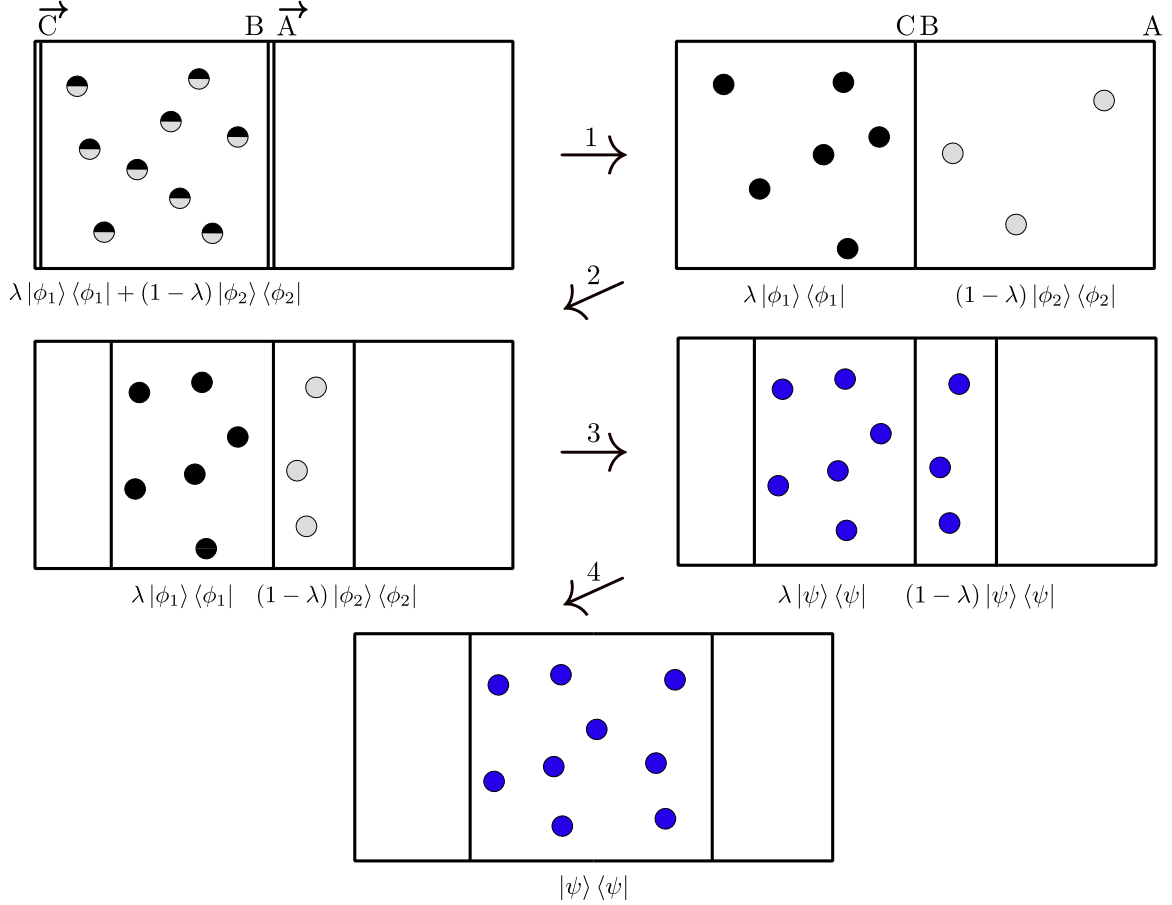


Figure 7.1: An overview of the gedanken experiment

gas of  $\phi_2$  particles on the left, and a gas of  $\phi_1$  particles on the right. No work was done and no heat was added, so the entropy should remain the same during this process. Now we will compress the two gases to their original specific densities,  $N/V$ , while coupled to a heat bath. We compress the  $\phi_1$  gas to a new volume of  $\lambda V$  and the  $\phi_2$  gas to a volume of  $(1-\lambda)V$ . Assuming the gas behaves like an ideal gas, we invoke the ideal gas law for an isothermal process

$$\Delta Q = N k_B T \ln \frac{V_2}{V_1}. \quad (7.4)$$

If we compress the  $\lambda N$  particles, using Clausius' definition of entropy we find

$$S = \frac{\Delta Q}{T} = k_B N \sum_i \lambda_i \ln \lambda_i, \quad (7.5)$$

where  $\lambda_i$  is the fraction of  $\phi_i$  in the total state. Thus, the term  $\lambda_i N$  in front of the logarithm should be interpreted as the number of particles that are compressed, and the  $\lambda_i$  term within the logarithm is the ratio of the final volume to the initial volume. This means the gas loses heat energy to the heat bath. When this is done, we reversibly change all the states to some state  $|\psi\rangle$ . Again, since the process is reversible, the entropy should not change. Then finally, we can remove the barrier between the two gases. Since the gases are indistinguishable, this mixing is also a reversible change. Now finally, we can add up the entropies so that we can relate the entropy of the first state, to the entropy of the final state. We denote the entropy of the initial mixture as  $S[\hat{\rho}; N]$  and the entropy of the final gas where all the particles are in state  $|\psi\rangle$ , as  $S[\psi; N]$ , so that we get

$$S[\hat{\rho}; N] + k_B N \lambda \ln \lambda + k_B (1-\lambda) N \ln(1-\lambda) = S[\psi; N]. \quad (7.6)$$

If we assume extensivity in the thermodynamic limit, we assume that entropy is proportional to the number of particles multiplied by the mean entropy. Then, dividing by  $N$  we obtain

$$S[\hat{\rho}] + k_B \lambda \ln \lambda + k_B (1 - \lambda) \ln(1 - \lambda) = S[\psi] \quad (7.7)$$

The same process may be imagined for a mixture of more states, where we first separate them all, and then transform them to some state  $|\psi\rangle$ . Since the choice of the state of the final gas is arbitrary and independent of the initial mixture, this should not contribute to the entropy. Thus we set  $S[\psi] = 0$ . We then we arrive at the formula

$$S[\hat{\rho}] = -k_B \sum_i \lambda_i \ln \lambda_i. \quad (7.8)$$

We see that this is indeed equal to the expression of the Von Neumann entropy in 7.2, only multiplied by the Boltzmann constant.

## 7.1.2 The Ensembles and Entropies of Quantum Statistical Physics

It is empirically known that statistical mechanics may be continued to the quantum domain, and still give useful predictions in a lot of different cases. However, as we will see the generalizations are not trivial, and concepts such as ergodicity for an isolated system must be abandoned.

In quantum statistical mechanics (QSM), the microcanonical average of some operator  $A$  is given by

$$\langle A \rangle = \frac{1}{W} \text{Tr} \left[ \hat{A} f_\delta(E - H_n) |n\rangle \langle n| \right] \quad (7.9)$$

where  $|n\rangle$  are the energy eigenstates. Moreover,  $f_\delta(E - H_n)$  is a function that is equal to 1 when the energy of the eigenstate  $|n\rangle$  is equal to the energy of the ensemble and 0 otherwise<sup>1</sup>[9, Ch. 8]. We see that, in analogy to the classical case, the thermodynamical expectation value of our operator, is given by a classical average over all states with energy  $E$ , where all states are equally likely. Ergodicity in this context may then be taken to mean that the expectation values are averaged equally over all states with energy  $E$ . Once we have established ergodicity, from similar arguments as in the classical case, we may derive the canonical ensemble. However, at this point it is reasonable to ask why this should give the correct averages. After all, why would a typical quantum system be described by a mixed density matrix. In general this is not an easy question to answer, so like in classical statistical mechanics, we rely on postulates to derive the QSM distributions.

Firstly, we postulate that in realistic macroscopic experiments, the pure state of the system itself is usually not what we are measuring. Instead, the central quantity becomes the time averaged density matrix  $\bar{\rho}_t$ , since the measurements are typically made over much larger time-scales than the phase oscillations. An important consequence of this is that as the averaging time  $T$  becomes sufficiently large, the state will behave like an incoherent state in the energy eigenbasis, provided there are no degeneracies (we will discuss this later). This is seen by considering that the density matrix can be written as follows

$$\hat{\rho}_t = \sum_{ij} e^{iE_i t} c_i e^{iE_j t} c_j^* |i\rangle \langle j|. \quad (7.10)$$

If we take  $\frac{1}{T} \int_0^T \hat{\rho}_t dt$ , as long as  $T$  is long enough, the initial phases of  $c_i$  and  $c_j^*$  become irrelevant and the phases will average to zero. Thus the density matrix will become

$$\bar{\rho}_t = \sum_n |c_n|^2 |n\rangle \langle n|, \quad (7.11)$$

where  $|n\rangle$  are the energy eigenstates. When the state is isolated and has degenerate energy eigenstates, the evolution of the phases of those states is the same, and thus cannot be argued to cancel

<sup>1</sup>We may also define the function  $f_\delta$  to be equal to 1 if  $H_n$  is within a small range of  $E$ . This ensures that there are always states in the ensemble, even if  $E$  is not in the spectrum of the Hamiltonian. For simplicity, we will just assume that  $E$  is in the spectrum.

out. As a result, we have to relinquish the requirement that the system underlying the microcanonical ensemble has to be isolated. In QSM, some coupling to an external system is always assumed [9, Ch. 8]. For instance, writing the total state of system  $S$  plus some environment  $E$  as

$$\hat{\rho} = \sum_{ijkl} c_i j c_{lk}^* |S_i\rangle |E_j\rangle \langle S_l| \langle E_k|, \quad (7.12)$$

we may invoke the so-called postulate of random phase [9, §8.1]<sup>2</sup>,  $\frac{1}{T} \int_0^T c_{ij}(t) c_{lk}^*(t) dt \rightarrow \delta_{ij, lk} |c_{ij}|^2$ . Then again, but now for all states, the time averaged density matrix of the system becomes completely incoherent.

### Microcanonical Ensemble

The (time averaged) density matrix of the microcanonical ensemble, at energy  $E$ , in QSM is given by

$$\bar{\rho} = \frac{1}{W} \sum_n f_\delta(E - H_n) |n\rangle \langle n|, \quad (7.13)$$

where  $f_\delta(E - H_n)$  has the same definition as for equation 7.9. Moreover,  $W = \text{Tr} [f_\delta(E - H_n) |n\rangle \langle n|]$ . The entropy is the same as in classical mechanics,

$$S(E) = k_B \ln W(E), \quad (7.14)$$

where, as defined above,  $W(E)$  is the amount of states in the ensemble at energy  $E$ . This is of course just equal to the Von Neumann entropy, since

$$S[\bar{\rho}] = -k_B \sum_n \frac{1}{W} \ln \frac{1}{W} = k_B \ln W. \quad (7.15)$$

### Canonical Ensemble

The density matrix of the canonical ensemble is given by

$$\bar{\rho} = \frac{1}{Q_N} e^{-\beta \hat{H}} = \frac{1}{Q_N} \sum_n e^{-\beta H_n} |n\rangle \langle n|, \quad (7.16)$$

where  $\hat{H}$  is the Hamiltonian and  $H_n$  are the energies of the energy eigenstates, and the  $N$  particle partition function  $Q_N = \text{Tr} [e^{-\beta \hat{H}}]$ . The entropy is defined in the exact same way as in the classical case, such that

$$S = -k_B \sum_n p_n \ln p_n = -k_B \sum_n \frac{e^{-\beta H_n}}{Q_N} \ln \frac{e^{-\beta H_n}}{Q_N} = \text{Tr} [-\bar{\rho} \ln \bar{\rho}], \quad (7.17)$$

which is thus again the Von Neumann entropy.

### 7.1.3 Properties

The Von Neumann entropy inherits many of its properties from the Shannon entropy, though there are a few differences [91]. Most of them can be derived from the identity that  $S[\hat{\rho}] = \sum_i \lambda_i \ln \lambda_i$ . However, the Von Neumann entropy has a more intricate dependence on the whole density matrix and the properties do not trivially generalize to the Von Neumann entropy. Because of this, we will show short derivations for most of the properties.

---

<sup>2</sup>Supporting the idea that external interference generates the microcanonical ensemble, Deutsch has shown that adding a small random matrix to the Hamiltonian, ergodicity is recovered [90]. This random matrix may represent noise generated by an environment

### Simple Properties

The Von Neumann entropy is a continuous function of the eigenvalues of the density matrix, since it is simply the BGS entropy of the eigenvalues.

It is defined for every density matrix and maps the set of all density matrices to the interval  $[0, \infty]$ . However, if  $\hat{\rho}$  is of finite rank, then necessarily  $S[\hat{\rho}] < \infty$ .

The Von Neumann entropy is invariant under unitary transformations

$$S[U^\dagger \hat{\rho} U] = S[\hat{\rho}], \quad (7.18)$$

which follows directly from the cyclic property of the trace. Analogously to the expansibility of the BGS entropy, it is invariant under the direct addition of an empty subspace,

$$S[\hat{\rho} \oplus 0] = S[\hat{\rho}] \quad (7.19)$$

Lastly, if we separate a Hilbert space into two parts,  $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ , and we have a state  $\hat{\rho}_{AB}$ , we then have for any state

$$S[\hat{\rho}_A] = S[\hat{\rho}_B]. \quad (7.20)$$

This is a consequence of the Schmidt decomposition which we discussed in section 4.3.

### Concavity

The Von Neumann entropy is concave

$$S\left[\sum_i p_i \hat{\rho}^{(i)}\right] \geq \sum_i p_i S[\hat{\rho}^{(i)}]. \quad (7.21)$$

This thus means that any density matrix which may be written as the sum of two other density matrix, is more mixed than its parts on average. This may be proven using the concavity of the function  $s(x) = -x \ln x$ . Consider the case where  $\hat{\rho} = p_1 \hat{\rho}_1 + p_2 \hat{\rho}_2$ . If we write  $\hat{\rho} = \sum_n \lambda_n |n\rangle \langle n|$ , then

$$\begin{aligned} S[\hat{\rho}] &= -\sum_n \lambda_n \ln \lambda_n = \sum_n s(\langle n|\hat{\rho}|n\rangle) = \sum_n s(p_1 \langle n|\hat{\rho}_1|n\rangle + p_2 \langle n|\hat{\rho}_2|n\rangle) \\ &\geq p_1 \sum_i s(\langle n|\hat{\rho}_1|n\rangle) + p_2 \sum_i s(\langle n|\hat{\rho}_2|n\rangle) \stackrel{(*)}{\geq} p_1 \sum_n \langle n|s(\hat{\rho}_1)|n\rangle + p_2 \sum_n \langle n|s(\hat{\rho}_2)|n\rangle \\ &= p_1 S[\hat{\rho}_1] + p_2 S[\hat{\rho}_2]. \end{aligned} \quad (7.22)$$

The marked inequality is obtained by considering the following: The density matrices  $\hat{\rho}_i$  are not necessarily diagonal in the same basis as  $\hat{\rho}$ . Thus  $\langle n|s(\hat{\rho}_i)|n\rangle = \sum_{m_i} s(\lambda_{m_i}) |\langle m_i|n\rangle|^2$ , where we defined  $\hat{\rho}_i = \sum_{m_i} \lambda_{m_i} |m_i\rangle \langle m_i|$ . Then, since  $s$  is a concave function we have, by definition  $s(\langle n|\hat{\rho}_1|n\rangle) = s(\sum_{m_i} \lambda_{m_i} |\langle m_i|n\rangle|^2) \geq \sum_{m_i} s(\lambda_{m_i}) |\langle m_i|n\rangle|^2 = \langle n|s(\hat{\rho}_1)|n\rangle$ .

### Schur Concavity

The Von Neumann is Schur concave. Through the dependence on the eigenvalues of the density matrix we may fill in the Von Neumann entropy into condition 4.26 to find

$$(\lambda_i - \lambda_j) \ln \left( \frac{\lambda_j}{\lambda_i} \right). \quad (7.23)$$

We have three cases,  $\lambda_i > \lambda_j$ ,  $\lambda_j > \lambda_i$  and  $\lambda_i = \lambda_j$ . In the case of equality, we trivially have  $(\lambda_i - \lambda_j) \ln \left( \frac{\lambda_j}{\lambda_i} \right) = 0$ . In both other cases the total expression is always negative. Thus we conclude that the Von Neumann is Schur concave and thus a measure of mixedness.

## Additivity

The Von Neumann entropy is additive. Let  $\hat{\rho}_{12} = \hat{\rho}_1 \otimes \hat{\rho}_2$ , then

$$S[\hat{\rho}_{12}] = S[\hat{\rho}_1] + S[\hat{\rho}_2]. \quad (7.24)$$

This means that, similar to the classical case, if both subsystems are unentangled, or separable, then the entropy of the eigenvalues of the total system, is the same as the sum of the entropies of the subsystems. This also again means that the entropy is extensive if all the subsystems are separable. In that case we may always assign  $N \cdot S_{mean}$  as the total entropy in the thermodynamic limit, which is finite as long as the entropy of the subsystems is finite.

For the proof we may simply use the additivity of the BGS entropy. If  $\hat{\rho}_{12} = \hat{\rho}_1 \otimes \hat{\rho}_2$  this means that

$$S[\hat{\rho}_{12}] = \sum_{ij} s(\hat{\rho}_{ijij}) = \sum_{ij} s(\lambda_i^{(1)} \lambda_j^{(2)}), \quad (7.25)$$

where  $\lambda_i^{(1)}$  and  $\lambda_j^{(2)}$  are the eigenvalues of  $\hat{\rho}_1$  and  $\hat{\rho}_2$  respectively. Then, due to the additivity of the BGS entropy, the Von Neumann entropy is additive. In other words, if two quantum systems are uncorrelated, the entropy separates into a sum of the entropies of the separate systems.

## Sub-Additivity

The Von Neumann entropy is subadditive. For any general  $\hat{\rho}_{12}$  we have that

$$S[\hat{\rho}_{12}] \leq S[\hat{\rho}_1] + S[\hat{\rho}_2]. \quad (7.26)$$

This means that if the subsystems are entangled, then the total system is less mixed than its subsystems together. Since the density matrix may be interpreted as a classical distribution over pure quantum states, we may interpret this statement as follows: We are always more certain about the pure state the total system is in, than we are about both the pure states the subsystems are in. This is obtained from the positivity of the relative entropy  $S[\hat{\rho}|\hat{\sigma}] = \text{Tr}[\hat{\rho}(\ln \hat{\rho} - \ln \hat{\sigma})] \geq 0$  [2, §II]. Take  $\hat{\rho} = \hat{\rho}_{12}$  and  $\hat{\sigma} = \hat{\rho}_1 \otimes \hat{\rho}_2$ . We then immediately have

$$\begin{aligned} S[\hat{\rho}_{12}|\hat{\rho}_1 \otimes \hat{\rho}_2] &= \text{Tr}[\hat{\rho}_{12}(\ln \hat{\rho}_{12} - \ln \hat{\rho}_1 \otimes \hat{\rho}_2)] = \text{Tr}[\hat{\rho}_{12} \ln \hat{\rho}_{12} - \hat{\rho}_{12} \ln(\hat{\rho}_1 \otimes I) - \hat{\rho}_{12} \ln(I \otimes \hat{\rho}_2)] \\ &= S[\hat{\rho}_{12}] - S[\hat{\rho}_1] - S[\hat{\rho}_2] \geq 0 \end{aligned} \quad (7.27)$$

## Strong Sub-additivity

The Von Neumann entropy is strongly subadditive. For any density matrix of three systems,  $\hat{\rho}_{123}$ , we have

$$S[\hat{\rho}_{123}] + S[\hat{\rho}_2] \leq S[\hat{\rho}_{12}] + S[\hat{\rho}_{23}] \quad (7.28)$$

Strong subadditivity is a very important property of the entropy that is not straightforwardly interpreted in terms of information theoretic concept. Still, it has been a crucial property in many proofs in quantum information and quantum communication [19] as well quantum entanglement. It also got a lot of attention in the earlier days of QSM. Strong subadditivity was already known to hold for the classical entropy and appeared in proofs establishing mean entropies for statistical lattice systems [92]. In fact, strong subadditivity comes up often in problems where we need to find scaling properties of the entropy, w.r.t. the underlying space or lattice coordinates, in quantum mechanics and field theories [93] and holography and AdS/CFT [94][95][96]. As an example of this, we will show that we can prove the existence of a mean entropy per unit length of a 1D translationally invariant system. Translationally invariant systems are systems for which the density may be defined on a region of space  $L$ ,  $\hat{\rho}_L$  such that a translation of this region does not change the density matrix. An example of such a system is the spin XY ground state, which we will treat in section 8.3.

First of all, the proof of strong subadditivity this depends crucially on the concavity of the functional  $S[\hat{\rho}_{12}] - S[\hat{\rho}_1]$  in  $\hat{\rho}_{12}$  [97]. Also, since the mapping  $\hat{\rho}_{123} \rightarrow \hat{\rho}_{12}$  is linear, which means that the above functional is concave in  $\hat{\rho}_{123}$  as well. Next we define

$$\Delta = (S[\hat{\rho}_{12}] - S[\hat{\rho}_1]) + (S[\hat{\rho}_{23}] - S[\hat{\rho}_2]), \quad (7.29)$$

which is thus concave in  $\hat{\rho}_{123}$ , as well, by construction. The minimum is clearly 0 for non entangled pure states, so together with the concavity we may state  $\Delta \geq 0$ . Equivalently we may then state

$$S[\hat{\rho}_1] + S[\hat{\rho}_2] - S[\hat{\rho}_{12}] - S[\hat{\rho}_{23}] \leq 0 \quad (7.30)$$

There always exists a pure state  $\hat{\rho}_{1234}$  such that  $\text{Tr}_4[\hat{\rho}_{1234}] = \hat{\rho}_{123}$ . We therefore have  $S[\hat{\rho}_{123}] = S[\hat{\rho}_4]$ , and also  $S[\hat{\rho}_{14}] = S[\hat{\rho}_{23}]$ , thus

$$S[\hat{\rho}_{123}] + S[\hat{\rho}_2] - S[\hat{\rho}_{12}] - S[\hat{\rho}_{23}] = S[\hat{\rho}_4] + S[\hat{\rho}_2] - S[\hat{\rho}_{12}] - S[\hat{\rho}_{14}] \leq 0, \quad (7.31)$$

where the inequality is obtained from equation 7.30. This establishes strong sub-additivity.

Next we will establish the mean entropy in translationally invariant systems. Consider the bounded regions  $V, V' \subset \mathbb{R}^d$  ( $\mathbb{Z}^d$ ), with Fock spaces  $\mathbb{F}(V), \mathbb{F}(V')$ , attached to it. If the regions have zero overlap,  $V \cup V' = \emptyset$ , we have that the Fock spaces separate  $\mathbb{F}(V \cup V') = \mathbb{F}(V) \otimes \mathbb{F}(V')$ . The density matrix of the region  $V$  is thus given by  $\text{Tr}_{V'}[\hat{\rho}]$ . We now aim to find if the limit  $\lim_{V \rightarrow \infty} S(V)/|V|$ , provided that the system is translationally invariant, meaning  $S[V + a] = S[V]$ . Due to the translational invariance, we may write  $S[V] = S(l)$ , where  $l = |V|$ . If we represent an interval with length  $l$  as the intersection of two intervals of length  $l'$ , thus also  $l \leq l'$ , SSA gives us the following inequality

$$S(l) \leq S(2l' - l) + S(l) \leq S(l') + S(l') \quad (7.32)$$

For clarity, the first inequality is obtained from the fact that  $S \geq 0$ , and for the second inequality we used SSA where we identified  $S[\hat{\rho}_{123}] \equiv S(2l' - l)$ ,  $S[\hat{\rho}_2] \equiv S(l)$  and  $S[\hat{\rho}_{12}] = S[\hat{\rho}_{23}] \equiv S(l')$ . Next define  $s = \inf_l \frac{S(l)}{l}$ , we see that

$$S(l) \leq 2S(l') - (2l' - l)s. \quad (7.33)$$

Next, define  $l'' = nl' + l$ , where  $n$  is an integer. By subadditivity we have that  $S(l'') \leq nS(l') + S(l)$ , and hence

$$s \leq \frac{S(l'')}{l''} \leq \frac{(2+n)S(l') - (2l' - l)s}{nl' + l}. \quad (7.34)$$

We may take the limit of  $n \rightarrow \infty$ , and thus simultaneously  $l'' \rightarrow \infty$  to get

$$s \leq \lim_{l'' \rightarrow \infty} \frac{S(l'')}{l''} \leq \frac{S(l')}{l'}. \quad (7.35)$$

By definition  $l' \leq l''$  and  $l' \geq l$ , so this means that, since we are free to choose  $l$ , we are free to choose  $l'$  as well. We may choose  $l'$  such that  $S(l') = l's$ . Thus we find that

$$\lim_{l \rightarrow \infty} \frac{S(l)}{l} = \inf_l \frac{S(l)}{l}. \quad (7.36)$$

In other words we have that in the thermodynamic limit, where we take  $l \rightarrow \infty$ , the mean entropy becomes the minimal value for any  $l$ .

#### 7.1.4 Entropy Maximization

In this section we derive the maximum entropy states for the Von Neumann entropy. In general, maximizing trace functionals is a very non trivial task, and often has to be done indirectly through inequalities. However, we will see that, due to a properties of the density matrix and the Von Neumann entropy, we may reduce the maximization problem to a smaller subset of density matrices, which effectively turns our problem into a standard optimization problem that may be solved using the Euler-Lagrange method.

### Deleting the Off-Diagonals

Let us investigate what happens to the entropy when we set the off diagonal terms of the density matrix to zero. We define the density matrix  $\hat{\rho}(\lambda) = \lambda\hat{\rho} + (1 - \lambda)\hat{\rho}^D$ , where  $\hat{\rho}_{ij}^D = \hat{\rho}_{ij}\delta_{ij}$ . We can then take the derivative of the entropy with respect to the parameter  $\lambda$ ,

$$\frac{d}{d\lambda}S[\hat{\rho}(\lambda)] = -\frac{d}{d\lambda}\text{Tr}[\hat{\rho}(\lambda)\ln\hat{\rho}(\lambda)] = -\text{Tr}\left[\dot{\hat{\rho}}(\lambda)\ln\hat{\rho}(\lambda)\right]. \quad (7.37)$$

This is proven by considering the following; Since  $\hat{\rho}(\lambda)$  and  $\dot{\hat{\rho}}(\lambda)$  commute we may write the derivative of the logarithm w.r.t.  $\lambda$  as follows  $\frac{d}{d\lambda}\ln\hat{\rho}(\lambda) = (\hat{\rho}(\lambda))^{-1}\dot{\hat{\rho}}(\lambda)$ . Then by the chain rule

$$\frac{d}{d\lambda}S[\hat{\rho}(\lambda)] = \text{Tr}\left[\dot{\hat{\rho}}(\lambda)\ln\hat{\rho}(\lambda) + \dot{\hat{\rho}}(\lambda)\right]. \quad (7.38)$$

Since the trace of  $\hat{\rho}(\lambda)$  is 1, for all  $\lambda$ , we have that  $\text{Tr}\left[\dot{\hat{\rho}}(\lambda)\right] = 0$ . Having established equation 7.37, we invoke the concavity property of the entropy in the following form

$$\frac{d^2}{d\lambda^2}S[\hat{\rho}(\lambda)] \leq 0. \quad (7.39)$$

Then since

$$\left.\frac{d}{d\lambda}S[\hat{\rho}(\lambda)]\right|_{\lambda=0} = -\text{Tr}\left[(\hat{\rho} - \hat{\rho}^D)\ln\hat{\rho}^D\right] = 0 \quad (7.40)$$

we see that the entropy is monotonically decreasing as  $\lambda$  goes from 1 to 0 and

$$S[\hat{\rho}(\lambda = 1)] \leq S[\hat{\rho}(\lambda = 0)]. \quad (7.41)$$

We conclude that, given a set of diagonal entries of some density matrix, the entropy always increases as the off-diagonal terms become smaller. We will use this result in section 10.3 to define an entropy quantity related to the Von Neumann entropy.

### The Microcanonical and Canonical States

We established in the preceding section that for a given set of diagonal values, the diagonal matrix has the largest entropy. We may therefore constrain our optimization problem to the set of diagonal density matrices. Using this, optimizing the Von Neumann entropy, given a single constraint of the form  $\text{Tr}\left[\hat{O}\hat{\rho}\right] = \langle\hat{O}\rangle$  becomes very simple. This is because we may perform the optimization problem in any basis. If we write the operator  $\hat{O}$  in diagonal form, only the diagonals contribute to the trace  $\text{Tr}\left[\hat{O}\hat{\rho}\right]$ . Thus writing the problem in this basis, we immediately see from the considerations of the previous section, that if we set all the off diagonals of  $\hat{\rho}$  to zero in this basis, it will have the maximum entropy given the expectation value. We have then effectively reduced our problem to maximizing the functional

$$-\sum\lambda_i\ln\lambda_i - \gamma\sum_i\lambda_i - \beta\sum_i o_i\lambda_i, \quad (7.42)$$

where  $\gamma$  and  $\beta$  are Lagrange multipliers and  $o_i$  are the eigenvalues of  $\hat{O}$ . This is just the classical BGS optimization problem to which we already found the answer in section 1.2.3,

$$\lambda_i = \frac{e^{-\beta o_i}}{\sum_i e^{-\beta o_i}}. \quad (7.43)$$

Without the constraint of the expectation value we find

$$\hat{\rho}_{\text{MaxEnt}} = \sum_i \frac{1}{N} |i\rangle\langle i|, \quad (7.44)$$



where  $|i\rangle$  form an orthonormal basis. If  $|i\rangle$  are energy eigenstates we recover the microcanonical ensemble. Given the constraint of an expectation value we find

$$\hat{\rho}_{\text{MaxEnt}} = \frac{1}{Z} \sum_i e^{-\beta o_i} |o_i\rangle \langle o_i| = e^{-\beta \hat{O}}, \quad (7.45)$$

where  $Z$  is the normalization constant. If the constraint is of the form  $\text{Tr} [\hat{H} \hat{\rho}] = E$ , we recover the canonical ensemble. This result nicely generalizes for multiple constraints when the operators commute. We then have, for the constraints  $\text{Tr} [\hat{O}_i \hat{\rho}] = \langle O_i \rangle$ , the maximum entropy state

$$\hat{\rho}_{\text{MaxEnt}} = \frac{1}{Z} e^{-\sum_i \beta_i \hat{O}_i}, \quad (7.46)$$

## 7.2 The Quantum Tsallis and Rényi entropies

We will briefly discuss the quantum generalizations of the Rényi and Tsallis entropies. Just like the Von Neumann entropy, they may be generalized in a straightforward way to density matrices. The quantum extension to the Rényi entropy is given by

$$R_\alpha[\hat{\rho}] = \frac{1}{1-\alpha} \ln \text{Tr} [\hat{\rho}^\alpha]. \quad (7.47)$$

and the quantum extension to the Tsallis entropy is given by

$$T_q[\hat{\rho}] = \frac{1}{q-1} (1 - \text{Tr} [\hat{\rho}^q]), \quad (7.48)$$

Like in the classical case, both reduce to the Von Neumann entropy for  $\alpha = q = 1$ . The Rényi as well as the Tsallis entropies may be checked to be Schur-concave for  $\alpha, q > 0$  and Schur-convex for  $\alpha, q < 0$ . As such, these measures are also often considered as measures of the mixedness/purity of the system. They are often used because of their much simpler dependence on the density matrix. In particular, take the cases where  $\alpha = 2$  and  $q = 2$ . The Rényi and Tsallis entropies take the simple form

$$R_2[\hat{\rho}] = -\ln \text{Tr} [\hat{\rho}^2], \quad T_2[\hat{\rho}] = 1 - \text{Tr} [\hat{\rho}^2]. \quad (7.49)$$

To evaluate either of these, no diagonalization of the density matrix is required, so in general these may be more readily computed. The Tsallis entropy with  $q = 2$  is again called the linear entropy, like we discussed in section 3.4.1.

