TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Neutron optical experiments studying uncertainty relations of fundamental quantum measurements

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Associate Prof. Dipl.-Ing. Dr. Yuji Hasegawa
E141 - Atominstitut

eingereicht an der Technischen Universität Wien
Fakultät für Physik

von

Dipl.-Ing. Bülent Demirel
Matr. Nr. 0725980
Liechtensteinstraße 101/4
1090, Wien

(Ort, Datum) (Unterschrift)

Abstract

The uncertainty principle is one of the cornerstones of quantum mechanics. It expresses the inability to jointly assign precise, definite values to two incompatible physical quantities. The best known form of this principle is expressed for the observables position and momentum mostly formulated by the product of standard deviation $\sigma(\hat{Q})\sigma(\hat{P}) \geq \frac{\hbar}{2}$. This expression conveys the complementary trade-off between statistical dispersions, which can be accessed by repetitions of measurements on identically prepared system states. This inequality does not treat the restriction of simultaneous determinability by an intervening quantum-mechanical measurement apparatus. The present dissertation deals with the experimental investigations of two new theories, which aim to confirm and establish new measurement uncertainty relations. The test object is the spin of neutrons for which no joint determination of different spin directions is possible. A new theory of Ozawa based on mean squared deviations of operators and entropic variants based on the Shannon information are being tested. Special features of this thesis include the treatment of error correction procedures and the study of the influence of mixed states and non-projective measurements. The results show that the new relations hold up to scrutiny, albeit the determination of minimal uncertainty presents great challenges.

Kurzfassung

Die Unschärferelation zählt zu den Grundpfeilern der Quantenmechanik. Sie drückt das Unvermögen aus zwei inkompatiblen physikalischen Größen gleichzeitig präzise, definite Werte zuzuordnen. Das bekannteste Beispiel dieses Prinzips ist gegeben durch die Observablen Ort und Impuls, meistens formuliert durch das Produkt von Standardabweichungen $\sigma(\hat{Q})\sigma(\hat{P}) \geq \frac{\hbar}{2}$. Dieser Ausdruck vermittelt den komplementären Austausch an statistischer Streuung, welche durch ständige Repetition an immer gleich präparierten Systemen gemessen wird. Was diese Ungleichung nicht beschreibt ist, wie sehr der Eingriff eines quantenmechanischen Messapparates die gleichzeitige Bestimmbarkeit begrenzt. In dieser Dissertation geht es um die experimentellen Nachforschungen von zwei neuen Theorien, die sich zur Aufgabe machen neue Messunschärferelationen zu etablieren. Als Testobjekt wird der Spin von Neutronen herangezogen, dessen unterschiedliche Richtungen nicht gemeinsam gemessen werden können. Getestet werden eine Theorie von Ozawa, basierend auf mittleren quadratischen Abweichungen von Operatoren, und entropische Varianten, beruhend auf der Shannon-Information. Zu den Besonderheiten zählen die Behandlung von Fehlerkorrekturprozeduren und die Studie des Einflusses gemischter Zustände und nicht projektiver Messungen. Die Ergebnisse zeigen, dass die neuen Relationen bestehen können, wobei die minimale Unschärfe zu bestimmen große Herausforderungen darstellt.

Contents

CHAPTER 1

Preface

The appearance of the uncertainty principle is closely related to the development of quantum mechanics and introduced a new concept to hitherto unknown physics. In 1925 Heisenberg achieved a first reformulation of the kinematic relationship of the Hamilton equation of motion which, after cooperation with Born and Jordan, led to the founding of matrix mechanics [1]. A consequence of the use of matrices is that interchanging the sequence of multiplication is in most cases not irrelevant. This non-commutation only dissolves when both matrices have the same system of eigenvectors, in other words can be determined at the same time. Purely mathematically, the uncertainty principle is due to the linear structure of Hilbert spaces, where vectors can be represented as linear combinations of eigenvectors of another operator or matrix, and thus the representation is 'uncertain' with respect to different bases. But apart from purely theoretical consequences, there are groundbreaking physical implications arising from this principle.

In Heisenberg's paper [2] "*Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*" published in 1927, the physical concept of the impossibility to simultaneously determine certain observables was conceptually introduced. Heisenberg's work was primarily concerned with demonstrating the trade-off of accuracies in measurements of incompatible observables, using understandable and tangible thought experiments. The textbook example is the gamma-ray microscope: in order to be able to measure the location of an electron, it must

be illuminated with light, which is then reflected and registered afterwards. The shorter the wavelength of the light used, the more accurate the location measurement will be. But with shorter wavelengths, the energy increases as well, thereby increasing the momentum that a photon gives to the electron. Heisenberg's explains this as a consequence of the commutator relation $[\hat{P}, \hat{Q}] = i\hbar$ and argues the accuracy of the measurement of the electron position q_1 and the 'discontinuous change' p_1 caused by the Compton effect gives a trade-off relation $q_1 p_1 \sim h$. The article escapes a specification of the indeterminacy and only qualitatively suggests that q_1 and p_1 correspond to 'some mean error'. This is all the more surprising because this relation is also conceived explicitly for a Gaussian probability distribution, therefore an identification with the standard deviation is somewhat obvious.

The uncertainty principle was a completely new phenomenon at that time which had not been known in classical physics before. At the end of the manuscript [2] is a comment that refers to reviews by Bohr which prompted Heisenberg to make revisions [3] and add another perspective to his discovery. Bohr himself has used the concept of complementarity in many ways without attempting to systemize or clarify it rigorously. For him, the principle of uncertainty was no more than a special case of the principle of complementarity. Similar to how the wave and particle description exclude each other, the uncertainty was due to a complementarity where the experiment decides which image (particle or wave) is chosen. Heisenberg, influenced by Bohr, agrees with his considerations and tolerates the ontic perception as being a valid alternative of his originally epistemological standpoint: observations are meaningful only in relation to an experiment and only by measurement can a value be assigned to the momentum. Historically, this can be considered as the first fork in the road of uncertainty principles. The concept of limited *joint determinability* of individual quantum objects and full knowledge acquisition is extended by the concept of *joint definability*, i.e. only above a specific range can a measurement give the system a real meaning much in agreement with the Copenhagen interpretation, but position and impulse cannot be equally well-defined [4].

The uncertainty principle quickly became a cornerstone of quantum physics. It is not surprising that a more rigorous expression was derived soon after by E.H. Kennard [5], i.e. the uncertainty principle given by the product of standard deviations of position and momentum. This inequality was soon regarded as a conformable representation of the uncertainty principle, even

by Heisenberg himself, who regarded the mathematical rigor a confirmation of his heuristic argumentation. This rooted in a confusion and misinterpretation of the matter over the next few decades, for this inequality reflects a different idea. It is not so much the validity of this relation that is questionable, but that there is no obvious connection to the microscope example. The standard deviation requires only one wave function and the corresponding operators, whereas Heisenberg's idea is based on the force exerted by the initial photon affecting the electron's position and the subsequent observation of a disturbed state. Equation and model are not compatible.

In the following years, the standard deviation was the preferred measure for quantifying the uncertainty principle. Between 1927 and 1933, in addition to Kennard, Weyl [6], Hardy [7], Robertson [8] and Schrödinger [9] developed a quantum mechanical and mathematical description of Heisenberg's uncertainty principle, the latter two in particular for arbitrary observables $\hat{A}$ and $\hat{B}$, that are not necessarily related to each other by a Fourier transformation.

The unfortunate mixture of different concepts lead to problems in the interpretation and raised criticisms early on. Popper [10] was one of the first to criticize the validity of the uncertainty principle on the level of individual particles since statistical laws are not directly transferable to single events. Looking at the uncertainty principle in its statistical form, one finds the following situation: using the wave function and an observable, one of many possible events in the system will occur. Random samplings provide a probability function, which will be scattered around a mean by frequent repetition of the same measurement. Ergo, an equally large number of identically prepared copies of the same quantum state is necessary to have an estimate of the appropriate distribution. The hypothetical value would then be obtained if infinite many samples are taken. Whether a well-defined position, momentum or energy is assigned to the individual particle, cannot be answered with such a formulation. This uncertainty is more concerned with the measured widths of distribution that results from many identically prepared ensembles. This type of uncertainty principle has hence got the name *preparation uncertainty relation* in today's scientific literature.

A change of perspective from joint measurements to the stochastic nature of statistically independent measurements persisted in the next decades after 1940. The standard deviation was

explored and questioned as a meaningful stochastic measure to represent statistical dispersions of probabilistic events. The first redefinitions altered the measures of concentration of a distribution [11] or relied on information entropies, first for position and momentum [12–14], later also for finite Hilbert [15, 16] spaces.

Since the standard formulation has little to do with a measurement inaccuracy and the associated disturbance of the conjugated quantity, consistent definitions of the measurement uncertainty principle was missing for a long time. Perhaps the first work dealing with joint measurements and the implication for the uncertainty principle is that of Arthurs and Kelly [17]. In their derivation, probe systems for coupling with position and momentum are taken into account and thus a joint measurement is modeled on both observables. As a measure of the inaccuracy, a *noise operator* is used to measure the quantum error. The work is later generalized by Arthurs and Goodman [18]. The problem of common measurability and the associated inaccuracies of incompatible observables is therefore only formed at this time with enough mathematical rigor. However, consideration of measurement disturbance are still missing at that time.

Approximately with the beginning of the new millennium the focus intensified on Heisenberg's originally conceptualized quantum principle, nowadays subsumed as *measurement uncertainty relation*. On the one hand this category of uncertainty includes the joint indeterminacy, on the other hand how much a variable is disturbed by the exact measurement of another. These categories are therefore often referred to as *noise/error* and *disturbance* uncertainty relations. The Arthurs-Kelly relation can be regarded as one of the first joint measurement inaccuracy trade-offs which served as an important template for establishing inaccuracy-disturbance trade-offs given by Appleby [19], Ozawa [20, 21] or Hall [22]. As has been the case for decades, debates and scientific controversies emerged on how meaningful or faithful these definitions based on operator differences are. As a result other concepts and methodologies have been developed which in turn brought forth a myriad of physical quantities that can be used to quantify uncertainty relations. For measurement uncertainty relations the most prominent alternatives are probabilistic distances [23–25] and different types of entropies [26–28].

Nowadays, there are many aspects to consider when forming uncertainty relations. The elementary components that make up quantum mechanical measurements are quantum states,

channels, and observables, whose properties can be modified which in turn can influence the inaccuracy of measurements. For quantum states, one must consider pure and mixed states or entangled and separable states. In turn, quantum channels can have noise or finite capacities, interact with the environment and alter correlations. With observables one can distinguish whether they are finite or infinite, sharp or fuzzy, existent as operators or not (time for example has no genuine operator representation). The uncertainty relations may ultimately be different in the number of considered observables (two or more) or can be state dependent or not, i.e. changing continuously depending on the state or remaining invariant for all input states. Perhaps the greatest challenge is to find an optimal estimate of inequalities as it is often of significant interest to show under which conditions the measurement inaccuracies are minimal.

The research on uncertainty relations is not only of purely academic interest, but also important for current topics in quantum physics. The relevance of the uncertainty principle for quantum cryptography is probably the most obvious since security protocols contain joint measurements of correlated states. The quality of the secrecy of a key distribution corresponds to high goodness of the parameter estimation. Uncertainty relations and error corrections pose natural limits to the assessment of the measurement quality. In the review [29] a long list of closely related topics are mentioned for which quantum indeterminacy play an important role, e.g. *quantum randomness*, *quantum key distribution*, *two-party cryptography*, *entanglement witnessing* and many more.

This thesis deals with experimental tests of new uncertainty relations in the field of neutron optics. Thermal neutrons outside the nucleus have low-energy (meV), are neutral and massive particles which makes them attractive for studying foundations of quantum mechanics. They can be described as matter waves according to the Schrödinger equation and share many phenomena with light optics. Experimental investigations with neutrons have therefore a long tradition in quantum physics comprising large areas of current and ongoing scientific topics [30, 31], including tests of complementarity [32] and position-momentum uncertainty relations [33–35]. In 2012, Erhart et al. [36] published a paper with the new findings testing Ozawa's theory in the two-dimensional case using the spin of the neutron. This is exactly the starting point of the studies described in this thesis, bringing together the progress that has been made since then. One focus here is on Ozawa's theory for which definitions, the first measurement procedures

and results will be presented in detail. In addition, the theoretical advancements that improve the lower limits for correction and mixed qubit states will be investigated. The other focus will be on the experimental analysis of the information-theoretic formulations defined via conditional Shannon entropies. On the one hand, an entropic noise-disturbance uncertainty relation is tested for this purpose, which is shown to be optimal in the case of projective measurements, while for more general observables measurable violations of the considered inequality are possible. Therefore, a less complex noise-noise inequality is analyzed as well that will highlight the importance of POVMs (positive operator valued measures) for the investigation of uncertainty relations.

The work is structured as follows. In chapter 2 necessary mathematical terms are introduced, followed by chapter 3 in which important quantum physical prerequisites are presented. Some of the terms encountered in these chapters are not analyzed in detail but rather illustrated by examples. Since only finite quantum systems are considered, the theory is presented almost exclusively in the finite case. Subsequently, in chapter 4, a brief overview of neutron optics can be found for the readers who are unfamiliar with the matter. Chapter 5 gives an overview of preparation uncertainty relations, which are quantified by means of the standard deviation and entropic quantities. Important intermediate results are presented here. At the heart of this thesis are chapter 6, which deals with Ozawa's root mean square (rms) error and disturbance and chapter 7, which uses conditional entropies. Everything is rounded off in the concluding chapter 8.

CHAPTER 2

Basics of probability and information theory

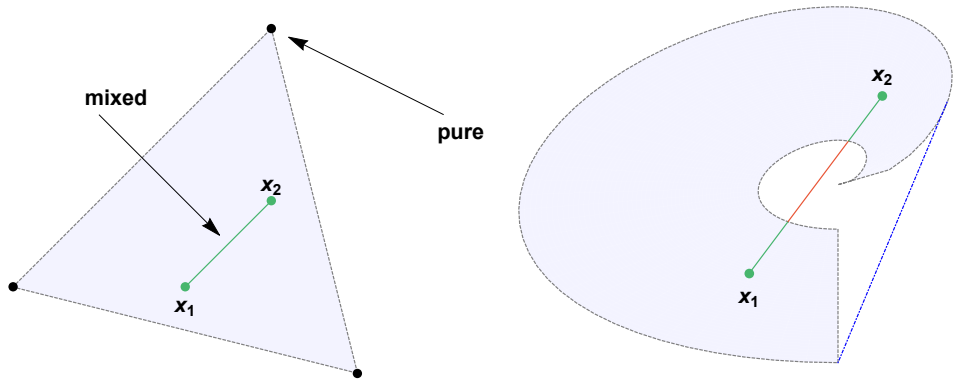
This chapter provides a quick overview of the elementary notions of convexity, probability and information theory that are indispensable for the understanding of theories introduced later on. Convex geometry will be encountered quite often since building a convex linear combination of states and operations is the most rudimentary property in quantum physics. Probability theory itself is a convex theory. Uncertainties or inaccuracies are in principle studies of joint probabilities, distributions and stochastic metrics, which need to be introduced as prerequisite. Furthermore, the relation of probabilities to Shannon and similar entropies is shown. Major sources for the following sections are [37–40].