However, the additivity properties differ of the Rényi and Tsallis entropies, as we have seen in chapter 3. Since the Rényi entropy is additive and thus possesses a similar dependence on the entropy of a product state. The Rényi entropy is in fact often considered alongside the Von Neumann entropy in cases where the interest lies in the scaling of the entropy which arises due to entanglement [98][99]. The similar scaling behavior is also shown in equations 8.68 and 8.69. Moreover, since it is a continuous generalization of the Von Neumann entropy, it may be used to calculate the Von Neumann entropy. We will show this in the following chapter. The Tsallis entropy has been used, among other things, in the context of non-extensive QSM (analog to non-extensive statistical mechanics [36], which we discussed in chapter 3, as a measure of mixedness/purity of a quantum system [100][101] and in the quantification of quantum coherence [102]. An application of the Tsallis entropy as a measure of mixedness will be explored in section 9.2.

## 7.3 Summary and Concluding Remarks

In this chapter, we have seen the quantum generalizations of the most common information theoretic measures, the BGS entropy and the Rényi and Tsallis entropies. First, we discussed the Von

Neumann entropy, which is the generalization of the BGS entropy. All the quantum generalizations of these measures are based on the entropy of the set of eigenvalues of the density matrix. However, they also have a functional form based on the trace operation. From their definition, we see that they are intimately tied to the density matrix formalism of quantum mechanics, which allows us to encode classical distributions in quantum theories. Since the eigenvalues encode the classical probability of our state to be in a particular pure state, we see that the entropies may be interpreted as measures for the classical uncertainty we have about the pure state our system is in<sup>3</sup>. Formally, this may be understood through the majorization relation we introduced in section 4.5, of which all the entropies we discussed are monotonic functions. In a sense, these entropies may thus be understood in the same way that we understood the classical entropies; through the relation with the eigenvalues of the density matrix. At the same time, from their functional form, we see that there is actually an intricate dependence on the density matrix and the underlying Hilbert space. Due to this, the properties of the entropies with respect to the probability distributions do not trivially generalize to density matrices. The fact that strong-subadditivity property generalizes to the Von Neumann entropy is even proclaimed a miracle by Witten [19], and it does not generalize to density matrices at all for the Tsallis entropy [103].

In discussing the Von Neumann entropy in particular, we have first shown how it can be derived from thermodynamical considerations. The main takeaway here is that when every particle of a system may be described as a mixture of states, then the Von Neumann entropy can be thought of as a thermodynamical entropy. Note that such a description is very similar to the statistical description of classical, ergodic systems in the thermodynamic limit, where all the sub-systems become approximately uncorrelated, and the entropy becomes extensive. In this case, we may ascribe to every subsystem an average entropy and, thus, an effective space of states. After that, we have discussed the conventional assumptions of quantum statistical mechanics, and where the Von Neumann entropy fits in. In the last part of that section, we discussed the properties of the Von Neumann entropy, and how they can be derived from the properties of the Shannon entropy. We also showed a trick that helped us to turn the maximization problem of the Von Neumann entropy over all density matrices, into a conventional maximization problem. The catch of this trick was that this only works when all operators, which were used for the constraints, must be simultaneously diagonalizable.

At the end of this chapter, we briefly discussed the generalizations of the Rényi and Tsallis entropies. For these quantum generalizations, the general rule is that we substitute the sum for the trace operation and the probabilities  $p_i$  for the density matrix  $\hat{\rho}$ . Moreover, both entropies are monotonic functions of the majorization relation, and may therefore be regarded as measures of mixedness, in the same way as the von Neumann entropy.

---

<sup>3</sup>In light of the next chapter, we remark the following: For this interpretation it does not matter if the mixture is created due to entanglement, as long as we only have access to and knowledge of one of the subsystems.

## Chapter 8

# Entanglement and the Entanglement Entropy

Entanglement is one of the most crucial properties of quantum theory and has been utilised in many fields of research, such as quantum information, quantum cryptography, quantum communication and quantum computing. Entanglement has also found a strong link with entropies, in multiple fields of research. With this chapter, we aim to give an overview of some of the uses of entropies when it comes to the study of entanglement.

First of all, we will look at classifying and quantifying entanglement. We have already seen the classification using the Schmidt rank. Moreover, historically, Bell's inequality is also used as a classification of entanglement, where states that violate Bell's inequality are considered entangled. From this definition we see that entanglement is not an all-or-nothing proposition. Of the two-qubit states we know that  $\frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle)$  and  $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$  violate Bell's inequality maximally, but there is a whole class of states that violate Bell's inequality, just not necessarily maximally. In more recent years a new quantitative classification of bi-partite<sup>1</sup> entanglement have been developed, commonly called the Local Operations and Classical Communication, or LOCC, paradigm. The idea behind this classification is that by agreeing upon what operations should not create entanglement, we can create an ordering on all quantum states. The link with entropy, comes from Nielsen's theorem, which we will define in the next sections. Qualitatively it states that the ordering of entangled states through LOCC is equivalent to the ordering of the partial traced states through majorization. Thus, we immediately find that all the Schur-concave or -convex quantities may be used as measures of entanglement. However, as we will make clear as well, the Von Neumann entropy actually has a special operational meaning within the paradigm.

In the first sections of this chapter, we will give an overview of the LOCC characterization of bi-partite entanglement. After that we will focus on the calculation of the Von Neumann entanglement entropy for the ground states of some specific systems. In these calculations, we will see a very particular behaviour of the entropy for some systems, namely the scaling of the entropy with the boundary of the entanglement region, instead of the volume. Such behavior is called an area law, and has sparked a lot of the interest due to a suggested connection to the famous Bekenstein-Hawking entropy of black holes, which scales along with the area of the horizon of a black hole. Since the calculation of the entanglement entropy is in general very difficult and unsolved for many systems, we focus on a few simpler systems, namely a particular spin chain model, called the spin XY model, and the free scalar field. The spin XY model is a well known example of a model with critical behaviour in certain parameter regions. What this means is that for different parameters on which the Hamiltonian depends, we will see discontinuous behavior for specific lines or points of the

---

<sup>1</sup>Bi-partite is a term often used quantum communication and means a system of two parts. This thus does not mean necessarily that the system is divided into particles, but generally into two Hilbert spaces.

parameter space. For this model, we have that the correlation length goes from exponential decay in the non-critical regions, to a power law decay in the critical regions. We will see that the behaviour of the entanglement entropy changes drastically as well. After that, we focus on calculations in field theory, and in particular the ground state of the free scalar field. We will show two different methods of calculation. First, we show the chronologically first method of calculation, which sparked the debate about the connection to the black hole entropy. In this method we first calculate the entanglement entropy in quantum mechanics, and then continue the results to quantum field theory. We conclude this method by showing the numerical results for a spherical volume, that show the area law. Lastly, we present a more novel way to calculate the entanglement entropy, using the Euclidean path integral. We will first discuss the general features of how to execute calculations in the Euclidean method and then apply this to the real scalar field. This will again reveal an area law.

## 8.1 Quantifying Entanglement: The LOCC Paradigm

In the following sections we will review a characterization of bi-partite entanglement in terms of the aforementioned LOCC operations, commonly called the LOCC paradigm or alternatively the distant lab paradigm. LOCC stands for local operations and classical communication, and in practice, LOCC operations include all unitary operations that can be done on a subsystem of an entangled system without access to the other subsystem<sup>2</sup>, and all measurements done on one of the subsystems. Moreover, we may communicate classically with someone in another lab who can perform actions on the other subsystem. We can for example discuss the measurement outcomes on a subsystem and use that knowledge to decide what other operations to perform. One of the most well known protocols which uses LOCC, is teleportation protocol [104]. The protocol roughly goes as follows: We have a state  $a|\uparrow\rangle + b|\downarrow\rangle$  in one lab. By coupling it to a maximally entangled pair, sending the second particle to the other lab and performing only LOCC operations, we can deterministically transform the state in the second lab to the state  $a|\uparrow\rangle + b|\downarrow\rangle$ . At the end of the protocol, the original state is collapsed, and the total state, containing three particles, is completely separable, meaning that all entanglement is destroyed. Later on we will give a full example of a similar LOCC protocol. There have been many more similar protocols, in which we, like for the quantum teleportation protocol, use an entangled state to perform operations we otherwise couldn't. In the process the entanglement is diminished or destroyed. As such, in the LOCC paradigm, entanglement is considered as a resource [105]–[107].

Setting up a mathematical definition of LOCC operations turns out to be more difficult, though what can be said is that the LOCC operations are a subset of all separable quantum operations [108]. The Kraus operators offer a convenient representations of LOCC protocols, and always take the form of separable operations, e.g. operations of the form  $K_i = K_i^A \otimes K_i^B$ . The intuition behind the use of LOCC as a characterization of entanglement, is that entanglement should not increase between two systems from operations that can be performed locally and from any classical communication. In what follows we will show that this allows us to view entanglement as a sort of resource, which is used up by LOCC operations. By identifying the maximally entangled state, as well as the minimally entangled state, this allows us to introduce an ordering in all bi-partite quantum states.

For pure states there is a clear interpretation for the entropies in the LOCC paradigm, using Nielsen's theorem [105, Thm. 1]. Let  $|\psi\rangle$  and  $|\phi\rangle$  be two bi-partite states with respective density matrices  $\hat{\rho}_\psi$  and  $\hat{\rho}_\phi$ . Nielsen's theorem states

$$\exists \Phi[|\psi\rangle] = |\phi\rangle, \text{ iff } \text{Tr}_{A/B}[\hat{\rho}_\psi] \succ \text{Tr}_{A/B}[\hat{\rho}_\phi], \quad (8.1)$$

where  $\Phi$  represents a quantum operation. In other words there exists a LOCC map  $\Phi[|\psi\rangle] = |\phi\rangle$  if and only if the marginal density matrix  $\text{Tr}_{A/B}[\hat{\rho}_\psi]$  majorizes  $\text{Tr}_{A/B}[\hat{\rho}_\phi]$ . From the Schur concavity it

<sup>2</sup>This includes adding a particle and then performing some unitary evolution on both the system and the added particle, thereafter discarding the extra particle. In the field such a particle is commonly called an ancilla particle, or just ancilla. Of course, such operations are most easily represented using Kraus operators.

then follows that a LOCC map from  $\hat{\rho}$  to  $\hat{\sigma}$  only exists if and only if  $S[\text{Tr}_{A/B} [\hat{\rho}_\psi]] < S[\text{Tr}_{A/B} [\hat{\rho}_\phi]]$ , or equivalently any other Schur concave function.

### 8.1.1 Maximally and Minimally Entangled States

In order to set up an ordering of all quantum states in terms of LOCC operations, it is instructive to define the extremal points of the ordering: the minimally and maximally mixed states. Though the actual ordering is something that might be dependent on definition, the extremal points should be something that everyone can agree upon. First of all the minimally entangled states are all the separable states. Separable states are states that can be written as follows

$$\hat{\rho}_{12\dots} = \sum_i p_i \hat{\rho}_A^i \otimes \hat{\rho}_B^i \otimes \dots \quad (8.2)$$

Though the state is mixed, it only represents classical correlations between the subsystems. LOCC operations always map these states onto themselves. The procedure to create this state from scratch may go as follows. The first party, let's call her Alice, performs a measurement on a state, that gives the local state  $\hat{\rho}_A^i$  with probability  $p_i$ . Next, she informs all other parties to prepare the respective states  $\hat{\rho}_X^i$ . Finally, we "forget", or in other words average over all the possibilities that Alice could've gotten for her initial measurement. In this sense the states of all parties are only correlated by the classical communication they had.

Following Nielsen's theorem we define the maximally entangled state, as the state that may be used to create any other state using only LOCC operations. For two d-dimensional subsystems the maximally entangled state is any state that is locally unitarily equivalent to

$$|\psi_{\max}\rangle = \frac{1}{\sqrt{d}} (|00\rangle + |11\rangle + \dots + |d-1d-1\rangle) \quad (8.3)$$

In the following section we show an example of how we can manipulate such a state to give an arbitrary state with the same dimensionality.

### 8.1.2 Locally Manipulating Two Qubits

We will now show a LOCC protocol in which we can manipulate a maximally entangled state to be any other bi-partite state. We share a maximally entangled state between two labs,  $A$  and  $B$ ,

$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B) \quad (8.4)$$

We will show that using this state, we can construct any other bi-partite state of the form

$$\alpha |0\rangle_A |0\rangle_B + \beta |1\rangle_A |1\rangle_B \quad (8.5)$$

using only LOCC operations. We first add an ancilla in the state  $|0\rangle$  to the first system, obtaining the state

$$\frac{1}{\sqrt{2}} (|00\rangle_A |1\rangle_B + |01\rangle_A |1\rangle_B). \quad (8.6)$$

Using the following unitary operations on the first system

$$|00\rangle \rightarrow \alpha |00\rangle + \beta |11\rangle, \quad (8.7)$$

$$|01\rangle \rightarrow \beta |01\rangle + \alpha |10\rangle, \quad (8.8)$$

to obtain the state

$$\frac{|0\rangle_A (\alpha |0\rangle_A |0\rangle_B + \beta |1\rangle_A |1\rangle_B) + |1\rangle_A (\beta |0\rangle_A |1\rangle_B + \alpha |1\rangle_A |0\rangle_B)}{\sqrt{2}} \quad (8.9)$$

A measurement may now be done on the ancilla particle, resulting in two different outcomes. If the outcome is  $|0\rangle$ , the desired state is obtained. If the outcome is  $|1\rangle$ , however, all we need to do is apply  $\sigma_x$  to the second particle, and the desired state is obtained as well. Communicating the result of the measurement of the ancilla, to then know how to obtain the desired state is in this example the classical communication part of LOCC.

By Nielsen's theorem, a conversion back to the initial state using LOCC is not possible since the final state majorizes the initial state for any  $|\alpha|^2 \neq |\beta|^2 \neq 1/2$ . It follows that the amount entanglement entropy of the final state is diminished as well. This of course happened during the measurement of the ancilla particle, since all other transformations were unitary.

It can be checked that this LOCC protocol may also directly be represented by the Kraus operators

$$K_0 = (\alpha |0\rangle \langle 0| + \beta |1\rangle \langle 1|) \otimes \mathbb{1} \quad (8.10)$$

and

$$K_1 = (\beta |1\rangle \langle 0| + \alpha |0\rangle \langle 1|) \otimes (|1\rangle \langle 0| |0\rangle \langle 1|). \quad (8.11)$$

They may be checked to satisfy the completeness relation  $K_0^\dagger K_0 + K_1^\dagger K_1 = \mathbb{1}$ .

### 8.1.3 The Distillable Entanglement and Entanglement Cost

We have seen the maximally entangled state and that using LOCC operations, entanglement can only diminish. We've also seen that entropies may be used to quantify relative entanglement. We have also seen that there exists a maximally entangled state, which may be viewed as the most valuable resource, i.e. the resource with which the most operations may be done. A reasonable question then is if and how we can compare the entanglement of a non-maximally entangled state with that of the maximally entangled state. Consider a non-maximally state  $|\phi\rangle$ , it turns out that if we consider a block of such states  $|\phi\rangle^{\otimes n}$ , then there always exists a conversion from this block of states to a smaller block of maximally entangled 2-qubit states,  $|\psi\rangle^{\otimes m}$  [106]. In a sense, we can thus "distill" the entanglement of a set of these  $n$  states, into a set of  $m$  maximally entangled states. The best achievable ratio  $r = m/n$  turns out to be dependent of  $n$  and achieves its maximal value when we let  $n \rightarrow \infty$ . In this limit two important and much researched quantities are defined, namely the entanglement cost and distillable entanglement. Denoting the density matrix of a block of  $|\psi\rangle$  states as  $\hat{\rho}^{\otimes n}$ , the LOCC map as  $\Phi(\cdot)$  and the maximally entangled state as  $\hat{\rho}_*$ , we define the entanglement cost<sup>3</sup>

$$E_C[\hat{\rho}] \equiv \inf \left\{ r : \lim_{n \rightarrow \infty} \left[ \inf_{\Phi} \text{Tr} [|\hat{\rho}^{\otimes n} - \Phi(\hat{\rho}_M^{\otimes rn})|] \right] = 0 \right\} \quad (8.12)$$

The distillable entanglement is defined as

$$E_D[\hat{\rho}] \equiv \sup \left\{ r : \lim_{n \rightarrow \infty} \left[ \inf_{\Phi} \text{Tr} [|\Phi(\hat{\rho}^{\otimes n}) - \hat{\rho}_M^{\otimes rn}|] \right] = 0 \right\}. \quad (8.13)$$

The entanglement cost is thus the smallest ratio of maximally entangled 2-qubit states to  $|\psi\rangle$ , such that we can create a block of  $\psi$  states from a block of maximally entangled states, using only LOCC in the asymptotic limit. Conversely the distillable entanglement is the largest ratio of maximally entangled 2-qubit states to  $|\psi\rangle$  states, such that we can create a block of maximally entangled 2-qubit states from a block of  $|\psi\rangle$  states, using only LOCC in the asymptotic limit. It turns out that for pure states, these two quantities are the same, and are in fact equal to the Von Neumann entropy of one of the subsystems, when using the base 2 logarithm,  $E_C = E_D = S[\hat{\rho}]$  [106]. So even though all Schur-concave/convex measures may be used to quantify relative entanglement, the Von Neumann entropy may be specifically interpreted, for pure states, as the entanglement cost or distillable entanglement.

<sup>3</sup>In literature sometimes other distance measures are used instead of the trace norm, but  $r$  is independent of the measure used as long as they are sufficiently similar to the trace norm distance. [109, §3]

### 8.1.4 Beyond Pure States

The above considerations becomes more difficult in the case of entanglement content of mixed states however, and the measures like the entanglement cost, distillable entanglement and Von Neumann entropy cease to be equivalent. It turns out that a total ordering of all states in terms of LOCC conversion is lost [110]. Moreover, the Von Neumann entropy ceases to be a good measure of entanglement, since it does not differentiate between the mixedness that is there in the initial state and the mixedness that is created by the entanglement. A way to solve this is by taking an axiomatic approach to entanglement measures, very similar to the axiomatic approaches to information measures in information theory. The aim is then to find a set of axioms that a “reasonable” measure for entanglement should quantify. An example of a set of axioms that may be used is given by Plenio *et al.* [110]

- **E1:** A bi-partite entanglement measure  $E[\hat{\rho}]$ , maps the density matrices onto the positive real numbers

$$E[\hat{\rho}] : \hat{\rho} \rightarrow \mathbb{R}^+ \quad (8.14)$$

- **E2:** The entanglement measure is zero if  $\hat{\rho}$  is separable,  $E[\hat{\rho}] = 0$
- **E3:** The entanglement measure does not increase under LOCC on average

$$E[\hat{\rho}] \geq \sum_i p_i E \left[ \frac{\hat{K}_i \hat{\rho} \hat{K}_i^\dagger}{\text{Tr} [\hat{K}_i \hat{\rho} \hat{K}_i^\dagger]} \right] \quad (8.15)$$

where  $K_i$  are the Kraus operators representing the LOCC protocol and  $p_i = \text{Tr} [\hat{K}_i \hat{\rho} \hat{K}_i^\dagger]$ .

- **E4:** If  $\hat{\rho}$  is pure,  $E$  is the Von Neumann entropy of entanglement.

Quantities that follow from these set of axioms are reviewed by Plenio *et al.* [110].

## 8.2 The Entanglement Entropy and Area Laws

In the following sections we will review aspects of the Von Neumann entanglement entropy, most often just referred to in literature as the entanglement entropy. As we have seen in the last sections, the Von Neumann entropy applied to the squares of the Schmidt coefficients gives a suitable measure of entanglement. The entanglement entropy shouldn't a-priori be thought of as a thermodynamical entropy, and more along the lines of an information theoretical entropy, though there are reasons to believe it may be connected to the well known black hole entropy. This is due the so called area law behavior that we have mentioned.

Let us partition some lattice or space into two regions,  $A$  and its complement  $B$ . The entropy is said to obey an area law if it scales with the size of the boundary  $|\partial A|$ . A substantial part of the interest comes from a suggested link to the famous Bekenstein-Hawking formula for the black hole entropy, which originate from papers by Bombelli [111] and Srednicki [112]. The Bekenstein-Hawking (BH) black hole entropy is given by

$$S_{BH} = \frac{k_B \text{Area}(\text{horizon})}{4L_p^2}, \quad (8.16)$$

where  $L_p$  is the Planck length,  $G\hbar/c^3$ , with Newton's gravitational constant  $G$ . The reason for this suggested correspondence is the peculiar scaling w.r.t. the properties of the black hole, that the BH entropy possesses. Indeed, the entropy does not depend extensively on the size of the black hole, but instead on the surface area. While black holes are of course still sources of many mysteries, this goes against classical expectations, where the entropy always scales with the amount of micro-states, which as we have seen is normally associated with the size of the total system<sup>4</sup>.

<sup>4</sup>How we define the size depends on the dimensionality of the system. Typically we associate a length with a 1D

However, in quantum mechanics, ground states of particular systems do in fact show the same behaviour for their entanglement entropy. This aspect of the entanglement entropy, which we will show later, may be a reason to believe that this entropy in fact originates from quantum properties of the black hole. In fact, there is a belief that the BH entropy may be the first order of the actual quantum mechanical entropy, to which quantum corrections apply.

In addition, interest in area laws has spread to other fields of research as well. Among other things, area laws have been studied in the context of quantum many-body systems, like spin lattice systems, specifically with regards to its relation to critical phenomena [114]. We will investigate this aspect of the entanglement entropy in the following section. Moreover, the area laws could be an indicator of complexity of simulation of many body systems. If a many body system obeys an area law, the state may be easier to approximate in numerical simulations. A theorem like this has been proven already for the Rényi entropies for  $\alpha \neq 0$  [115][116, Thm. 18]. Lastly, area laws have been studied in the intersection of topology and quantum theory as a measure of topological order, characterizing quantum phases that are not distinguishable by their symmetries [117]. For an overview the paper by Eisert *et al.* [116] is recommended.

In the following sections we will derive the ground state entanglement entropy in the context of XY spin lattice model, as well as for the free scalar field.

### 8.3 The Spin XY Model and the Ground State Entropy

Calculating the entanglement entropy for states of infinitely large systems is a notoriously difficult task and as such, systems that have been solved analytically are not easy to come by. However, spin lattice XY model is one of such a system, which we will use as an example to show how the ground state entanglement entropy, between a region of the lattice  $L$  and the rest of the lattice may be derived. In this model we deal with  $N$  spins, each at lattice site  $i$ . The spin XY model is characterized by the following Hamiltonian

$$H_{XY} = -\frac{1}{2} \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \left( \frac{1+\gamma}{2} \sigma_i^x \sigma_{i+1}^x + \frac{1-\gamma}{2} \sigma_i^y \sigma_{i+1}^y \right) - \frac{\lambda}{2} \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sigma_i^z. \quad (8.17)$$

where  $\sigma_i^{(\cdot)}$  are the Pauli matrices acting on the spin at lattice site  $i$ . We see that there is a nearest neighbor interaction, in the  $xy$  plane, between the spins and a coupling of the spin in the  $z$  direction to an external field. The strength of this interaction is characterized by  $\lambda$ . The energy spectrum of this Hamiltonian, in the limit of large  $N$ , is given by

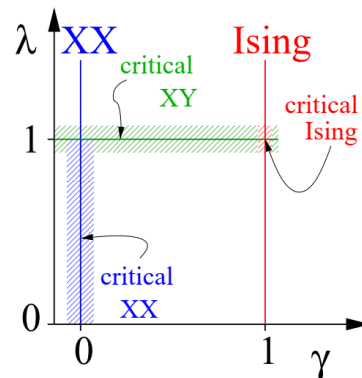
$$\Lambda_\phi = \sqrt{(\lambda - \cos \phi)^2 + \gamma^2 \sin^2 \phi} \quad (8.18)$$

where  $\phi \in [-\pi, \pi]$  is a momentum label. Moreover, the ground state energy is set to zero. The spectrum for different values of  $\lambda$  and  $\gamma$  is shown in figure 8.2 [114]. The spin XY model is actually a well know model for which the ground state exhibits phase transitions for particular values of the

---

system, an area with a  $2D$  system etc. If we see the black hole as a  $3D$  system, then the characteristic size of the system would be the volume. However, there may be a sense in which we may ignore the thermodynamically frozen part of the system, and redefine the thermodynamic system to have an amount of degrees of freedom that scales like the surface[113].

Figure 8.1: The critical regions of the spin XY model in  $\gamma$  and  $\lambda$  parameter space. The image is taken from Latorre *et al.* [114]





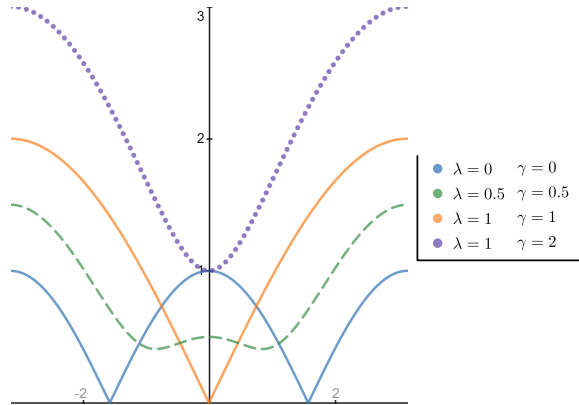


Figure 8.2: The spectrum of the XY Hamiltonian for different values of  $\lambda$  and  $\gamma$ .

parameters  $\lambda$  and  $\gamma$ . The critical regions of the model are shown in figure 8.1. These phase transitions may be characterized by the scaling of two characteristics of the model. We have correlation length

$$\langle \sigma_i \sigma_{i+L} \rangle - \langle \sigma_i \rangle \langle \sigma_{i+L} \rangle \sim e^{-\frac{L}{\xi}} \quad (8.19)$$

and the energy gap

$$\Delta = \Lambda_0. \quad (8.20)$$

Near the critical regions these quantities scale like

$$\xi \sim |\lambda - \lambda_c|^{-\nu} \quad (8.21)$$

$$\Delta \sim |\lambda - \lambda_c|^s. \quad (8.22)$$

In the critical region the correlation function suddenly scales like a power law

$$\langle \sigma_i \sigma_{i+L} \rangle - \langle \sigma_i \rangle \langle \sigma_{i+L} \rangle \sim L^{-q} \quad (8.23)$$

Before we go on with the calculation of the entanglement entropy, we will give a brief overview of the procedure. We will first re-write the Hamiltonian in terms of Majorana operators,  $\check{a}_i$ , where we have two for every lattice site. Using these operators, we can define the so called correlation matrix  $\langle \check{a}_i \check{a}_j \rangle$ . This correlation matrix is important, because for ground state entropies, we can calculate the entanglement entropy of a subset of  $L$  spins from it, by diagonalizing “only” a  $2L \times 2L$  block from it. In contrast, if we were to calculate the entanglement entropy from the density matrix, we would need to diagonalize a  $2^L \times 2^L$  matrix. To calculate this correlation matrix, we diagonalize the Hamiltonian in terms of transformed Majorana operators  $\check{b}$ . Using those operators, the correlation matrix has a very simple form. However, the block we need to calculate the entanglement entropy is only well defined in terms of the indices of  $\check{a}$ . Using the simple expression of the correlation matrix in terms of  $\check{b}$ , the correlation matrix in terms of  $\check{a}$  can be calculated. We then take the block of  $2L \times 2L$ , which we must diagonalize on its own. Using this, the entanglement entropy may basically be read off.

We will now write the Hamiltonian in a more convenient form using the Majorana operators. We may introduce two operators for each lattice site  $i$ <sup>5</sup>,

$$\check{a}_{2i-1} \equiv \left( \prod_{j<i} \sigma_j^z \right) \sigma_i^x, \text{ and } \check{a}_{2i} \equiv \left( \prod_{j<i} \sigma_j^z \right) \sigma_i^y. \quad (8.24)$$

<sup>5</sup>An alternative approach, which will be familiar to anyone who has studied spin systems is taking the Jordan-Wigner transform of the spin operators, followed by a Fourier transform. The end product is the same, which is a diagonalized Hamiltonian.

The Pauli matrices are of course Hermitian so we immediately have  $\check{a}_i^\dagger = \check{a}_i$ . Moreover they obey the anti-commutation relations. Let for convenience  $i < j$

$$\begin{aligned} \{\check{a}_i, \check{a}_j\} &= \prod_{k<i} \sigma_k^z \sigma_i^\mu \prod_{k'<j} \sigma_{k'}^x \sigma_j^\nu + \prod_{k<j} \sigma_k^z \sigma_j^\nu \prod_{k<i} \sigma_k^z \sigma_i^\mu = \sigma_{i-1}^z \dots \sigma_{j-1}^z (\sigma_i^\mu \sigma_j^\nu - \sigma_j^\nu \sigma_i^\mu) \\ &= \begin{cases} 0, & i \neq j \\ 2, & i = j \end{cases} = 2\delta_{ij}. \end{aligned} \quad (8.25)$$

for  $\mu, \nu = x, y$ . This is easily seen from  $\{\sigma_i^\mu, \sigma_i^\nu\} = 2\delta_{\mu\nu}$ . We now have the following,

$$\check{a}_{2i} \check{a}_{2i+1} = \sigma_i^y \sigma_i^z \sigma_{i+1}^x = i \sigma_i^x \sigma_{i+1}^x \quad (8.26)$$

$$\check{a}_{2i-1} \check{a}_{2i+2} = \sigma_i^x \sigma_i^z \sigma_{i+1}^y = -i \sigma_i^y \sigma_{i+1}^y \quad (8.27)$$

$$\check{a}_{2i-1} \check{a}_{2i} = \sigma_i^x \sigma_i^y = i \sigma_i^z. \quad (8.28)$$

Using this, we re-write the Hamiltonian as

$$H_{XY} = \frac{i}{2} \sum_i (c_+ 2\check{a}_{2i} \check{a}_{2i+1} - c_- \check{a}_{2i-1} \check{a}_{2i+2} + \lambda \check{a}_{2i-1} \check{a}_{2i}), \quad (8.29)$$

where  $c_\pm = \frac{1 \pm \lambda}{2}$ . Due to the anti-commutation relation we may write the Hamiltonian in a convenient matrix form

$$\begin{aligned} H_{XY} &= \frac{i}{4} \sum_i (c_+ (\check{a}_{2i} \check{a}_{2i+1} - \check{a}_{2i+1} \check{a}_{2i}) + c_- (\check{a}_{2i-1} \check{a}_{2i+2} - \check{a}_{2i+2} \check{a}_{2i-1}) + \lambda (\check{a}_{2i-1} \check{a}_{2i} - \check{a}_{2i} \check{a}_{2i-1})) \\ &= \frac{i}{4} \sum_{i,j} A_{ij} \check{a}_i \check{a}_j, \end{aligned} \quad (8.30)$$

where  $A$  is a real, skew-symmetric matrix

$$\begin{bmatrix} A_0 & A_1 & \cdots & A_N \\ -A_1^T & A_0 & A_1 & \\ \vdots & & \ddots & \\ A_N^T & -A_1^T & A_0 & A_1 \\ & -A_1^T & A_0 & \end{bmatrix}, \quad (8.31)$$

with

$$A_0 = \begin{bmatrix} 0 & \lambda \\ -\lambda & 0 \end{bmatrix}, \text{ and } A_1 = \begin{bmatrix} 0 & -c_- \\ c_+ & 0 \end{bmatrix}. \quad (8.32)$$

$A$  is a skew-symmetric matrix, and thus allows for a block diagonal form  $B$ , such that  $B = WAW^T$ , where  $W$  is a special orthogonal matrix and

$$B = \oplus_k \left( \Lambda_k \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right). \quad (8.33)$$

The Hamiltonian may then be written as

$$H_{XY} = \frac{i}{4} \sum_{ijkl} \check{a}_k W_{ik} B_{ij} W_{jl} \check{a}_l \equiv \frac{i}{4} \sum_{ij} B_{ij} \check{b}_i \check{b}_j, \quad (8.34)$$

where we defined new Majorana operators  $\check{b}_i = \sum_j W_{ij} \check{a}_j$ . Since  $W$  is a special orthogonal matrix, it preserves the anti-commutation relations

$$\check{b}_i^\dagger = \check{b}_i, \text{ and } \{\check{b}_i, \check{b}_j\} = 2\delta_{ij}. \quad (8.35)$$

We will now briefly switch back to Fermionic operators, as it then should become clear how the ground state of the Hamiltonian should behave. We define  $N$  Fermionic operators

$$\hat{b}_i \equiv \frac{\check{b}_{2i-1} + i\check{b}_{2i}}{2} \quad (8.36)$$

We check the anti-commutation relations

$$\{\hat{b}_i^\dagger, \hat{b}_j\} = \frac{1}{4} (\{\check{b}_{2i-1}, \check{b}_{2j-1}\} - i\{\check{b}_{2i}, \check{b}_{2j-1}\} + i\{\check{b}_{2i-1}, \check{b}_{2j}\} + \{\check{b}_{2i}, \check{b}_{2j}\}) = \frac{1}{4} (2\delta_{ij} + 0 + 0 + 2\delta_{ij}) = \delta_{ij}, \quad (8.37)$$

and

$$\{\hat{b}_i, \hat{b}_j\} = \frac{1}{4} (\{\check{b}_{2i-1}, \check{b}_{2j-1}\} + i\{\check{b}_{2i}, \check{b}_{2j-1}\} + i\{\check{b}_{2i-1}, \check{b}_{2j}\} - \{\check{b}_{2i}, \check{b}_{2j}\}) = \frac{1}{4} (2\delta_{ij} + 0 + 0 - 2\delta_{ij}) = 0, \quad (8.38)$$

which are the usual Fermionic relations. We can now write the Hamiltonian in diagonal form, using the inverse transformations

$$\check{b}_{2i-1} = \hat{b}_i + \hat{b}_i^\dagger \quad (8.39)$$

$$\check{b}_{2i} = -i(\hat{b}_i - \hat{b}_i^\dagger) \quad (8.40)$$

$$\begin{aligned} H_{XY} &= \frac{i}{4} \sum_i \Lambda_i (\check{b}_{2i-1}\check{b}_{2i} - \check{b}_{2i}\check{b}_{2i-1}) = \frac{i}{4} \sum_i 2(\check{b}_{2i-1}\check{b}_{2i}) = \frac{1}{2} \sum_i \Lambda_i \left( (\hat{b}_i)^2 + \hat{b}_i^\dagger \hat{b}_i - \hat{b}_i \hat{b}_i^\dagger - (\hat{b}_i^\dagger)^2 \right) \\ &= \sum_i \Lambda_i \hat{b}_i^\dagger \hat{b}_i - 1/2 \end{aligned} \quad (8.41)$$

This was obtained using the anti-commutation relations, from which is seen that  $(\hat{b}_i)^2 = 0$ . We may ignore the constant contribution of  $-1/2$ . In this representation, we see that the Hamiltonian is diagonal in the Fermionic operators, so it is clear that the ground state should be annihilated by all  $\hat{b}_i$

$$\hat{b}_i |\Psi_g\rangle = 0 \quad (8.42)$$

and that

$$\hat{b}_i \hat{b}_i^\dagger |\Psi_g\rangle = \left(1 - \hat{b}_i^\dagger \hat{b}_i\right) |\Psi_g\rangle = |\Psi_g\rangle. \quad (8.43)$$

For these systems, the ground state has the particular behavior, that all the expectation values of products of  $\hat{b}$  and  $\hat{b}^\dagger$ , can be written using only sums of  $\langle \hat{b}_i \hat{b}_j \rangle$  and  $\langle \hat{b}_i \hat{b}_j^\dagger \rangle$ . This may be seen from the anti-commutation relations and the rules stated above, but is more generally expressed in Wick's theorem. Such states are called Gaussian states. Denoting the ground state expectation values by  $\langle \hat{O} \rangle_g$ , we may express all ground state expectation values using

$$\langle \hat{b}_i \hat{b}_i^\dagger \rangle_g = \frac{1}{4} \left( \langle (\check{b}_{2i-1})^2 \rangle_g + i \langle \check{b}_{2i} \check{b}_{2i-1} \rangle_g - i \langle \check{b}_{2i-1} \check{b}_{2i} \rangle_g + \langle (\check{b}_{2i})^2 \rangle_g \right) \quad (8.44)$$

$$= \frac{1}{4} \left( 2 + i \langle \check{b}_{2i} \check{b}_{2i-1} \rangle_g - i \langle \check{b}_{2i-1} \check{b}_{2i} \rangle_g \right) = 1 \quad (8.45)$$

$$\implies \langle \check{b}_{2i} \check{b}_{2i-1} \rangle_g - \langle \check{b}_{2i-1} \check{b}_{2i} \rangle_g = -2i. \quad (8.46)$$

Since the Majorana operators anti-commute we must have that

$$\langle \check{b}_{2i-1} \check{b}_{2i} \rangle_g = -i, \text{ and } \langle \check{b}_{2i} \check{b}_{2i-1} \rangle_g = i \quad (8.47)$$

Moreover, we have trivially that

$$\langle (\check{b}_i)^2 \rangle_g = 1 \quad (8.48)$$

from the anti-commutation relation. We may write the correlation matrix as  $\langle \check{b}_i \check{b}_j \rangle = \delta_{ij} + i\Gamma_{ij}^B$ , with

$$\Gamma_{ij}^B = \oplus \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (8.49)$$

The convenience of using the Majorana operators is that it includes all the information of the expectation values of  $\langle \tilde{b}_i \tilde{b}_j \rangle_g$  and  $\langle \tilde{b}_i \tilde{b}_j^\dagger \rangle_g$ . The transformation from  $\tilde{a}$  to  $\tilde{b}$ ,  $W_{ij}$ , commutes with the trace, so we may transform the whole correlation matrix back in terms of  $\tilde{a}_i$ , such that  $\langle \tilde{a}_i \tilde{a}_j \rangle_g = \delta_{ij} + i\Gamma_{ij}^A$  with  $\Gamma^A = W^T \Gamma^B W$ . The analytical results were found by Latorre *et al.* [114, Appendix B], who found for the correlation matrix

$$\Gamma^A = \begin{bmatrix} \Pi_0 & \Pi_1 & \cdots & \Pi_{N-1} \\ -\Pi_1 & \Pi_0 & & \\ \vdots & \ddots & & \\ -\Pi_{N-1} & & & \Pi_0 \end{bmatrix}, \text{ with } \Pi_i = \begin{bmatrix} 0 & g_i \\ -g_i & 0 \end{bmatrix}, \quad (8.50)$$

where the coefficients  $g_i$ , in the limit where we take the number of spins  $N \rightarrow \infty$ , are given by

$$g_i = \frac{1}{2\pi} = \int_0^{2\pi} d\phi e^{i\phi} \frac{\cos \phi - \lambda - i\gamma \sin \phi}{|\cos \phi - \lambda - i\gamma \sin \phi|}. \quad (8.51)$$

Next we turn to the actual computation of the entanglement entropy of a subsection of the spin lattice with length  $L$ . Note that we can write the density matrix in the following form

$$\hat{\rho}_L = 2^{-L} \sum_{\mu_1, \dots, \mu_L=0,x,y,z} \rho_{\mu_1 \dots \mu_L} \sigma_1^{\mu_1} \cdots \sigma_L^{\mu_L}, \quad (8.52)$$

where  $\sigma^0 = \mathbb{1}$ . Moreover, the Hamiltonian possesses the following symmetry,

$$\left( \prod_i \sigma_i^z \right) H_{XY} \left( \prod_i \sigma_i^z \right) = H_{XY} \quad (8.53)$$

due to that  $\sigma^\mu \sigma^\nu \sigma^\mu = -\sigma^\nu$  if  $\mu \neq \nu$  and  $\sigma^\mu \sigma^\mu \sigma^\mu = \sigma^\mu$ . The term containing the  $\sigma^z$  in the Hamiltonian is then immediately invariant, and the terms containing the  $\sigma^x$  and  $\sigma^y$  are products of two Pauli matrices, and thus are also invariant. This symmetry has as consequence that  $\rho_{\mu_1 \dots \mu_L} = 0$  when the sum of  $x$  and  $y$  is odd, due to the cyclic property of the trace

$$\begin{aligned} E = \text{Tr} [H_{XY} \hat{\rho}_L] &= \text{Tr} \left[ \left( \prod_i \sigma_i^z \right) H_{XY} \left( \prod_i \sigma_i^z \right) \hat{\rho}_L \right] \\ &= \text{Tr} \left[ H_{XY} \left( \prod_i \sigma_i^z \right) \hat{\rho}_L \left( \prod_i \sigma_i^z \right) \right]. \end{aligned} \quad (8.54)$$

This is because the terms with an odd number of  $\sigma^x$ 's and  $\sigma^y$ 's would have a negative contribution to the sum after the symmetry transformation, which would change the expectation value. As such they have to be equal to zero. This shows us that we can fully describe the complete density matrix  $\hat{\rho}_L$  by specifying the following expectation values

$$\rho_{\mu_1 \dots \mu_L} = \langle \sigma^{\mu_1} \cdots \sigma^{\mu_L} \rangle_g. \quad (8.55)$$

Since these may only contain even numbers of  $\sigma^x$  and  $\sigma^y$ , we may write all expectation values needed to fully reconstruct  $\hat{\rho}_L$ , using even products of Majorana operators, as seen from the inverse transformations expressed in equations 8.26-8.28. It follows that all the reduced density matrix may be reconstructed from the correlation matrix  $\langle \tilde{a}_i \tilde{a}_j \rangle_g = \delta_{ij} + i\Gamma_{ij}^A$ . In fact, let the region  $L = [l, l + L - 1]$ , then we only need the expectation values  $\langle \tilde{a}_i \tilde{a}_j \rangle_g$  for the indices  $i, j = 2l - 1, \dots, 2(l + |L|)$ . We denote the correlation matrix denoting only these terms  $\langle \tilde{a}_i \tilde{a}_j \rangle_g^L = \delta_{ij} + i\Gamma_A^L$ ,  $i, j = 0, \dots, 2|L|$

$$\Gamma_L^A = \begin{bmatrix} \Pi_0 & \cdots & \Pi_{L-1} \\ \vdots & \ddots & \\ -\Pi_{L-1} & & \Pi_0 \end{bmatrix}. \quad (8.56)$$

$\Gamma_L^A$  can again be put in block diagonal form,  $\Gamma_L^C = V\Gamma_L^A V^T$ , since it is skew-symmetric, such that

$$\Gamma_L^C = \oplus \begin{bmatrix} 0 & \xi_i \\ -\xi_i & 0 \end{bmatrix} \quad (8.57)$$

We can again define  $L$  Fermionic operators

$$\hat{c}_i \equiv \frac{\check{c}_{2i-1} + i\check{c}_{2i}}{2} \quad (8.58)$$

such that

$$\{\hat{c}_i, \hat{c}_j\} = 0, \text{ and } \{\hat{c}_i^\dagger, \hat{c}_j\} = \delta_{ij}. \quad (8.59)$$

However, most importantly, they obey by construction of the correlation matrix

$$\langle \hat{c}_i \hat{c}_j \rangle = \frac{1}{4} (\langle \check{c}_{2i-1} \check{c}_{2j-1} \rangle + i \langle \check{c}_{2i-1} \check{c}_{2j} \rangle + i \langle \check{c}_{2i} \check{c}_{2j-1} \rangle - \langle \check{c}_{2i} \check{c}_{2j} \rangle) = 0 \quad (8.60)$$

and

$$\langle \hat{c}_i^\dagger \hat{c}_j \rangle = \frac{1}{4} (\langle \check{c}_{2i-1} \check{c}_{2j-1} \rangle + i \langle \check{c}_{2i-1} \check{c}_{2j} \rangle - i \langle \check{c}_{2i} \check{c}_{2j-1} \rangle + \langle \check{c}_{2i} \check{c}_{2j} \rangle) = \delta_{ij} \frac{1 + \xi_i}{2}, \quad (8.61)$$

$$\langle \hat{c}_i \hat{c}_j^\dagger \rangle = \delta_{ij} \frac{1 - \xi_i}{2}. \quad (8.62)$$

Since these Fermionic operators carry all the information of the density matrix, we may reconstruct the density matrix using the above correlators. Since the operators are uncorrelated unless they have the same index, the density matrix must be in a product state of the  $L$  Fermionic modes,

$$\hat{\rho}_L = \otimes_i \hat{\rho}_i. \quad (8.63)$$

The density matrix has the eigenvalues

$$\lambda_{i_\pm} = \frac{1 \pm \xi_i}{2}, \quad (8.64)$$

and the ground state entropy of a single spin at a lattice site is

$$S[\hat{\rho}_i] = - \sum_{\pm} \lambda_{i_\pm} \ln \lambda_{i_\pm}. \quad (8.65)$$

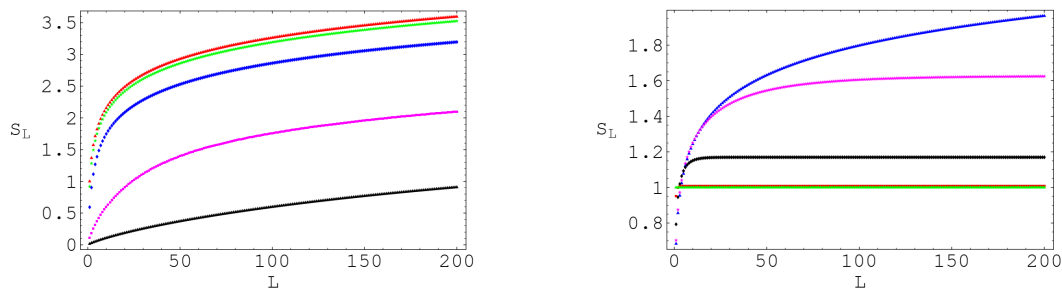
Since the state is a product state, due to the additivity of the Von Neumann entropy, the total entropy is the sum over all lattice sites

$$S_L = - \sum_{i_\pm} \lambda_{i_\pm} \ln \lambda_{i_\pm} \quad (8.66)$$

Numerical results were obtained by Latorre *et al.* [114, §D] and fitted as a function of  $L$ , for multiple values of  $\lambda$  and  $\gamma$ . For  $\lambda = \gamma = 0$ , the model is critical, and reduces to the spin XX model without an external magnetic field. The scaling behavior is found to be

$$S_L^{XX} = \frac{1}{3} \ln L + c, \quad (8.67)$$

where  $c \approx \pi/3$ . Latorre *et al.* also find that when the system is not critical, the entropy always converges to some saturation value  $S^*$  as  $|L| \rightarrow \infty$ , though the actual value depends on the parameters. Thus in these cases the entropy may be said to follow an area law since the size of the boundary is constant. The entanglement entropy for different parameters of  $\gamma$  and  $\lambda$ , as a function of the length of the block  $L$ , may be seen in figure 8.3



(a) The entanglement entropies for the spin XX model, i.e.  $\gamma = 0$ , for different values of  $\lambda \leq 1$ . The entropy decreases as  $\lambda$  increases. These parameter regions are part of the critical region, and we see that the entropies are unbounded.

(b) The entanglement entropies for the Ising model, i.e.  $\gamma = 1$ , for different values of  $\lambda \in \{0, 1\}$ . The entanglement entropy grows with  $\lambda$ , and is thus maximal at  $\lambda = 1$ . This once again corresponds to a critical point and we see that the entropy grows unboundedly, whereas for the other parameters it saturates.

Figure 8.3: Both images are taken from Latorre *et al.* [114]

### Concluding Remarks

We have seen that the entanglement entropy for the spin XY chain obeys an area law when the system is not critical, but diverges logarithmically when  $L \rightarrow \infty$  when the system is critical. We found this result by diagonalizing a  $2L \times 2L$  block of the correlation matrix  $\langle \tilde{a}_i \tilde{a}_j \rangle$ .

First of all from figure 8.3 we see that the increase in the saturation point of the entropy for non-critical spin chains is in line with the increase of the correlation length  $\xi$ . When the correlation length diverges, so as well does the saturation point. Latorre *et al.* [114] remark that this could explain the difficulty of simulation of critical spin chains using the so called DMRG method (density matrix renormalization group), which is a numerical and approximate way to simulate such systems. We may see how this makes sense by interpreting the entanglement entropy as a measure for the number of pure states that are relevant for the behavior of the system. If the probability of a pure state occurring in the total density matrix is very low, we may suspect it is less important for numerical calculation. At the same time these states contribute very little to the density matrix. As such, the unbounded entropy implies that as  $L$  grows, more and more eigenstates of  $\hat{\rho}_L$  must be considered to sustain a given accuracy. The saturation point of the entropy for non-critical spin chains implies that there is a given amount of pure states that contribute in a significant way to the density matrix  $\hat{\rho}_L$ , even as  $L \rightarrow \infty$ .

Secondly, the result that the entropy of the spin chain is given by the eigenvalues of the correlation matrix turns out to be a more general result. This result turns out to hold for any 1D Fermionic chain. Moreover, a similar result is found for Bosonic harmonic chains [116, p. III].

Lastly, there is a link between critical spin chains and conformal field theories. In fact, there is a well known result in  $(1+1)D$  conformal field theory, for which the entanglement entropy is given by [118]

$$S[\hat{\rho}_L] = \frac{c}{3} \ln \frac{|L|}{a} + \mathcal{O}(1), \quad (8.68)$$

where  $c$  is the central charge and  $a$  is an ultra violet cutoff. This result has also been generalized to the Rényi entropy which has

$$R_\alpha[\hat{\rho}_L] = \frac{c}{6} \left(1 + \frac{1}{\alpha}\right) \ln \frac{|L|}{a} + \mathcal{O}(1) \quad (8.69)$$

## 8.4 Entanglement Entropy in Field Theory

In this section we will study two different methods to calculate the ground state entropy in quantum field theory, namely the so-called real time and euclidean approaches. The real time consists of finding the entanglement entropy directly from the Hamiltonian, while the Euclidean method employs the Euclidean path integral approach, with imaginary time. In recent literature the so called Euclidean method is typically used more often than the real time approach. However, the real time approach was employed in the pivotal papers by Bombelli *et al.* [111] and Srednicki [112], which sparked the interest in the connection between the entanglement and black hole entropy. As such we feel it is important to give an overview of their work and conclusions. Though, more than that, the real time approach may still be more useful for particular generalizations, such as entanglement regions that are not separated by a single hypersurface or generalizations to interacting field theories [119],[120].

### 8.4.1 The Real Time Approach

In what follows, we will find the entanglement entropy of the ground state of the free scalar field using the real time approach. In this so called real time approach, the goal is first to find an  $N$  body quantum mechanical system that can represent the field theory we're interested in. We mean with this that the quantum mechanical system becomes equivalent to the field theory the limit where we take  $N \rightarrow \infty$ , such that the index of a subsystem becomes the space coordinate. We then find an expression for the entanglement entropy for the  $N$  body quantum mechanical system, which we relate to the general form of the Hamiltonian. This result is then used to find the entanglement entropy of the field theory.

#### Two Coupled Harmonic Oscillators

Consider the Hamiltonian of two coupled harmonic oscillators

$$H = \frac{1}{2} (p_1^2 + p_2^2 + k(x_1^2 + x_2^2) + l(x_1 - x_2)^2). \quad (8.70)$$

The constant  $k$  is related to the mass of the oscillators and  $l$  to the coupling. We may diagonalize this Hamiltonian by defining the following transformations  $x_{\pm} = (x_1 \pm x_2)/\sqrt{2}$ ,  $p_{\pm} = (p_1 \pm p_2)/\sqrt{2}$ , and defining  $\omega_+ = k^{\frac{1}{2}}$  and  $\omega_- = (k + 2l)^{\frac{1}{2}}$ . The Hamiltonian is now

$$H = \frac{1}{2} (p_+ + p_- + \omega_+^2 x_+^2 + \omega_-^2 x_-^2) \quad (8.71)$$

with the ground state wavefunction

$$\psi_g(x_+, x_-) = \frac{(w_+ w_-)^{\frac{1}{4}}}{\sqrt{\pi}} \exp\left(-\frac{1}{2} (\omega_+ x_+^2 + \omega_- x_-^2)\right). \quad (8.72)$$

The energy of the ground state is  $E_g = (\omega_+ + \omega_-)$ . In analog to field theory, the index of the position operators will be taken to represent the spatial degree of freedom of the field. To trace out a part of space is then to trace out one of the oscillators,  $x_1$  or  $x_2$ . We will choose to trace out  $x_2$ . We obtain the partially traced density matrix from

$$\begin{aligned} \hat{\rho}_1(x_1, x'_1) &= \int_{-\infty}^{\infty} \psi_g(x_1, x_2) \psi_g^*(x'_1, x_2) dx_2 \\ &= \sqrt{\frac{\gamma - \beta}{\pi}} \exp\left(-\frac{\gamma}{2} (x_1^2 + x_1'^2) + \beta x_1 x_1'\right) \end{aligned} \quad (8.73)$$

where the parameters  $\alpha$  and  $\beta$  are given by [112, p. 2]

$$\gamma = \frac{2\omega_+ \omega_-}{\omega_+ + \omega_-} + \beta, \quad \beta = \frac{(\omega_+ - \omega_-)^2}{4(\omega_+ + \omega_-)}. \quad (8.74)$$

We then need to find the eigenfunctions  $\phi_n$ , with eigenvalue  $\lambda_n$ , which satisfy

$$\hat{\rho}_1(x, x') f_n(x') dx = \lambda_n \phi_n. \quad (8.75)$$

The eigenfunctions are found by Srednicki to be

$$f_n(x) = H_n(\alpha^{\frac{1}{2}} x) e^{-\frac{\alpha}{2} x^2} \quad (8.76)$$

where  $H_n$  are the Hermite polynomials and  $\alpha = \sqrt{w_+ w_-}$ . These may be found by expanding the density matrix in terms of the Hermite polynomials [121, §III A]. The eigenvalues are

$$\lambda_n = (1 - \xi) \xi^n, \quad \xi = \frac{\beta}{\alpha + \gamma}. \quad (8.77)$$

The entanglement entropy is given by

$$S[\hat{\rho}_1] = - \sum_{n=0} \lambda_n \ln \lambda_n = (1 - \xi) \sum_n (\xi^n \ln(1 - \xi) + n \xi^n \ln \xi). \quad (8.78)$$

Since the eigenvalues of the density matrix are smaller than one, we may use the geometric series to obtain

$$S[\hat{\rho}_1] = - \ln(1 - \xi) - \frac{\xi}{1 - \xi} \ln \xi \quad (8.79)$$

From this we may conclude that as  $l$  gets larger, the entanglement in the ground state between the two modes gets larger. At the same time, as  $k$ , gets larger the entropy is diminished. When  $l$  is equal to zero, the oscillators are uncoupled and thus there is no entanglement.

### **$N$ Coupled Harmonic Oscillators**

The previous result may be generalized to  $N$  coupled harmonic oscillators, with Hamiltonian

$$H = \frac{1}{2} \sum_{i=1}^N p_i^2 + \frac{1}{2} \sum_{i,j=1}^N x_i V_{ij} x_j, \quad (8.80)$$

where  $\hat{V}$  is a real and symmetric matrix and  $p_i$  the conjugate momentum of the oscillator  $x_i$ . The ground state wavefunction has a similar form [119]

$$\psi_g(x_1, \dots, x_N) = \prod_i \left( \frac{\omega_i}{\pi} \right) e^{-\frac{1}{2} \omega_i x_i^2} \quad (8.81)$$

$$= \left( \frac{\det \hat{W}}{\pi} \right)^{\frac{1}{4}} e^{-\frac{1}{2} \vec{x}^T \hat{W} \vec{x}} \quad (8.82)$$

with  $\hat{W} = \sqrt{\hat{V}}$ , which is also a symmetric matrix. The density matrix is given by

$$\hat{\rho}(\vec{x}, \vec{x}') = \sqrt{\frac{\det \hat{W}}{\pi}} e^{-\frac{1}{2} \vec{x}^T \hat{W} \vec{x}} e^{-\frac{1}{2} \vec{x}'^T \hat{W} \vec{x}'}. \quad (8.83)$$

Next, we wish to split up the space of harmonic oscillators into two sub-spaces. Defining the region  $A$  as all the oscillators from 1 to  $n$  and  $B$  as the oscillators  $n+1$  to  $N$ , we decompose  $W$  into blocks, corresponding to these regions and their interactions

$$\hat{W} = \begin{bmatrix} \hat{W}_1 & \hat{W}_2 \\ \hat{W}_2^T & \hat{W}_3 \end{bmatrix}, \quad (8.84)$$



and we define  $\hat{U} = \hat{W}^{-1}$  and  $\bar{U}_i = \hat{U}_i^{-1}$ , for notational convenience. The reduced density matrix obtained by tracing out  $B$  is then [111]

$$\hat{\rho}_A(\vec{x}_A, \vec{x}'_A) = \sqrt{\frac{\det \bar{U}_1}{\pi}} \exp\left(-\frac{1}{2}(\vec{x}_A^T \hat{\Gamma} x_A + \vec{x}'_A{}^T \hat{\Gamma} \vec{x}'_A) + \vec{x}'_A{}^T \Delta \vec{x}_A\right), \quad (8.85)$$

with

$$\hat{\Delta} = \frac{1}{2} \hat{W}_2 \hat{U}_3 \hat{W}_2^T, \text{ and } \hat{\Gamma} = \hat{W}_1 - \hat{\Delta}. \quad (8.86)$$

Next, we will try to separate the  $x$  coordinates. First we define the coordinate transformation

$$\vec{y}_A = \hat{\Gamma}_D^{\frac{1}{2}} \hat{O} \vec{x}_A, \quad (8.87)$$

with the diagonal matrix  $\Gamma = A^T \hat{\Gamma}_D A$ , where  $\hat{A}$  is an orthogonal matrix. The density matrix then becomes

$$\hat{\rho}_A(\vec{x}_A, \vec{x}'_A) = \sqrt{\frac{\det \bar{U}_1 \hat{\Gamma}}{\pi}} \exp\left(-\frac{1}{2}(\vec{y}_A^2 + \vec{y}'_A{}^2) + \vec{y}'_A{}^T \hat{\Lambda} \vec{y}_A\right) \quad (8.88)$$

with  $\hat{\Lambda} = \hat{\Gamma}_D^{-\frac{1}{2}} \hat{A} \hat{\Delta} \hat{A}^T \hat{\Gamma}_D^{-\frac{1}{2}}$ . The extra prefactor of the determinant of  $\hat{\Gamma}$  comes from the normalization condition of the density matrix. Finally, we can define a final coordinate transformation  $\vec{z}_A = \hat{B}^T \vec{y}_A$ , where  $\hat{B}$  is the orthogonal matrix that diagonalizes  $\hat{\Lambda}$ . We see that the density matrix may be written as

$$\hat{\rho}_A(\vec{x}_a, \vec{x}'_a) = \sqrt{\frac{\det \bar{U}_1 \hat{\Gamma}}{\pi}} \prod_a \exp\left(-\frac{1}{2}(z_a^2 + z'_a{}^2) + \lambda_a z_a z'_a\right), \quad (8.89)$$

where  $\lambda_a$  are the eigenvalues of  $\Lambda$ . Much in the same way as in the case of two harmonic oscillators, the entropy of one term in the product is

$$S_a = -\ln(1 - \zeta_a) - \frac{\zeta_a}{1 - \zeta_a} \ln(1 - \zeta_a), \quad (8.90)$$

where  $\zeta_a$  is related to  $\lambda_a$  by [112]

$$\zeta_a = \frac{\lambda_a}{1 + \sqrt{1 - \lambda_a}}. \quad (8.91)$$

Since the state is a product state of density matrices, the entropy becomes the sum of the entropies of the products due to the additivity property

$$S[\hat{\rho}_A(\vec{x}_a, \vec{x}'_a)] = \sum_a S_a \quad (8.92)$$

## The Real Scalar Field

In this last section we show how the above results may be generalized to a field theory. We will first show an example where we take region  $A$  to be the half of space for which  $x_1 > 0$ , and  $B$  its complement. We will later on repeat this example in the Euclidean method. Then, in the last part of this section we show numerical results obtained by Srednicki [112] for a spherical entanglement region, showing the area law.