2.1 Convex sets

A subset K of a real or complex vector space is called *convex* if for every pair of points ($\mathbf{x}_1, \mathbf{x}_2 \in K$) within the set the straight line, parametrized by $\lambda \in \mathbb{R}$, that joins them is also within the set, ergo

$$\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in K \ , \quad 0 \leq \lambda \leq 1 \ . \tag{2.1}$$

Illustrations of a two dimensional convex and non-convex set is given in Fig. 2.1, in which the



(a) A convex set. (b) Not a convex set.

Fig. 2.1: Examples of two dimensional sets, which are either convex or not. The line segment is color-coded green when it remains within the region of the set and red when it is outside. Changing the boundary indicated by the blue dashed lines encloses the area and gives the convex hull.

two points $\mathbf{x}_1 = \mathbf{x}(\lambda = 0)$ and $\mathbf{x}_2 = \mathbf{x}(\lambda = 1)$ are elements of the set. The connection $\mathbf{x}$ is also called *mixture* and must be entirely within the set to be convex as shown on the left side in Fig. 2.1(a), otherwise the set is not convex as shown in Fig. 2.1(b). A point that is not an interior point of any line segment lying entirely in the subset K is called *extreme point* or *pure*. Regular polygons like the triangle are all examples of two dimensional convex bodies. In three dimensions concrete examples of convex bodies would be the Platonic solids, whose extremal points are vertices, but also the ball whose pure points make up the entire sphere.

Consider a finite set of points in the plane, then the *convex hull* is the smallest convex polygon in which all points are contained, see Fig. 2.1(b). For a mathematically more rigorous explanation the term *convex combination* needs to be introduced first, which for the vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ is defined by the sum of the form

$$\mathbf{x} = \sum_i^n \lambda_i \mathbf{x}_i \ , \quad 0 \leq \lambda_i \leq 1 \ , \quad \sum_i \lambda_i = 1 \ . \tag{2.2}$$

For the line segment of Eq. (2.1) the point $\mathbf{x}$ is on the line and can be written as a convex combination of it's endpoints $\mathbf{x}_1, \mathbf{x}_2$. For the triangle Fig. 2.1(a) any point $\mathbf{x}$ in the set can be written as a convex combination of its three vertices $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$. The convex hull $\text{conv}(S)$ of

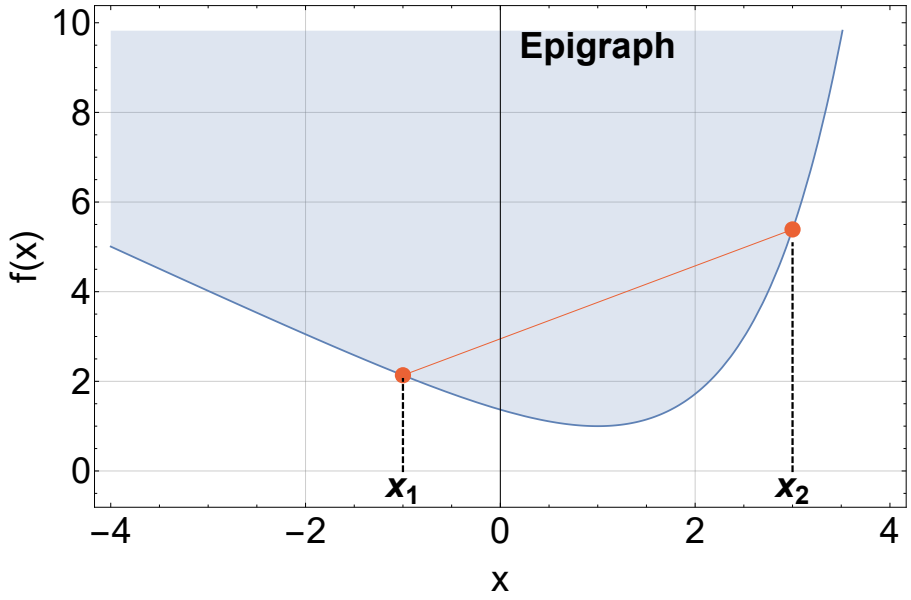


Fig. 2.2: Example of a convex function. The shaded area above the function is called the epigraph and is a convex set. Additionally, a secant from $f(x_1)$ to $f(x_2)$ is drawn, which is always above the function except for the endpoints and gives a visual interpretation of the convexity condition of functions Eq. (2.4).

a set S is identical to the set of all their convex combinations

$$\mathrm{conv}(S) = \left\{ \sum_i^n \lambda_i \mathbf{x}_i \,\middle|\, \mathbf{x}_i \in S, n \in \mathbb{N}, \sum_i \lambda_i = 1, \lambda_i \geq 0 \right\} . \tag{2.3}$$

The term convex and concave function is linked to the shape of the *epigraph*, the area above the function. A function $f(x)$ is convex if the epigraph is a convex set or, mathematically more precise, for a function $f : K \to \mathbb{R}$ if the line $\lambda f(x_1) + (1-\lambda) f(x_2)$ is always above or equal to the corresponding value of the function

$$f\left(\lambda x_1 + (1-\lambda) x_2\right) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \;, \quad x_1, x_2 \in K \;, \quad \lambda \in [0,1] \;. \tag{2.4}$$

An illustration of a convex function is given in Fig. 2.2 with a visualization of the convexity criterion of Eq. (2.4). Since the coefficients λ_i have to be positive numbers between 0 and 1 and sum to unity the convexity of sets is linked to probability theory.

2.2 Some aspects of probability theory

Quantum physics makes statements about the probability for the occurrence of events or about the physical laws of quantum mechanical expectation values. Accordingly, it is necessary to introduce some stochastic entities. The definition of probability was axiomatized in 1931 by the *Kolmogorov axioms* [41] and is closely interlinked with the notion of *measure* or *measurable space* . The necessity for the terms can be motivated by a problem called the *Banach-Tarski paradox* [40, 42]. The mathematical technicalities of its nature are irrelevant here, but the quintessence of the paradox is important: it is possible to take a solid ball in $\mathbb{R}^3$, slice it into pieces, reassemble the pieces by using rigid transformations only, and end up with two balls of precisely the same size and shape. The conclusion is that sets can be counter-intuitive to the natural understanding of areas or volumes, since in the paradox it appears that a new object is being created out of nothing.

Mathematically, there exist non-measurable sets without any meaningful 'size'. A measurable set is therefore a set that very much complies with the common sense of divisible bodies and a measure is the mapping $\mu(X)$ that attributes a number like a volume, weight or a probability to an object X indicating the 'intensity' of that particular property. To define a measure a few termini are required first. A *power set* $P(X)$ of a set X is the set of all subsets of X.

Example 2.2.1. The power set $P(X)$ of the set containing three elements $X = \{a, b, c\}$ is $P(X) = \{\{\emptyset\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$. ◊

The next necessary definition is that of a *σ-algebra*. Let X be a set and $\Sigma \subset P(X)$. The subset Σ is a σ-algebra on X if and only if

1. $X \in \Sigma$,
2. it is closed under complementation: $Y \in \Sigma$, then so is its complement $Y^c = X \backslash Y$,
3. it is closed under countable unions: $Y_i \in \Sigma$, then $\bigcup_i^\infty Y_i \in \Sigma$.

Example 2.2.2. For a die with the six usual outcomes $X = \{\boxdot, \boxdot, \boxdot, \boxdot, \boxdot, \boxdot\}$ the following examples are typical σ-algebras

1. $\Sigma_1 = \{\emptyset, X\}$,

2. $\Sigma_2 = \{\emptyset, \{\boxdot, \boxdot, \boxdot\}, \{\boxdot, \boxdot, \boxdot\}, X\}$,

3. $\Sigma_3 = P(X)$.

The first example is called the trivial σ-algebra. The example Σ_2 allows to conclude whether the outcome of the die roll has an even or odd number of pips, it is however not possible to measure a specific event, unlike for the power set Σ_3, the largest possible σ-algebra on X with which every individual outcome is inquirable. $\Diamond$

The pair (X, Σ) consisting of the set X and the σ-algebra Σ over X is called a measurable space. Given a measurable space (X, Σ) a function μ is called a measure, if it is

1. non negative for all $Y_i \in \Sigma$, $\mu(Y_i) \geq 0$,

2. the empty set gives $\mu(\emptyset) = 0$,

3. it is σ-*additive* , that is, for all pairwise disjoint $\{Y_i\} \in \Sigma$ the measure fulfills $\mu\left(\bigcup_{i=1}^{\infty} Y_i\right) = \sum_{i=1}^{\infty} \mu(Y_i)$.

The triple (X, Σ, μ) is referred to as measure space. Let Ω be the sample space, i.e. the set of all possible outcomes, $\mathcal{F}$ be called the event space (the sigma-algebra) and p be a measure with $p(\Omega) = 1$, then p is called the probability (measure) in the *probability space* $(\Omega, \mathcal{F}, p)$.

2.2.1 Joint, conditional and marginal probabilities

Naturally, it is possible to ask for the likeliness of two events or how a particular outcome depends on another result. From the probability axioms it can be deduced that for two subspaces A, B of the event space the following holds

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \ , \tag{2.5}$$

a fact that is easily comprehensible by looking at the *Venn diagram* shown in Fig. 2.3. The union in Eq. (2.5) is logically related to *combined events*, i.e. 'either/or' relations, while the intersection corresponds to *joint events* of A 'and' B. For joint probabilities the usage of commas $p(A \cap B) = p(A, B)$ is more common and will be applied frequently from now on. From Eq. (2.5) a law called the *subadditivity* of the probability measure is attainable

$$p(A \cup B) \leq p(A) + p(B) \ . \tag{2.6}$$

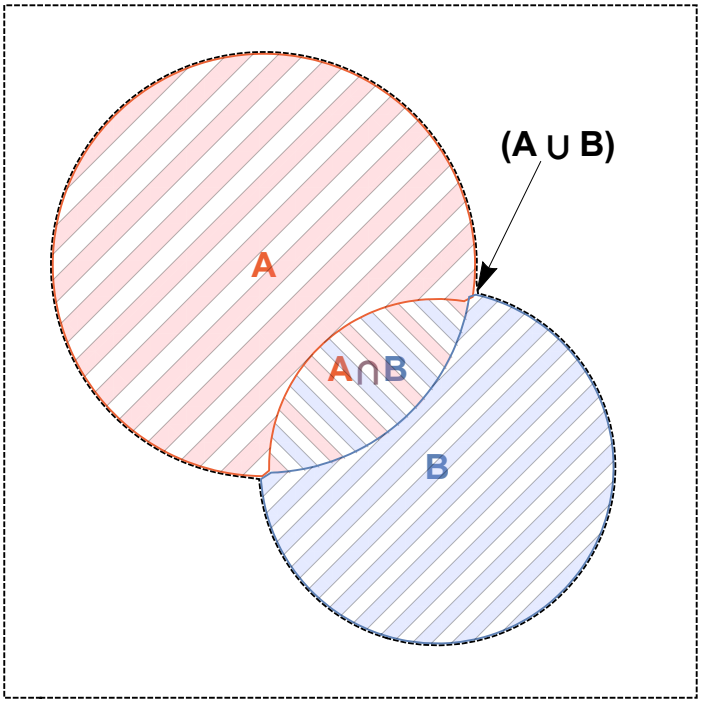


Fig. 2.3: Simple Venn diagram with two sets A, B and their intersection $A \cap B$ enabling simple illustration of probability relations. The space created by merging the disks is given by the union $A \cup B$.

The combined probability is less or equal to the sum of the single probabilities where equality only holds if the last term in Eq. (2.5) vanishes, i.e. when the events are mutually exclusive. Given the same two events with the assumption $p(B) > 0$ it is possible to define the *conditional probability*

$$p(A|B) := \frac{p(A,B)}{p(B)} \ , \tag{2.7}$$

where event A is said to be conditioned on B. It is to be understood as the probability that event A occurs given the fact that B has occurred or that its outcome is known. The definition is chosen such that if two events are *statistically independent*, i.e. $p(A,B) = p(A)p(B)$, then $p(A|B) = p(A)$, which indicates that no matter the outcome of B, the probability of A is unaffected by it. From the symmetry of the intersection $A \cap B = B \cap A$ (or $(A,B) = (B,A)$) in Eq. (2.7) a very useful formula, called *Bayes' theorem*, can be deduced

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \ . \tag{2.8}$$

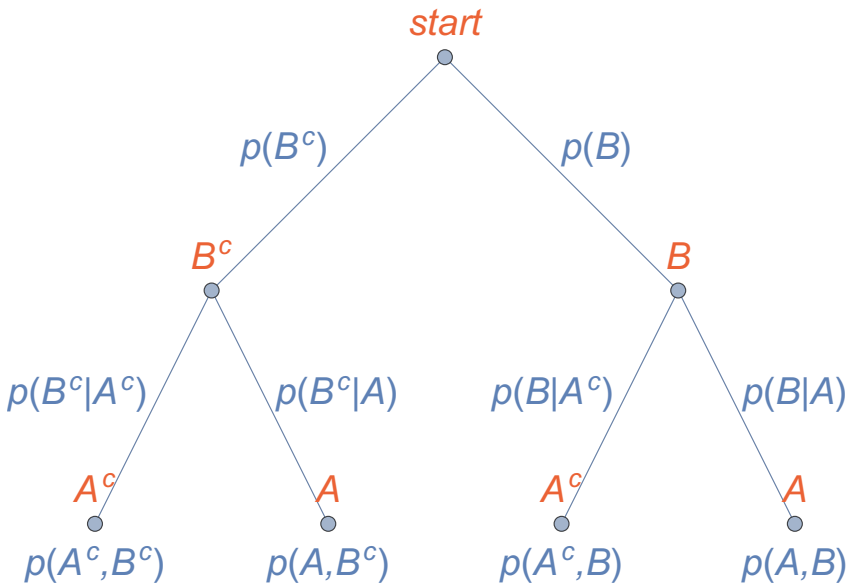


Fig. 2.4: Binary tree diagram with the root and all events colored in red and the probabilities leading to the events colored in blue. The definition of the conditional probability appears naturally as the connecting branches.