First of all, from similar considerations as above, the parameter  $\zeta_a$  may be related to the eigenvalues  $\mu_a$  of the following operator

$$\hat{M} = -\hat{U}_2 \hat{W}_2^T, \quad (8.93)$$

through the following relation

$$\zeta_a = 1 + \frac{2}{\mu_a} - 2\sqrt{\frac{1 + \lambda}{\lambda^2}} \quad (8.94)$$

as shown by Bombelli [111, eq. 29]<sup>6</sup>. To find the entanglement entropy in field theory, we thus must find the analogs of the operators  $\hat{V}_2$  and  $\hat{W}_2^T$  in the continuum limit. Assuming the Klein Gordon equation in the continuum limit, we make the correspondence

$$\frac{1}{2} \sum_{i,j} x_i V_{ij} x_j \rightarrow \int \left( \frac{1}{2} (\nabla \phi)^2 + \frac{1}{2} m^2 \phi^2 \right) d^3 x \quad (8.95)$$

Analogous to the spectral decomposition of matrix operators, we may write the potential in terms of its eigenfunctions, which are plane waves

$$V(x, y) = \int \frac{d^3 k}{(2\pi)^3} (k^2 + m^2) e^{ik(x-y)}. \quad (8.96)$$

We then have that the square root of the operator is simply the square root of the coefficients

$$W(x, y) = \int \frac{d^3 k}{(2\pi)^3} (k^2 + m^2)^{\frac{1}{2}} e^{ik(x-y)} \quad (8.97)$$

and since the operator is diagonal

$$W^{-1}(x, y) = \int \frac{d^3 k}{(2\pi)^3} (k^2 + m^2)^{-\frac{1}{2}} e^{ik(x-y)} \quad (8.98)$$

We may write equation 8.93 a little more transparently using Einstein notation. Equation 8.93 becomes

$$\Lambda^a_b = -W^{a\alpha} W_{\alpha b}, \quad (8.99)$$

where we defined the indices  $a$  and  $b$  to correspond to all oscillators in region  $A$ , and  $\alpha$  to corresponds to  $B$ . Moreover,  $W^{a\alpha}$  is the inverse of  $W_{a\alpha}$ . Thus we see that the operator  $\Lambda^a_b$  only works on oscillators in region  $A$  and was obtained by summing over the index  $\alpha$ . We see that, in the continuum limit this operator then becomes

$$\Lambda(x, y) = - \int_B W^{-1}(x, z) W(z, y) d^3 z, \quad x, y \in A. \quad (8.100)$$

Thus in the end, the calculation of the entanglement entropy comes down to the calculation of the eigenvalue problem

$$\int_A \Lambda(x, y) f(y) d^3 y = \lambda f(x). \quad (8.101)$$

These eigenvalues then again relate to the eigenvalues of the density matrix through equation 8.91.

For our example, we will partition  $\mathbb{R}^3$  as a whole into two cubic volumes, where region  $A$  is the region with  $x_1 > 0$  and  $B$  is its complement. Thus the entanglement area is the plane given by  $x_1 = 0$ . We can now decompose the operator  $W(x, y)$  into normal modes along the boundary and into the perpendicular modes.  $W$  becomes

$$W(x, y) = \int \frac{dk_{\perp}}{2\pi} \int \frac{d^2 k_{\parallel}}{(2\pi^2)} \sqrt{k_{\perp}^2 + k_{\parallel}^2 + m^2} e^{ik_{\perp}(x-y)} e^{i\vec{k}_{\parallel} \cdot (\vec{x}_{\parallel} - \vec{y}_{\parallel})} \quad (8.102)$$

In the same way  $\Lambda(x, y)$  becomes

$$\Lambda(x, y) = \int_B d^3 z \int \frac{dk_{\perp}}{2\pi} \int \frac{d^2 k_{\parallel}}{(2\pi^2)} \int \frac{dp_{\perp}}{2\pi} \int \frac{d^2 p_{\parallel}}{(2\pi^2)} \sqrt{k_{\perp}^2 + k_{\parallel}^2 + m^2} \sqrt{p_{\perp}^2 + p_{\parallel}^2 + m^2} \\ \times e^{ik_{\perp}(x-z)} e^{i\vec{k}_{\parallel} \cdot (\vec{x}_{\parallel} - \vec{z}_{\parallel})} e^{ip_{\perp}(z-y)} e^{i\vec{p}_{\parallel} \cdot (\vec{z}_{\parallel} - \vec{y}_{\parallel})} \quad (8.103)$$

---

<sup>6</sup>Bombelli obtains this relation by casting the density matrix of  $N$  harmonic oscillators in a slightly different form than we have shown earlier. Other than that, the procedure is mostly the same

To find the eigenvalues, we can make the ansatz that the eigenfunctions have the form [111]

$$f(x) = e^{i\vec{\nu}_{\parallel} \cdot \vec{x}_{\parallel}} f(x_{\perp}). \quad (8.104)$$

so that we obtain

$$\int d^3y \Lambda(x, y) e^{i\nu_{\parallel} y_{\parallel}} f(y_{\perp}) = \frac{1}{(2\pi)^2} \int_B dz_{\perp} \int_A dy_{\perp} \int dk_{\perp} dp_{\perp} \sqrt{k_{\perp}^2 + \nu_{\parallel}^2 + m^2} \sqrt{p_{\perp}^2 + \nu_{\parallel}^2 + m^2} \\ \times e^{ik_{\perp}(x_{\perp} - z_{\perp})} e^{ip_{\perp}(z_{\perp} - y_{\perp})} e^{i\nu_{\parallel} x_{\parallel}} f(y_{\perp}) \quad (8.105)$$

Thus finally, we see that due to the symmetry of the boundary of our entanglement region, the eigenvalue problem reduces to a 1D problem

$$\lambda f(x_{\perp}) = - \int_0^{\infty} dy_{\perp} \int_{-\infty}^{-\epsilon^2} dz_{\perp} \int_{-\infty}^{\infty} \frac{dk_{\perp} dp_{\perp}}{(2\pi)^2} \sqrt{k_{\perp}^2 + \nu_{\parallel}^2 + m^2} e^{ik_{\perp}(x_{\perp} - z_{\perp})} \\ \times \sqrt{p_{\perp}^2 + \nu_{\parallel}^2 + m^2} e^{ip_{\perp}(z_{\perp} - y_{\perp})} f(y_{\perp}), \quad (8.106)$$

with effective mass  $m_e = \sqrt{\nu_{\parallel}^2 + m^2}$ . The parameter  $\epsilon^2$  in the integration boundary of  $z_{\perp}$  was added as a regulator to remove the UV divergence [111, §III B]. The above eigenvalue problem adds a discrete spectrum of eigenvalues  $\lambda_n$ . We can plug this in to equation 8.94, to obtain the contribution of the entropy per momentum mode  $\nu_{\parallel}$ . If the entangling surface would be finite, then we would have a discrete number of states. Thus, we see that in the continuum limit we can replace the sum by an integral multiplied by a density of states. This density of states has to be proportional to the area [111]. In other words

$$S \propto A \int d^2\nu_{\parallel} s(\vec{\nu}_{\parallel}), \quad (8.107)$$

where  $s$  is the entropy density per mode. This shows explicitly that the entropy scales according to the area of the entangling surface.

The above results are generalizable much in the same way if the system has a spherical symmetry, such as in the case where region  $B$  represents the volume of a black hole. Then, the Hamiltonian may be cast in a similar form to the Hamiltonian in equation 8.80 [112, p.3]. It turns out that the entanglement entropy is actually UV divergent, so direct computation requires some sort of regulation. Srednicki calculated the entanglement entropy numerically on a lattice, for a sphere with radius  $R$ , the following relation was found, which clearly shows the area law

$$S_E = 0.3 \frac{R^2}{a^2} \quad (8.108)$$

where  $a$  is the lattice spacing.

## 8.4.2 The Euclidean Method

Analogously to the quantum mechanical wavefunction, QFT may be thought of in terms of the wavefunctional  $\Psi[\phi(\vec{x})]$ , which assigns an amplitude to every classical field configuration. These may be thought of as eigenfunctions of the field operator  $\hat{\phi}(t, \vec{x})$  at  $t = 0$ , such that  $\hat{\phi}(t, \vec{x}) |\phi(\vec{x})\rangle = \phi(\vec{x}) |\phi(\vec{x})\rangle$ . In the path integral formalism, we may express the ground state wavefunctional from the Euclidean path integral [121, §IV. C]

$$\Psi_0[\phi(\vec{x})] = \int_{\tau=-\infty}^{\tau=0, \phi(\tau=0, \vec{x})=\phi(\vec{x})} \mathcal{D}\phi(t, \vec{x}) e^{A_E[\phi(\tau, \vec{x})]} \equiv \langle \phi(\vec{x}) | \Psi \rangle, \quad (8.109)$$

where we denoted the Euclidean action by  $A_E$ , to avoid confusion with the entropy. Its conjugate is given by

$$\Psi_0^*[\phi(\vec{x})] = \int_{\tau=0, \phi(\tau=0, \vec{x})=\phi(\vec{x})}^{\tau=\infty} \mathcal{D}\phi(t, \vec{x}) e^{A_E[\phi(\tau, \vec{x})]} \equiv \langle \Psi | \phi(\vec{x}) \rangle. \quad (8.110)$$

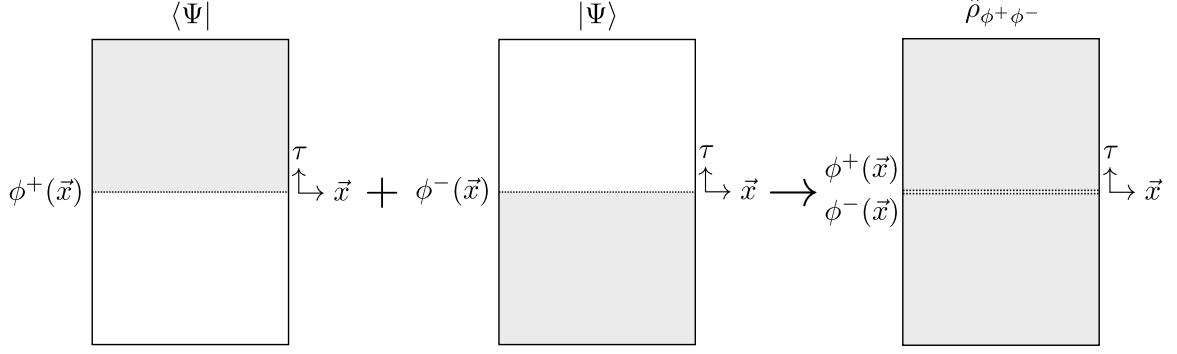


Figure 8.4: A representation of the ground state wavefunctional and its conjugate. The shaded regions are integrated over, and the dotted line indicates that the field always is the same spatial field at  $\tau = 0$ .

Notice the action is integrated over  $\tau = (-\infty, 0]$  for the wavefunctional and  $\tau = [0, \infty)$  for its conjugate. Upon multiplication of the exponential factors, we see that we may equivalently “glue” the regions of  $\tau = [-\infty, 0)$  and  $\tau = [0, \infty)$  together, to obtain a new path integral over a field, defined over all of  $\tau$ ,  $\phi(\tau, \vec{x})$ . The new field must still be equal to  $\phi_a$  at  $\tau = 0^-$  and to  $\phi_b(\vec{x})$  when  $\tau = 0^+$ , thus we have the following expression for the density matrix

$$\begin{aligned}
 (\hat{\rho})_{\phi_a\phi_b} &= \frac{1}{Z} \int \mathcal{D}\phi(\tau, x) e^{A_E[\phi(\tau, \vec{x})]} \\
 &\times \prod_{\vec{x}} \delta(\phi(0^-, \vec{x}) - \phi_a(\vec{x})) \prod_{\vec{x}} \delta(\phi(0^+, \vec{x}) - \phi_b(\vec{x})).
 \end{aligned} \tag{8.111}$$

The partition function  $Z$  is given by

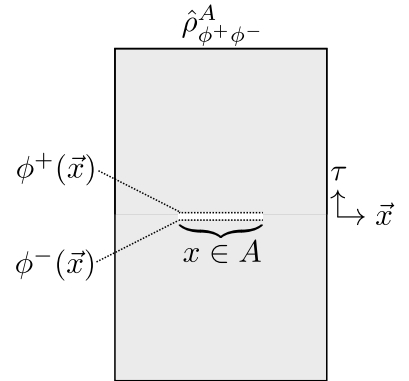
$$Z = \text{Tr} [\hat{\rho}_{\phi_a\phi_b}] = \int \mathcal{D}\phi_a(\vec{x}) \hat{\rho}_{\phi_a\phi_a} \tag{8.112}$$

Let us now cut up space into two regions,  $A$  and its complement  $B$ . To obtain the marginal density matrix  $(\hat{\rho}_A)_{\phi_a(x \in A)\phi_b(x \in A)}$ , we impose that the Euclidean field is equal to the field  $\phi_a(\vec{x})$  at  $\tau = 0^-$  and  $\phi_b(\vec{x})$  at  $\tau = 0^+$  in the region  $x \in A$ , so that we sum over all fields with different configurations for  $x \in B$ . We obtain the marginal density matrix of the region  $A$

$$(\hat{\rho}_A)_{\phi_a^A\phi_b^A} = \frac{1}{Z} \int \mathcal{D}\phi(\tau, x) e^{A_E[\phi(\tau, \vec{x})]} \prod_{\vec{x} \in A} \delta(\phi(0^-, \vec{x}) - \phi_a(\vec{x})) \prod_{\vec{x} \in A} \delta(\phi(0^+, \vec{x}) - \phi_b(\vec{x})). \tag{8.113}$$

In the following, we will denote the density matrix  $(\hat{\rho}_A)_{\phi_a^A\phi_b^A}$  simply as  $(\hat{\rho}_A)_{ab}$ .

Figure 8.5: A representation of the reduced density matrix of region  $A$ , showing the volume of the path integral that is integrated over, as well as the discontinuity at  $\tau = 0$ . The white space in the middle represents the region  $A$  which is not integrated over. The actual shape depends on how the region  $A$  is chosen.



### 8.4.3 The Replica Trick

Normally to find the Von Neumann entropy, we need to diagonalize the density matrix, which is a formidable task in field theory. Instead, we can use a clever trick called the replica trick [121]. For this trick, we exploit that the Rényi entropy goes to the Von Neumann entropy in the limit  $n \rightarrow 1$

$$\begin{aligned} S[\hat{\rho}] &= - \lim_{n \rightarrow 1} \frac{\ln \text{Tr}_A [\hat{\rho}_A^n]}{n-1} \\ &= - \lim_{n \rightarrow 1} \partial_n (\ln \text{Tr}_A [\hat{\rho}_A^n]) \end{aligned} \quad (8.114)$$

where we have used l'Hôpital's rule. Though technically the Rényi entropy is only defined for integer  $n$ , through analytic continuation we may obtain the derivative. First of all we then need to find a way to express  $\text{Tr}_A [\hat{\rho}_A^n] = \sum_{a_1, a_2, \dots, a_n} \hat{\rho}_{a_1 a_2} \hat{\rho}_{a_2 a_3} \cdots \hat{\rho}_{a_n a_1}$ . We see that in this expression, every field occurs twice, once on the positive side of  $\tau = 0$  and once on the negative side. We can exploit this fact in a similar fashion to how we constructed the density matrix. We construct a Riemannian surface, also called an  $n$ -fold cover, over space time,  $\mathcal{M}_n$ , by gluing together the  $n$  copies of the field such that they meet where they are equal. For a visual representation, see figure 8.6. In this way, we may represent the trace over  $\hat{\rho}^n$  in terms of the partition function of the field  $\mathcal{M}_n$ , which we will call  $Z_n$ . The trace becomes  $\text{Tr}_A [\hat{\rho}_A^n] \equiv \frac{1}{Z^n} Z_n$ , where  $Z$  is still the partition function of one field. Plugging this into equation 8.113 we see that we must evaluate

$$- \lim_{n \rightarrow 1} \partial_n (\ln Z_n - n \ln Z), \quad (8.115)$$

to evaluate the entanglement entropy.

### 8.4.4 The Entanglement Entropy of the Free Scalar Field

We will proceed to apply the techniques above to the example of the scalar field. For convenience, we choose the region  $A$  to be the region such that  $x_1 > 0$ . The entangling surface is thus the plane at  $x_1 = 0$ .  $\mathcal{M}_n$  can now be constructed by defining a polar angle  $\theta_n$ , with  $0 \leq \theta_n \leq 2\pi n$ , in the  $\tau, x_1$  plane. We can enter the region of  $\phi_{a_i}$  by rotating our angle  $2\pi i$  times, beginning at  $\phi_{a_0}$ . The metric on every separate sheet is initially Euclidean, so by “sewing” the sheets together, the metric on  $\mathcal{M}_n$  may be written as

$$ds^2 = dr^2 + r^2 d\theta_n^2 + dx_2^2 + dx_3^2. \quad (8.116)$$

where the angle  $\theta_n$  goes from 0 to  $2\pi n$ . By defining  $\theta = \frac{1}{n} \theta_n$ , the first two terms of the metric become  $dr^2 + \frac{1}{n^2} r^2 d\theta^2$ . We therefore see that the  $x_0$  and  $x_1$  coordinates actually are geometrically equivalent to a cone<sup>7</sup>. It is well known that the partition function of the scalar field is given by the determinant  $(\det(\partial^2 + m^2))^{-\frac{1}{2}}$  [122, §9.2]. Thus similarly here, finding the entropy ultimately amounts to finding

$$\ln Z_n = -\frac{1}{2} \ln \det(-\nabla^2 + m^2). \quad (8.117)$$

Since the determinant is the product of the eigenvalues, we may equivalently write

$$\ln Z_n = -\frac{1}{2} \text{Tr} [\ln(-\nabla^2 + m)], \quad (8.118)$$

Figure 8.6: A representation of the trace over the  $n$ -fold cover  $\mathcal{M}_n$ .

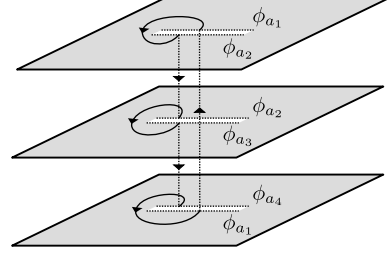
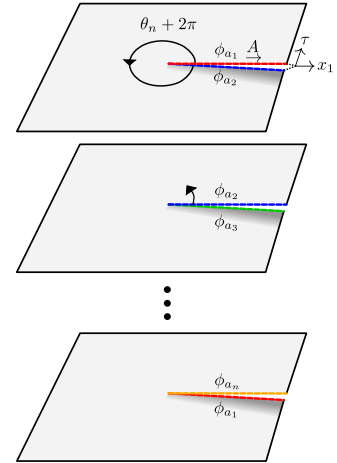


Figure 8.7: The  $N$ -fold  $\mathcal{M}_n$  in the case the half space entanglement. Every subsequent surface is accessed by rotation of  $\theta_n$  by  $2\pi$



<sup>7</sup>With that we mean that the relation between the radial and angular component are the same of that of a cone embedded in  $3D$  Euclidean space.

and finally using an integral representation of the logarithm

$$\ln Z_n = \frac{1}{2} \int_{\epsilon^2}^{\infty} \frac{1}{s} \text{Tr} \left[ e^{-s(-\nabla^2+m^2)} - e^{-s} \right] ds. \quad (8.119)$$

Normally, the lower integration boundary would be zero in the definition of the logarithm. The parameter  $\epsilon^2$  is introduced as a regulator for the UV divergence occurring at  $s = 0$ . Since  $\mathcal{M}_n$  has a direct product structure of a cone and euclidean space  $\mathcal{C}_n \times \mathbb{R}^2$ , the Laplacian can be decomposed into the sum of Laplacians  $\nabla^2 = \nabla_{\mathcal{C}_n}^2 + \nabla_{\mathbb{R}^2}^2$ . For the space  $\mathbb{R}^2$  we know that the eigenfunctions of the Laplacian operators are plane waves. For the eigenfunctions on the cone, we recognize that the Laplacian for the cone is just the Laplacian in polar coordinates

$$\nabla_{\mathcal{C}_n}^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta_n^2}. \quad (8.120)$$

The eigenfunctions are well known and easily found by separation of variables. The only difference here, is that we impose the eigenfunctions to be periodic over  $2\pi n$ , in contrast to  $2\pi$ . Solving for

$$\nabla_{\mathcal{C}_n}^2 \phi_{k,l}(r, \theta_n) = -k^2 \phi_{k,l}(r, \theta_n), \quad (8.121)$$

we find the eigenfunctions, including normalization so that they form an orthonormal basis on  $\mathcal{C}_n$ ,

$$\phi_{k,l}(r, \theta_n) = \sqrt{\frac{k}{2\pi n}} e^{il\theta_n/n} J_{|l/n|}(kr), \quad (8.122)$$

where  $J_n$  is the Bessel function of the first kind. We see that  $l$  must be integer while  $k$  is continuous. The trace  $\text{Tr} \left[ e^{-s(-\nabla^2+m^2)} \right]$  may then be solved for, by evaluating the following

$$\begin{aligned} \text{Tr} \left[ e^{-s(-\nabla^2+m^2)} \right] &= \int_{\mathcal{C}_n} r dr d\theta_n \sum_{l=-\infty}^{\infty} \left( \int_0^{\infty} e^{-s(k^2+m^2)} \phi_{k,l}(r, \theta_n) \phi_{k,l}^*(r, \theta_n) dk \right) \\ &\times \int_{\mathbb{R}^2} dx^2 \int d^2 k_{\perp} e^{-sk_{\perp}^2} \phi_{k_{\perp}}(x) \phi_{k_{\perp}}^*(x) \end{aligned} \quad (8.123)$$

We will now first evaluate the integrals corresponding to the cone  $\mathcal{C}_n$ . For this we use the following identities [123, Eq. 155, 156]

$$\int_0^{\infty} k e^{-sk^2} J_{\alpha}(kr)^2 dk = \frac{e^{-\frac{r^2}{2s}}}{2s} I_{\alpha} \left( \frac{r^2}{2s} \right) \quad (8.124)$$

$$\int_0^{\infty} r e^{-r} I_{\alpha}(r^2) = -\frac{\alpha}{2} + \int_0^{\infty} \frac{1}{\sqrt{2\pi}} dr. \quad (8.125)$$

We find that

$$\begin{aligned} &\int_{\mathcal{C}_n} r dr d\theta_n \sum_{l=-\infty}^{\infty} \left( \int_0^{\infty} e^{-s(k^2+m^2)} \phi_{k,l}(r, \theta_n) \phi_{k,l}^*(r, \theta_n) dk \right) \\ &= \frac{1}{n\sqrt{2s}} e^{-sm^2} \sum_{l=-\infty}^{\infty} \left( -\frac{|l|}{2} + n \int_0^{\infty} \frac{1}{2\pi} dr \right). \end{aligned} \quad (8.126)$$

This term contains both IR and UV divergent terms. The UV divergence comes from the integral over  $r$ . Luckily, the term is not proportional to  $n$ , so it does not contribute to equation 8.115. However, the term is also UV divergent due to the sum over the modes  $l$ . We may use the zeta function regularization

$$\sum_{l=-\infty}^{\infty} |l| = 2\zeta(-1) = -\frac{1}{6} \quad (8.127)$$

to obtain the the final expression for equation 8.126

$$\frac{1}{12n\sqrt{2s}}e^{-sm^2}. \quad (8.128)$$

The other integral is easily evaluated, since the plane waves and their conjugates cancel out to 1 at every point in space. We have [121]

$$\begin{aligned} \int_{\mathbb{R}^2} dx^2 \int d^2k_{\perp} e^{-sk_{\perp}^2} \phi_{k_{\perp}}(x) \phi_{k_{\perp}}^*(x) &= \\ &= \text{Area}(\mathbb{R}^2) \frac{\pi}{s}. \end{aligned} \quad (8.129)$$

Putting this together one may find

$$\text{Tr} \left[ e^{-s(-\nabla^2+m^2)} \right] = \frac{\pi \text{Area}(\mathbb{R}^2)}{12n} \frac{e^{-sm^2}}{\sqrt{2s^3}}. \quad (8.130)$$

Next, we would have to evaluate the kernel  $e^{-s}$ . However, the term is directly proportional to  $n$  due to the volume of the cone  $\mathcal{C}_n$ . Since  $\ln Z$  contains the same term, only not proportional to  $n$ , is cancels out in equation 8.115. What is left is to evaluate  $\text{Tr} \left[ e^{-\nabla^2+m^2} \right]$  on the surface  $\mathcal{M}_{n=1}$ , to find  $\ln Z$ . Much in the same way one may find that this is equal to

$$\frac{\pi \text{Area}(\mathbb{R}^2)}{12} \frac{e^{-sm^2}}{\sqrt{2s^3}}. \quad (8.131)$$

Putting this all together, we find that the entanglement entropy is given by

$$S[\hat{\rho}_A] = \frac{\pi \text{Area}(\mathbb{R}^2)}{3} \int_{\epsilon^2}^{\infty} \frac{e^{-sm^2}}{\sqrt{8s^5}} ds. \quad (8.132)$$

Thus we see that the area of the entangling surface is a prefactor for the entanglement entropy of the scalar field.

## 8.5 Summary and Concluding Remarks

In this chapter, we have discussed the entropies as measures of bi-partite entanglement of pure states. We properly defined entanglement in the LOCC framework. The basic idea is that when we only have access to the LOCC, which include all local operations and classical communication, then entanglement can not increase. If we take this as basic principle, then we can induce an ordering on all density matrices, based on if we can transform one into the other through LOCC. Mathematically, this was expressed using the majorization relation, and Nielsen's theorem 8.1. At the same time, this allowed us to use the Von Neumann and quantum Rényi and Tsallis entropies as measures of entanglement, since they are all monotonic in the majorization relation. We then discussed an example of a LOCC protocol, which showed explicitly that the entropy increased under such operations. Finally, we briefly mentioned how the entanglement measures may be defined on mixed states, rather than pure states.

In the sections after this, we discussed the application of the Von Neumann entropy as a measure of entanglement, commonly called the entanglement entropy. First, defined a particular property that the entanglement entropy can have for particular systems, namely an area law. An area law is a specific manifestation of non-extensivity of the entropy, where it asymptotically scales with the boundary of an entangled region of a system. Interest in area laws comes partly from the fact that the well known Bekenstein-Hawking black hole entropy shows the same behavior, and partly from the information theoretic idea that this behavior may be used to characterize certain systems. After that we have calculated the entanglement entropy for the ground states of a few systems. The first

system we discussed was the spin XY model, which is a spin lattice model with a nearest neighbor interaction. This model has critical parameter regions, where in the vicinity of those regions the typical exponential correlation length diverges, and in those regions becomes a power law. For this system, we calculated the entanglement entropy for a block of spins with length  $L$ . We saw that for non-critical systems, the area law is manifested in the saturation of the entropy. This is because the system is one-dimensional, and the boundary that interacts with the environment stays constant. As the block size becomes large, most of the spins within the block barely interact with the environment, so at some point making the block any larger does not cause any extra entanglement with the environment. In other words, the entanglement mostly happens at the boundary. On the other hand, when the system is critical, the entropy scales as the logarithm of the block size. The reason for this is that due to the longer range interactions, there is a non-negligible amount of entanglement between all the spins of the block and the environment. However, the entropy is still not extensive, which means there is some form of classical correlations between the spins in the block. Thus the block itself should not be considered a thermodynamic system in the BGS sense.

Next we discussed the entanglement entropy in the ground state of the scalar field, and presented two methods of calculation. First is the real time method. This method is based on the continuation of a quantum mechanical lattice system, to a quantum field in the continuum limit. In our case, we modeled the scalar field as a lattice of quantum mechanical oscillators, with a linear coupling to all other oscillators. Our first main result was the entanglement entropy for system of  $N$  coupled harmonic oscillators, by dividing the system into two blocks. Next, we connected these results to the continuous functional forms of the operators in field theory, in which an area law was found due to the counting of states. Then finally, we showed results by Srednicki, based upon the same method, only now with space cut up into a sphere and its complement. His results also explicitly showed the area law for a sphere.

After this, we discussed the more modern Euclidean method. In this method we can calculate the density matrix by evaluating a path integral over of Euclidean space. In this path integral, we integrate over all Euclidean field configurations, only with all fields fixed at two different spatial configurations at  $\tau = 0^+$  and  $\tau = 0^-$ . The spatial configuration at  $\tau = 0^-$  determines the first index we evaluate the density matrix at, and configuration at  $\tau = 0^+$  determines the second index. To evaluate the marginal density matrix, we only hold the spatial configuration fixed within the spatial region, and integrate over all fields with different configurations outside of that region. Next, the evaluation of the entropy is done using the so called replica trick. For this trick, we construct a Riemannian surface by gluing  $n$  fields together, such that two field meet at the side of the discontinuity where they have the same spatial configuration. By gluing the ends together, and performing an integration over all configurations located at the regions where the fields are glued together, we perform the trace. Finally the entanglement entropy is evaluated by analytically continuing the result for integer  $n$  to all  $n$  and taking the derivative. The example we showed was again in the case where we separate space into two cubic regions. The Riemannian surface  $\mathcal{M}_n$  in this case became a cone, and evaluating the entropy, essentially comes down to finding the eigenmodes of the Laplacian on  $\mathcal{M}_n$ . This result again shows explicitly that the area of the region in between the two cubic volumes is obtained as a prefactor. Of course, since this region is just an infinite plane, the entanglement entropy formally diverges. However, the results do again exemplify the area dependence of the entropy.

Of course, the methods that we have discussed are in the simplest case of the scalar field theory. In more recent works, the entanglement entropy has applied in more difficult scenarios such as in AdS/CFT and holography [121][96], and there has been a lot of research of the effect of renormalization on the entropies [124][125]. As an example, versions of the c-theorem have been proven for the entanglement entropy and rely heavily on the strong-subadditivity property that we discussed in section 4.5 [126]

Another path to explore for further research, is the use of the generalized quantum entropies for systems for which the Von Neumann entropy is not extensive, like the systems we've seen here [127].