In case $p(A,B) \neq p(A)p(B)$ the events are statistically dependent on each other and are called either correlated $p(A,B) > p(A)p(B)$ or anti-correlated $p(A,B) < p(A)p(B)$. For the conditional probability this implies, that knowledge of the outcome B can therefore both increase and decrease the probability of the occurrence of A. Joint and conditional probabilities can be nicely illustrated by tree diagrams which describe mutually exclusive events. As an example, a binary tree diagram is shown in Fig. 2.4.

Assuming there are two discrete events, where $\bigcup_i^k B_i = \Omega$ partitions the sample space into pairwise mutually exclusive sets, then the joint probability is given by $p(A, B_i)$. It is often of interest to remove the dependency of one set of events from the probability function. In that case, averaging over these events relinquishes the dependency on it and the probability received thereby is called the *marginal probability*, given as

$$p(A) = \sum_i^k p(A, B_i) \ . \tag{2.9}$$

2.2.2 Random variables and distributions

A real valued *random variable* $X : \Omega \to \mathbb{R}$ is a measurable function on a probability space $(\Omega, \mathcal{F}, p)$ that assigns a real number to elements $\omega_i \in \Omega$.

A real random variable does not describe the actual outcome of a particular experiment, but rather demands that events are quantified in the domain of real numbers. It is more suitable in many ways to ask for the outcome of the random variable than asking in terms of sets corresponding to the result. For a coin with $A = \{$'heads', 'tails'$\}$ natural choices for real valued random variables would be $\{0, 1\}$ or $\{-1, 1\}$.

A *probability mass function* $f_X(x_i)$ of a discrete random variable X with possible value x_i is defined as

$$f_X(x_i) = p(X = x_i) = p(\{\omega_i \in \Omega | X(\omega_i) = x_i\}) \ , \tag{2.10}$$

while for a continuous random variable the *probability density function* $f_X(x)$ is related to the probability by

$$p(X \in [a, b]) = \int_a^b f_X(x) dx \ . \tag{2.11}$$

The finite outcomes n of a random variables X can be put together into the *probability vector* $\mathbf{p}_X = (p(X = x_1), p(X = x_2), ..., (X = x_n)) = (p_1, p_2, ..., p_n)$, which is commonly referred to as *discrete probability distribution*.

The *mean* or *expected value* of a random discrete variable X with possible values x_i is defined as

$$E(X) = \langle X \rangle = \sum_i^N x_i p(x_i) \ , \tag{2.12}$$

The expected value weights all $\{x_i\}$ with the respective probabilities $p(x_i)$. For a uniform distribution $p(x_i) = 1/N$ it coincides with the arithmetic mean $\langle X \rangle = \bar{X}$. The expected value is monotone $X \leq Y \to E(X) \leq E(Y)$ and linear $E(aX + bY) = aE(X) + bE(Y)$.

Another important statistical measure is the *variance*. The variance $\mathrm{var}(X)$ is a measure of the width of a distribution. It is non-negative and can only be zero if X equals exactly the expected

value $\langle X \rangle$. The square root of the variance is named *standard deviation* $\sigma(X)$

$$\sigma(X) = \sqrt{\text{var}(X)} = \sqrt{\langle (X - \langle X \rangle)^2 \rangle} = \sqrt{\langle X^2 \rangle - \langle X \rangle^2} \ . \tag{2.13}$$

The area $E(X) - \sigma(X) < X < E(X) + \sigma(X)$ is regarded as the confidence interval of 1σ in a normal distribution. While the variance gives the deviation around a mean, the *root mean square error (rms)* indicates how well a function curve or a prediction $\tilde{x}_i$ deviates on average from the actual observed values x_i

$$rms = \sqrt{\frac{1}{n} \sum_{i}^{n} (\tilde{x}_i - x_i)^2} \ . \tag{2.14}$$

Example 2.2.3. A random variable X is *Poisson distributed* if for $\lambda > 0$

$$p(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \ , \qquad k \in \mathbb{N}_0 \ . \tag{2.15}$$

The expected value and its square yield

$$\langle X \rangle = \sum_{k=0}^{\infty} k \left(\frac{\lambda^k}{k!} e^{-\lambda} \right) = \lambda, \qquad \langle X^2 \rangle = \sum_{k=0}^{\infty} k^2 \left(\frac{\lambda^k}{k!} e^{-\lambda} \right) = \lambda^2 + \lambda \ . \tag{2.16}$$

Utilizing Eq. (2.13) it turns out that for the Poisson distribution the variance $\text{var}(x) = \langle X \rangle = \lambda$ is equal to the expected value. This distribution is particularly important for the counting statistic of a particle detector. ♢

The variance of a linear combination of random variables does not generally yield the sum of the single variances. Instead the rule is $\text{var}(aX + bY) = a^2 \text{var}(X) + 2\ ab\ \text{cov}(X, Y) + b^2 \text{var}(X)$, which also contains a mixed term known as *covariance*

$$\text{cov}(X, Y) = E(X - E(X))E(Y - E(Y)) \ . \tag{2.17}$$

The covariance is a measure of the statistical independence of two event spaces and gives, when normalized, the so-called (Pearson) *correlation*. It is possible to combine variance and covariance with the notation $\text{cov}(X, X) = \text{var}(X)$. Assuming that X_i are random variables,

each with a well-defined variance then the *covariance matrix* is

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) \ . \tag{2.18}$$

The *n-th moment of a distribution*[1], where $n \in \mathbb{N}$, is calculated in close resemblance to Eq. (2.12) by

$$E(X^n) = \langle X^n \rangle = \sum_i x_i^n p(x_i) \ . \tag{2.19}$$

The variance of a random variable is therefore given as the difference between the second moment and the square of the first moment of the distribution. This term will also be encountered later as the n-th moment operator in quantum physics.

2.2.3 Majorization

Assuming there are two probabilistic vectors $\mathbf{p}$ and $\mathbf{q}$, it is natural to ask which of the two vectors is less uniform, more precisely which events are more equally distributed and which are more unbalanced. A comparison of standard deviations provides one option, an alternative method is offered by *majorization*. For that purpose an order relation must be introduced: let $\mathbf{p} = (p_1, ..., p_m)^T$ be the original distribution of size m, then

$$\mathbf{p}^{\downarrow} = (p_1^{\downarrow}, ..., p_m^{\downarrow}) \ , \qquad p_1^{\downarrow} \geq p_2^{\downarrow} \geq .... \geq p_m^{\downarrow} \tag{2.20}$$

is the vector with the same entries but sorted in descending order from largest to smallest. The order on probability vectors allows to characterize the degree of their randomness by majorization

$$\sum_{i=1}^{k} p_i^{\downarrow} \leq \sum_{i=1}^{k} q_i^{\downarrow} \ , \qquad \forall k \in [1, m] \ . \tag{2.21}$$

If the inequality holds, then the equivalence relation is written as $\mathbf{p} \prec \mathbf{q}$ and one says that $\mathbf{q}$ majorizes $\mathbf{p}$. Consequently, Eq. (2.21) is a measure of uncertainty, which implies that a vector

[1] The word moment is very fitting as the quantity is reminiscent of the moments of inertia in classical mechanics.

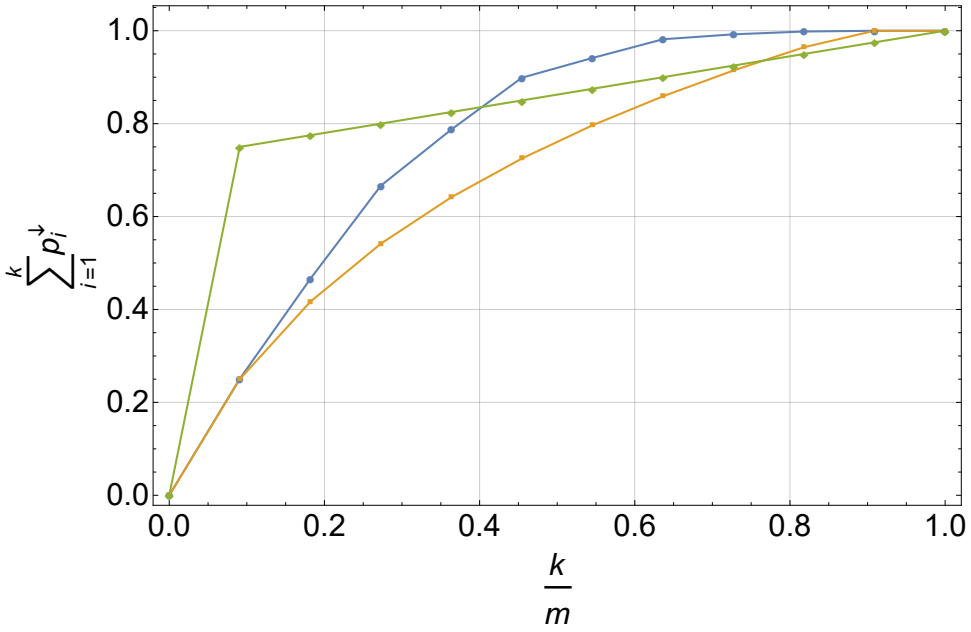


Fig. 2.5: Visualization of majorization, where the blue curve majorizes the yellow one, while no such relation can be given for the green curve, making this distribution incomparable by means of majorization. Here m is the sample size and k the index of summation.

is less spread out if it preserves the partial order at every step

$$\begin{aligned} p_1^{\downarrow} &\leq q_1^{\downarrow} \\ p_1^{\downarrow} + p_2^{\downarrow} &\leq q_1^{\downarrow} + q_2^{\downarrow} \\ &\vdots \\ \sum_{i=1}^{m} p_i^{\downarrow} &= \sum_{i=1}^{m} q_i^{\downarrow} = 1 \ . \end{aligned} \tag{2.22}$$

Example 2.2.4. For the three probability vectors $\mathbf{p}_2 = (0,1,0)^T$, $\mathbf{p}_2 = (\frac{1}{2},\frac{1}{6},\frac{1}{3})^T$ and $\mathbf{p}_3 = (\frac{1}{3},\frac{1}{3},\frac{1}{3})^T$ a quick test shows that $\mathbf{p}_2 \succ \mathbf{p}_1 \succ \mathbf{p}_3$. Generally, a distribution consisting only of a single certain event majorizes every other distribution, while the uniformly distributed events are majorized by every other probability vector. ◊

If the partial sums in definition Eq. (2.21) are represented by the points $(\frac{k}{m}, \sum_{i=1}^{k} p_i^{\downarrow})$ as coordinates, where the upper bound of summation k goes from 0 to m, and points are connected by straight lines, then the curve obtained that way graphically indicates majorization. As illustrated in Fig. 2.5 a distribution that is constantly below the other one is accordingly majorized,

while a curve that is intersected between the endpoints is not.

Note the following immensely important attribute related to majorization: a function $f : \mathbb{R}^d \to \mathbb{R}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ is called *Schur concave* when it satisfies

$$f(\mathbf{x}) \geq f(\mathbf{y}) \ , \qquad \text{if} \quad \mathbf{x} \prec \mathbf{y} \ . \tag{2.23}$$

So f is Schur concave means: the less spread out the distribution $\mathbf{x}$ is compared to $\mathbf{y}$, the larger the value of the function gets. The standard deviation is an example of a Schur convex function. If $\mathbf{x} \prec \mathbf{y}$, then the standard deviation is also smaller $\sigma(\mathbf{x}) \leq \sigma(\mathbf{y})$.

2.2.4 Statistical distances

For a comparison of two probability distributions the distance between the two stochastic objects can be regarded. In mathematics, a *metric* on a non-empty set X is a *distance* $d : X \times X \to \mathbb{R}$ between a pair of elements x and y if the following conditions for $x, y, z \in X$ hold

1. $d(x, y) \geq 0$ with equality $d(x, y) = 0$ for $x = y$.
2. $d(x, y) = d(y, x)$, the metric is symmetric.
3. $d(x, z) \leq d(x, y) + d(y, z)$, the triangular inequality is satisfied.

In close analogy to the *p-norm* of vector spaces there exits an L^p metric [43] for probability measures $p(x), q(x)$ which, for a finite set with $p \geq 1$, are given by

$$(\sum_x |p(x) - q(x)|^{\frac{1}{p}}) \ . \tag{2.24}$$

This metric for $p = 1$ is twice the *total variation distance* $d_{TV}(\mathbf{p}, \mathbf{q}) = \frac{1}{2}\|\mathbf{p}-\mathbf{q}\|_1$, which gives the maximum difference between the two discrete probability distributions. The *Hellinger distance* is related to the L^2-difference for the square root of the probability distributions $\sqrt{\mathbf{p}}, \sqrt{\mathbf{q}}$

$$d_H(\mathbf{p}, \mathbf{q}) = \frac{1}{\sqrt{2}}\|\sqrt{\mathbf{p}} - \sqrt{\mathbf{q}}\|_2 = \sqrt{1 - \mathrm{BC}(\mathbf{p}, \mathbf{q})} \ , \tag{2.25}$$

where $\mathrm{BC}(\mathbf{p}, \mathbf{q}) = \sum_i^n \sqrt{p_i q_i}$ is known as *Bhattacharyya coefficient* or *fidelity*. The list of stochastic distances is much longer than the few presented just now, but those mentioned here

have in a generalized form found their way into quantum physics.

Another common statistical distance is called the *Wasserstein metric*. For its motivation the *transportation problem*, formalized by French mathematician Gaspard Monge, shall be regarded. Assume there is a certain amount of earth excavated from the ground from a wide area X. For further processing the soil is locally put into many piles x and transported to different factories F. The original question is: how must the transport be set-up, so that the monetary costs for the delivery are the lowest. If the transport is given by a one-to-one map $T : X \to F$ from each pile of earth $x \in X$ to each facility $T(x)$ and the particular cost function for the deliveries is $c(x, T(x))$, then one seeks to minimize the total cost $\sum_{x \in X} c(x, T(x))$.

The transport problem was later rigorously formalized in measure theory and therefore has immediate significance to probability measures. Given two (separable, metric) probability spaces (X, μ) and (Y, ν) and a cost function $c : X \times Y \to \mathbb{R}$, Monge's optimal transport is given as the infimum

$$\inf \left(\int_X c(x, T(x)) d\mu(x) \right) \tag{2.26}$$

for all transport plans. The issue with Eq. (2.26) is that the map T does not in general exist. The problem was simplified by Kantorovich who considered the probability measure π to be from the product space $X \times Y$ such that π admits μ and ν as marginals: for all measurable sets $A \subset X$, $B \subset Y$; $\pi(A \times Y) = \mu(A)$, $\pi(X \times B) = \nu(B)$. The collection of probability measures π with marginals μ, ν is said to the set of all admissible *transference plans* and is denoted by $\Pi(\mu, \nu)$. This additional requirement obviates the problem that a transported unit of mass is not splittable. Monge's transport problem Eq. (2.26) is then recast to

$$\inf \left(\int_{X \times Y} c(x, y) d\pi(x, y) \, | \pi \in \Pi(\mu, \nu) \right) . \tag{2.27}$$

The *Wasserstein p-distance* between two probability measures appears as the optimal transport cost from one to the other, when the 'cost' to be minimized is the distance function

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{X \times Y} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}} . \tag{2.28}$$

So, the Wasserstein 2-distance is a measure of the most efficient way to transport one probability

distribution onto the other, where the measure of cost to be minimized is the Euclidean distance. At the end of chapter 6 an uncertainty relation is treated briefly, which uses the Wasserstein 2-metric as an accuracy measure. Although it is not a topic of the thesis, it should be noted that there are other classical distances for the purpose of quantifying uncertainty relations, see for example [44].

2.3 Entropy in information theory

Information is a frequently used term in everyday life. Many have a good understanding what information is, however, scientifically it remains very vague in itself without a rigorous mathematical formalism. Information must explain different ideas like transference, acquisition and communication of knowledge and assess some quality to the data, e.g ordered or disordered, new or outdated, copied, compressed, redundant or knowledgeable. The following subsections introduce the basics of information theory.

2.3.1 Self-information

Information is naturally related to probability theory. When a person receives news the reaction usually depends on the value of the content. So, if someone replied to a new message for example with "that doesn't come as a surprise" that person basically expresses that there is almost no gain of information because the event was highly expected or very probable, while an opposite reaction would indicate that the imparted facts are very useful as a consequence of being unlikely.

Putting it mathematically, if $p(x_i)$ is the probability for the event x_i and $I(x_i)$ the information, then

1. $I(x_i)$ should be non-negative,

2. approach zero for the certain event $p(x_i) \to 1 \Rightarrow I(x_i) \to 0$,

3. increase for $p(x_i) \to 0$ and

4. for $x_1 \neq x_2$, less probable events $p(x_1) < p(x_2)$ should be more informative $I(x_1) > I(x_2)$.

These features can be modeled by making information a continuous function of the probability $I(x) = f(p(x_i))$. Another property usually demanded is that the total information associated with statistically independent joint events $p(x_1, x_2) = p(x_1)p(x_2)$ should be additive, i.e.

$$p(x_1, x_2) = p(x_1)p(x_2) \quad \Rightarrow \quad I(x_1, x_2) = I(x_1) + I(x_2) \ . \tag{2.29}$$

But, because of the functional relation to the probability the following equation has to hold

$$f(p(x_1, x_2)) = I(x_1, x_2) = I(x_1) + I(x_2) = f(p(x_1)) + f(p(x_2)) \ . \tag{2.30}$$

The function that has the property $f(x\,y) = f(x) + f(y)$ is the logarithm. The so-called *self-information* or *surprisal* $I(x)$ is therefore defined as

$$I(x_i) = -\log(p(x_i)) \ , \tag{2.31}$$

where the minus guarantees that information remains a positive quantity. The choice for the logarithmic base is arbitrary, but by convention the base is 2 making *1 bit* the basic unit of information. To understand the implication of the base choice, the following example is of help.

Example 2.3.1. For a fair coin the information of getting heads or tails is $I(\text{heads}) = I(\text{tails}) = -\log_2(1/2) = 1$ bit and for a particular singular outcome x_i of a die $I(x_i) = -\log_2(1/6) \approx 2.585$ bits. The second result seems a bit odd, but it tells us that the result can be coded by a string of 3 bits. On the other hand for the base six $I(\boxdot) = -\log_6(1/6) = 1$ the result is a nice integer, but in that case all elementary information must be coded by six different symbols instead of only 0's and 1's. ♢

2.3.2 Shannon entropy

In sec. 2.2.2 the expected value and standard deviation were introduced as two important means of condensing usually large data and information about the distributions to a few characterizing numbers. Consequently, such procedures also exist for $I(x)$ and indeed the renown *Shannon entropy* $H(X)$ is nothing but the weighted average of the self-information:

$$H(X) = \langle I(x_i) \rangle = -\sum_{i=1}^{N} p(x_i) \log(p(x_i)) \ . \tag{2.32}$$

The resemblance with Eq. (2.12) is apparent. It is good to know the properties of the Shannon entropy $H(X)$ of one random variable

- The entropy vanishes $H = 0$ for the certain event $p(x_i) = 1$ and likewise for the impossible event, $p(x_i) = 0$ for $\forall x_i \in X$, the second case being consistent with the limit $\lim_{p \to 0+} p \log(p) = 0$. If there is no uncertainty about the outcome of an event there is no gain of information.

- The highest value H_{top} is attained for the uniform distribution $p(x_i) = 1/N$ and is identical to the self-information in that case: $H_{top} = -\sum_{i=1}^{N} \frac{1}{N} \log(1/N) = \log(N) = I(x_i)$, where N is the number of elements of the sample space. Consequently, the entropy is a positive, continuous function with $0 \leq H(X) \leq \log(N)$.

- The entropy does not depend on the sequence with which the probabilities occur $H(p_i, p_j) = H(p_j, p_i)$ and is also invariant under expansion by impossible events $H(p1, ..., p_N, 0) = H(p1, ..., p_N)$.

- Schur Concavity. The entropy inherits the properties of majorization (see sec. 2.2.3). If the probability distribution $\mathbf{x}$ is 'less spread out' than $\mathbf{y}$, to be more accurately $\mathbf{x} \prec \mathbf{y}$, then its Shannon entropy will be larger $H(\mathbf{x}) \geq H(\mathbf{y})$.

- The *binary entropy function* describes the entropy for a random variable with only two possible outcomes $p(X = 1) = x$ and $p(X = 0) = (1 - x)$ and is mostly written as $\tilde{H}_b(x) = -x \log(x) - (1 - x) \log(1 - x)$. An alternative, more symmetric parametrization of the binary entropy for $p(X = 1) = \frac{1+x}{2}$ and $p(X = -1) = \frac{1-x}{2}$ is given by

$$H_b(x) = -\left(\frac{1+x}{2}\right) \log_2 \left(\frac{1+x}{2}\right) - \left(\frac{1-x}{2}\right) \log_2 \left(\frac{1-x}{2}\right) \tag{2.33}$$

  and its derivative is

$$\frac{d}{dx} H_b(x) = -\log_2 \left(\frac{1+x}{1-x}\right) \ . \tag{2.34}$$

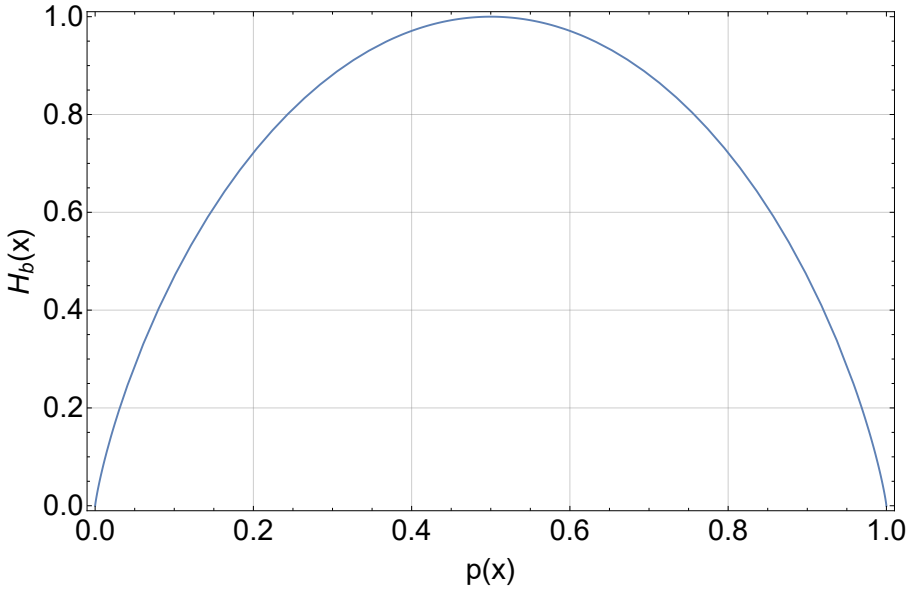


Fig. 2.6: Binary Entropy function. The maximum is given at equal likeliness $p(x) = 1/2$ and the minima at absolute outcome certainty, i.e. either 100 % possibility or impossibility. The further away the result from the center the less equiprobable the original probability distribution is.

Example 2.3.2. The entropy of a coin toss is given by the binary Shannon entropy $H_b(x)$ which is plotted in Fig. 2.6. In this case heads and tails occur equally frequently, the coin toss is fair and the entropy is maximal, while if the resulting entropy is less then 1 bit it can be said that the coin toss is rigged. ◊

2.3.3 Joint, conditional, mutual and relative entropy

To gain further insights it is necessary to have a closer look at the Shannon entropy depending on at least two random variables. Since entropy and probabilities are functionally linked, similar expression as in sec. 2.2.1 can be obtained. First, the definition of the *joint entropy*, associated with the joint probability $p(x_i, y_j)$ of the random variables X,Y, is given by

$$H(X,Y) = -\sum_i \sum_j p(x_i, y_j) \log(p(x_i, y_j)) \tag{2.35}$$

and from this definition the *conditional entropy* is deducible

$$H(X|Y) = H(X,Y) - H(Y) = -\sum_i \sum_j p(x_i, y_j) \log(p(x_i|y_j)) \ . \tag{2.36}$$

Proceeding with the list of properties of the Shannon entropy the following statements hold

- The joint entropy is symmetric $H(X,Y) = H(Y,X)$ and subadditive $H(X,Y) \leq H(X) + H(Y)$. Subadditivity was introduced in Eq. (2.6) and, as was the case for probabilities, equality only holds for statistically independent events. For correlated random variables it can be said that knowledge about the outcome decreases the information.

- Since the conditional probability is generally not symmetric, the conditional entropy is also not symmetric $H(X|Y) \neq H(Y|X) = H(X,Y) - H(X)$. It is possible to define a Bayes' theorem also for the Shannon entropy which follows from the symmetry of the joint entropy

$$H(Y|X) = H(X|Y) - H(X) + H(Y) \ . \tag{2.37}$$

The joint entropy tells something about the information of two random processes altogether. Naturally, there should be an entropy that quantifies how much information two events have in common. A measure of information coupling is given by the *mutual information* defined as

$$I(X;Y) = \sum_{i=1} \sum_{j=1} p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{p(x_i) p(y_j)} \right) \ . \tag{2.38}$$

- The mutual information is symmetric $I(X;Y) = I(Y;X)$, non-negative $I(X;Y) \geq 0$ and vanishes exactly when the two samples are independent, which reflects that this quantity is a measure of the overlap of the joint probability distributions with its marginals.

- Mutual information can be connected to the other entropies in many ways, e.g.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \ . \tag{2.39}$$

A nice feature of all these entropies is that they can be visualized in a Venn-Diagram, see Fig. 2.7, in the same way as probability sets. In this diagram the two marginal entropies $H(X)$ and $H(Y)$ are subsets of the joint entropy $H(X,Y)$, the conditional entropies $H(X|Y), H(Y|X)$ are given by the relative complements and the mutual information $I(X;Y)$ corresponds to the intersection of the two disks. The formula Eq. (2.36), Eq. (2.37) and Eq. (2.39) are very intuitive

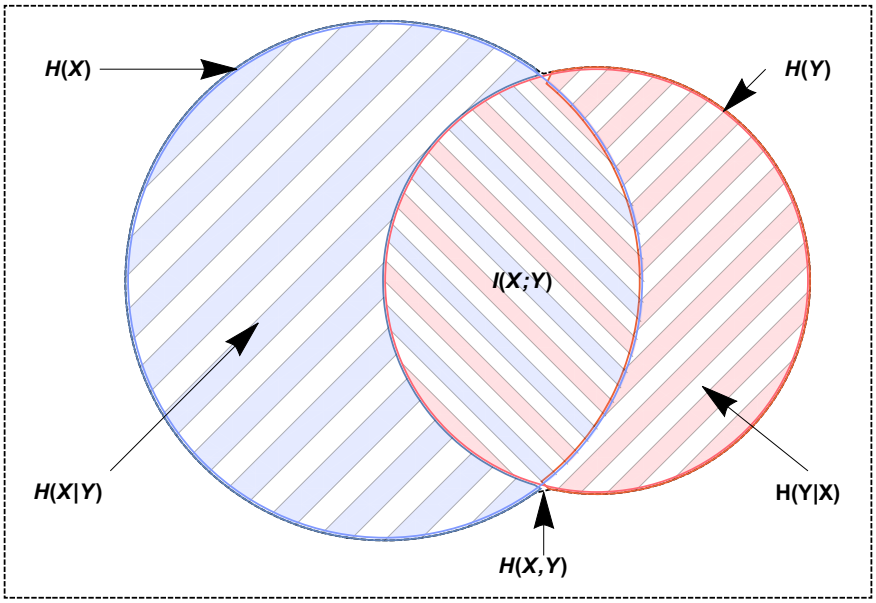


Fig. 2.7: Venn-Diagram for bivariate entropies which in set algebra terms can be written as $H(X,Y) \leftrightarrow H(X \cup Y)$ for the joint entropy, $I(X;Y) \leftrightarrow H(X \cap Y)$ for the mutual information and $H(X|Y) \leftrightarrow H(X \cap Y^c)$ for the conditional entropy, where $X \cap Y^c$ is the relative complement of the sets.

in this particular picture.

As a logical continuation a useful quantity will be presented, namely the *Kullback-Leibler divergence (KL-divergence)*, which also goes by some other names like *relative entropy* or information divergence. Formally, it is defined as

$$D_{KL}(P||Q) = \sum_i p(x_i) \log \left(\frac{p(x_i)}{q(x_i)} \right) = -H(X) + H(X;P;Q) \tag{2.40}$$

and is frequently used to measure the difference between two probability distributions **p** and **q**. The mutual information $I(X;Y) = D_{KL}\left(p(x_i,y_j)||p(x_i)p(y_j)\right)$ for example can be written in terms of the KL-divergence. The quantity $H(X;P;Q)$ is known as *cross entropy*. The relative entropy compares distributions, but since Eq. (2.40) is neither symmetric nor obeys the triangle inequality it is thoroughly distinguished from distances by the term (statistical) divergence.

p.d.f	std. dev. $\sigma(f)$	diff. entr. $h(x)$
$f(x) = \frac{1}{b-a}$	$\frac{b-a}{2\sqrt{3}}$	$\log(b-a)$
$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	σ	$\ln(\sigma\sqrt{2\pi e})$
$f(x) = \frac{1}{\pi}\frac{a}{a^2+x^2}$	undef.	$\ln(2\pi a)$

Tab. 2.1: The table lists the uniform, the normal and the Cauchy distributions along with their respective standard deviations, if existent, and their differential entropies.