An area law or logarithmic scaling with the system size indicates non-extensivity. Thus, the quantum generalized entropies may be useful to use as statistical entropies in these cases [113]

# Chapter 9

## Phase Space Entropies

The phase space description of systems has had great success in classical physics, while at the same time being at odds with quantum mechanics. Whereas classical systems may always be described in a continuous space of states, finite quantum systems always are always described from a discrete set of states and most importantly, can fundamentally not occupy a point in both position and momentum space, due to the uncertainty principle. However, as we have seen in chapter 6, there is a sense in which quantum mechanical systems may be pictured in a phase space using the Wigner distribution, analogously to the description of classical systems. The analogy is not perfect however, since the quantum description is always probabilistic, whereas the classical description need not be. Moreover, the Wigner distribution should not be considered a probability distribution itself. This is of course since no one point in phase space may actually be measured due to the uncertainty principle. Moreover the Wigner distribution generally is not positive definite for arbitrary states, though it was specifically for the examples we have seen already. The probabilistic description and proper probability distributions are only recovered when we integrate over either the position or momentum degree of freedom. Since the entropy for classical systems was intimately tied to the phase space description of such systems, naturally there has been research in entropies that may be applied to the phase space description of quantum mechanical systems. In this chapter we will look at two phase space representations in particular. First the Wigner representation which we have discussed, and secondly the Husimi-Q representation, which is related to the Wigner representation by a course graining operation using a minimal uncertainty Gaussian. The interesting property of the Husimi representation is that a proper probabilistic interpretation is recovered, in terms of the overlap with coherent states with means at the specific phase space coordinates.

We will first very briefly discuss the phase space representations themselves, and then discuss the entropies which are used in these settings. For an introduction which is more in depth, yet still brief we recommend the book by Ballentine [128]. For a more in depth introduction, consider the books by Schleich [129] or Leonhardt [76].

### 9.1 The Wigner and Husimi Representations

First of all, the Wigner representation of any density matrix  $\hat{\rho}$  is defined as follows

$$W(x, p) \equiv \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \langle x + \frac{y}{2} | \hat{\rho} | x - \frac{y}{2} \rangle e^{ipy/\hbar} dy, \quad (9.1)$$

The probability densities of  $x$  and  $p$  are obtained by integrating over  $p$  or  $x$  respectively

$$\rho(x) = \int_{-\infty}^{\infty} W(x, p) dp \quad (9.2)$$

$$\rho(p) = \int_{-\infty}^{\infty} W(x, p) dx \quad (9.3)$$

Analogously to the classical Liouville equation 1.3, the time evolution of the Wigner distribution is given by the Wigner equation

$$\frac{\partial W}{\partial t} + \frac{p}{m} = \frac{i}{2\pi\hbar^2} \int [V(x - \frac{x'}{2}) - V(x + \frac{x'}{2})] \exp(-\frac{i}{\hbar}(p - p')x') W(x, p', t) dx' dp', \quad (9.4)$$

where  $V$  is just the potential. To express the expectation values, we may use the Wigner transforms of operators

$$\langle \hat{A} \rangle = \text{Tr} [\hat{\rho} \hat{A}] = \int W(x, p) \hat{A}(x, p) dx dp, \quad (9.5)$$

with

$$\hat{A}(x, p) = \int \langle x - \frac{y}{2} | \hat{A} | x + \frac{y}{2} \rangle e^{\frac{ipy}{\hbar}} dy. \quad (9.6)$$

In the same way we may express the squared overlap of states. Let  $W$  and  $W'$  denote the Wigner representations of two states with density matrices  $\hat{\rho}$  and  $\hat{\rho}'$ , then

$$\text{Tr} [\rho \rho'] = 2\pi\hbar \int W(x, p) W'(x, p) dx dp. \quad (9.7)$$

The other phase space representation of a state is the so called Husimi-Q representation. The Husimi distribution is defined by

$$Q(\alpha) = \frac{1}{2\pi\hbar} \langle \alpha | \hat{\rho} | \alpha \rangle, \quad (9.8)$$

where  $|\alpha\rangle$  is a coherent state like we discussed in section 6.3. This definition ensures that the distribution is always positive definite. We identify the points in phase space with the means of the coherent states. In other words we have  $\alpha = x + ip$ . This representation does have a sort of probabilistic interpretation in phase space<sup>1</sup>. Interestingly, the Husimi distribution may itself be interpreted as a Wigner distribution of a mixed state [130, §IV]. The Husimi distribution may in fact be viewed as a coarse grained version of the Wigner distribution of a state. Consider that the Husimi distribution may be written in terms of Wigner functions

$$Q(\alpha) = \int W_\alpha(x', p') W(x', p') dx' dp'. \quad (9.9)$$

Recall that the Wigner distribution for the coherent state is given by a 2D Gaussian

$$W_\alpha(x', p') = \frac{1}{\pi\hbar} \exp\left(-\frac{(x - x')^2}{s^2} - \frac{s^2}{\hbar^2}(p' - p)^2\right). \quad (9.10)$$

In other words, the Husimi distribution may be written as the convolution of a minimal uncertainty 2D Gaussian and the Wigner representation. The Husimi distribution does not reproduce the correct probability density upon integration, but instead goes to the smoothed probability densities. In general, it does also not produce the normal quantum mechanical expectation values, though it does reproduce the correct expectation values specifically for position and momentum. In other words we have

$$\langle x \rangle = \langle \psi | x | \psi \rangle = \int x Q(x, p) dx dp, \quad (9.11)$$

---

<sup>1</sup>At the same time these “probabilities” should not be interpreted in the standard sense, since all points are not mutually exclusive, like is the case for conventional probability distributions.

and

$$\langle p \rangle = \langle \psi | p | \psi \rangle = \int p Q(x, p) dx dp. \quad (9.12)$$

The standard deviations computed by the Husimi distribution also satisfy an uncertainty relation [128, Ch. 15]

$$\Delta x_H \Delta p_H \geq \hbar. \quad (9.13)$$

The reason the lower limit is higher than normal, is that the smoothing adds extra uncertainty.

## 9.2 The Linear Entropy And Wigner Distributions

Since the Wigner distributions are in general not positive definite, there in general is a problem in directly applying the standard entropy quantities. However, all Rényi and Tsallis entropies with even integer parameter could be directly applied, though the Rényi, and in particular the Tsallis entropy with parameter  $\alpha = q = 2$  are especially interesting. This is because we have that the square of the Wigner distribution has the following relation with the density matrix

$$\begin{aligned} \int W^2(x, p) dx dp &= \frac{1}{(2\pi\hbar)^2} \int \langle x + \frac{y}{2} | \hat{\rho} | x - \frac{y}{2} \rangle \langle x + \frac{z}{2} | \hat{\rho} | x - \frac{z}{2} \rangle e^{\frac{i}{\hbar} p(y+z)} dy dz dx dp \\ &= \frac{1}{2\pi\hbar} \int \langle x + \frac{y}{2} | \hat{\rho} | x - \frac{y}{2} \rangle \langle x - \frac{y}{2} | \hat{\rho} | x + \frac{y}{2} \rangle dx dy \\ &= \frac{1}{2\pi\hbar} \int \langle a | \hat{\rho} | b \rangle \langle b | \hat{\rho} | a \rangle da db \\ &= \frac{1}{2\pi\hbar} \text{Tr} [\hat{\rho}^2]. \end{aligned} \quad (9.14)$$

Thus we see that we may write the quantum generalization of the linear Tsallis entropy, which we discussed in section 7.2 and in section 3.3.2 for the classical case,

$$T_2[\hat{\rho}] = 1 - \text{Tr} [\hat{\rho}^2] = 1 - 2\pi\hbar \int W^2(x, p) dx dp. \quad (9.15)$$

Moreover,  $\int W dx dp$  and  $\int W^2 dx dp$  are both invariant under the Wigner equation 9.4 [130]. This is expected however, since the Wigner equation of course represents unitary time evolution. Because of this, the total probability, nor the mixedness of a state should change. As such the Linear entropy is also time invariant.

Wigner distributions are often employed in quantum optics applications. In such applications, the Linear entropy has been employed as a measure of the mixedness of a state, or as a measure for how much a state has decohered [131]–[133]. It is also sometimes called the decoherence parameter.

We will only cover the Tsallis entropy with  $q = 2$  in this section, since the Rényi entropy with  $\alpha = 2$  a monotonic function of the Tsallis entropy. Moreover, the Tsallis entropy is more easily calculated and, as such, used more often.

Interestingly, formulating the linear entropy this way gives a sort of local character to the entropy. We can define the local spatial entropy

$$\sigma(x, t) = \int W dp - 2\pi\hbar \int W^2 dp, \quad (9.16)$$

where simply we have that  $\int \sigma dx = T_2$ . It can then be proven that

$$\frac{\partial \sigma(x, t)}{\partial t} + \frac{\partial J(x, t)}{\partial x} = 0 \quad (9.17)$$

where

$$J(x, t) = \int \frac{p}{m} W dp - 2\pi\hbar \int \frac{p}{m} w^2 dp. \quad (9.18)$$

This can be proven by substituting the Wigner equation into equation 9.17. One may find that this gives two integrals that both integrate to zero. Thus, this equation shows that while the entropy is conserved overall, it may be transferred over spatial regions.

### 9.2.1 Properties

In this section we will very briefly discuss the properties of the linear entropy for Wigner distributions.

#### Concavity

The linear entropy is concave. If  $W$  is a mixed state  $W = \sum \lambda_i W_i$  (the  $W_i$ 's do not have to be orthogonal) then the following holds [130, §III]

$$T_2[W] \geq \sum \lambda_i T_2[W_i]. \quad (9.19)$$

This is expected, since a sum of Wigner distributions represents a sum of density matrices. Of course, this then represents mixing enhancement, which we discussed in section 4.5.

#### Additivity

The linear entropy is not additive, but pseudo-additive, which follows from the composition law of the Tsallis entropy. If a Wigner distribution for a composite system can be separated such that

$$W(x_A, p_A, x_B, p_B) = W_A(x_A, p_A)W_B(x_B, p_B), \quad (9.20)$$

then

$$T_2[W] = T_2[W_A] + T_2[W_B] - T_2[W_A]T_2[W_B]. \quad (9.21)$$

In the limit where the subsystems become approximately pure,  $T_2[W_{A,B}] \ll 1$  and, thus, the second term becomes increasingly small. In this limit the entropy  $T_2$  is approximately additive.

#### Subadditivity

To our knowledge, no sub-additivity property has been proven for the linear entropy for Wigner distributions directly. This may be because the linear entropy is in principle defined on any distribution  $W(x, p)$ , not only the ones that represent a density matrix. In contrast, subadditivity has been proven for the linear entropy for density matrices [134]. As such, due to the correspondence that  $\int W^2 dx dp = 2\pi\hbar \text{Tr} [\hat{\rho}^2]$ , if the distribution  $W$  represents a density matrix  $\hat{\rho}$ , then the linear entropy is subadditive,

$$T_2[W_{12}(x_1, p_1, x_2, p_2)] \leq T_2[W_1(x_1, p_1)] + T_2[W_2(x_2, p_2)]. \quad (9.22)$$

### 9.2.2 Entropy Maximization

By now the power of the maximum entropy method has been demonstrated on multiple occasions. However, since Wigner distributions are not proper probability distributions so it's not obvious whether the maximum entropy principle should give any useful results. In fact, if we optimize using the microcanonical constraint, e.g. normalization, we get the equidistribution for an area in phase space  $\Omega$ ,

$$W_{mc}(x, p) = \frac{1}{\Omega}. \quad (9.23)$$

This result may in general not be seen as a proper Wigner function, since we could choose  $\Omega < \frac{\hbar^2}{4}$ , which would violate the uncertainty principle. Maximizing with the canonical constraints,  $\int W dx dp = 1$  and  $\int WH(x, p) dx dp = E$ , on the other hand, we get an interesting result. Here,

$H(x, p)$  is the representation of the Hamiltonian in phase space  $H(x, p) = p^2/2m + V(x)$ . We get the power law distribution, which is typical for the Tsallis type entropies,

$$W = -\frac{1}{2}(\lambda - \beta' E(x, p)) \quad (9.24)$$

Dividing by  $\lambda$ , setting  $\beta = \beta'/\lambda$  we obtain

$$W_c(x, p) = \frac{1}{Z}(1 - \beta E(x, p)), \quad \beta E < 1 \quad (9.25)$$

$$W_c(x, p) = 0, \quad \beta E \geq 1. \quad (9.26)$$

where  $Z$  is the normalization constant. To make the function integrable it is needed to introduce the cutoff in energy. Remarkably, this distribution turns out to be a stationary solution of the Wigner equation [130, §III]. In other words, it is a stationary state under unitary time evolution. In chapter 3 we have discussed the significance of the maximum entropy distributions, and that they may give the distributions produced by complex systems. Thus, the fact that this distribution is also invariant under time evolution is remarkable. However, we don't know if there is any quantum process that produces such a distribution.

### 9.2.3 Entropy of Course graining

In this section we will investigate the effects of smoothing operations on the linear entropy. We will also show that, when using Gaussians as a smoothing kernel, the resulting distribution may also still be considered a Wigner function, in that it represents a density matrix of a quantum state [130, §IV]. This shows that the Husimi-Q distributions may also be considered to be Wigner distributions, since they may be obtained from the Wigner distribution by the same smoothing operator. We denote the smoothed Wigner function as

$$\overline{W}(x, p) \equiv \int W(x', p') K(x - x', p - p') dx' dp' = W * K, \quad (9.27)$$

where we can use any Wigner function as the kernel  $K$ . The smoothing of a Wigner function erases some of the fluctuations that are small compared to the size of the smoothing kernel. The fluctuations of a Wigner function signify quantum coherences, thus after smoothing, we expect an entropy quantity to increase. To prove this we first define the double Fourier transform

$$W(k, \lambda) = \int W(x, p) \exp(-ikx - i\lambda p) dx dp. \quad (9.28)$$

Using it we can write a smoothed Wigner function in Fourier space as

$$\overline{W}(k, \lambda) = W(k, \lambda) K(k, \lambda), \quad (9.29)$$

due to the fact that the Fourier transform of a convolution becomes a multiplication of the Fourier representations. Moreover, as a consequence of Parseval's identity, we have that  $\int |W|^2 dx dp = \int |W|^2 dk d\lambda$ . We will call  $I[W] = 2\pi\hbar \int |W|^2 dx dp$  the quantum information quantity corresponding to  $T_2$ . We then have

$$I[\overline{W}] = \frac{\hbar}{2\pi} \int |W(k, \lambda)|^2 |K(k, \lambda)|^2 dk d\lambda \quad (9.30)$$

$$\leq \sup_{k, \lambda} |K(k, \lambda)|^2 \frac{\hbar}{2\pi} \int |W(k, \lambda)|^2 dk d\lambda \quad (9.31)$$

$$= \sup_{k, \lambda} |K(k, \lambda)|^2 2\pi\hbar \int |W|^2 dx dp \leq 2\pi\hbar \int |W|^2 dx dp = I[W]. \quad (9.32)$$

This directly implies

$$S_2[\overline{W}] \geq S_2[W], \quad (9.33)$$

which means that smoothing does indeed increase the entropy.

## 9.3 Wehrl Entropy and The Husimi-Q Representation

Interest in the Husimi distribution comes partly from the fact that it is one of the closest concepts to a classical phase space that we know of in quantum mechanics. By smoothing the Wigner distribution, we have a representation of phase space that is always positive, allows some kind of probabilistic interpretation, and is continuous in  $x$  and  $p$ . Given these facts, it is perhaps somewhat unsurprising that Wehrl introduced a classical-like entropy on these distributions. This entropy is now commonly called the Wehrl entropy is defined as the differential entropy of the Husimi distribution

$$S^W[\hat{\rho}] = - \int Q_\rho(\alpha) \ln(Q_\rho(\alpha)) d^2\alpha. \quad (9.34)$$

This entropy may be seen as an approximation to the Von Neumann entropy, and may alternatively be written as

$$S^W[\hat{\rho}] = - \int \frac{d^2\alpha}{2\pi\hbar} \langle \alpha | \hat{\rho} | \alpha \rangle \ln \langle \alpha | \rho | \alpha \rangle. \quad (9.35)$$

The factor  $1/2\pi\hbar$  is added because the coherent states are an overcomplete set, which causes so called “overcounting”<sup>2</sup>. This entropy serves as an upper bound of the Von Neumann entropy, due to the concavity of  $s(x) = -x \ln x$ . We showed in section 7.1.3 that this implies that  $s(\langle \alpha | \hat{\rho} | \alpha \rangle) \geq \langle \alpha | s(\hat{\rho}) | \alpha \rangle$ . Thus since the Von Neumann entropy may be written alternatively as

$$S[\hat{\rho}] = - \int \frac{d^2\alpha}{2\pi\hbar} \langle \alpha | \hat{\rho} \ln \hat{\rho} | \alpha \rangle, \quad (9.36)$$

we have that

$$S \leq S_W. \quad (9.37)$$

We will not go over all the properties of this entropy, since most are inherited directly from the differential entropy. We will note two properties specific to the Wehrl entropy however. First of all, the Wehrl entropy has a minimum of 1, which happens when  $\hat{\rho}$  represents a pure coherent state [135]. Secondly, we have shown that the Wehrl entropy gives an upper bound to the Von Neumann entropy. In general, the two quantities may be very far apart. However, if the characteristic features, e.g. fluctuations, of the Husimi distribution span much larger than  $\hbar$ , then the Wehrl entropy becomes a good approximation of the Von Neumann entropy [136].

The Wehrl entropy is often also called a classical entropy. This is because it approximates the von Neumann entropy [2]. Thus since we know that the von Neumann entropy is a thermodynamical entropy, as well should the Wehrl entropy be for systems where the difference between the two is negligible. However, to our knowledge no direct correspondence has been shown between the Wehrl entropy and the thermodynamical entropy from the microscopic behavior of a many body quantum system, or even from assumptions about the microscopic behavior (e.g. something akin to ergodicity).

More recently, the Wehrl entropy has also gotten some interest within resource theories as a measure of non-classicality [137][138]. It turns out that in some optical settings, non-classicality may be viewed as a resource [139]. This field has some overlap with the resource theory of entanglement [140], which we discussed in section 8.1.

### 9.3.1 Quantum-Optical States

The Husimi distribution, like the Wigner distribution, is often studied in the context of so called optical states. Moreover, as we have remarked, the Wehrl entropy sometimes used as a measure of non-classicality. In this section we will discuss the Wehrl entropies of a few optical states, which where

---

<sup>2</sup>This means that we can perform a trace integrating over all inner products with coherent states, but we would have to multiply by an extra normalization factor of  $1/2\pi\hbar$ .

calculated by Orłowski [141], to see how it behaves in this setting. Moreover, we will use the results to discuss the differences and similarities between the von Neumann and Wehrl entropies.

The Hamiltonian that generates the quantum optical states may be formally found by quantizing the amplitude of a monochromatic wave of light. What we then essentially find the Hamiltonian of the harmonic oscillator. States like the coherent and coherent squeezed states are states of special interest in quantum optics, which we have discussed in section 6.3. In quantum optics, adding a photon to a state is represented by the creation operator. As such, since coherent states may be written as superposition of all energy eigenstates (see equation 6.55) they are made up of an indeterminate number of photons. However, we may associate an average number of photon's to them, which is represented by the number operator  $a^\dagger a$ . For example, we have for the coherent state

$$\langle N \rangle = \langle \alpha | a^\dagger a | \alpha \rangle = \langle 0 | D^\dagger(\alpha) a^\dagger a D(\alpha) | 0 \rangle = |\alpha|^2 \quad (9.38)$$

### Two photon Coherent State (TPCS)

The so called two photon coherent states are the squeezed states that we have discussed in section 6.3. Recall that they are defined as the eigenstates of the Bogoliubov transform of the annihilation and creation operator;  $\hat{b} = \mu \hat{a} - \nu \hat{a}^\dagger$ , with  $|\mu|^2 - |\nu|^2 = 1$ , so that  $\hat{b} |\mu, \nu; \beta \rangle = \beta |\mu, \nu, \beta \rangle$ . The Husimi distribution of these states is given by [142]

$$Q(\alpha) = \frac{1}{|\mu|} \exp(-|\alpha|^2 - |\beta|^2) \exp\left(-\frac{\nu}{2\mu} \alpha^{*2} + \frac{\nu^*}{2\mu} \beta^2 + \frac{1}{\mu} \alpha^* \beta + c.c.\right) \quad (9.39)$$

where *c.c.* in the second exponent stands for the complex conjugate of the foregoing term in that exponent. The entropy then simply becomes

$$S_W = \ln |\mu| + 1. \quad (9.40)$$

Now consider that  $\langle \hat{N} \rangle = |\nu|^2 + |\beta|^2$  which may be checked either from the identities 6.63, 6.70 and 6.71, or from the inverse Bogoliubov transformations and the definition of the two photon coherent state. We may write the entropy in terms of the average number of photons,

$$S_W = 1 + \frac{1}{2} \ln(1 + \langle \hat{N} \rangle - |\beta|^2). \quad (9.41)$$

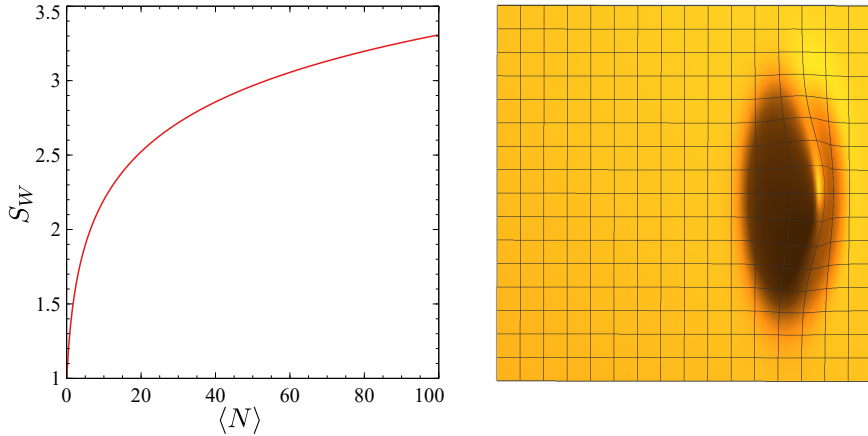


Figure 9.1: The Wehrl entropy of the TPCS with  $\beta = 0$ , and its Husimi-Q representation.



## Photon Number States

The photon number states are simply the eigenstates of the Hamiltonian,  $|n\rangle$ . The Husimi-Q representation of these states is given by the Poisson distribution

$$Q(\alpha) = \frac{|\alpha|^{2n}}{n!} \exp(-|\alpha|^2), \quad (9.42)$$

which is easily seen from the number state representation of the coherent state. We also have simply  $n = \langle \hat{N} \rangle$ . The entropy of these states is given by [141]

$$S_W = 1 + n + \ln(n!) - n(-\gamma + \sum_{i=1}^n \frac{1}{i}), \quad (9.43)$$

where  $\gamma$  is the Euler-Gamma constant.

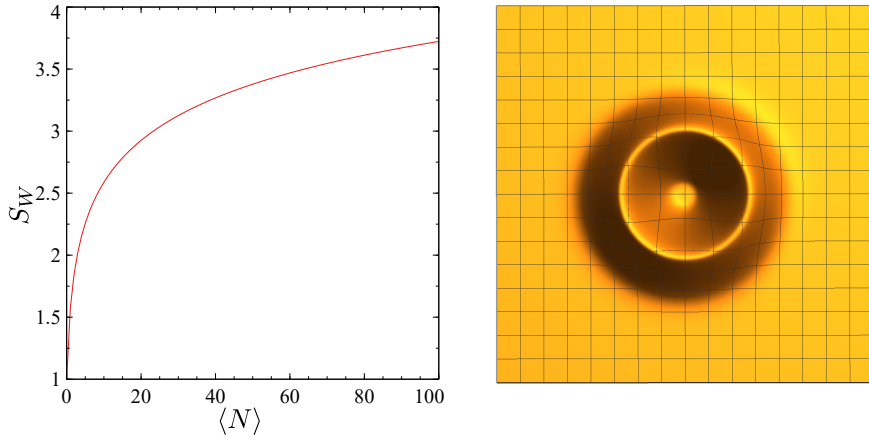


Figure 9.2: The Wehrl entropy of the number eigenstates and the Husimi representation for  $n = 2$ .

## Thermal Radiation

Thermal radiation is produced when a photon source is coupled to a heat bath. The resulting state is the incoherent sum of energy eigenstates, weighed by the Boltzmann factor

$$\hat{\rho} = \frac{1}{Z} \sum_n e^{-\beta n \omega} |n\rangle \langle n|, \quad \text{with } Z = \frac{1}{1 - e^{-\beta \omega}}, \quad (9.44)$$

where we used  $\beta_T$  as the thermodynamic temperature. The Q representation of the state is [143]

$$Q(\alpha) = (1 - \xi) \exp(1 - (1 - \xi)|\alpha|^2) \quad (9.45)$$

where  $\xi = \exp(-\beta_T \omega) = \langle N \rangle / (\langle N \rangle + 1)$ . The entropy is then

$$S = 1 + \ln(1 + \langle \hat{N} \rangle). \quad (9.46)$$

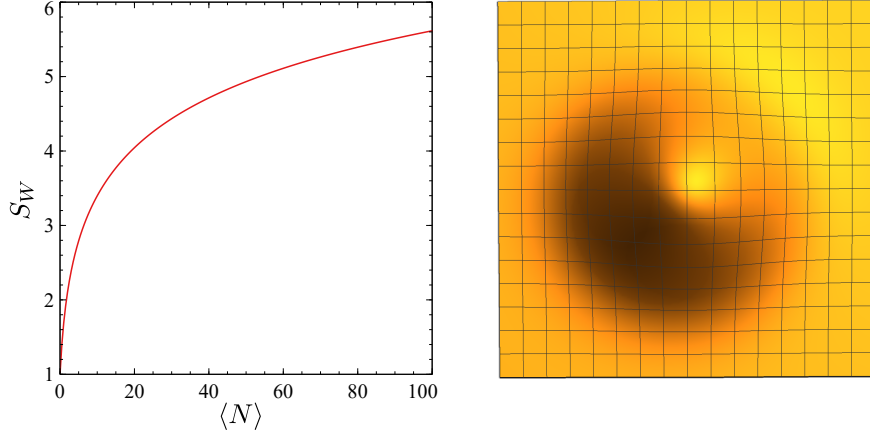


Figure 9.3: The Wehrl entropy of the thermal radiation and its Husimi-Q representation.

### Displaced Photon-Number States

Such a state is defined using the displacement operator  $D(\beta)$  as follows

$$|\alpha', n\rangle = \exp(\beta\hat{a}^\dagger - \beta^*\hat{a})|n\rangle. \quad (9.47)$$

Given that

$$\langle\beta, n|\alpha\rangle = e^{i\text{Im}(\alpha\beta^*)} \langle n|D(\alpha - \beta)|0\rangle, \quad (9.48)$$

which may be seen from the Baker–Campbell–Hausdorff formula, the Q representation is given by [144]

$$Q(\alpha) = \frac{|\alpha - \beta|^{2n}}{n!} \exp(-|\alpha - \beta|^2). \quad (9.49)$$

We see then that  $\beta$  just becomes a shift in phase space, w.r.t. the original Husimi distribution. The entropy is, however, calculated from an integral over the whole complex plane, and will thus be independent of this shift  $\beta$ . Thus the entropy is again given by

$$S_W = 1 + n + \ln(n!) - n\left(-\gamma + \sum_{i=1}^n \frac{1}{i}\right). \quad (9.50)$$

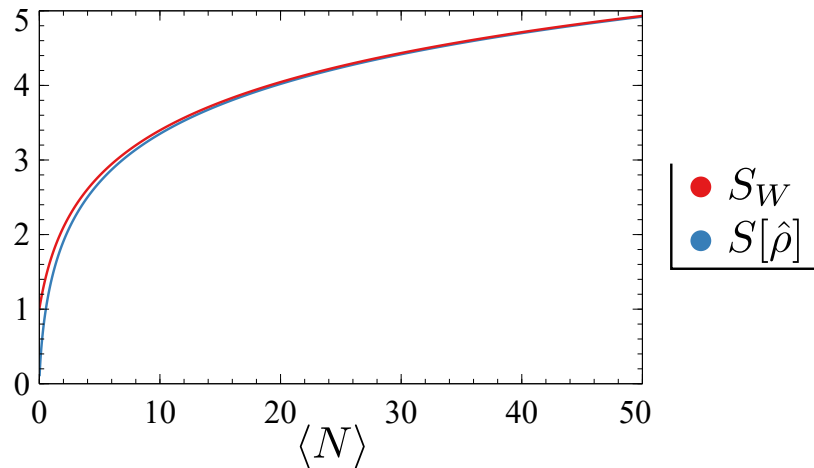


Figure 9.4: The Wehrl and the Von Neumann entropy of the thermal state.

However, we now have that  $\langle N \rangle = n + |\beta|^2$ , which we may substitute into the entropy to obtain

$$S_W = 1 + \langle \hat{N} \rangle - |\beta|^2 + \ln \left( \left( \langle \hat{N} \rangle - |\beta|^2 \right)! \right) - \left( \langle \hat{N} \rangle - |\beta|^2 \right) \left( -\gamma + \sum_{i=1}^n \frac{1}{i} \right). \quad (9.51)$$

From the plots shown in the figures 9.2-9.4, we see that the Wehrl entropy of the thermal distribution is the highest for any  $\langle N \rangle$ . We also saw that for the displaced states, the entropy is diminished as the coherent component grows. At the same time, the entropy grows as  $\langle N \rangle$  grows, provided we keep  $|\beta|^2$  constant. This implies that  $|\nu|^2$  gets bigger, which controls the squeezing of the TPCS. The displacement of a state is generally associated with classical behavior of the state, e.g. the sinusoidal oscillation of the mean, while squeezing of the uncertainty is strictly non-classical behavior. We see similar behavior for the displaced number states, where the coherent component decreases the entropy at a given  $\langle N \rangle$ , while the non-classical behavior increases with  $\langle N \rangle$  at fixed  $|\beta|^2$ . From these results, it seems that in this setting the Wehrl entropy may indeed be a reasonable measure for non-classicality, where a higher Wehrl entropy at a given  $\langle N \rangle$  denotes more non-classical behavior.

Orlowski concludes based on these results that the Wehrl entropy is a good measure for the strength of the coherent component [141]. However,  $|\beta|^2$ ,  $\langle N \rangle$  and  $S_W$  depend on each other is different for displaced squeezed states and displaced number states. As such, it is not possible to extract the coherent component from just the Wehrl entropy and average number of photons  $\langle N \rangle$ , without additional knowledge of the underlying state.