2.3.4 Rényi and differential entropy

Are there other continuous, Schur concave functions that maximize in the case of a uniform/normal probability distribution, are invariant under extension by null events and also satisfy the product rule Eq. (2.30) for independent random variables? The class of *Rényi entropies*

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_{x\in X} p^\alpha(x)\right), \qquad \alpha \geq 0, \alpha \neq 1 \tag{2.41}$$

is the generalized version which meets all the requirements. The value α influences the weighing of the events: $\alpha = 0$ gives lower weights to less surprising events while larger α's put more priority to events with high surprisal. The Shannon entropy $H(X) = \lim_{\alpha\to 1} H_\alpha(X)$ is special among Rényi entropies as it is the only one with a commonly accepted notion of conditional entropy [45] and therefore has a well defined chain rule (l.h.s of Eq. (2.36)).

In contrast to a discrete, a continuous source emits a symbol continuum, i.e the information in a signal is capable of having uncountable many values. The theoretic model of this source is in the simplest case a one-dimensional random variable X which can assume arbitrary values x in an interval and is described by a continuous probability density $f_X(x)$. Instead of the Shannon entropy, the *differential entropy*

$$h(x) = -\int_{\mathrm{X}} f_X(x) \log(f_X(x)) dx \tag{2.42}$$

is used for continuous probability distributions.

Since every selected value x occurs with typically zero probability and arbitrarily close values are practically indistinguishable, it is meaningless to ask for the information content of such a

value. No entropy can be given in the sense of the mean information content per symbol, as is considered for the discrete source. Differential entropy should hence not be considered as a straightforward extension of the Shannon entropy [46].

Despite some dissimilarities with the Shannon entropy, the differential entropy is very helpful in describing probability distributions, because it is closely related to standard deviations. The table Tab. 2.1 contains three different probability distribution functions along with the respective entropies. The example of a uniform probability density already implies a few important facts: the entropy $h(x)$ becomes negative when $b - a < 1$ and diverges for a Dirac delta distribution $b - a \to 0$. For the normal distribution the entropy grows as the standard deviation increases. It is particularly interesting that for a given variance a Gaussian provides the probability density with the maximal entropy, therefore for all $h(x)$

$$h(x) \leq \ln(\sqrt{2\pi e\sigma^2}) \tag{2.43}$$

with equality for a normal distribution. The last function in Tab. 2.1 is an example which has no finite standard deviation, but a definite entropy.

CHAPTER 3

Quantum formalism

The mathematical formalism of quantum mechanics has not only been constantly complemented but also conceptually developed further since its foundation. An overview of the theoretical concepts of quantum physics important in this thesis is given. In general, no proofs are provided, but only references to standard textbooks are cited. The key sources for the following sections in this chapter are [21, 47–51].

3.1 Linear operators in Hilbert space

A Hilbert space $\mathcal{H}$ over complex numbers is a vector space in which a scalar product $\langle\cdot|\cdot\rangle$ is defined and for all $\phi, \chi, \psi \in \mathcal{H}$ and $c_1, c_2 \in \mathbb{C}$: (i) the inner product is anti-linear in its first argument and linear its second $\langle c_1\phi|c_2\chi + \psi\rangle = c_1^*(c_2\langle\phi|\chi\rangle + \langle\phi|\psi\rangle)$, (ii) complex conjugation is given by $\langle\phi|\chi\rangle^* = \langle\chi|\phi\rangle$, (iii) $\langle\phi|\phi\rangle \geq 0$ and (iv) $\langle\phi|\phi\rangle > 0$ if $\phi \neq 0$. Every vector in a Hilbert space can be written uniquely as a *linear combination* of orthogonal basis vectors $\{|e_n\rangle\}$

$$|\psi\rangle = \sum_n c_n |e_n\rangle = \sum_n \langle e_n|\psi\rangle |e_n\rangle \tag{3.1}$$

with the coefficients retrieved by $\langle e_m|\psi\rangle = \sum_n c_n \langle e_m|e_n\rangle = \sum_n c_n\delta_{nm} = c_m$. From this Eq. (3.1) *Parseval's identity* can be deduced $\|\psi\|^2 = \langle\psi|\psi\rangle = \sum_n |\langle e_n|\psi\rangle|^2$. This identity holds for all inner-product spaces, but the name is mostly mentioned for the special case of

the Fourier theorem, where $|e_n\rangle = \{e^{-inx}\}_{n\in\mathbb{N}}$ is an orthonormal basis for $L^2[-\pi,\pi]$ and the coefficients

$$c_n = \langle f|e_n\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx \tag{3.2}$$

are exactly given by the *Fourier transform*. The Fourier transform is a linear, unitary transformation and in its discrete form is simply a change of basis. As mentioned in the preface, the quantum superposition Eq. (3.1) is mathematically the least common attribute rooted in all uncertainty relations. Explicit applications of unitary transformations for uncertainty relations are found in [52, 53].

Operators map elements from one vector or function space to another. If $\psi \in \mathcal{H}$ and every ϕ belongs to the *domain* of observable $\hat{A}$ then the corresponding mapping is typically written as $|\psi\rangle = \hat{A}\,|\phi\rangle$. A mapping $\hat{A}$ is linear when $\hat{A}(c\,|\phi\rangle + |\chi\rangle) = c\hat{A}\,|\phi\rangle + \hat{A}\,|\chi\rangle\,,\ \forall\,|\phi\rangle\,,|\chi\rangle \in \mathcal{H}\,,\forall c \in \mathbb{C}$ and is a bounded operator if there exists a real number $r \geq 0$ such that $\|\hat{A}\phi\| \leq r\|\phi\|$. The *Banach space* $\mathcal{L}(\mathcal{H})$, fulfills the axioms of vector spaces and is a normed space with the *operator norm* $\|\hat{A}\| := \sup\|\hat{A}\phi\|\ ,\|\phi\| = 1$. A bounded operator is *Hermitian* if it is self-adjoint $\hat{A} = \hat{A}^\dagger$, write $\hat{A} \in \mathcal{L}_s(\mathcal{H})$ (the space of self-adjoint operators). Since Hermitian operators have a real-valued spectrum a bounded operator is regarded as self-adjoint if $\langle\psi|\hat{A}|\psi\rangle$ is a real number and it is called positive if $\langle\psi|\hat{A}|\psi\rangle \geq 0$ (or simply $\hat{A} > 0$) for all $\psi \in \mathcal{H}$. It is useful to observe that any operator that satisfies $\hat{A} = \hat{B}^\dagger\hat{B}$ is positive, because $\langle\psi|\hat{A}|\psi\rangle = \langle\psi|\hat{B}^\dagger\hat{B}|\psi\rangle = \langle\hat{B}\psi|\hat{B}\psi\rangle = \|\hat{B}\psi\| \geq 0$. In that case the operator might be identified with its *square root* $\hat{B} = \sqrt{\hat{A}}$.

For positive bounded operators the *trace* operation is well-defined as summing of the diagonal entries

$$\mathrm{Tr}(\hat{A}) := \sum_i^d \langle\phi_i|\hat{A}|\phi_i\rangle\,, \tag{3.3}$$

which does not depend on the chosen orthonormal basis $\{|\phi_i\rangle\} \in \mathbb{C}^d$. With the definition of the trace at hand it is possible to define: an arbitrary operator in $\mathcal{L}(\mathcal{H})$ is called a trace-class operator if $\|\hat{A}\|_{\mathrm{Tr}} := \mathrm{Tr}(\sqrt{\hat{A}^\dagger\hat{A}}) < \infty$ and shall be denoted by $\mathcal{T}(\mathcal{H})$. For two trace class operators $\hat{A}$, $\hat{B}$ it is possible to introduce the *Hilbert-Schmidt inner product* given as

$$\langle\hat{A}|\hat{B}\rangle = \mathrm{Tr}(\hat{A}^\dagger\hat{B}) \tag{3.4}$$

and the induced *Hilbert-Schmidt norm* $\|\hat{A}\|_{\mathrm{HS}} := \sqrt{\langle \hat{A}|\hat{A}\rangle}$. The operators satisfying $\|\hat{A}\|_{\mathrm{HS}} < \infty$ are called *Hilbert-Schmidt operators*.

A self-adjoint operator $\hat{P}$, element of a set of projections $\mathcal{P}(\mathcal{H})$, is an *orthogonal projection* onto a closed subspace $M \subset \mathcal{H}$ if it is *idempotent* $\hat{P}^2 = \hat{P}$. The norm of a projector is $\|\hat{P}\| = 1$ for $P \neq 0$ and it's eigenvalues are 1 and 0. For two distinct projection operators $\hat{P}_1$ and $\hat{P}_2$ it can be said that $\hat{P}_1\hat{P}_2$ is zero if the projectors are in orthogonal subspaces $M, M^{\perp}$. The sum $\hat{P}_1 + \hat{P}_2$ is a projector only when they belong to orthogonal subspaces, the product $\hat{P}_1\hat{P}_2$ of two projectors is generally not a projector unless they commute $[\hat{P}_1, \hat{P}_2] = 0$. A one-dimensional projection of the form $\hat{P} = |\phi_i\rangle \langle\phi_i|$, where $\{|\phi_i\rangle\}$ is a complete orthonormal system, is a *rank-1* operator. To every projection operator $\hat{P}$ exists a *complementary projection operator* $\hat{P}^{\perp} = \mathbb{1} - \hat{P}$. A projector decomposes a linear Hilbert space to $\mathcal{H} = M^{\perp} \oplus M$.

A bounded linear operator $\hat{U}$ is called *unitary* if $\hat{U}^{\dagger}\hat{U} = \hat{U}\hat{U}^{\dagger} = \mathbb{1}$, where the adjoint can also be identified with it's inverse $\hat{U}^{\dagger} = \hat{U}^{-1}$, and it is called *isometric* because it preserves the norm $\|\hat{U}\phi\| = \|\phi\|$. The operator norm yields $\|\hat{U}\| = 1$ and eigenvalues of a unitary operator are of the form e^{ir}, $r \in \mathbb{R}$. Unitary operators are related to self-adjoint operators, that is if $\hat{A} = \hat{A}^{\dagger}$ then $e^{i\hat{A}}$ is unitary and the adjoint exponential operator is simply given by $(e^{i\hat{A}})^{\dagger} = e^{-i\hat{A}}$. A crucial property for the product of two unitary operators is given by the *Baker-Campbell-Hausdorff formula*, which indicates that the exponential map for non-commuting operators $[\hat{A}, \hat{B}] \neq 0$ is generally not homomorphic $e^{i\hat{A}}e^{i\hat{B}} \neq e^{i(\hat{A}+\hat{B})}$.

3.1.1 Common operator decompositions

In the following, some frequently used operator and matrix factorizations are introduced which appear later on. For a purely discrete spectrum the *spectral decomposition* works out in the same way as the *eigenvalue decomposition* of Hermitian matrices, which for *non-degenerate eigenvalues* is

$$\hat{A} = \sum_i a_i \, |a_i\rangle\langle a_i| = \sum_i a_i \hat{P}(a_i) \ . \tag{3.5}$$

The parenthesis after the projector is meant to emphasize the association with the operator by its eigenvalue decomposition. Every self-adjoint operator has a complete orthonormal system of eigenvectors $|a_i\rangle$, that span a linear subspace. The eigenvectors of a Hermitian matrix with

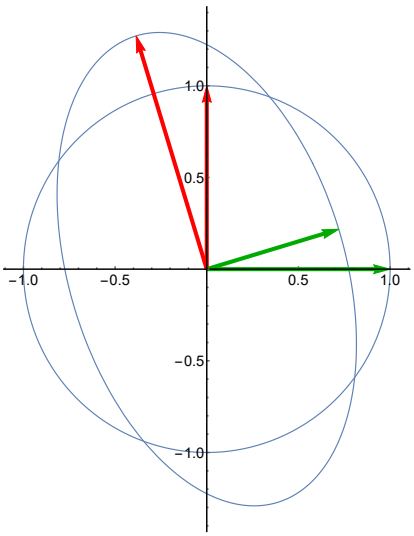


Fig. 3.1: Interpretation of the matrix polar decomposition, where a rotation is followed by a stretch along the green and red arrows to end up in the final configuration.

real eigenvalue $a_i = \langle a_i|\hat{A}|a_i\rangle$ constitute an orthogonal basis and the projection operators that arise that way sum to unity

$$\sum_i \hat{P}(a_i) = \sum_i |a_i\rangle\langle a_i| = \mathbb{1} \ . \tag{3.6}$$

In analogy to the polar form of complex numbers $z = re^{i\phi}$, any linear operator has a *polar decomposition*

$$\hat{A} = |\hat{A}|\hat{U} = \sqrt{\hat{A}^\dagger \hat{A}}\ \hat{U} \ , \tag{3.7}$$

where $|\hat{A}|$ is the absolute and $\hat{U}$ is a unitary operator. Geometrically a matrix polar decomposition splits the action of the deformation into a shape conserving rotation and a stretching, which is illustrated in Fig. 3.1.

Let $\hat{A}$ be an $m \times n$ matrix. Then there exists a unitary $m \times m$ matrix $\hat{U}$, a unitary $n \times n$ matrix $\hat{V}$ and an $m \times n$ diagonal matrix $\hat{\Sigma}$ with nonnegative, decreasing real entries σ_i on the diagonal such that

$$\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\dagger \ . \tag{3.8}$$

The form Eq. (3.8) is called the *singular value decomposition (SVD)* of $\hat{A}$. Because of the *matrix similarity* $\hat{A}^\dagger\hat{A} = \hat{V}(\hat{\Sigma}^\dagger\hat{\Sigma})\hat{V}^{-1}$, the square root of the eigenvalues of the Hermitian,

positive matrix $\hat{A}\hat{A}^\dagger$ are the diagonal elements of $\hat{\Sigma}$ and the σ_i are called the *singular values*.