Moreover, what we have seen from the above calculations is that for the thermal state, the Wehrl entropy is maximum for any  $\langle N \rangle$ , which would be expected from thermodynamical considerations. Interestingly however, is that for all states we have discussed, except the thermal state, the Von Neumann entropy is zero, and yet the Wehrl entropy may be made arbitrarily large by increasing  $\langle N \rangle$ . Still, for the thermal state the entropies behave similarly. This may be seen from the following: Consider that the Von Neumann entropy is given by

$$S\left[\frac{1}{Z}e^{-\beta n\omega} |n\rangle \langle n|\right] = \frac{\beta\omega e^{-\beta\omega}}{1 - e^{-\beta\omega}}. \quad (9.52)$$

Using again that  $\exp(-\beta\omega) = \langle N \rangle / (1 + \langle N \rangle)$ , we get

$$S\left[\frac{1}{Z}e^{-\beta n\omega} |n\rangle \langle n|\right] = \ln(1 + \langle N \rangle) + \langle N \rangle \ln \frac{1 + \langle N \rangle}{\langle N \rangle}. \quad (9.53)$$

For large  $\langle N \rangle$ , the second term goes to 1, which means that both entropies have the same limiting behavior w.r.t.  $\langle N \rangle$ , which may also be seen in figure 9.4. Thus, for thermal states with large  $\langle N \rangle$ , the Wehrl entropy does seem to be a good approximation of the Von Neumann entropy.

## 9.4 Summary and Concluding Remarks

In this chapter, we have discussed entropies of quantum phase space distributions. The first entropy we have discussed is the linear entropy for the Wigner entropy. There were a few reasons for introducing this particular entropy for Wigner distributions. First, the Wigner distributions are not positive definite, so immediately, applying the BGS entropy and most Rényi entropies is not meaningful due to the logarithm. The linear entropy fixes this, since it depends on the square of the Wigner distribution, it is invariant under unitary time evolution and it has a direct connection with the linear entropy of the density matrix representing the same state. This correspondence shows us that the linear entropy is a measure of the mixedness of a state. Interest in it comes mainly from quantum optics applications, which are usually most naturally discussed using phase space representations. In this field, it may be used, for example as a decoherence measure. Finally, we have discussed its behavior under entropy maximization and course graining.

After that, we have discussed the Wehrl entropy, also sometimes denoted the classical entropy. This entropy is defined specifically on the Husimi distribution, which may be obtained by smoothing the Wigner distribution of a state with a minimal uncertainty Gaussian. Due to this smoothing, the Husimi distributions are positive definite, and as such, a direct application of the differential entropy is possible. The Wehrl entropy was originally introduced as a classical approximation of the von Neumann entropy [2]. Although, interestingly, it has also found applications as a measure of non-classicality. We have discussed applications of the Wehrl entropy in particular for optical states. This is because there is a particular interest in non-classical behavior of states in this field. We have seen that, as expected, the Wehrl entropy gives non-zero values even for pure states, since they are inherently delocalized in phase space. For the displaced states we discuss, we saw that the Wehrl entropy for a given average number of particles, is lower when the state is displaced, which causes "classical" behavior. Moreover, we found that specifically for thermal states, the Wehrl entropy does indeed approximate the Von Neumann entropy well, as  $\langle N \rangle$  gets larger.

## Chapter 10

# Entropy and Parton Distributions

The structure of protons, and hadrons in general, is a topic that has enjoyed extensive research over the years. Many of the difficulties of this field originate from the fact that quantum field theories are at the moment best understood perturbatively in terms of interacting particles using Feynman diagrams, which require a small (running) coupling constant to work. In our current model of hadrons, the standard model, the coupling constant for the relevant forces within the proton, described by quantum chromodynamics (QCD), vanishes asymptotically for very high energies. As such, we can only cleanly probe hadrons at exceedingly high energies, where at the same time their internal interactions become negligible. At the moment our description of hadrons relies on a few separate models which are assumed to hold in certain parameter regimes, based on fits to experimental data.

One such model is the parton model, which was initially developed by Bjorken and Feynman. Its inception was a consequence of scattering experiments at the time. In high energy proton proton collisions, where pions are readily produced, it was found that these pions mostly had their momentum collinear with the collision axis. This led to the idea of protons being made up of cloud-like constituents that by assumption could not easily absorb a large invariant momentum transfer, which would be needed for larger transverse momenta [122, Ch. 14]. Subsequent experiments of high energy electron scattering, also called deep inelastic scattering (DIS) were conducted. The benefit of using electrons to probe the proton is that the total energy exchange to the proton may easily be calculated by measuring the momentum of the outgoing electron. These experiments revealed that in these setting, protons instead showed behavior that was in line with what would be expected from the scattering of elementary particles. To unify both views, the parton model postulated that the proton was made up of loosely bound partons which could interact with the electrons as elementary particles through the electromagnetic interaction, while interacting with each other mainly through the strong force. Since the strong force only is “strong” at large distances or low energy, they are generally unable to give each other large transverse momentum. In this model, there only is a dependence on a single kinematic, dimensionless parameter called the Bjorken  $x$  or  $x_B$ . This dependence accounts for, for example, scattering rates for different angles. The dependence on this dimensionless parameter, while being independent of the separate energy scales in the process, is called Bjorken scaling. The dependence on  $x_B$  is contained in the so called parton distribution function, which is a classical distribution that determines the probability for the electron to interact with a parton with collinear momentum fraction  $x_B$ .

In the modern day, the parton model is understood through QCD, in which the parton model may be seen as the zeroth order. QCD offers subsequent quantum corrections in the high energy limit. This also predicts corrections to the so called scaling behavior, where the parton distribution function is

no longer only a function of  $x_B$ , but is also dependent on the energy transfer from the electron to a quark.

The parton distribution functions are in a sense a peculiar feature of QCD. If we boost to the so called infinite momentum frame then the interactions between the partons become negligible, and they may be described using classical distributions. In this way it seems like the partons may be characterized by an incoherent sum of states, which would indicate a mixed state. However, going from the hadron rest frame, where it is in a pure state, to the infinite momentum frame, is done through the Lorentz boost, which does not change the pureness of a state. Quite recently, there has been an increased interest in entropies in the context of QCD and the parton model, in hope of getting a better understanding of the properties of partons and their distribution functions. Since the hadrons in DIS are probed with quite short wavelength, it may be possible that the incoherent description of partons emerges as a consequence of the partial probing of the hadron, e.g. as a consequence of entanglement, which would be characterized by non-zero Von Neumann entropy. This is Kharzeev *et al.*'s view [145], which we will discuss and contrast to other works in this chapter. However, first we will discuss the parton model in a bit more detail. We will first discuss it, mostly as it was formulated originally. After that we will discuss qualitatively how the parton model fits within QCD.

## 10.1 DIS and the Parton Model

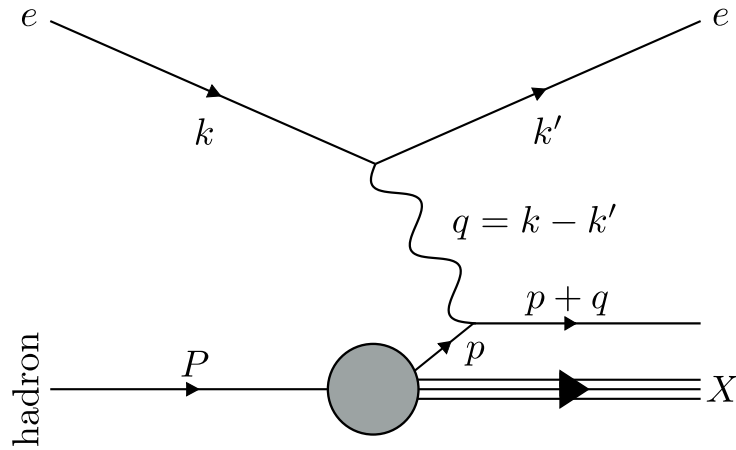


Figure 10.1: The Feynman diagram representing DIS

In DIS we use the electron to probe, most importantly because it interacts with the partons through the electrodynamic interaction. This allows us to calculate the energy transfer from the electron to the parton, from the momentum of the outgoing electron which may easily be measured. Since, by definition, the energy exchange is very high in DIS, we may assume the particles to be massless. The cross sections in QED can be conveniently written using the Mandelstam variables,  $\hat{s}, \hat{t}, \hat{u}$ , which obey  $\hat{s} + \hat{t} + \hat{u} = 0$  for collisions of massless particles. They are given by

$$\begin{aligned}\hat{s} &= (k + p)^2 \\ \hat{t} &= (k - k')^2 \\ \hat{u} &= (k - p')^2\end{aligned}\tag{10.1}$$

We can write the invariant matrix element that characterizes the QED scattering process in terms of the Mandelstam variables [122, Ch. 14]

$$\frac{1}{4} \sum_{\text{spins}} |\mathcal{M}|^2 = \frac{8e^4 Q_i^2}{\hat{t}^2} \left( \frac{\hat{s}^2 + \hat{u}^2}{4} \right),\tag{10.2}$$

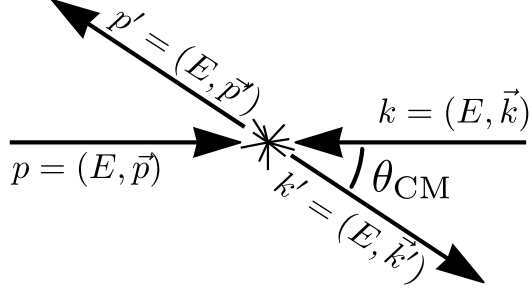


Figure 10.2: Diagram of the collision in the CM frame.

where  $Q_i$  is the quark charge in units of  $e$ . The differential cross section in the center of mass frame may be calculated from this matrix element and is given by

$$\frac{d\sigma}{d\cos\theta_{\text{CM}}} = \frac{\pi\alpha^2 Q_i^2}{\hat{s}} \left( \frac{\hat{s}^2 + (\hat{s} + \hat{t})^2}{\hat{t}^2} \right). \quad (10.3)$$

The constant  $\alpha$  is the fine structure constant. It can be seen that  $q^2$ , from figure 10.1, is space-like, since

$$\begin{aligned} k - k' &= q \\ \implies -2k \cdot k' &= q^2. \end{aligned} \quad (10.4)$$

Then, since all particles are assumed to be massless, the invariant four momenta in the center of mass frame must be equally split between the parton and electron, before and after the collision. We find that

$$k \cdot k' = k_0^2(1 - \cos\theta_{\text{CM}}) \geq 0, \quad (10.5)$$

and therefore  $q^2$  is negative or equal to zero. We will then write  $q^2 = -Q^2 = \hat{t}$ . We can write the cross section in terms of  $Q^2$  using  $Q^2 = \hat{s}(1 - \cos\theta_{\text{CM}})$ ,

$$\frac{d\sigma}{d\hat{t}} = \frac{2\pi\alpha^2 Q_i^2}{\hat{s}^2} \left( \frac{\hat{s}^2 + (\hat{s} + \hat{t})^2}{\hat{t}^2} \right) \quad (10.6)$$

In the model we assume that the momentum of the parton is sufficiently collinear with the momentum of the hadron, which translates to saying  $\vec{p} = x\vec{P}$  with  $0 < x < 1$ . Then

$$\hat{s} = 2p \cdot k = 2xP \cdot k = xs \quad (10.7)$$

where  $s$  denotes the center of mass energy squared. If we assume elastic scattering of the electron from the parton with which it reacts, we can determine  $x$  from the electron momentum. If we again assume massless particles, then

$$0 \approx p'^2 = (p + q)^2 = 2p \cdot q + q^2 = 2xP \cdot q - Q^2. \quad (10.8)$$

Thus we have

$$x = \frac{Q^2}{2P \cdot q} \equiv x_B, \quad (10.9)$$

which is the aforementioned Bjorken  $x$ . We also assume that the proton has some structure independent of the energy it is probed, thus we should be able to write equation 10.6 at constant  $x$  as

$$\frac{d^2\sigma}{dx dQ^2} = \sum_i f_i(x) Q_i^2 \frac{2\pi\alpha}{Q^4} \left( 1 + \left( 1 - \frac{Q^2}{xs} \right) \right) \quad (10.10)$$

The so called Bjorken scaling was confirmed in experiment to hold reasonably well to energies down to about 1 GeV [146]. Qualitatively Bjorken made the argument that for high enough energies,

the interaction between the photon and constituent parton happens so quickly compared to the characteristic time over which the partons interact with each other, that the inner interactions of the proton can be ignored. Thus, the partons in this picture seem like free particles for the very small duration over which the interactions happen.

However, this parton model picture raises new questions. The forces in quantum field theory always have an effect, and do not just vanish for small enough time scales. As such, the picture of free partons at first seemed at odds with the earlier known principles of field theory, being that forces generally seemed to become stronger at shorter distances. This is because the strong interaction was instead known to be stronger than the EM interaction at short distances. The resolution to this problem came in the form of the renormalization group in non-abelian gauge theories. These theories can possess the quality that the running coupling constant goes to zero when the energy becomes infinite, which is called asymptotic freedom. The leading theory we now have for the nuclear force, the strong force, is quantum chromodynamics, which exactly shows this behavior. In the next section we will briefly investigate this theory and its relation to the parton model.

### 10.1.1 QCD and Corrections to Bjorken Scaling

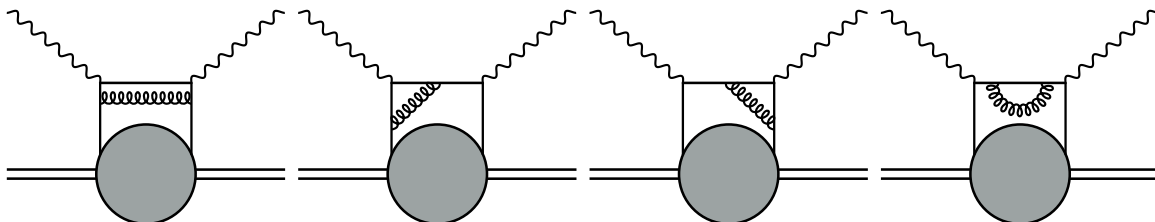


Figure 10.3: The so called handbag diagrams representing the corrections to DIS. The curled lines represent gluons, so these diagrams are first order in  $\alpha_s$ .

The parton model, and corrections to it, are now understood through the field of quantum chromodynamics. We will very briefly explain where this model came from. To explain all the different structure of the strongly interacting particles Gell-Mann and Zweig in 1963 proposed an explanation in terms of the now familiar Fermions called quarks [122, Ch. 17]. They first introduced the up ( $u$ ), ( $d$ ) and strange ( $s$ ). To explain further hadrons found in experiment later the three species charm ( $c$ ), bottom ( $b$ ) and top ( $t$ ) were added. To explain the integer charge of baryons they had to have fractional charge,  $+2/3$  for the  $u, c, t$  quarks and  $-1/3$  for the  $d, s, b$  quarks. Later on, an extra ad-hoc quantum number called color, was added. This was needed to make the baryon wavefunctions totally anti-symmetric. Without this symmetry, the Fermionic quarks would thus not obey Fermi-Dirac statistics. However, the small effect of the strong force at large energies was still not explained. The answer turned out to be the  $SU(3)$  symmetry into a gauge symmetry, which is now how the standard model expresses this color symmetry. This gives rise to the force carriers of the strong force; the vector bosons called gluons. Indeed, the QCD predicted evolution of the strong force coupling constant  $\alpha_s$  has been shown to possess asymptotic freedom. Thus the usual Feynman diagram method only works at high energies, where the coupling constant is small. It has since been shown for strong coupling, that the only finite energy states are those that are singlets of color. This has as a consequence that the only simple (no more than three quark) states that are of finite energy must be three quarks or anti-quarks or a quark and an anti-quark. Thus, the initially ad-hoc assumption of quarks, to explain the spectrum of hadrons, eventually turned out to consequence of the  $SU(3)$  gauge color coupling in QCD.

Since the base interaction of DIS is through the exchange of a photon, in QCD, the parton model naturally finds its home as the zeroth order in  $\alpha_s$ . Thus the quarks and gluons are identified as the partons that make up the hadrons. Moreover, due to the smallness of  $\alpha_s$ , exchanging a large transverse momentum is suppressed, since this has to be done through gluon exchange between the constituents. Therefore the assumption that  $p = xP$  remains a good assumption. However, QCD



predicts additional corrections in orders of  $\alpha_s$ . These contributions come from the Feynman diagrams of figure 10.3. These terms violate the Bjorken scaling of the parton model, such that the actual parton distribution may be written as  $f(x, Q^2)$ , where  $x = f_x(x_B, Q^2)$ ,  $f_x$  being some function. In DIS,  $x_B$  scales like  $x_B \sim Q^2/(Q^2 + s)$ . The parton model may be considered the limit where  $x$  is fixed and of the order of  $\sim 10^{-1}$ ,  $s$  is finite and fixed and  $Q^2 \rightarrow \infty$  [147]. For finite but large  $Q^2$ , the  $Q^2$  dependence of the parton distribution is described by the DGLAP equation. However,  $x$  must not be close to zero. This is because QCD also describes interactions with gluons through quark pair creation. As  $x$  becomes smaller, the gluon distribution, denoted  $xG(x)$ , grows. This  $x$  dependence for small  $x$  is given by the linear BFKL equation. As the gluon distribution grows, the interactions with the gluons become more and more important. However, the growth predicted by the BFKL equation is too high as  $x$  becomes even smaller, so it is believed that there is a need for non-linear growth equations, one of which is the BK equation. Due to the nonlinear growth, these equations are expected to lead to a moderation of the growth of the parton distributions. When this happens, we have to consider scatterings off of multiple partons rather than one. We will discuss a model that is derived from the BK equation in the next section.

Moreover, it is expected that finally the gluon distribution saturates due to the non-linear growth, as  $x$  goes to zero while  $Q^2$  is fixed and finite (thus  $s \rightarrow \infty$ ). When this happens we enter the so called color glass condensate (CGC) regime. In this region there is no sense in talking about interaction with single gluons at all, and the interactions instead become interactions with a coherent color field. These models have been studied extensively, but all the details won't be necessary for the rest of the considerations in this section. As such, we will only develop these ideas further where needed, to find and understand the entropies of these models.

## 10.2 The Entropy of the Parton Distribution functions

In DIS only a region of the proton is probed, approximately a tube of radius  $1/Q$  and length  $1/(mx)$ . If the entropy of the parton could be interpreted as entanglement entropy of the unprobed region with the probed region, then the non-zero entropy of the parton distribution would explain the apparent incoherent nature of the partons that are described by a classical distribution. It is claimed by Kharzeev *et al.* [145] that the resulting incoherent distribution is indeed the product of entanglement, and the entropy should be interpreted as the Von Neumann entropy.

To discuss this view, we will work with a model that is derived from the BK equation [148]. It may be solved in the so called dipole picture. Though we will not go to deep into the specifics of the model, we will briefly explain qualitatively how it emerges. At larger  $x$ , the energy is higher and thus the partons are effectively frozen in time, while at the same time the coupling constant of the strong force is small. As such it often suffices to consider scatterings of single partons. However, as  $x$  becomes smaller, the assumption that the electron always scatters off of a single parton may not hold anymore. After all, there may be many particles at lower momentum fractions. As such, while the gluon distribution  $xG(x)$  may still be calculated, we should not assume that this can be done from the assumption that the electron scatters from a single parton. Instead, we may look at the problem from the perspective of the emitted photon, which may split into dipoles of the color charge, which may in turn interact with the gluons in the proton [147, §2.5]. Moreover, many subsequent splittings into dipoles may occur, with the probabilities predicted by QCD. This will result in the discrete probabilities  $P_n$  to interact with  $n$  dipoles. We will start from these discrete probabilities and soon show how the gluon distribution may be calculated from it. In general the dipoles may have a size associated with them, but this model emerges when that size is fixed.

First of all, we will introduce the so called rapidity  $Y$ , which is given by  $-\ln x$ , and we will use  $\omega$  to denote the so called BFKL intercept which is related to the growth rate (this may be seen later on in equation 10.15). The following equation describes the growth of the probability to find  $n$  dipoles,  $P_n$ , w.r.t.  $Y$

$$\frac{dP_n(Y)}{dY} = -\omega n P_n(Y) + (n-1)\omega P_{n-1}(Y) \quad (10.11)$$

This equation is a so called cascade equation and may be interpreted as follows. During the cascade at any point there is a probabilistic chance for a dipole to split into two other dipoles, which depends on  $Y$ . The total probability for one splitting to happen obviously depends on some inherent factor, times the number of dipoles present. If we change  $Y$  by an amount of  $dY$ , then  $P_n$  then there are two factors that contribute to the change of  $P_n$ . First of all we have the splitting of  $n-1$  dipoles to  $n$  dipoles, which should be proportional to the number of dipoles present,  $n-1$ , and the probability to have  $n-1$  dipoles at  $Y$  in the first place. This is represented by the positive second term. Second, we have the splitting of  $n$  to  $n+1$  dipoles, which, in the same vein, should be proportional to  $nP_n$ . This is represented by the negative second term. The BFKL intercept gives the correct splitting probabilities for this problem. Figure 10.4 shows this process schematically.

Next, we can solve equation 10.11 by introducing the function

$$Z(Y, u) = \sum_n P_n(Y) u^n, \quad (10.12)$$

with which we will soon show that we can re-write the differential equation into a continuous version. The function  $Z$  is related to the gluon distribution. In fact, if we write

$$Z(Y, u) = \int \frac{dr^2}{2\pi} e^{\vec{u} \cdot \vec{r}} \frac{\bar{Z}(r, Y)}{r^2}, \quad (10.13)$$

such that  $u$  and  $r$  are the lengths of the vector quantities  $\vec{u}$  and  $\vec{r}$ , then we have

$$\bar{N}(r, Y) \sim r^2 x G(x, Y). \quad (10.14)$$

Moreover, it is also easy to show that the gluon distribution may be computed from  $Z$  and  $P_n$  as follows

$$xG(x) = \sum_n n P_n(Y) = \left. \frac{dZ}{du} \right|_{u=1} = e^{\omega Y}. \quad (10.15)$$

We will now re-write equation 10.11 using the definition of  $Z(u, Y)$ , in the following way

$$\frac{\partial Z(Y, u)}{\partial Y} = -\omega u(1-u) \frac{\partial Z(Y, u)}{\partial u} \quad (10.16)$$

We can rewrite this, using the initial conditions, as

$$\frac{\partial Z}{\partial Y} = \omega(Z - Z^2) \quad (10.17)$$

which shows that this model implies non-linear growth of the gluon distribution.

We can see that equation 10.16 is equivalent to 10.11 by writing it in terms of  $P_n$

$$\begin{aligned} \sum_n \frac{dP_n(Y)}{dY} u^n &= -\omega(1-u) \sum_n n P_n(Y) u^n \\ &= \omega \sum_n (-n P_n(Y) u^n + (n+1) P_{n+1}(Y) u^{n+1}) \end{aligned} \quad (10.18)$$

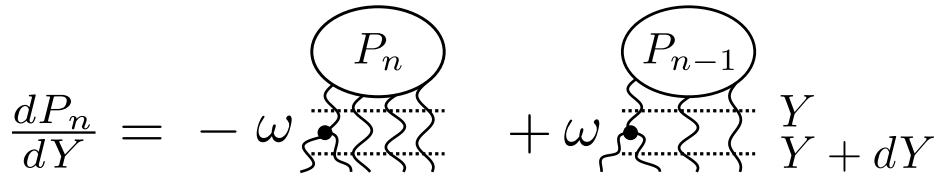


Figure 10.4: A schematic representation of the dipole cascade. Every curled line represents a dipole.

We thus have an equality of polynomials in  $u$ , which means that we can equate the coefficients. The last term having the factor  $u^{n+1}$  thus provides a  $P_{n-1}$  to each order polynomial term of order  $u^n$ . When  $Y = 0$ , or equivalently  $x = 1$  we only have one dipole so  $P_1(0) = 1$  and all other  $P_n$  are zero. If  $u = 1$ , we must have, by definition,  $Z(Y, 1) = 1$ . Using this and equation 10.15, it may be found that the solution to this differential equation is

$$Z(Y, u) = ue^{-\omega Y} \frac{1}{1 + u(e^{-\omega Y} - 1)}. \quad (10.19)$$

This can be written in polynomial form using the geometric series

$$Z(Y, u) = ue^{-\omega Y} \sum_{n=0}^{\infty} u^n (1 - e^{-\omega Y})^n. \quad (10.20)$$

Equating the coefficients of the powers of  $u$ , we find

$$P_n(Y) = e^{-\omega Y} (1 - e^{-\omega Y})^{n-1}. \quad (10.21)$$

The Shannon entropy of this distribution is given by

$$\begin{aligned} S &= - \sum_{n=1} e^{-\omega Y} (1 - e^{-\omega Y})^{n-1} (-\ln(e^{\omega Y} - 1) + n \ln(1 - e^{-\omega Y})). \\ &= \ln(e^{\omega Y} - 1) e^{-\omega Y} \sum_{n=0} (1 - e^{-\omega Y})^n + \ln\left(\frac{1}{1 - e^{-\omega Y}}\right) e^{-\omega Y} \sum_{n=0} (n+1)(1 - e^{-\omega Y})^n \\ &= \ln(e^{\omega Y} - 1) Z(Y, u = 1) + \ln\left(\frac{1}{1 - e^{-\omega Y}}\right) \frac{dZ}{du} \Big|_{u=1} \end{aligned} \quad (10.22)$$

Using the boundary condition that  $Z(Y, u = 1) = 1$  and that  $dZ/du = \exp(\omega Y)$  we obtain the simple form

$$S = \ln(e^{\omega Y} - 1) + e^{\omega Y} \ln\left(\frac{1}{1 - e^{-\omega Y}}\right) \quad (10.23)$$

As  $Y \gg 1$ , or equivalently  $x \ll 1$ , the second term in equation 10.23 goes like  $\sim 1$ , so we can write the entropy as

$$S \sim \omega Y = \ln(xG(x)) \quad (10.24)$$

This shows the very simple relation between the entropy of the distribution  $P_n$  and the gluon distribution. Moreover, Kharzeev *et al.* argue that since the distance probed of the proton is approximately  $L = 1/(mx)$  [145], using again that  $xG(x) = \exp(\omega Y) = (1/x)^\omega$ , we may write the entropy

$$S(x) = \omega \ln\left(\frac{1}{x}\right) = \omega \ln(mL) = \omega \ln\left(\frac{L}{\epsilon}\right), \quad (10.25)$$

where  $m$  is the proton mass and  $\epsilon \equiv 1/m$  is the proton's Compton wavelength. We thus see that we obtain a very similar looking result as those found for the critical spin XY chains, and the result that is known to hold for conformal field theory, all of which we discussed in section 8.3.

Kharzeev *et al.* also derive a result for the  $(3+1)D$  case. We will not derive it ourselves, and focus mostly on the result for the  $(1+1)D$  case, but we will state it here for completeness. The entropy is given by

$$S = \bar{\alpha}_S - e^{-z\bar{\alpha}_s(Y)} \int_0^{\bar{\alpha}_s z Y} \frac{t - te^{-t}}{t} e^{\Xi(t)} dt, \quad (10.26)$$

where  $\Xi(t) = \gamma\Gamma(0, t) + \ln t$  and  $z \equiv \ln(r(Q)^2 Q_s^2)$ . Here  $\gamma$  is the Euler constant,  $\Gamma(0, t)$  is the incomplete gamma function,  $r(Q^2)$  is the characteristic dipole size set by the energy scale and  $Q_s^2$  is the momentum scale at which non-linear corrections become important [149]. Moreover,  $Q_s^2$  is a function of  $x$ . In the large  $Y$  regime this result reduces to

$$S = \bar{\alpha}_S Y \ln(r^2(Q) Q_s^2), \quad (10.27)$$

The result is qualitatively similar, we have again a logarithm of a length scale. However, this time the simple interpretation of the entropy being equal to the logarithm of the probed length is lost, since the prefactor now also depends on  $x$  through  $Y$  and even the  $Q_s^2$  is dependent on  $x$  itself.

Now that we have seen the derivation of the entanglement entropy and the expression itself, we can discuss its implications. To recap, we have calculated the Shannon entropy of a classical distribution that gives the probability to interact with a certain number of gluons. Kharzeev *et al.* go a step further and describe this immediately as the Von Neumann entropy, and many other authors have followed suit. Formally, to show that this is indeed a Von Neumann entropy, we would need to show two things. First of all we would need to show that we can somehow write the state of the proton as a pure state in the following way

$$|\psi\rangle = \sum_n c_n |X_n\rangle |n\rangle \quad (10.28)$$

where  $|n\rangle$  denotes the state of  $n$  dipoles. Tracing over the states  $|X_n\rangle$  would give us a completely mixed density matrix, where the eigenvalues give the probabilities  $P_n$ . Then we would also need to show that the trace over a spatial region within the proton would indeed correspond exactly to tracing over the states  $|X_n\rangle$ . This seems like a formidable task, and not something we will resolve in this work. However, Duan *et al.* have worked on the first task in the related color glass model.

Assuming both of the above conditions can be proven to be true, then the hadrons that are created after the collision happens, would need to directly correspond to the states  $|X_n\rangle$ . We have discussed in chapter 7 that the Von Neumann entropy is the same if we trace over either of the sub-spaces. Essentially what this means is that we have the same uncertainty about the number of dipoles we find, as about the particular final hadron pure state we find. Of course, some interactions with the environment may occur, so the distribution may be different in general. However, if these interactions are limited, or if they are (sufficiently) unitary on the level of the outgoing particles, then the entropy (and the distribution itself) of the outgoing state should be approximately the same as that of the parton distribution. This hypothesis was tested by Kharzeev *et al.* in their original paper, but also very recently by the H1 collaboration. Kharzeev *et al.* report that at center of mass energies of  $\sqrt{s} \approx 7$  TeV, the results are in good agreement with results of the LHC [145, §II.C]. On the other hand, the tests done by H1 collaboration conclude that, for the experiments they have done at  $\sqrt{s} = 319$  GeV and multiple values for  $x_B$  and  $Q^2$ , the parton entropy does not correctly predict the entropy of the final hadron state [150, §6.4]. Their results show that in the lab frame, the hadronic entropy seems to be independent of  $x_B$ , in the region of  $x \sim 10^{-3}$  and a bit below, which agrees with theoretical results from a Monte-Carlo simulation. On the other hand the parton entropy increases as  $x_B$  decreases.