Example 3.1.1.

$$\hat{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} , \qquad \hat{A}^\dagger \hat{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} . \tag{3.9}$$

The square root of the eigenvalues of $\hat{A}^\dagger\hat{A}$ are $\sigma_1 = \sqrt{3}, \sigma_2 = \sqrt{1}$ and the eigenvectors $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1)^T$ and $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(-1,1)^T$ constitute the unitary matrix $\hat{V} = (\mathbf{v}_1, \mathbf{v}_2)$. The second unitary matrix is calculated by $\mathbf{u}_i = \frac{1}{\sigma_i}\hat{A}\mathbf{v}_i$ and is therefore given by $\hat{U} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_1 \times \mathbf{u}_2)$. The singular value decomposition in the example is therefore

$$\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\dagger = \begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} . \tag{3.10}$$

$\Diamond$

The SVD for square matrices has a similar geometric interpretation as the matrix polar decomposition containing one additional rotation, but it also extends to matrices with unequal row and column sizes. Another very important feature of the SVD is that it is a common method to calculate the *Moore - Penrose pseudoinverse* $\hat{A}^+$, the inverse of rectangular or singular matrices. To be more precisely, a matrix $\hat{A} \in \mathbb{C}^{m\times n}$ has a pseudoinverse $\hat{A}^+ \in \mathbb{C}^{n\times m}$ when satisfying the following four criteria: (i) $\hat{A}\hat{A}^+\hat{A} = \hat{A}$, (ii) $\hat{A}^+\hat{A}\hat{A}^+ = \hat{A}^+$ and Hermiticity in both directions (iii) $(\hat{A}\hat{A}^+)^\dagger = \hat{A}\hat{A}^+$ (iv) $(\hat{A}^+\hat{A})^\dagger = \hat{A}^+\hat{A}$. For a matrix with SVD as in Eq. (3.8) the pseudoinverse is $\hat{A}^+ = \hat{V}\hat{\Sigma}^+\hat{U}^\dagger$, where $\hat{\Sigma}^+$ is attained by inverting the nonzero (diagonal) entries and transposing the matrix. The inverse of the matrix from the previous example is calculated as

Example 3.1.2.

$$\hat{\Sigma}^+ = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \Rightarrow \qquad \hat{A}^+ = \frac{1}{3}\begin{pmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix} . \tag{3.11}$$

$\Diamond$

3.2 Density operator

The *density operator* $\hat{\rho}$ (a.k.a *statistical operator*) on a Hilbert space $\mathcal{H}$ is defined as a Hermitian, positive, bounded trace-class operator with unit trace $\mathrm{Tr}(\hat{\rho}) = 1$, formally

$$\mathcal{S}(\mathcal{H}) := \{\hat{\rho} \in \mathcal{T}(\mathcal{H}) \mid \hat{\rho} \geq 0, \mathrm{Tr}(\hat{\rho}) = 1\} \ . \tag{3.12}$$

The set $\mathcal{S}(\mathcal{H})$ of all density operators is convex, which means that $\hat{\rho} = \lambda\hat{\rho}_1 + (1-\lambda)\hat{\rho}_2 \in \mathcal{S}(\mathcal{H})$ for all $\hat{\rho}_1, \hat{\rho}_2 \in \mathcal{S}(\mathcal{H})$ and $\lambda_i \in [0,1]$; The extreme points of density operators are the one-dimensional orthogonal projections $\hat{P}_\psi = |\psi\rangle\langle\psi|$, where $\|\psi\| = 1$, which correspond to pure states of the Hilbert space. A criterion for pureness of the density matrix is therefore given by idempotence $\hat{\rho}^2 = \hat{\rho}$, from which follows

$$\begin{aligned} \mathrm{Tr}(\hat{\rho}^2) &= 1 \qquad \text{pure states} \ , \\ \mathrm{Tr}(\hat{\rho}^2) &< 1 \qquad \text{mixed states} \ . \end{aligned} \tag{3.13}$$

One-dimensional orthogonal projections have *rank 1*, which therefore also holds for pure states, while convex combination of two orthogonal rank 1 elements can be used to construct higher ranked elements. The rank of the operator in a convex set is given as the minimum number of projectors that are necessary to express it as a convex combination.

Another property of mixes states is that they can be written as a mixture of pure states in many ways if they are not extremal points of the convex set, for example $\hat{\rho} = \sum_i \psi_i \hat{P}_{\psi_i} = \sum_i \phi_i \hat{P}_{\phi_i}$, where $|\psi_i\rangle, |\phi_i\rangle$ need not be orthogonal unit vectors. Even for orthogonal vectors the convex combination is generally not unique. A visualization of the situation is given in Fig. 3.2 . For $\dim(\mathcal{H}) \geq 3$ it can be shown that a density matrix belongs to the boundary if one of the eigenvalues in the diagonalized form Eq. (3.5) equals zero [49]. For dimension 2 the state space is spherical as indicated in Fig. 3.2 and the boundary coincides with the set of all pure states. In quantum physics, if the same density matrix is given for ensembles consisting of different pure states then physically these states cannot be distinguished by any operational procedure.

3.2.1 Bloch vectors

The *Bloch sphere* is very useful to visualize the geometry of a two-dimensional quantum system. It emerges from the symmetry group of the two-dimensional quantum system $SU(2)$ which is a the set of matrices in $\mathbb{C}^2$ with unit determinant

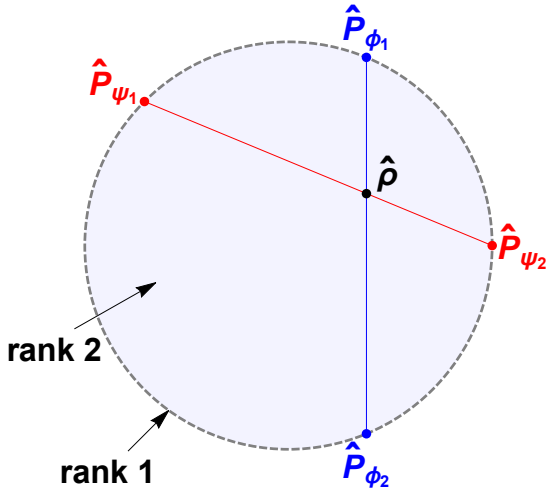


Fig. 3.2: Any interior point in the convex disk corresponds to a mixed state that can be realized by an arbitrary convex combination of two extremal points. The projectors that lie on the exterior circle have rank 1, while the rank increases in the interior.

$$\hat{\mathcal{U}} = \begin{pmatrix} z_1 & -z_2 \\ z_2^* & z_1^* \end{pmatrix} , \qquad \det(\hat{\mathcal{U}}) = |z_1|^2 + |z_2|^2 = 1 \ . \tag{3.14}$$

By decomposing the complex entries of the matrix into real and imaginary part $z_1 = x_0 + ix_1$ and $z_2 = x_2 + ix_3$ it becomes obvious that Eq. (3.14) describes a 3 dimensional sphere $x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1$ embedded in the four dimensional real space $\mathbb{R}^4$. The reason Bloch states can be drawn in Euclidean space is due the fact that the global phase $e^{i\alpha}$ ($S^1 \cong U(1)$) is typically removed from the formalization of two-level quantum states[2]. Taking the modulo $S^2 \cong S^3/S^1 \cong SU(2)/U(1)$ exactly gives the surface of a ball in $\mathbb{R}^3$ as shown in Fig. 3.3, whose Hilbert state can be parametrized by a polar angle ϑ and an azimuthal angle φ

$$|\chi(\vartheta,\varphi)\rangle = \begin{pmatrix} \cos\left(\frac{\vartheta}{2}\right) \\ e^{i\varphi}\sin\left(\frac{\vartheta}{2}\right) \end{pmatrix} = \cos\left(\frac{\vartheta}{2}\right)|+z\rangle + e^{i\varphi}\sin\left(\frac{\vartheta}{2}\right)|-z\rangle \ . \tag{3.15}$$

Notice that there is only half the angle in Eq. (3.15), because the *generators* of $SU(2)$ are given by $\hat{S}_i = \frac{1}{2}\hat{\sigma}_i$, with the Pauli observables

[2]These phase circles S^1 are still related to geometric phases.

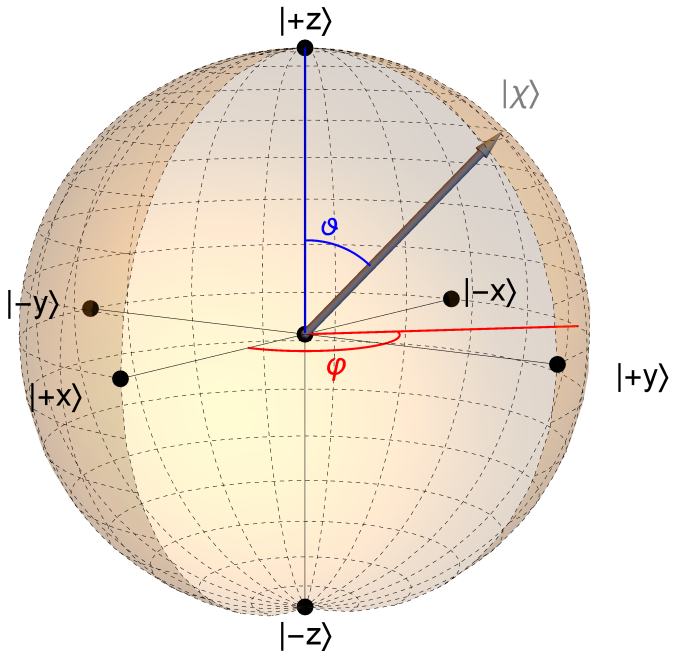


Fig. 3.3: Bloch sphere. The state vector Eq. (3.15) of a two-dimensional system can be geometrically illustrated in a sphere with polar angle ϑ and azimuthal angle φ.

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad \hat{\sigma}_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{3.16}$$

so the exponential map reads $e^{ix\hat{S}_i} = e^{i\frac{x}{2}\hat{\sigma}_i}$. Consequently, orthogonal states belong to antipodal points on the sphere and any state reverts back to its origin after a rotation by $\vartheta = 4\pi$. This causes a small peril that should be kept in mind when plotting vectors in the Bloch sphere. It is safe to use the spherical representation of the Hilbert vectors as long as one is aware that all signs change $|\chi(2\pi, \varphi)\rangle = -|\chi(0, \varphi)\rangle$ after a full rotation of the Bloch ball.

In order to see how two dimensional states can be formulated in terms of a density matrix it is good to know that the Pauli matrices $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$ given in Eq. (3.16) together with the identity matrix $\mathbb{1} = \hat{\sigma}_0$ form, when endowed with the Hilbert-Schmidt inner product Eq. (3.4), an orthogonal basis for the Hilbert space $\mathbb{C}^2$. Therefore any 2x2 matrix $\hat{A}$ over the complex field must be factorisable as a linear combination of the form

$$\hat{A} = \sum_{i=0}^{3} r_i \hat{\sigma}_i = r_0 \mathbb{1} + \sum_{i=1}^{3} r_i \hat{\sigma}_i \ , \qquad r_i = \frac{1}{2} \langle \sigma_i | \hat{A} \rangle = \frac{1}{2} \mathrm{Tr}(\sigma_i \ \hat{A}) \ , \tag{3.17}$$

where the factor $1/2$ comes in, because $\langle \sigma_i | \sigma_j \rangle = 2\delta_{ij}$.

Example 3.2.1. Let $\hat{A}$ be a some random matrix of the form

$$\hat{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3.18}$$

then the coefficients are: $r_0 = (a+d)/2$, $r_1 = (b+c)/2$, $r_2 = (ib - ic)/2$, $r_3 = (a-d)/2$. $\Diamond$

In case the operator in Eq. (3.17) is the pure density matrix given by the projector $\hat{\rho}_\chi = |\chi\rangle\langle\chi|$ of Eq. (3.15) calculation will yield: $r_0 = \mathrm{Tr}(\hat{\rho}_\chi \hat{\sigma}_0)/2 = 1/2$ and for $i > 0$, $r_{i>0} = \mathrm{Tr}(\hat{\rho}_\chi \hat{\sigma}_i)/2 = \langle \chi | \sigma_i | \chi \rangle /2$. Usually, the two-level density matrix is then written out as

$$\hat{\rho} = \frac{\mathbb{1} + \mathbf{r} \cdot \hat{\boldsymbol{\sigma}}}{2} \ , \tag{3.19}$$

where $\mathbf{r}$ in case of pure states is exactly the expectation value of the Pauli matrices $\mathbf{r} = (r_1, r_2, r_3)^T = \langle \chi | \hat{\boldsymbol{\sigma}} | \chi \rangle$. The *Bloch vector* $\mathbf{r}$ is element of the real space $\mathbb{R}^3$ symmetric under rotations of the group $SO(3)$. Checking out the trace of the squared density matrix, the result is

$$\mathrm{Tr}(\hat{\rho}^2) = \mathrm{Tr}(\frac{\mathbb{1} + 2\mathbf{r} \cdot \hat{\boldsymbol{\sigma}} + \mathbb{1}\mathbf{r} \cdot \mathbf{r}}{4}) = \frac{1 + \|\mathbf{r}\|^2}{2} \ , \tag{3.20}$$

which according to Eq. (3.13) implies that the density matrix is pure if the state has unit length $\|\mathbf{r}\| = \|\chi\| = 1$ and consequently $\hat{\rho}_\chi$ is a projector. On the other hand, allowing the vector length to be $\|\chi\| < 1$ implies that all inner points of the Bloch sphere are mixed states. Another way to see this feature is to calculate the eigenvalues of Eq. (3.19) which are $\lambda_\pm = (1 \pm \|\mathbf{r}\|)/2$, therefore for a unit length vector one eigenvalue equals 0 and the state cannot be written as convex combination of extreme points. The totally mixed state is represented by $\hat{\rho} = \mathbb{1}/2$ and is acquired at $\|\mathbf{r}\| = 0$. A Bloch sphere as in Fig. 3.3 is therefore only centered at $|\chi\rangle = 0$ for the state space, while for density matrix the origin is shifted by $\mathbb{1}/2$.

With the two-dimensional quantum system analyzed above, it is a good opportunity to introduce

the term of *mutually unbiased states*. Mathematically, the orthonormal eigenstates $|a_i\rangle, |b_j\rangle$ belonging to pairs of observables over a d-dimensional Hilbert space which satisfy

$$|\langle b_j|a_i\rangle|^2 = \frac{1}{d}\ , \qquad \forall i,j \in \{1,2....d\} \tag{3.21}$$

are called mutually unbiased. The eigenvectors $|\pm x\rangle, |\pm y\rangle, |\pm z\rangle$ of the Pauli matrices are the easiest example of orthonormal bases that have the property of Eq. (3.21). Physically these state imply the maximal trade-off in the distinguishability, i.e. perfect knowledge about one observable implies maximal ignorance for the other. This interpretation makes these states particularly important in quantum information theory.

3.2.2 Composite two-state systems

Quantum theory contains a lot of interesting new phenomena when several particles are involved. Suppose, $\hat{T} \in \mathcal{T}(\mathcal{H})$ is a general trace class operator on a composite Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$, where the subscripts refer to separate subsystems in the sense that experiments in system 1 and system 2 can be performed individually (not necessarily causally independent). The *partial trace* over the system 1 for such an operator is defined by the mapping

$$\mathrm{Tr}_1 : \mathcal{T}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{T}(\mathcal{H}_2) \tag{3.22}$$

and vice-versa for the partial trace Tr_2 over system 2. Assuming that the operator is given by a tensor product $\hat{T} = \hat{T}_1 \otimes \hat{T}_2$ the following relations hold

$$\begin{aligned} \mathrm{Tr}_1(\hat{T}_1 \otimes \hat{T}_2) &= \mathrm{Tr}(\hat{T}_1)\ \hat{T}_2\ , \\ \mathrm{Tr}_2(\hat{T}_1 \otimes \hat{T}_2) &= \hat{T}_1\ \mathrm{Tr}(\hat{T}_2)\ . \end{aligned} \tag{3.23}$$

The partial trace can also be written in terms of Hilbert states; for $\{|\phi_m\rangle\}$ a basis in $\mathcal{H}_1$ and $\{|\psi_n\rangle\}$ a basis in $\mathcal{H}_2$ it follows from the unity of the projector sum Eq. (3.6) of $\hat{T}$

$$\begin{aligned} \hat{T} &= \left(\sum_{k,l} |\phi_k\psi_l\rangle\langle\phi_k\psi_l|\right)\hat{T}\left(\sum_{m,n} |\phi_m\psi_n\rangle\langle\phi_m\psi_n|\right) \\ &= \sum_{k,l,m,n} \langle\phi_k\psi_l|\hat{T}|\phi_m\psi_n\rangle\, |\phi_k\psi_l\rangle\langle\phi_m\psi_n|\ , \end{aligned} \tag{3.24}$$

then the partial trace over system 1 is given as

$$\begin{aligned} \mathrm{Tr}_1(\hat{T}) &= \sum_{k,l,m,n} \langle \phi_k \psi_l | \hat{T} | \phi_m \psi_n \rangle \ \ \delta_{km} \ \ |\psi_l\rangle \langle \psi_n| \\ &= \sum_{l,m,n} \langle \phi_m \psi_l | \hat{T} | \phi_m \psi_n \rangle \ \ |\psi_l\rangle \langle \psi_n| \ . \end{aligned} \tag{3.25}$$

After taking the partial trace on a bipartite system it is possible to perform a trace operation on the second part yielding the total trace. A quick exercise reveals that the order with which the total trace is calculated is not relevant $\mathrm{Tr}(\hat{T}) = \mathrm{Tr}(\mathrm{Tr}_1(\hat{T})) = \mathrm{Tr}(\mathrm{Tr}_2(\hat{T}))$. Another important property is that for $\hat{T} \geq 0$ the partial traces of each system $\mathrm{Tr}_1(\hat{T}) \geq 0$ (resp. $\mathrm{Tr}_2(\hat{T}) \geq 0$) remains positive. This is an example of a complete positive map (detailed description is given in sec. 3.4). The formula about the partial trace are significantly important for the class of density operators. Regard a joint state $\hat{\rho}_{12} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ in the composite system, then the partial trace $\mathrm{Tr}_1(\hat{T})$ gives the *reduced state* of subsystem 2 and vice-versa $\mathrm{Tr}_2(\hat{T})$ gives the reduced density operator of subsystem 1. In case the joint state is not entangled and given by a tensor product of two reduced states Eq. (3.23) yields $\hat{\rho}_1 = \mathrm{Tr}_2(\hat{\rho}_1 \otimes \hat{\rho}_2)$.

Interestingly, reduced states, emerging in the process of partial tracing of a pure joint state $\hat{\rho}_{12}$ are not pure in general. Mixed states are therefore often considered as a product of state-reduction. Following this track of consideration one might ask if it is possible to purify a mixed state: adding an extra system to a mixed state, can the resulting larger system become pure? This task is referred to as *state purification* and can be though of the inversion of state reduction. Here is how the purification works: Let $\hat{\rho}_1$ be a state on $\mathcal{H}_1$ and add an aiding space $\mathcal{H}_{anc}$, which is called *ancilla*, then the joint pure state $\hat{\rho}_{1,anc}$ is a purification of $\hat{\rho}_1$ if $\mathrm{Tr}_{anc}(\hat{\rho}_{1,anc}) = \hat{\rho}_1$. Formally, assume a mixed state $\hat{\rho}_1 = \sum_i^n \lambda_i |\phi_i\rangle \langle \phi_i|$ in space $\mathcal{H}_1$ where $\{|\phi_i\rangle\}$ is orthogonal, for the ancillary system take $\{|\psi\rangle_i\}$ as orthonormal basis of $\mathcal{H}_{anc}$, then

$$|\chi\rangle = \sum_i \sqrt{p_i} \, |\phi_i\rangle \otimes |\psi_i\rangle \ , \tag{3.26}$$

gives $\hat{\rho}_{1,anc} = |\chi\rangle \langle \chi|$, the purification of $\hat{\rho}_1$ with $\sum_i \lambda_i = 1$. From this, one deduces that a mixed subsystem can, when combined with an ancilla, be composed to yield a pure state in the extended space. Any vectorial form $|\chi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ as in Eq. (3.26) with two respective orthonormal bases is also known as the *Schmidt decomposition* of $|\chi\rangle$.

3.3 Effects

In the context of quantum physics, the term of *effect* [54] can be associated with a measurement on a physical object. A prepared ensemble of particles that passes a measurement yields a certain result and the effect assigns a probability to the ensembles associated with that outcome. From the framework of states $\hat{\rho}$, the effect operators $\hat{\Pi}$ are exactly the maps from the set of states $\mathcal{S}(\mathcal{H})$ to the unit interval $[0, 1]$, ergo to the range of probabilities. Consequently, for the expectation value $\mathrm{Tr}(\hat{\rho}\hat{\Pi})$ to be in the unit interval, the effect $\hat{\Pi}$ must be a positive, self-adjoint operator with the property $0 \leq \hat{\Pi} \leq \mathbb{1}$. The set of effects shall be denoted by

$$\mathcal{E}(\mathcal{H}) := \left\{ \hat{\Pi} \in \mathcal{L}_s(\mathcal{H}) \mid 0 \leq \hat{\Pi} \leq \mathbb{1} \right\} . \tag{3.27}$$

The most prominent example of effects that comply with Eq. (3.27) and Eq. (3.6) are normalized projection operators: such effects $\hat{\Pi}_m = \hat{P}_m = \hat{P}_m^\dagger \hat{P}_m$ are called *sharp observables*. It is known from previous considerations that projectors are extremal elements of the convex set, so that they can be regarded a special subclass of all effects.

Let $(\Omega, \mathcal{F})$ be a measurable space (see sec. 2.2) and $\mathcal{P}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ be the set of orthogonal projectors of the complex Hilbert space $\mathcal{H}$ then a *projection-valued measure (PVM)* maps events (outcomes) to operators $\hat{\Pi} : \mathcal{F} \to \mathcal{L}(\mathcal{H})$ with the properties

1. $\hat{\Pi}(\emptyset) = O$.
2. $\hat{\Pi}(\Omega) = I_{\mathcal{H}}$, with $I_{\mathcal{H}}$ the identity on $\mathcal{H}$.
3. $\hat{\Pi}(\cup_i X_i) = \sum \hat{\Pi}(X_i)$ (series convergence in the weak operator topology) for any sequence of disjoint sets $X_i \in \mathcal{F}$.
4. For all $X \in \mathcal{F}$ it holds $\hat{\Pi}(X) \in \mathcal{P}(\mathcal{H})$.

These conditions connect measurable results and their probabilities to operators of quantum physics. It is necessary for the effects to be positive, sum up to unity and be additive since they are related to measures in probability space $(\Omega, \mathcal{F}, p)$.

More generally, a complete set $\{\hat{\Pi}_m \in \mathcal{E}(\mathcal{H})\}$ that sums to unity, but is not a collection of projectors and does not meet requirement 4, is called a *positive operator-valued measure (POVM)*.

These effects are in contrast to their sharp PVM counterparts referred to as *unsharp observables*. As shown in the definition in sec. 3.1 positive operators can be decomposed to

$$\hat{\Pi}_m := \hat{M}_m \hat{M}_m^\dagger \ , \tag{3.28}$$

where $\hat{M}_m$ in this context will be called *measurement operators*. The measurement operators are only unique up to a unitary transformation $\hat{M}_m = \sqrt{\hat{\Pi}_m}\hat{U}$ (this is nothing but the polar decomposition Eq. (3.7)), as inserting into Eq. (3.28) makes it immediately apparent. As a consequence POVMs alone do not yield the form of the measurement operators.

The set of sharp and unsharp observables $\hat{\Pi} : \mathcal{F} \to \mathcal{L}(\mathcal{H})$ constitutes the set of all quantum observables $\mathcal{O}(\Omega, \mathcal{F}, \mathcal{H})$, where $\hat{\Pi}$ is a POVM if it satisfies requirements 1-3 of the list, or equivalently if $\mathcal{F} \ni X \to \mathrm{Tr}(\hat{\rho}\hat{\Pi}(X))$ is a probability for all $\hat{\rho} \in \mathcal{S}(\mathcal{H})$. The set of observables form a convex set which allow to mix up two observables $\hat{\Pi}_1, \hat{\Pi}_2 \in \mathcal{O}(\Omega, \mathcal{F}, \mathcal{H})$ in the form $\lambda\hat{\Pi}_1 + (1-\lambda)\hat{\Pi}_2$. Any effect can be be decomposed by the spectral theorem into n mutually orthogonal projectors

$$\hat{\Pi} = \sum_{i=1}^{n \leq \dim(\mathcal{H})} \alpha_i \hat{P}_i \ . \tag{3.29}$$

The upper bound of summation n determines the rank, if $n = 1$ then the effect will be rank-1. To understand POVMs better, consider the following example.

Example 3.3.1. One of the simplest illustration of a valid two-dimensional POVM can be constructed as follows. Using the state notation of Eq. (3.15) the set of three states given by

$$|t_1\rangle = \sqrt{\frac{2}{3}}\,|\chi\left(\frac{\pi}{2}, 0\right)\rangle \ , \quad |t_2\rangle = \sqrt{\frac{2}{3}}\,|\chi\left(\frac{\pi}{2}, \frac{2\pi}{3}\right)\rangle \ , \quad |t_3\rangle = \sqrt{\frac{2}{3}}\,|\chi\left(\frac{\pi}{2}, \frac{4\pi}{3}\right)\rangle \tag{3.30}$$

allow to construct three effects $\hat{\Pi}_i$ that constitute a POVM

$$\hat{\Pi}_1 = |t_1\rangle\langle t_1| = \frac{2}{3}\hat{P}(\mathbf{r}_1) \ , \quad \hat{\Pi}_2 = |t_2\rangle\langle t_2| = \frac{2}{3}\hat{P}(\mathbf{r}_2) \ , \quad \hat{\Pi}_3 = |t_3\rangle\langle t_3| = \frac{2}{3}\hat{P}(\mathbf{r}_3) \ . \tag{3.31}$$

The states in Eq. (3.30) are called *trine states*, the collection of effects in Eq. (3.31) in turn are known as *trine POVM*. Each matrix is positive and the sum of the three matrices yields the identity matrix. Because of the $\frac{2}{3}$-factor the effects are not projectors $\hat{\Pi}_i^2 \neq \hat{\Pi}_i$, but they are proportional to the projector $\hat{P}$ pointing in direction $\mathbf{r}_i$. Because of their proportionality the

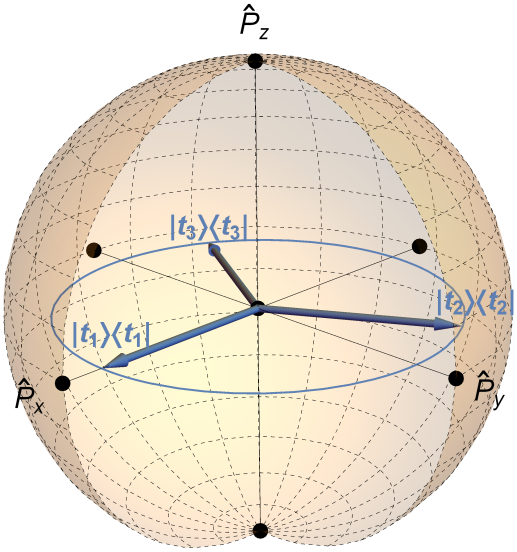


Fig. 3.4: The 3-outcome trine POVM illustrated as non-normalized projectors in an operator sphere confined to the xy-plane. The effects have Hilbert-Schmidt norm $\|\hat{\Pi}_i\|_{HS} = \frac{2}{3}$ and are separated by a constant angle of 120°.

effects retain their rank, i.e. they are rank-1 operators in the same way as projectors. Another consequence from this plain example is that the measurement operators are easily related to the effects by $\hat{M}_i = \hat{U}\sqrt{\frac{2}{3}}\hat{P}(\mathbf{r}_i)$, which are unique up to an arbitrary unitary operator. An illustration of the trine state is given in Fig. 3.4. ◊

In the two-dimensional case, if a POVM is proportional to PVMs of the form in Eq. (3.19) then all effects sum to unity if the total displacement of the Bloch vectors $\mathbf{r}_i$ is zero, that is

$$\sum_{i=1}^{n} \mathbf{r}_i = \mathbf{0}, \quad \|\mathbf{r}_i\| = 1 \qquad \Rightarrow \qquad \sum_{i=1}^{n} \hat{\Pi}_i = \frac{2}{n} \sum_{i=1}^{n} \hat{P}(\mathbf{r}_i) = \mathbb{1} \ . \tag{3.32}$$

This formula Eq. (3.32) comprises the projective case given for $n = 2$ and the trine state example for $n = 3$. The PVM with only two elements must necessarily be collinear showing to antipodal points on the sphere, while the three-elemental POVM forms an equilateral triangle. For $n = 4$ the rigidness of the shape decreases and it is possible to have 4 vectors pointing to a vertices of a parallelogram, but also to the vertices of a regular tetrahedron.

Up to now all effects considered were proportional to projectors $\hat{\Pi} = \lambda \hat{P}(\mathbf{r})$ with $\lambda \in [0, 1]$. Relaxing the unit length restriction of the qubit POVM ads an additional degree of freedom. Any arbitrary, finite, two-dimensional quantum effect can be written in the form

$$\hat{\Pi}_i(\varrho, \mathbf{r}) = \frac{\varrho_i \mathbb{1} + \mathbf{r}_i \cdot \hat{\boldsymbol{\sigma}}}{2} , \qquad \|\mathbf{r}_i\| \leq \varrho_i \leq 2 - \|\mathbf{r}_i\| . \tag{3.33}$$

The inequality in Eq. (3.33) is a necessary condition that ensures that all $\hat{\Pi}_i$, which have eigenvalues $\frac{1}{2}(\varrho \pm \|\mathbf{r}_i\|)$, satisfy the range condition $0 \leq \hat{\Pi}_i \leq \mathbb{1}$ according to Eq. (3.27) . The effect is rank 1 if $\varrho_i = \|\mathbf{r}_i\|$ and equal to a projector exactly for $\|\mathbf{r}_i\| = 1$.