### 10.3 The Entropy of Ignorance

Partly in response to the paper by Kharzeev *et al.*, Duan *et al.* [99] set out to calculate the entanglement entropy in the color glass condensate (CCG) model. In their paper they also go into an important aspect of Kharzeev *et al.*'s paper which we did not yet touch upon. Namely, in DIS only the average number of particles at a certain transverse momentum  $k$  is measured,

$$\text{Tr} \left[ \int \frac{d^2k}{(2\pi)^2} a^\dagger(k) a(k) \hat{\rho} \right], \quad (10.29)$$

where  $Q^2 = k^2$ . Even in more general, where we would measure multi-parton distributions<sup>1</sup> our set of observables would still consist of operators of the type  $a^\dagger(k_1) a(k_1) \cdots a^\dagger(k_n) a(k_n)$ , which are all diagonal in the number basis. Thus these operators only give us information about the

---

<sup>1</sup>With this we mean, instead of measuring only the number of particles at a single momentum, we simultaneously measure multiple different momenta. E.g. we measure the average number of particles with momenta  $k_1, \dots, k_n$  respectively.

diagonal entries of the density matrix. We will elaborate on this claim in this section. The resulting distribution and entropies are similar to the ones we discussed in chapters 5 and 6. As such, it is expected that for most states, even pure states, the entropy of the resulting distribution of outcomes is non-zero. However, this does not mean that this cannot be a Von Neumann entropy. If the state describing the partons is diagonal in the number basis, then this measurement entropy is the Von Neumann entropy.

Duan *et al.* define entropy property similar to the measurement entropy, which they name the entropy of ignorance (EOI). They define it as follows: Let  $\hat{O}_i$  denote a set of observables which is incomplete. Incomplete means that under independent measurements of only these observables, we cannot deduce the complete quantum state. We will show what this means in an example after this. Equivalently, if the set is incomplete, that means there exists a family of density matrices, with parametrization variables  $\alpha_1, \dots, \alpha_n$ , which reproduce the same set of expectation values for that given set of operators. Duan *et al.* then define the entropy ignorance as

$$S_I = \sup_{\alpha_1, \dots, \alpha_n} -\text{Tr} [\hat{\rho}(\alpha_1, \dots, \alpha_n) \ln \hat{\rho}(\alpha_1, \dots, \alpha_n)] \quad (10.30)$$

i.e. as the maximum Von Neumann entropy w.r.t. variation of all  $\alpha_i$ .

The entropy of ignorance is most easily understood when all the operators  $\hat{O}_i$  are simultaneously diagonalizable, like in the case of DIS, and when they probe every diagonal entry in the basis in which they are diagonal<sup>2</sup>. If we think of the problem in the basis where the operators are diagonal, then in that basis, only the diagonal entries of  $\hat{\rho}$ , but all of them, contribute to the trace. Thus, we see that the parameters  $\alpha_i$  correspond to all the off diagonal entries of the density matrix. We know from section 7.1.3, that the Von Neumann entropy is maximal for a given set of diagonal entries, when all the off diagonals are set to zero. Thus the entropy of ignorance is equal to the Von Neumann entropy of the diagonal matrix that produces the correct probability distribution. As an example we will show the entropy of ignorance in a  $2D$  Hilbert space setting

### 10.3.1 Entropy of Ignorance for a Spin System

In the case of the  $2D$  Hilbert space, every density matrix may be parameterized as

$$\hat{\rho} = \begin{bmatrix} a & b \\ b^\dagger & 1-a \end{bmatrix} \quad (10.31)$$

where  $a$  is real and  $0 \leq a \leq 1$  and  $|b| \leq 0$ . In this space, a complete set of observables would be the set of Pauli matrices. In fact, there is a convenient way to parametrize all density matrices on and within the Bloch sphere. In this case we parametrize the density matrices as

$$\hat{\rho} = \frac{1}{2}I + \vec{r} \cdot \vec{\sigma}, \quad (10.32)$$

where  $\vec{r}$  denotes a point within or on the Bloch sphere. We may again use the orthogonality of the Pauli matrices w.r.t. the trace operation so that we see that

$$\text{Tr} [\hat{\rho}\sigma_i] = r_i. \quad (10.33)$$

In other words, the expectation value of the  $i$ -th Pauli matrix probes the  $i$ -th component of the decomposition into Pauli matrices. We immediately see that to determine the density matrix, probing all three Pauli matrices is necessary and sufficient. This shows what is meant with a complete set of operators.

Next, if we only probe  $\sigma_z$  then the off diagonals characterize the expectation values of  $\sigma_x$  and  $\sigma_y$ . The entropy of ignorance is equal to the largest Von Neumann entropy upon variation of the off

---

<sup>2</sup>What we mean by this is that there are no entries on the diagonal that we could tweak, that do not change any of the expectation values of the operators.

diagonals, so this again occurs when they are set to zero. In other words, the entropy of ignorance for any matrix of the form of equation 10.31, given that  $\{\hat{O}_i\} = \{\sigma_z\}$ , is equal to

$$S_I = -a \ln a - (1 - a) \ln(1 - a), \quad (10.34)$$

even if the state is initially pure. In this simple case, where our set only consists of one operator, it is just equal to the measurement entropy  $S[\sigma_z, \hat{\rho}]$  of chapter 5.

### 10.3.2 The Entropy of Ignorance vs the Von Neumann entropy in the CCG model

In this section, we will show the results of the calculation of the entropy of entanglement and ignorance, obtained by Duan *et al.* [99] and Kovner *et al.* [151]. The actual calculations are beyond the scope of this paper, however, and for that we refer to the aforementioned works. We will, however, qualitatively explain where the entropies come from, and how they affect Kharzeev *et al.*'s conclusion. In the color glass condensate, we describe scattering of a hadron at high energies, but small  $x$ . In this case a large fraction of the hadron momentum is still carried by the so called valence partons (be it quark or gluon). They may radiate gluons of low energy that evolve much faster than the valence partons, and have smaller lifetimes [99]. Thus the high energy (hard) valence partons may be seen as approximately static sources of these low energy (soft) gluons. These soft quarks should be thought of more along the lines of a radiation field than as separate particles. In this sense, the state of the color glass condensate has the form

$$|\psi\rangle = |s\rangle \otimes |v\rangle \quad (10.35)$$

where  $|s\rangle$  denotes the soft gluon degrees of freedom and  $|v\rangle$  the valence degrees of freedom. Though it seems so from the notation, the state is not actually a direct product of the valence and soft degrees of freedom, since the soft gluon field has a dependence on the valence quark state [151]. The state  $|s\rangle$  is in fact modeled very similar to the coherent states we have already seen,

$$|s\rangle = \hat{C} |0\rangle = \exp\left(-2 \sum_a \int_k g \rho_a(k) \frac{k_i}{k^2} (a_i^\dagger(k) + a_i(-k))\right) |0\rangle, \quad (10.36)$$

where  $g$  is the coupling constant,  $k_i$  denote the transverse momenta and  $\rho_a(k)$  is the density of gluons per momentum. Moreover, the index  $i$  denotes the vector index going over the transverse momenta and  $a$  is the color index. Duan *et al.* model the wavefunction  $|v\rangle$  in the so-called McLerran-Venugopalan model, the specifics of which we will not go over.

As we have mentioned, the calculation of the CCG entanglement entropy is meant by Duan *et al.* as a proxy to study the questions posed by Kharzeev *et al.*, though they do concede that the comparison is not perfect. Qualitatively they make the following links between the two models. First of all, in both models the wavefunction of the proton is naturally partitioned into two parts, the probed part which consists of degrees of freedom at a certain momentum scale, and the environment that contains the rest of the momentum of the proton. The difference between the two models is that in the parton model the higher momentum degrees of freedom are probed, while in the CCG, the lower momentum degrees of freedom within the proton are probed. Still, the underlying idea of probing the proton at some length scale seems intact, and thus while the analogy is not perfect, the model in this way seems to capture the relevant part of Kharzeev *et al.*'s model, while being analytically calculable.

Alongside the Von Neumann entropy and the EOI, Duan *et al.* also investigate the Rényi entropy of entanglement and the Rényi EOI, both with  $\alpha = 2$ . The Rényi EOI is defined in the same way w.r.t. the Rényi entropy as the regular EOI. We will denote them  $R_2$  and  $R_2^I$  respectively. For the analytic expressions we refer to Duan *et al.* [99]. The Von Neumann entropy may be calculated using the

replica trick that we discussed in chapter 8, and is given by [151]

$$S(k) = \frac{1}{2} (N_c^2 - 1) A_\perp \int \frac{d^2k}{(2\pi)^2} \left[ \ln \left( \frac{\pi Q_s^2}{g^2 k^2} \right) + \sqrt{1 + 4 \frac{\pi Q_s^2}{g^2 k^2}} \ln \left( 1 + \frac{g^2 k^2}{2\pi Q_s^2} + \frac{g^2 k^2}{2\pi Q_s^2} \sqrt{1 + 4 \frac{\pi Q_s^2}{g^2 k^2}} \right) \right]. \quad (10.37)$$

Here, the factor  $N_c^2$  comes from the trace over all colors,  $A_\perp$  is an area factor that arises from putting the system in a box and  $q$  denotes the transverse momenta of the soft gluon fields. On the other hand, the EOI is given by

$$S_I = -\frac{1}{2} (N_c^2 - 1) A_\perp \int \frac{d^2k}{(2\pi)^2} \sum_{m,n} \left[ (1-B) \frac{(m+n)}{m!n!} \left( \frac{B}{2} \right)^{m+n} \ln \left( (1-B) \frac{(m+n)}{m!n!} \left( \frac{B}{2} \right)^{m+n} \right) \right], \quad (10.38)$$

where

$$B = \left( 1 + \frac{g^2 k^2}{\pi Q_s^2} \right). \quad (10.39)$$

The ratios between the Von Neumann EOI and Von Neumann entanglement entropy densities,  $\partial S^I / \partial q$ , are plotted in figure 10.5, alongside the same ratio for the Rényi entropies. We see that in general the entanglement entropies may be very different compared to their ignorance counterparts, though the ratio goes to 1 in the UV. Duan *et al.* conclude that this is due to the fact that at higher momenta, the state is mostly dominated by the vacuum. Thus there is mostly only one pure state, which is inherently diagonal in the number basis, which contributes to the entropies. Thus, in a sense, this is only a trivial correspondence. In this region the entropy density also goes to zero, so there is only a small contribution from the UV regions.

Lastly Duan *et al.* also discuss an interesting special case where the gluon color distribution  $\rho_a(k)$  is fixed, and there is no entanglement at all. Here the Von Neumann entropy is just zero but the entropy of ignorance is non-zero [99, §V].

Of course, these results do not prove that Kharzeev *et al.*'s parton entropy isn't an entanglement entropy. In regions of higher  $q$ , the difference entanglement entropy density and EOI density starts to become negligible. In general though, they are different. What this does show is that one has to be careful when applying entropy quantities, since we have shown that some may have quite different interpretations, even if their appearance is similar.

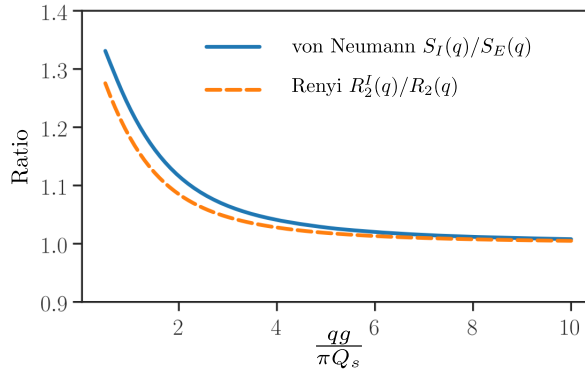


Figure 10.5: The ratios between the entropy densities of ignorance and their entanglement counterparts. For high momenta we see that they approximately become equal. Moreover, the Rényi entropy density with  $\alpha = 2$ ,  $R_2(k)$ , shows very similar behavior compared to its ignorance counterpart,  $R_2^I(k)$ . This image is taken from Duan *et al.* [99].

## 10.4 Summary and Concluding Remarks

In this section, we have discussed the entropy of parton distributions as proposed by Kharzeev *et al.* [145]. The entropy is found by considering the scattering of a set number of gluons, which allows us to formulate a probability distribution of the number of gluons we interact with at a certain momentum. In other words, we find the particle number distribution at a certain momentum, and its entropy. Kharzeev *et al.* interpret this entropy as an entanglement entropy, originating from the entanglement of the probed region within a hadron and the unprobed region. This solves the conceptual problem of the partons being fully described by a classical distribution. Moreover, the unprobed region at the time of interaction, has the same amount of entanglement with the probed region. Thus if there is little entanglement with the environment between the time of the collision and measurement, we may expect that the distribution of outgoing hadrons is the same as that of the partons, due to the Schmidt decomposition. From what we have discussed, it seems that this last claim may not hold, at least for small  $x$ , based on the results obtained by the H1 group.

We have also discussed Duan *et al.*'s view on the entropy of parton distributions [99], along with numerical results they have obtained within the CCG. First of all, they introduce a different entropy, similar to the entropies we discussed in chapters 5 and 6. This entropy, which they call the entropy of ignorance, shows that we may assign an entropy to a set of measurements, based on our ignorance of the actual quantum state, rather than an inherent entropy like the entanglement entropy. Since the DIS measurement probes particle number, which is represented by a set of simultaneously diagonalizable set of operators, we see that we only probe the diagonals of the density matrix. This means that the entropy of the parton distributions should only be interpreted as the von Neumann entropy, if the density matrix of the state is already completely diagonal in that basis. From this, we conclude that it is not possible yet to say that the parton distribution entropy is an entanglement entropy. However, at the same time, Kharzeev *et al.*'s conclusion is not ruled out. If the unprobed part of the hadrons typically undergo a lot of entanglement with the environment before they are measured, then it is expected that their final distribution would change. Moreover, Duan *et al.*'s calculations are in the related and similar, but still different, CCG modes. Because of this it is not possible to say with certainty if the results do indeed generalize to the parton model.



# Conclusion

In this thesis, we set out to understand the different forms and interpretations of the entropy. We started out with the classical entropies of thermodynamics, statistical mechanics and information theory. In their most basic form, entropies are functions of the probabilities of a probability distribution. However, the entropies have been adopted into many different fields, and as a consequence have gotten a lot of different interpretations. In particular, the fields of statistical mechanics and information theory can be seen as the origins of modern interpretations of the entropy. In the former, the entropy is viewed as an extensive measure of the size of phase space that a dynamical system can visit. In the latter, the entropy is a measure of the uncertainty that we have about the outcome of a random process. Here, the entropy is based on a set of properties that characterize a reasonable measure of uncertainty. We have also seen that both of these interpretations leave room for other quantities, based on the role that they are meant to fulfill. In statistical mechanics, complex systems may behave in a way that breaks the extensivity of the entropy, which at the same time breaks the thermodynamic description. In information theory, the entropy may not be useful if the properties that it satisfies are not desirable for specific applications, like expressing the uncertainty, or in statistical inference. In both cases, different entropies may be defined. We discussed a set of entropies that satisfy only the first three Shannon-Khinchin axioms, called the generalized entropies. These entropies may be extensive where the BGS is not, like the Tsallis entropy, or may, for example employ a different averaging technique, to give different weights to different probabilities, like the Rényi entropy. However, there are so many information measures, that we did not get to review all of them, like, for example, the Fisher-information measure or the Kolmogorov-Sinai entropy for dynamical systems. We also did not discuss the spin-off quantities of the entropies, like the relative and conditional entropies, and their significance and applications in information theory and other theories. Thus, there are still more entropy quantities to explore.

After this, we have used the entropy quantities in different fields of quantum physics. We started off with the entropies as an uncertainty measure in quantum mechanics, similar to how the standard deviation is currently used. In discrete Hilbert spaces, we have seen a clear advantage, mostly due to the uncertainty principles that are based on the sum of entropies, which are state independent. In continuous systems, the advantage of the entropy was not always clear. For example, for the harmonic oscillator, the standard deviations were generally easier to compute, and could even be computed analytically without analytic solutions for the wavefunctions. The entropy on the other hand, requires those solutions and, at the same time, requires numerical integration. In multiple dimensions, the entropy does seem a more transparent measure of overall uncertainty, but we have not found any clear uses for it ourselves.

After that, we discussed the quantum generalizations of the entropies. Generally, in quantum mechanical terms, these measures may be thought of as measures of mixedness, analogously to the measures of uncertainty in the classical theories. Here, we specifically measure the classical uncertainty in outcome of a quantum measurement. Within the field of quantum mechanics, we have discussed several applications of the entropies, and in particular of the von Neumann entropy. They may be used as measures of the mixedness of states, as measures of entanglement, measures of criticality, and the von Neumann entropy in particular may be used as a thermodynamical entropy for thermal quantum systems, and perhaps even for complex systems such as black holes. Most of these

applications are considered as part of the field now commonly referred to as quantum information theory. Although we have seen and discussed some of the applications of the entropies within this field, we have only scratched the surface. Among other things, quantum information theory governs quantum communication and cryptography, general quantum resource theories and its tools intensively used in other fields like quantum computation or within theoretical physics, like in the study of complex systems and black hole physics. Hopefully, this thesis has served as a stepping stone into the field of quantum information theory and in one's further research into the entropies.

# Bibliography

- [1] M. Tribus and E. C. McIrvine, “ENERGY AND INFORMATION,” *Scientific American*, vol. 225, no. 3, pp. 179–190, Nov. 1971, ISSN: 00368733, 19467087. [Online]. Available: [http://www.esa.lq.usp.br/lepse/imgs/conteudo\\_thumb/Energy-and-Information.pdf](http://www.esa.lq.usp.br/lepse/imgs/conteudo_thumb/Energy-and-Information.pdf).
- [2] A. Wehrl, “General properties of entropy,” *Reviews of Modern Physics*, vol. 50, no. 2, pp. 221–260, 1978, ISSN: 00346861. DOI: 10.1103/RevModPhys.50.221.
- [3] J. Uffink, “Compendium of the Foundations of Classical Statistical Physics,” *Philosophy of Physics*, pp. 923–1074, 2007. DOI: 10.1016/B978-044451560-5/50012-9.
- [4] P. G. Wright, “Entropy and disorder,” *Contemporary Physics*, vol. 11, no. 6, pp. 581–588, Nov. 1970, ISSN: 0010-7514. DOI: 10.1080/00107517008202196. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00107517008202196>.
- [5] D. Styer, “Entropy as Disorder: History of a Misconception,” *The Physics Teacher*, vol. 57, no. 7, pp. 454–458, 2019, ISSN: 0031-921X. DOI: 10.1119/1.5126822.
- [6] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948, ISSN: 15387305. DOI: 10.1002/j.1538-7305.1948.tb00917.x.
- [7] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical Review*, vol. 106, no. 4, pp. 620–630, May 1957, ISSN: 0031899X. DOI: 10.1103/PhysRev.106.620. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRev.106.620>.
- [8] F. Mandl, *Statistical Physics*. Wiley, 2013, ISBN: 9780471566588. [Online]. Available: <https://books.google.nl/books?id=ZMSNPFrfaU8C>.
- [9] K. Huang, *Statistical Mechanics*, 2nd editio. John Wiley & Sons, 2008, ISBN: 0471815187. [Online]. Available: <https://books.google.nl/books?id=ZH18HLk-K3AC>.
- [10] C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems*. Cambridge University Press, Jul. 1993, ISBN: 9780521433679. DOI: 10.1017/CB09780511524585. [Online]. Available: <https://www.cambridge.org/core/product/identifier/9780511524585/type/book>.
- [11] C. R. de Oliveira and T. Werlang, “Ergodic hypothesis in classical statistical mechanics,” *Revista Brasileira de Ensino de Física*, vol. 29, no. 2, pp. 189–201, 2007, ISSN: 01024744. DOI: 10.1590/s1806-11172007000200003.
- [12] S. Hilbert, P. Hänggi, and J. Dunkel, “Thermodynamic laws in isolated systems,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 90, no. 6, 2014, ISSN: 15502376. DOI: 10.1103/PhysRevE.90.062116. arXiv: 1408.5382.
- [13] P. Buonsante, R. Franzosi, and A. Smerzi, “On the dispute between Boltzmann and Gibbs entropy,” *Annals of Physics*, vol. 375, pp. 414–434, 2016, ISSN: 1096035X. DOI: 10.1016/j.aop.2016.10.017. arXiv: 1601.01509. [Online]. Available: <http://dx.doi.org/10.1016/j.aop.2016.10.017>.
- [14] R. Frigg and C. Werndl, “Entropy: A Guide For the Perplexed,” pp. 29–37, 2011.
- [15] E. T. Jaynes, “Gibbs vs Boltzmann Entropies,” *American Journal of Physics*, vol. 33, no. 5, pp. 391–398, 1965, ISSN: 0002-9505. DOI: 10.1119/1.1971557.

- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, X. Wiley, Sep. 2005, pp. 229–256, ISBN: 9780471241959. DOI: 10.1002/047174882X. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.fmatter>.
- [17] E. T. Jaynes, “Information Theory and Statistical Mechanics. II,” *Physical Review*, vol. 108, no. 2, pp. 171–190, Oct. 1957, ISSN: 0031-899X. DOI: 10.1103/PhysRev.108.171. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRev.108.171>.
- [18] D. Tse, *Information theory Lecture 15*. [Online]. Available: [https://tselab.stanford.edu/mirror/ee376a\\_winter1617/Lecture\\_15.pdf](https://tselab.stanford.edu/mirror/ee376a_winter1617/Lecture_15.pdf).
- [19] E. Witten, “A mini-introduction to information theory,” *Rivista del Nuovo Cimento*, vol. 43, no. 4, pp. 187–227, 2020, ISSN: 18269850. DOI: 10.1007/s40766-020-00004-5. arXiv: arXiv:1805.11965v5.
- [20] P. Ehrenfest and T. Ehrenfest, “The Conceptual Foundations of the Statistical Approach in Mechanics,” *Physics Bulletin*, vol. 11, no. 7, pp. 191–191, Jul. 1960, ISSN: 0031-9112. DOI: 10.1088/0031-9112/11/7/008. [Online]. Available: <https://iopscience.iop.org/article/10.1088/0031-9112/11/7/008>.
- [21] A. Alonso-Serrano and M. Visser, “Coarse graining Shannon and von Neumann entropies,” *Entropy*, vol. 19, no. 5, pp. 1–22, 2017, ISSN: 10994300. DOI: 10.3390/e19050207. arXiv: 1704.00237.
- [22] A. Rényi, “On Measures of Entropy and Information,” in *Proceedings of the fourth Berkeley symposium on mathematics, statistics and probability, volume 1, pages 547–561 (1961)*, Hungarian Academy of Sciences. [Online]. Available: <https://projecteuclid.org/euclid.bsmsp/1200512181>.
- [23] A. Rényi and A. Rényi, “On the Foundations of Information Theory,” *Revue de l’Institut International de Statistique / Review of the International Statistical Institute*, vol. 33, no. 1, p. 1, 1965, ISSN: 03731138. DOI: 10.2307/1401301. [Online]. Available: <https://www.jstor.org/stable/1401301?origin=crossref>.
- [24] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, Jul. 1988, ISSN: 0022-4715. DOI: 10.1007/BF01016429. [Online]. Available: <http://link.springer.com/10.1007/BF01016429>.
- [25] B. C. Brookes and A. I. Khinchin, *Mathematical Foundations of Information Theory*, 346. 1959, vol. 43, p. 309. DOI: 10.2307/3610679.
- [26] J. M. Amigó, S. G. Balogh, and S. Hernández, “A brief review of generalized entropies,” *Entropy*, vol. 20, no. 11, pp. 1–21, 2018, ISSN: 10994300. DOI: 10.3390/e20110813.
- [27] H. Suyari, “Generalization of Shannon–Khinchin Axioms to Nonextensive Systems and the Uniqueness Theorem for the Nonextensive Entropy,” *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1783–1787, Aug. 2004, ISSN: 0018-9448. DOI: 10.1109/TIT.2004.831749. [Online]. Available: <http://ieeexplore.ieee.org/document/1317122/>.
- [28] S. Abe, “Tsallis entropy: How unique?” *Continuum Mechanics and Thermodynamics*, vol. 16, no. 3, pp. 237–244, 2004, ISSN: 09351175. DOI: 10.1007/s00161-003-0153-1.
- [29] I. Csiszár, “Axiomatic characterizations of information measures,” *Entropy*, vol. 10, no. 3, pp. 261–273, 2008, ISSN: 10994300. DOI: 10.3390/e10030261.
- [30] R. Hanel and S. Thurner, “A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions,” *Epl*, vol. 93, no. 2, 2011, ISSN: 02955075. DOI: 10.1209/0295-5075/93/20006. arXiv: 1005.0138.
- [31] H. Touchette, “When is a quantity additive, and when is it extensive?” *Physica A: Statistical Mechanics and its Applications*, vol. 305, no. 1-2, pp. 84–88, 2002, ISSN: 03784371. DOI: 10.1016/S0378-4371(01)00644-6. arXiv: 0201134 [cond-mat].
- [32] B. Lesche, “Instabilities of Rényi entropies,” *Journal of Statistical Physics*, vol. 27, no. 2, pp. 419–422, 1982, ISSN: 00224715. DOI: 10.1007/BF01008947.

- [33] S. Abe, B. Lesche, and J. Mund, “How should the distance of probability assignments be judged?” *Journal of Statistical Physics*, vol. 128, no. 5, pp. 1189–1196, 2007, ISSN: 00224715. DOI: 10.1007/s10955-007-9344-7.
- [34] P. Jizba and T. Arimitsu, “Observability of Rényi’s entropy,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2 2, pp. 1–12, 2004, ISSN: 1063651X. DOI: 10.1103/PhysRevE.69.026128.
- [35] T. Matolcsi and P. Ván, “On the continuity and Lesche stability of Tsallis and Rényi entropies and q-expectation values,” 2009. arXiv: 0910.1918.
- [36] C. Tsallis, *Introduction to nonextensive statistical mechanics: Approaching a complex world*. 2009, pp. 1–382, ISBN: 9780387853581. DOI: 10.1007/978-0-387-85359-8.
- [37] R. Hanel and S. Thurner, “Maximum configuration principle for driven systems with arbitrary driving,” *Entropy*, vol. 20, no. 11, pp. 1–13, 2018, ISSN: 10994300. DOI: 10.3390/e20110838. arXiv: 1809.03601.
- [38] S. Thurner, P. Klimek, and R. Hanel, *Introduction to the Theory of Complex Systems*, 9. Oxford University Press, Nov. 2018, vol. 1, pp. 1689–1699, ISBN: 9780198821939. DOI: 10.1093/oso/9780198821939.001.0001. arXiv: arXiv:1011.1669v3.
- [39] J. Havrda and F. Charvát, “Quantification Method of Classification Processes,” *Kybernetika*, vol. 3, no. 1, pp. 30–35, 1967. [Online]. Available: <http://www.kybernetika.cz/content/1967/1/30/paper.pdf>.
- [40] P. Douglas, S. Bergamini, and F. Renzoni, “Tunable Tsallis distributions in dissipative optical lattices,” *Physical Review Letters*, vol. 96, no. 11, pp. 1–4, 2006, ISSN: 00319007. DOI: 10.1103/PhysRevLett.96.110601.
- [41] C. Tsallis, J. C. Anjos, and E. P. Borges, “Fluxes of cosmic rays: A delicately balanced stationary state,” *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 310, no. 5-6, pp. 372–376, 2003, ISSN: 03759601. DOI: 10.1016/S0375-9601(03)00377-3. arXiv: 0203258 [astro-ph].
- [42] C. Beck, “Dynamical foundations of nonextensive statistical mechanics,” *Physical Review Letters*, vol. 87, no. 18, pp. 180601–1–180601–4, 2001, ISSN: 10797114. DOI: 10.1103/PhysRevLett.87.180601. arXiv: 0105374 [cond-mat].
- [43] I. Bediaga, E. M. Curado, and J. M. De Miranda, “Nonextensive thermodynamical equilibrium approach in  $e+e\rightarrow$ hadrons,” *Physica A: Statistical Mechanics and its Applications*, vol. 286, no. 1, pp. 156–163, 2000, ISSN: 03784371. DOI: 10.1016/S0378-4371(00)00368-X.
- [44] S. Furuichi, “Information Theoretical Properties of Tsallis entropies,” *Journal of Mathematical Physics*, vol. 47, no. 2, pp. 1–20, 2006, ISSN: 00222488. DOI: 10.1063/1.2165744. arXiv: 0405600 [cond-mat].
- [45] S. Abe, G. Kaniadakis, and A. M. Scarfone, “Stabilities of generalized entropies,” *Journal of Physics A: Mathematical and General*, vol. 37, no. 44, pp. 10 513–10 519, 2004, ISSN: 03054470. DOI: 10.1088/0305-4470/37/44/004. arXiv: 0401290 [cond-mat].
- [46] S. Abe, “General pseudoadditivity of composable entropy prescribed by the existence of equilibrium,” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 63, no. 6, pp. 8–10, 2001, ISSN: 1063651X. DOI: 10.1103/PhysRevE.63.061105.
- [47] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, “Generalized privacy amplification,” *IEEE International Symposium on Information Theory - Proceedings*, vol. 41, no. 6, p. 350, 1994, ISSN: 21578095. DOI: 10.1109/ISIT.1994.394668.
- [48] M. Skórski, “Shannon Entropy Versus Rényi Entropy from a Cryptographic Viewpoint,” in, 2015, pp. 257–274. DOI: 10.1007/978-3-319-27239-9\_16.
- [49] P. Jizba and T. Arimitsu, “The world according to Rényi: Thermodynamics of multifractal systems,” *Annals of Physics*, vol. 312, no. 1, pp. 17–59, 2004, ISSN: 00034916. DOI: 10.1016/j.aop.2004.01.002. arXiv: 0207707 [cond-mat].

- [50] A. S. Parvan and T. S. Biró, “Rényi statistics in equilibrium statistical mechanics,” *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 374, no. 19-20, pp. 1951–1957, 2010, ISSN: 03759601. DOI: 10.1016/j.physleta.2010.03.007. [Online]. Available: <http://dx.doi.org/10.1016/j.physleta.2010.03.007>.
- [51] J. F. Lutsko, J. P. Boon, and P. Grosfils, “Is the Tsallis entropy stable?” *Epl*, vol. 86, no. 4, 2009, ISSN: 02955075. DOI: 10.1209/0295-5075/86/40005. arXiv: 0902.4579.
- [52] S. Thurner and R. Hanel, “The Entropy of Non-Ergodic Complex Systems — a Derivation From First Principles,” *International Journal of Modern Physics: Conference Series*, vol. 16, pp. 105–115, 2012, ISSN: 2010-1945. DOI: 10.1142/s2010194512007817.
- [53] S. G. Balogh, G. Palla, P. Pollner, and D. Czégel, “Generalized entropies, density of states, and non-extensivity,” *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, ISSN: 20452322. DOI: 10.1038/s41598-020-72422-8. [Online]. Available: <https://doi.org/10.1038/s41598-020-72422-8>.
- [54] M. Nielsen and I. Chuang, “Quantum Computation and Quantum Information,” *Contemporary Physics*, vol. 52, no. 6, pp. 604–605, Nov. 2010, ISSN: 0010-7514.
- [55] A. Chefles, “Quantum operations, state transformations and probabilities,” *Physical Review A*, vol. 65, no. 5, p. 052314, May 2002, ISSN: 1050-2947. DOI: 10.1103/PhysRevA.65.052314. arXiv: 0109060 [quant-ph]. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.65.052314>.
- [56] J. Pečarić, F. Proschan, and Y. L. Tong, “CONVEX FUNCTIONS, PARTIAL ORDERINGS, AND STATISTICAL APPLICATIONS,” in, 1992, p. 463, ISBN: 9780125492508. [Online]. Available: <https://www.elsevier.com/books/convex-functions-partial-orderings-and-statistical-applications/peajcariac/978-0-12-549250-8>.
- [57] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, “Entropic uncertainty relations and their applications,” *Reviews of Modern Physics*, vol. 89, no. 1, 2017, ISSN: 15390756. DOI: 10.1103/RevModPhys.89.015002. arXiv: 1511.04857.
- [58] E. Schrödinger, “About Heisenberg Uncertainty Relation,” vol. 2, pp. 296–303, 1930. arXiv: 9903100 [quant-ph]. [Online]. Available: <http://arxiv.org/abs/quant-ph/9903100>.
- [59] L. Maccone and A. K. Pati, “Stronger uncertainty relations for all incompatible observables,” *Physical Review Letters*, vol. 113, no. 26, pp. 1–5, 2014, ISSN: 10797114. DOI: 10.1103/PhysRevLett.113.260401.
- [60] D. Deutsch, “Uncertainty in quantum measurements,” *Physical Review Letters*, vol. 50, no. 9, pp. 631–633, 1983, ISSN: 00319007. DOI: 10.1103/PhysRevLett.50.631.
- [61] H. Maassen and J. B. Uffink, “Generalized entropic uncertainty relations,” *Physical Review Letters*, vol. 60, no. 12, pp. 1103–1106, 1988, ISSN: 00319007. DOI: 10.1103/PhysRevLett.60.1103.
- [62] G. C. Ghirardi, L. Marinatto, and R. Romano, “An optimal entropic uncertainty relation in a two-dimensional Hilbert space,” *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 317, no. 1-2, pp. 32–36, 2003, ISSN: 03759601. DOI: 10.1016/j.physleta.2003.08.029. arXiv: 0310120 [quant-ph].
- [63] J. I. De Vicente and J. Sánchez-Ruiz, “Improved bounds on entropic uncertainty relations,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 77, no. 4, pp. 1–5, 2008, ISSN: 10502947. DOI: 10.1103/PhysRevA.77.042110.
- [64] A. Rudnicki, Z. Puchała, and K. Życzkowski, “Strong majorization entropic uncertainty relations,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 89, no. 5, pp. 1–8, 2014, ISSN: 10941622. DOI: 10.1103/PhysRevA.89.052115. arXiv: 1402.0129.
- [65] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, “The uncertainty principle in the presence of quantum memory,” *Nature Physics*, vol. 6, no. 9, pp. 659–662, 2010, ISSN: 17452481. DOI: 10.1038/nphys1734. arXiv: 0909.0950.

- [66] A. Mohajeri and M. Alipour, “Shannon information entropy of fractional occupation probability as an electron correlation measure in atoms and molecules,” *Chemical Physics*, vol. 360, no. 1-3, pp. 132–136, 2009, ISSN: 03010104. DOI: 10.1016/j.chemphys.2009.04.016.
- [67] M. Ohya and D. Petz, *Quantum entropy and Its Use*, 3. Nov. 1993, vol. 19, ISBN: 3-540-54881-5.
- [68] I. Białynicki-Birula and J. Mycielski, “Uncertainty relations for information entropy in wave mechanics,” *Communications in Mathematical Physics*, vol. 44, no. 2, pp. 129–132, 1975, ISSN: 00103616. DOI: 10.1007/BF01608825.
- [69] R. L. Frank and E. H. Lieb, “Entropy and the Uncertainty Principle,” *Annales Henri Poincare*, vol. 13, no. 8, pp. 1711–1717, 2012, ISSN: 14240637. DOI: 10.1007/s00023-012-0175-y. arXiv: 1109.1209.
- [70] I. Białynicki-Birula, “Formulation of the uncertainty relations in terms of the Rényi entropies,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 74, no. 5, 2006, ISSN: 10502947. DOI: 10.1103/PhysRevA.74.052101. arXiv: 0608116 [quant-ph].
- [71] V. Majerník, R. Charvot, and E. Majerníková, “The momentum entropy of the infinite potential well,” *Journal of Physics A: Mathematical and General*, vol. 32, no. 11, pp. 2207–2216, 1999, ISSN: 03054470. DOI: 10.1088/0305-4470/32/11/013.
- [72] A. I. Aptekarev, J. S. Dehesa, P. Sánchez-Moreno, and D. N. Tulyakov, “Rényi entropy of the infinite well potential in momentum space and Dirichlet-like trigonometric functionals,” *Journal of Mathematical Chemistry*, vol. 50, no. 5, pp. 1079–1090, 2012, ISSN: 02599791. DOI: 10.1007/s10910-011-9950-0. arXiv: 1305.4743.
- [73] W. Van Assche, “Entropy of Hermite polynomials with application to the harmonic oscillator,” 1991.
- [74] J. S. Dehesa, A. Martínez-Finkelshtein, and V. N. Sorokin, “Quantum-information entropies for highly excited states of single-particle systems with power-type potentials,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 66, no. 6, p. 7, 2002, ISSN: 10941622. DOI: 10.1103/PhysRevA.66.062109.
- [75] V. Majerník and T. Opatrný, “Entropic uncertainty relations for a quantum oscillator,” *Journal of Physics A: Mathematical and General*, vol. 29, no. 9, pp. 2187–2197, 1996, ISSN: 03054470. DOI: 10.1088/0305-4470/29/9/029.
- [76] U. Leonhardt, *Measuring the Quantum State of Light (Cambridge Studies in Modern Optics)*. 2005, p. 208, ISBN: 0521023521. [Online]. Available: <http://www.amazon.co.uk/Measuring-Quantum-Cambridge-Studies-Modern/dp/0521023521>.
- [77] A. Furusawa, “Quantum States of Light,” in, 2015, pp. 1–67, ISBN: 9784431559580. DOI: 10.1007/978-4-431-55960-3\_1. [Online]. Available: [http://link.springer.com/10.1007/978-4-431-55960-3%7B%5C\\_%7D1](http://link.springer.com/10.1007/978-4-431-55960-3%7B%5C_%7D1).
- [78] Fan Hong-Yi and J. VanderLinde, “Squeezed-state wave functions and their relation to classical phase-space maps,” *Physical Review A*, vol. 40, no. 8, pp. 4785–4788, Oct. 1989, ISSN: 0556-2791. DOI: 10.1103/PhysRevA.40.4785.
- [79] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1957, ISBN: 978-3-662-12871-8. DOI: 10.1007/978-3-662-12869-5.
- [80] L. G. Jiao, L. R. Zan, Y. Z. Zhang, and Y. K. Ho, “Benchmark values of Shannon entropy for spherically confined hydrogen atom,” *International Journal of Quantum Chemistry*, vol. 117, no. 13, 2017, ISSN: 1097461X. DOI: 10.1002/qua.25375.
- [81] P. K. Jakobsen and V. V. Lychagin, “Maximum entropy wave functions,” *Lobachevskii Journal of Mathematics*, vol. 23, pp. 29–56, 2006, ISSN: 19950802.

- [82] I. Białynicki-Birula and L. Rudnicki, “Entropic Uncertainty Relations in Quantum Physics,” in *Statistical Complexity*, Dordrecht: Springer Netherlands, 2011, pp. 1–34, ISBN: 9789048138906. DOI: 10.1007/978-90-481-3890-6\_1. arXiv: arXiv:1001.4668. [Online]. Available: [http://link.springer.com/10.1007/978-90-481-3890-6\\_1](http://link.springer.com/10.1007/978-90-481-3890-6_1).
- [83] J. S. Dehesa, A. Martínez-Finkelshtdein, and J. Sánchez-Ruiz, “Quantum information entropies and orthogonal polynomials,” *Journal of Computational and Applied Mathematics*, vol. 133, no. 1-2, pp. 23–46, 2001, ISSN: 03770427. DOI: 10.1016/S0377-0427(00)00633-6.
- [84] S. R. Gadre, S. B. Sears, S. J. Chakravorty, and R. D. Bendale, “Some novel characteristics of atomic information entropies,” *Physical Review A*, vol. 32, no. 5, pp. 2602–2606, 1985, ISSN: 10502947. DOI: 10.1103/PhysRevA.32.2602.
- [85] E. Romera and J. S. Dehesa, “The Fisher-Shannon information plane, an electron correlation tool,” *Journal of Chemical Physics*, vol. 120, no. 19, pp. 8906–8912, 2004, ISSN: 00219606. DOI: 10.1063/1.1697374.
- [86] R. Nalewajski, *Information Theory of Molecular Systems*. 2006, ISBN: 9780080459745. [Online]. Available: <https://www.elsevier.com/books/information-theory-of-molecular-systems/nalewajski/978-0-444-51966-5>.
- [87] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*. 1955, ISBN: 9781400889921.
- [88] D. Petz, “Entropy, von Neumann and the von Neumann Entropy,” *John von Neumann and the Foundations of Quantum Physics*, pp. 83–96, 2001. DOI: 10.1007/978-94-017-2012-0\_7. arXiv: 0102013 [math-ph].
- [89] E. Chua, “Does Von Neumann Entropy Correspond to Thermodynamic Entropy?” *Philosophy of Science*, 2020, ISSN: 0031-8248. DOI: 10.1086/710072.
- [90] J. M. Deutsch, “Quantum statistical mechanics in a closed system,” *Physical Review A*, vol. 43, no. 4, pp. 2046–2049, 1991, ISSN: 10502947. DOI: 10.1103/PhysRevA.43.2046.
- [91] E. H. Lieb and M. B. Ruskai, “A fundamental property of quantum-mechanical entropy,” *Physical Review Letters*, vol. 30, no. 10, pp. 434–436, 1973, ISSN: 00319007. DOI: 10.1103/PhysRevLett.30.434.
- [92] D. W. Robinson and D. Ruelle, “Mean entropy of states in classical statistical mechanics,” *Communications in Mathematical Physics*, vol. 5, no. 4, pp. 288–300, 1967, ISSN: 00103616. DOI: 10.1007/BF01646480.
- [93] H. Casini, “Geometric entropy, area and strong subadditivity,” *Classical and Quantum Gravity*, vol. 21, no. 9, pp. 2351–2378, 2004, ISSN: 02649381. DOI: 10.1088/0264-9381/21/9/011. arXiv: 0312238 [hep-th].
- [94] T. Hirata and T. Takayanagi, “AdS/CFT and strong subadditivity of entanglement entropy,” *Journal of High Energy Physics*, vol. 2007, no. 2, 2007, ISSN: 10298479. DOI: 10.1088/1126-6708/2007/02/042.
- [95] E. Caceres, A. Kundu, J. F. Pedraza, and W. Tangarife, “Strong Subadditivity, Null Energy Condition and Charged Black Holes,” 2013. DOI: 10.1007/JHEP01(2014)084. arXiv: 1304.3398.
- [96] A. Prudenziati, “Strong subadditivity and holography,” *Physical Review D*, vol. 93, no. 10, pp. 1–26, 2016, ISSN: 24700029. DOI: 10.1103/PhysRevD.93.106003.
- [97] E. H. Lieb and M. B. Ruskai, “Proof of the strong subadditivity of quantum-mechanical entropy,” *Journal of Mathematical Physics*, vol. 14, no. 12, pp. 1938–1941, 1973, ISSN: 00222488. DOI: 10.1063/1.1666274.
- [98] X. Dong, “The gravity dual of Rényi entropy,” *Nature Communications*, vol. 7, no. May, 2016, ISSN: 20411723. DOI: 10.1038/ncomms12472.
- [99] H. Duan, C. Akkaya, A. Kovner, and V. V. Skokov, “Entanglement, partial set of measurements, and diagonality of the density matrix in the parton model,” *Physical Review D*, vol. 101, no. 3, p. 36017, 2020, ISSN: 24700029. DOI: 10.1103/PhysRevD.101.036017. [Online]. Available: <https://doi.org/10.1103/PhysRevD.101.036017>.



- [100] J. Batle, M. Casas, A. R. Plastino, and A. Plastino, “Entanglement, mixedness, and q-entropies,” *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 296, no. 6, pp. 251–258, 2002, ISSN: 03759601. DOI: 10.1016/S0375-9601(02)00271-2.
- [101] S. Abe, “Quantum q-divergence,” *Physica A: Statistical Mechanics and its Applications*, vol. 344, no. 3-4 SPEC. ISS. Pp. 359–365, 2004, ISSN: 03784371. DOI: 10.1016/j.physa.2004.06.003.
- [102] A. E. Rastegin, “Quantum-coherence quantifiers based on the Tsallis relative  $\alpha$  entropies,” *Physical Review A*, vol. 93, no. 3, 2016, ISSN: 24699934. DOI: 10.1103/PhysRevA.93.032136. arXiv: arXiv:1512.06652v3.
- [103] D. Petz and D. Virosztek, “Some inequalities for quantum tsallis entropy related to the strong subadditivity,” *Mathematical Inequalities and Applications*, vol. 18, no. 2, pp. 555–568, 2015, ISSN: 13314343. DOI: 10.7153/mia-18-41. arXiv: 1403.7062.
- [104] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, “Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels,” *Physical Review Letters*, vol. 70, no. 13, pp. 1895–1899, 1993, ISSN: 00319007. DOI: 10.1103/PhysRevLett.70.1895.
- [105] M. A. Nielsen, “Conditions for a class of entanglement transformations,” *Physical Review Letters*, vol. 83, no. 2, pp. 436–439, 1999, ISSN: 10797114. DOI: 10.1103/PhysRevLett.83.436. arXiv: 9811053 [quant-ph].
- [106] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, “Concentrating partial entanglement by local operations,” *Physical Review A*, vol. 53, no. 4, pp. 2046–2052, Apr. 1996, ISSN: 1050-2947. DOI: 10.1103/PhysRevA.53.2046. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.53.2046>.
- [107] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, “Exact and asymptotic measures of multipartite pure-state entanglement,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 63, no. 1, pp. 012307–012301, 2001, ISSN: 10502947. DOI: 10.1103/PhysRevA.63.012307.
- [108] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, “Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask),” *Communications in Mathematical Physics*, vol. 328, no. 1, pp. 303–326, 2014, ISSN: 14320916. DOI: 10.1007/s00220-014-1953-9. arXiv: 1210.4583.
- [109] P. M. Hayden, M. Horodecki, and B. M. Terhal, “The asymptotic entanglement cost of preparing a quantum state,” *Journal of Physics A: Mathematical and General*, vol. 34, no. 35, pp. 6891–6898, 2001, ISSN: 03054470. DOI: 10.1088/0305-4470/34/35/314. arXiv: 0008134 [quant-ph].
- [110] M. B. Plenio and S. Virmani, “An introduction to entanglement measures,” *Quantum Information and Computation*, vol. 7, no. 1-2, pp. 1–51, 2007, ISSN: 15337146. arXiv: 0504163 [quant-ph].
- [111] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, “Quantum source of entropy for black holes,” *Physical Review D*, vol. 34, no. 2, pp. 373–383, 1986, ISSN: 05562821. DOI: 10.1103/PhysRevD.34.373.
- [112] M. Srednicki, “PHYSICAL REVIEW LETTERS Entropy and Area,” vol. 71, no. 5, 1993.
- [113] A. M. Souza, P. Rapčan, and C. Tsallis, “Area-law-like systems with entangled states can preserve ergodicity,” *European Physical Journal: Special Topics*, vol. 229, no. 5, pp. 759–772, 2020, ISSN: 19516401. DOI: 10.1140/epjst/e2020-900003-3.
- [114] J. I. Latorre, E. Rico, and G. Vidal, “Ground state entanglement in quantum spin chains,” *Quantum Information and Computation*, vol. 4, no. 1, pp. 48–92, 2004, ISSN: 15337146. arXiv: 0304098 [quant-ph].

- [115] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, “Entropy scaling and simulability by matrix product states,” *Physical Review Letters*, vol. 100, no. 3, pp. 1–4, 2008, ISSN: 00319007. DOI: 10.1103/PhysRevLett.100.030504. arXiv: 0705.0292.
- [116] J. Eisert, M. Cramer, and M. B. Plenio, “Colloquium: Area laws for the entanglement entropy,” *Reviews of Modern Physics*, vol. 82, no. 1, pp. 277–306, 2010, ISSN: 00346861. DOI: 10.1103/RevModPhys.82.277.
- [117] M. Levin and X. G. Wen, “Detecting topological order in a ground state wave function,” *Physical Review Letters*, vol. 96, no. 11, pp. 3–6, 2006, ISSN: 00319007. DOI: 10.1103/PhysRevLett.96.110405. arXiv: 0510613 [cond-mat].
- [118] C. Holzhey, F. Larsen, and F. Wilczek, “Geometric and renormalized entropy in conformal field theory,” *Nuclear Physics, Section B*, vol. 424, no. 3, pp. 443–467, 1994, ISSN: 05503213. DOI: 10.1016/0550-3213(94)90402-2. arXiv: 9403108 [hep-th].
- [119] H. Casini and M. Huerta, “Entanglement entropy in free quantum field theory,” *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 50, 2009, ISSN: 17518113. DOI: 10.1088/1751-8113/42/50/504007. arXiv: 0905.2562.
- [120] T. Barthel, S. Dusuel, and J. Vidal, “Entanglement Entropy beyond the Free Case,” *Physical Review Letters*, vol. 97, no. 22, p. 220402, Nov. 2006, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.97.220402. arXiv: 0606436 [cond-mat]. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.97.220402>.
- [121] T. Nishioka, “Entanglement entropy: Holography and renormalization group,” *Reviews of Modern Physics*, vol. 90, no. 3, p. 35007, 2018, ISSN: 15390756. DOI: 10.1103/RevModPhys.90.035007. arXiv: 1801.10352. [Online]. Available: <https://doi.org/10.1103/RevModPhys.90.035007>.
- [122] M. E. Peskin, *An Introduction To Quantum Field Theory*. CRC Press, May 2018, p. 368, ISBN: 9780429503559. DOI: 10.1201/9780429503559. [Online]. Available: <https://www.taylorfrancis.com/books/9780429503559>.
- [123] T. Nishioka, “Entanglement entropy: Holography and renormalization group,” *Reviews of Modern Physics*, vol. 90, no. 3, p. 35007, 2018, ISSN: 15390756. DOI: 10.1103/RevModPhys.90.035007. arXiv: 1801.10352.
- [124] G. Vidal, “Entanglement renormalization,” *Physical Review Letters*, vol. 99, no. 22, pp. 1–4, 2007, ISSN: 00319007. DOI: 10.1103/PhysRevLett.99.220405. arXiv: 0512165 [cond-mat].
- [125] M. Taylor and W. Woodhead, “Renormalized entanglement entropy,” *Journal of High Energy Physics*, vol. 2016, no. 8, 2016, ISSN: 10298479. DOI: 10.1007/JHEP08(2016)165. arXiv: 1604.06808.
- [126] H. Casini and M. Huerta, “A c-theorem for entanglement entropy,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 25, pp. 7031–7036, 2007, ISSN: 17518113. DOI: 10.1088/1751-8113/40/25/S57. arXiv: 0610375 [cond-mat].
- [127] F. Caruso and C. Tsallis, “Extensive nonadditive entropy in quantum spin Chains,” *AIP Conference Proceedings*, vol. 965, pp. 51–59, 2007, ISSN: 0094243X. DOI: 10.1063/1.2828759. arXiv: 0711.2641.
- [128] L. E. Ballentine, “Quantum Mechanics: A Modern Development,” *American Journal of Physics*, Dec. 1998, ISSN: 0002-9505.
- [129] W. P. Schleich, *Quantum optics in the phase space*, 1. 2012, vol. 203, pp. 3–24, ISBN: 352729435X.
- [130] G. Manfredi and M. R. Feix, “Entropy and Wigner functions,” *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 62, no. 4, pp. 4665–4674, 2000, ISSN: 1063651X. DOI: 10.1103/PhysRevE.62.4665. arXiv: 0203102 [quant-ph].
- [131] W. H. Zurek, S. Habib, and J. P. Paz, “Coherent states via decoherence,” *Physical Review Letters*, vol. 70, no. 9, pp. 1187–1190, 1993, ISSN: 00319007. DOI: 10.1103/PhysRevLett.70.1187.

- [132] P. Facchi, A. Mariano, and S. Pascazio, “Decoherence versus entropy in neutron interferometry,” *Physical Review A. Atomic, Molecular, and Optical Physics*, vol. 63, no. 5, pp. 521081–521089, 2001, ISSN: 10502947. DOI: 10.1103/PhysRevA.63.052108.
- [133] J. Řeháček, Z. Hradil, J. Peřina, S. Pascazio, P. Facchi, and M. Zawisky, “Advanced Neutron Imaging and Sensing,” *Advances in Imaging and Electron Physics*, vol. 142, no. 05, pp. 53–157, 2006, ISSN: 10765670. DOI: 10.1016/S1076-5670(05)42002-9.
- [134] K. M. R. Audenaert, “Subadditivity of q-entropies for  $q_i1$ ,” *Journal of Mathematical Physics*, vol. 48, no. 8, p. 083507, Aug. 2007, ISSN: 0022-2488. DOI: 10.1063/1.2771542. arXiv: 0705.1276. [Online]. Available: <http://aip.scitation.org/doi/10.1063/1.2771542>.
- [135] E. H. Lieb, “Proof of an entropy conjecture of Wehrl,” *Communications in Mathematical Physics*, vol. 62, no. 1, pp. 35–41, 1978, ISSN: 00103616. DOI: 10.1007/BF01940328.
- [136] A. Wehrl, “On the relation between classical and quantum-mechanical entropy,” *Reports on Mathematical Physics*, vol. 16, no. 3, pp. 353–358, 1979, ISSN: 00344877. DOI: 10.1016/0034-4877(79)90070-3. arXiv: 1806.08827.
- [137] J. S. Ivan, M. S. Kumar, and R. Simon, “A measure of non-Gaussianity for quantum states,” *Quantum Information Processing*, vol. 11, no. 3, pp. 853–872, 2012, ISSN: 15700755. DOI: 10.1007/s11128-011-0314-2. arXiv: 0812.2800.
- [138] S. Bose, “Wehrl Entropy Based Quantification of Nonclassicality for Single Mode Quantum Optical States,” *arXiv*, 2016. arXiv: 1701.00064.
- [139] K. C. Tan and H. Jeong, “Resource Theories of Nonclassical Light,” *Quantum Reports*, vol. 1, no. 2, pp. 151–161, 2019. DOI: 10.3390/quantum1020014.
- [140] W. Vogel and J. Sperling, “Unified quantification of nonclassicality and entanglement,” *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 89, no. 5, pp. 1–6, 2014, ISSN: 10941622. DOI: 10.1103/PhysRevA.89.052302. arXiv: 1401.5222.
- [141] A. Orłowski, “Classical entropy of quantum states of light,” *Physical Review A*, vol. 48, no. 1, pp. 727–731, 1993, ISSN: 10502947. DOI: 10.1103/PhysRevA.48.727.
- [142] H. P. Yuen, “Two-photon coherent states of the radiation field,” *Physical Review A*, vol. 13, no. 6, pp. 2226–2243, 1976, ISSN: 10502947. DOI: 10.1103/PhysRevA.13.2226.
- [143] H. M. Nussenzveig, *Introduction to Quantum Optics*, ser. Documents on Modern Physics: Gordon and Breach. Gordon and Breach Science Publishers, 1973, ISBN: 9780677039008. [Online]. Available: <https://books.google.nl/books?id=GRMTKt6wJLYC>.
- [144] F. A. M. de Oliveira, M. S. Kim, P. L. Knight, and V. Buek, “Properties of displaced number states,” *Physical Review A*, vol. 41, no. 5, pp. 2645–2652, Mar. 1990, ISSN: 1050-2947. DOI: 10.1103/PhysRevA.41.2645. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.41.2645>.
- [145] D. E. Kharzeev and E. M. Levin, “Deep inelastic scattering as a probe of entanglement,” *Physical Review D*, vol. 95, no. 11, pp. 1–12, 2017, ISSN: 24700029. DOI: 10.1103/PhysRevD.95.114008. arXiv: 1702.03489.
- [146] J. S. Poucher, M. Breidenbach, R. Ditzler, J. I. Friedman, H. W. Kendall, E. D. Bloom, R. L. A. Cottrell, D. H. Coward, H. DeStaebler, C. L. Jordan, H. Piel, and R. E. Taylor, “High-Energy Single-Arm Inelastic e-p and e-d scattering at 6 and 10°,” *Physical Review Letters*, vol. 32, no. 3, pp. 118–121, Jan. 1974, ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.32.118. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.32.118>.
- [147] E. Wessels, “Signatures of gluon saturation in high energy scattering,” Ph.D. dissertation, Vrije Universiteit Amsterdam, 2009. [Online]. Available: <https://research.vu.nl/en/publications/signatures-of-gluon-saturation-in-high-energy-scattering>.
- [148] E. Levin and M. Lublinsky, “A linear evolution for non-linear dynamics and correlations in realistic nuclei,” *Nuclear Physics A*, vol. 730, no. 1-2, pp. 191–211, Jan. 2004, ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2003.10.020. arXiv: 0308279v1 [arXiv:hep-ph]. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0375947403018062>.

- [149] E. Levin and K. Tuchin, “Solution to the evolution equation for high parton density QCD,” *Nuclear Physics B*, vol. 573, no. 3, pp. 833–852, 2000, ISSN: 05503213. DOI: 10.1016/S0550-3213(99)00825-1.
- [150] H. collaboration, “Measurement of charged particle multiplicity distributions in DIS at HERA and its implication to entanglement entropy of partons,” Nov. 2020. arXiv: 2011.01812.
- [151] A. Kovner and M. Lublinsky, “Entanglement entropy and entropy production in the color glass condensate framework,” *Physical Review D - Particles, Fields, Gravitation and Cosmology*, vol. 92, no. 3, pp. 1–12, 2015, ISSN: 15502368. DOI: 10.1103/PhysRevD.92.034016. arXiv: 1506.05394.

# Acknowledgements

First of all, I wish to thank my supervisor Daniël Boer for his prolonged support over the course of my thesis.

I also want to thank my parents, Frank IJpelaar and Margit Hendriks, my girlfriend Marthe de Roo and her mother Iris de Roo-Kwant, for their emotional support. Over the long span of this project, they have helped me immensely in going through some of the tougher periods. I really couldn't have done this without them.

Last but not least, I want to thank my good friend Giovanni van Marion for all the interesting and enlightening discussions, and for his many insights. I wish him all the best in his further endeavors in the field of physics.

# Appendices

# Appendix A

## Hydrogen Atom Entropies

$l \backslash n$	1	2	3	4	5	6	7	8	9	10	11
0	1.614	5.580	7.895	9.544	10.827	11.877	12.768	13.540	14.223	14.834	15.388
1		5.166	7.707	9.435	10.755	11.827	12.730	13.511	14.200	14.815	15.372
2			7.305	9.221	10.618	11.731	12.659	13.456	14.156	14.780	15.342
3				8.840	10.403	11.586	12.553	13.375	14.091	14.727	15.299
4					10.039	11.374	12.406	13.265	14.005	14.657	15.241
5						11.023	12.198	13.118	13.894	14.569	15.168
6							11.857	12.915	13.748	14.457	15.078
7								12.582	13.548	14.313	14.967
8									13.221	14.116	14.825
9										13.795	14.631
10											14.313

Table A.1: The values for the momentum differential entropy of the radial part,  $R_{nl}$  of the hydrogen atom.

$l \backslash n$	1	2	3	4	5	6	7	8	9	10	11
0	-0.109	-3.289	-4.719	-5.677	-6.400	-6.982	-7.469	-7.888	-8.256	-8.583	-8.861
1		-2.057	-3.988	-5.162	-6.005	-6.663	-7.201	-7.658	-8.053	-8.403	-8.733
2			-3.274	-4.636	-5.598	-6.333	-6.926	-7.422	-7.848	-8.221	-8.571
3				-4.169	-5.199	-6.001	-6.646	-7.181	-7.637	-8.034	-8.390
4					-4.879	-5.690	-6.371	-6.940	-7.425	-7.845	-8.216
5						-5.468	-6.124	-6.710	-7.217	-7.657	-8.046
6							-5.972	-6.511	-7.022	-7.476	-7.878
7								-6.412	-6.861	-7.310	-7.719
8									-6.802	-7.180	-7.577
9										-7.153	-7.472
10											-7.472

Table A.2: The values for the momentum differential entropy of the radial part,  $P_{nl}$  of the hydrogen atom.

m \ l	1	2	3	4	5	6	7	8	9	10	11
1	2.531	2.099	2.041	2.021	2.011	2.005	2.001	1.998	1.996	1.994	1.993
2		2.406	2.263	2.192	2.150	2.122	2.101	2.087	2.075	2.066	2.058
3			2.263	2.241	2.206	2.177	2.153	2.135	2.120	2.108	2.097
4				2.147	2.191	2.185	2.171	2.157	2.143	2.131	2.121
5					2.050	2.136	2.153	2.153	2.147	2.139	2.131
6						1.969	2.084	2.118	2.129	2.130	2.128
7							1.899	2.035	2.082	2.103	2.111
8								1.837	1.989	2.048	2.076
9									1.782	1.946	2.014
10										1.733	1.907
11											1.688

Table A.3: The values for the entropies of the spherical part of the entropy,  $Y_{lm}$ . The entropy is the same for  $m$  and  $-m$ . The Entropy is also the same for the position and momentum representations.

m \ l	1	2	3	4	5	6	7	8	9	10	11
1	6.840	8.725	9.902	10.751	11.396	11.940	12.400	12.799	13.151	13.485	13.770
2		8.328	9.805	10.750	10.890	11.456	11.930	12.338	12.697	13.582	13.871
3			9.323	10.544	10.945	11.564	12.068	12.495	13.275	13.604	13.900
4				10.037	10.696	11.414	11.971	12.430	13.206	13.549	13.855
5					10.236	11.149	11.796	12.305	13.104	13.468	13.789
6						10.721	11.534	12.124	12.966	13.361	13.333
7							11.138	11.870	12.783	13.222	13.226
8								11.504	12.536	13.042	13.088
9									12.196	12.806	12.913
10										12.489	12.688
11											12.390

Table A.4: The values for the logarithm of the entropy product  $\ln(\pi^3 e^3 \Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z)$ , all with  $m = 0$ .