3.3.1 Measurement outcomes and outputs

Measurements in quantum physics are often described by postulates, which are rules telling how to acquire statistics in an ensemble of identical measurements. Here, the measurement process is explained without looking at the physical and operational details of the measurement apparatus.

The two essential ingredients that characterize the process of measurement are the state describable by a Hilbert state $|\psi\rangle$ or density matrix $\hat{\rho}$, and an measurement operator that makes the observation given by the effect $\hat{\Pi}$. The treatment of the measuring process was for a long time limited to projective *von-Neumann measurements*. Such a measurement is described by an Hermitian observable which has the spectral decomposition $\hat{A} = \sum_m m\hat{P}_m$, where the eigenvalues m correspond to the measurement outcomes. *Born's rule* then supplies the probability for that particular outcome to occur for the initial state $\hat{\rho}_i$

$$p(m) = \mathrm{Tr}(\hat{\rho}_i \hat{P}_m) = \langle \hat{\rho}_i | \hat{P}_m \rangle \tag{3.34}$$

and furthermore the expectation value is given by $\langle \hat{A} \rangle = \sum_m m\mathrm{Tr}(\hat{\rho}_i \hat{P}_m)$. The final state after a particular outcome m is described by the density operator

$$\hat{\rho}_f(m) = \frac{\hat{P}_m \hat{\rho}_i \hat{P}_m^\dagger}{\mathrm{Tr}(\hat{P}_m \hat{\rho}_i \hat{P}_m^\dagger)} , \tag{3.35}$$

while the full ensemble after the measurement will be of the form $\hat{\rho}_f = \sum_m \langle \hat{\rho}_i | \hat{P}_m \rangle \, \hat{\rho}_f(m) = \sum_m \hat{P}_m \hat{\rho}_i \hat{P}_m^\dagger$. The key property of von Neumann measurements is *repeatability*, i.e. after a

projective measurement produces outcome m, the result is reproduced when the measurement is repeated. Repeatability in quantum theory is not strictly necessary and for many important measurements can be regarded a dispensable restriction. The employment of POVMs allows to access a wider spectrum of quantum measurements. The previous result Eq. (3.34) changes to

$$p(m) = \mathrm{Tr}(\hat{\rho}_i \hat{\Pi}_m) = \mathrm{Tr}(\hat{M}_m \hat{\rho}_i \hat{M}_m^\dagger) \ , \tag{3.36}$$

where the resolution of identity must hold for $\hat{\Pi} \in \mathcal{O}(\Omega, \mathcal{F}, \mathcal{H})$, hence $\hat{\Pi}(\Omega) = \sum_{\omega_m \in \Omega} \hat{\Pi}(\{\omega_m\}) = \sum_m \hat{\Pi}_m = \mathbb{1}$. The outcome m now corresponds no longer to the eigenvalue, but some output variable, and the state after measurement is

$$\hat{\rho}_f(m) = \frac{\hat{M}_m \hat{\rho}_i \hat{M}_m^\dagger}{\mathrm{Tr}(\hat{M}_m \hat{\rho}_i \hat{M}_m^\dagger)} \neq \frac{\hat{\Pi}_m \hat{\rho}_i \hat{\Pi}_m^\dagger}{\mathrm{Tr}(\hat{\Pi}_m \hat{\rho}_i)} \ . \tag{3.37}$$

The right hand side has been deliberately added to emphasize that the post-measurement state is not given by the operative action of the effects $\hat{\Pi}_m$, but by the action of the operators $\hat{M}_m$. A more detailed description about quantum measurements follows in sec. 3.6.

3.4 Quantum operations

A substantial component which has not been discussed until now is the concept of a quantum channel, the 'pipeline' that connects states and observations. The function of a channel is to take input states and guide them, permitting all operations that map quantum ensembles to other valid quantum ensembles.

A quantum operation is a linear map $\mathcal{E}$ transforming a set of states $\mathcal{S}(\mathcal{H})$ to another set of density operators $\tilde{\mathcal{S}}(\tilde{\mathcal{H}})$. Since density operators can always be written as convex combination of pure states, the map $\mathcal{E}$ has to be convexity preserving [49], i.e.

$$\mathcal{E}\left(\sum_i \lambda_i \hat{\rho}_i\right) = \sum_i \lambda_i \mathcal{E}\left(\hat{\rho}_i\right) \ . \tag{3.38}$$

The map in Eq. (3.38) should meet further mathematical requirements. First, the map has to preserve positivity $\mathcal{E}(\hat{\rho}) \geq 0$ for all inputs $\hat{\rho} \in \mathcal{S}(\mathcal{H})$. A second important requirement is that the unit trace of input states is preserved or at least not increased: $\mathrm{Tr}(\mathcal{E}(\hat{\rho})) \leq \mathrm{Tr}(\hat{\rho}) = 1$. Depending on the strictness of the inequality $\mathrm{Tr}(\mathcal{E}(\hat{\rho})) \leq 1$ there are two distinguished terms

1. *Quantum operations*, correspond to trace non increasing maps.

2. *Quantum channels*, correspond specifically to trace preserving maps, $\mathrm{Tr}(\mathcal{E}(\hat{\rho})) = 1$.

The physical difference between the two is that quantum channels are *non-selective*, i.e. they will either map unitarily or pure states to mixed states deterministically, while trace decreasing operators describe nondeterministic, *selective* mappings of states as is the case in information providing quantum measurements. Such quantum operations transform pure states to either pure or mixed states [55].

Another property of quantum operations and channels can only be comprehended when looking at composite quantum systems. Let $\mathcal{H}_1 \otimes \mathcal{H}_2$ be the Hilbert space of a composite system and consider a map as in Eq. (3.38) where $\mathcal{E}_1$ acts on $\mathcal{H}_1$ extended by the identity map $\mathbb{1}_2$ acting on system $\mathcal{H}_2$, which indicates that the second subsystem is left untouched. Then the action of the map on the composite density matrix is given by

$$(\mathcal{E}_1 \otimes \mathbb{1}_2)(\hat{\rho}_1 \otimes \hat{\rho}_2) = \mathcal{E}_1(\hat{\rho}_1) \otimes \hat{\rho}_2 \ . \tag{3.39}$$

In general the extended map $(\mathcal{E}_1 \otimes \mathbb{1}_2)$ should also transform quantum ensembles appropriately. Positivity of the submap $\mathcal{E}_1$ is not enough for the composite map $(\mathcal{E}_1 \otimes \mathbb{1}_2)$ to be positive. Generally speaking, if a map $\mathcal{E}_1$ between the sets of trace class operators is such that its extended map $\mathcal{E}_1 \otimes \mathbb{1}_2$ is positive on the composite Hilbert space for all finite dimensional extensions of $\mathcal{H}_2$ then $\mathcal{E}_1$ is called *completely positive (CP)* . Hence, a quantum channel is rigorously defined as a linear map $\mathcal{E}$ that has to be *completely positive, trace-preserving (CPTP).*

Evolutions of quantum systems are usually considered in the more prevalent Schrödinger picture, which puts the entire dynamics into the quantum states leaving the quantum operators unchanged, in opposition to the *Heisenberg picture* which transforms the observables keeping the states in their static form. This switch between the pictures can be accomplished for quantum channels $\mathcal{E}$ on $\mathcal{L}(\mathcal{H})$ by the dual map $\mathcal{E}^*$. Suppose, $\hat{\rho} \in \mathcal{S}(\mathcal{H})$ and $\hat{A} \in \mathcal{L}(\mathcal{H})$, then irrespective of which pictures is used the expectation values have to be the same

$$\mathrm{Tr}\left(\mathcal{E}(\hat{\rho})\hat{A}\right) = \mathrm{Tr}\left(\hat{\rho}\ \mathcal{E}^*(\hat{A})\right) \ , \tag{3.40}$$

where the mapping $\mathcal{E}^*$ is now acting on the observable. From the following equation

$$\mathrm{Tr}\left(\mathcal{E}(\hat{\rho})\right) = \mathrm{Tr}\left(\mathcal{E}(\hat{\rho})\mathbb{1}\right) = \mathrm{Tr}\left(\hat{\rho}\ \mathcal{E}^*(\mathbb{1})\right) \tag{3.41}$$

the trace non-increasing property $\mathrm{Tr}(\hat{\rho}) \geq \mathrm{Tr}(\mathcal{E}(\hat{\rho}))$ of quantum operations translates to $\mathcal{E}^*(\mathbb{1}) \leq \mathbb{1}$ for the dual quantum operation $\mathcal{E}^*$. When equality holds $\mathcal{E}^*(\mathbb{1}) = \mathbb{1}$ the operation is called *unital*. The most common examples of quantum channels are

1. Unitary maps $\mathcal{E}_U(\hat{\rho}) = \hat{U}\hat{\rho}\hat{U}^\dagger$.

2. Partial tracing $\mathcal{E}_p(\hat{\rho}) = \mathrm{Tr}_2(\hat{\rho})$, $\hat{\rho} \in \mathcal{H}_1 \otimes \mathcal{H}_2$ as has already been introduced (sec. 3.2.2).

3. Let $\hat{\rho}_{anc}$ be a fixed state of $\mathcal{H}_{anc}$, the ancillary Hilbert space, then the extension of the system by an uncorrelated ancilla is also a valid and substantial quantum channel $\mathcal{E}_{anc}(\hat{\rho}) = \mathrm{Tr}(\rho \otimes \hat{\rho}_{anc})$.

The distinguishing feature of the examples mentioned above is that unitary transformations only describe the deterministic evolution of a closed subsystem. The other kind of channels can serve as physical models for taking environmental influences into account. If the state $\hat{\rho} \in \mathcal{S}(\mathcal{H}_S)$ is considered and $\hat{\rho}_E$ is a defined state of the environment Hilbert space $\mathcal{H}_E$, then their common global evolution can be described as an unitary operator $\hat{U}$ on the system plus environment (see Fig. 3.5 for a pictograph of an open quantum system). This leads to a further important quantum channel, emerging as a combination of tensoring with an ancilla $\mathcal{E}_{anc}$, common unitary evolution $\mathcal{E}_U$ and eventually using a partial trace $\mathcal{E}_p$ to extract information about the changes of the system state.

The following corollary can be provided: Let $\mathcal{E} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ be a CPTP map between states in a Hilbert state $\mathcal{H}$, then there exists a Hilbert space $\mathcal{H}_E$, a pure state $\hat{\rho}_E = |e\rangle\langle e| \in \mathcal{H}_E$ and a unitary operator $\hat{U} \in \mathcal{H} \otimes \mathcal{H}_E$ such that any channel can be represented by

$$\mathcal{E}_E(\hat{\rho}) = \mathrm{Tr}_E\left(\hat{U}(\hat{\rho} \otimes \hat{\rho}_E)\hat{U}^\dagger\right) \tag{3.42}$$

for all $\hat{\rho} \in \mathcal{H}$. The channel in Eq. (3.42) is a chained composition of three channels: $\mathcal{E}_{anc} \circ \mathcal{E}_U \circ \mathcal{E}_p$. The formula of Eq. (3.42) can be regarded a special version of a more general from of the *Stinespring theorem* and guarantees that each channel can be expressed in this form; the triplet

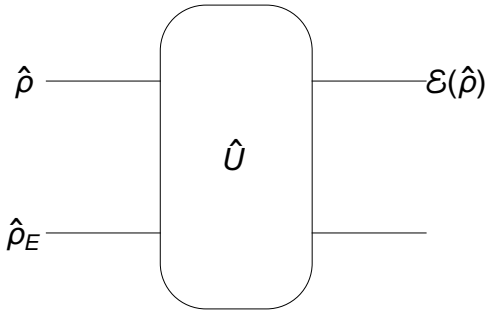


Fig. 3.5: Model of the interaction of a unitary operator acting on the dilated space resulting in a channel $\mathcal{E}(\hat{\rho})$ that describes the input state's change by coupling with an ancilla and interaction through a global unitary. The resulting state is obtained after discarding the environment $\hat{\rho}_E$ by partial tracing.

$\langle \mathcal{H}_E, \hat{U}, \hat{\rho}_E \rangle$ is usually called *Stinespring dilation*[3].

It is possible to make a change in the appearance of Eq. (3.42) to arrive at the *operator sum representation*, also *Kraus representation*, of quantum operations and channels. For the purpose of reformulation the orthonormal eigenstates of the environment is chosen $\{|e_i\rangle\}$ and the initial state fixed as $\hat{\rho}_E = |e_0\rangle \langle e_0|$. Thus

$$\begin{aligned} \mathcal{E}_E(\hat{\rho}) &= \mathrm{Tr}_E \left(\hat{U}(\hat{\rho} \otimes |e_0\rangle\langle e_0|)\hat{U}^\dagger \right) \ , \\ &= \sum_k \langle e_k|\hat{U}|e_0\rangle \, \hat{\rho} \, \langle e_0|\hat{U}^\dagger|e_k\rangle \ , \\ &= \sum_k \hat{E}_k \hat{\rho} \hat{E}_k^\dagger \ , \end{aligned} \tag{3.43}$$

where $\hat{E}_k = \langle e_k|\hat{U}|e_0\rangle$ is called *Kraus operator*. From the trace-preservation of quantum channels, one obtains

$$\mathrm{Tr}(\mathcal{E}_E\,(\hat{\rho})) = \mathrm{Tr}\left(\sum_k \hat{E}_k \hat{\rho} \hat{E}_k^\dagger \right) = \mathrm{Tr}\left(\sum_k \hat{E}_k^\dagger \hat{E}_k \hat{\rho} \right) = 1 \ . \tag{3.44}$$

It can be further deduced that for a CPTP map the resolution of the identity holds

$$\sum_k \hat{E}_k^\dagger \hat{E}_k = \mathbb{1} \ . \tag{3.45}$$

[3]The Stinespring dilation is more general statement on the theory of CP maps between C^*-algebras [48].

The operator sum representation puts the interaction with the environment into the operators, such that the details about the ambient are concealed and the dynamic of the system's state emerges without explicitly considering the properties of the environment. The description of quantum channels with the operator sum representation satisfying Eq. (3.45) is equivalent to the dilation Eq. (3.42). For trace non increasing quantum operations the equality in Eq. (3.45) must be replaced with $\sum_k \hat{E}_k^\dagger \hat{E}_k \leq \mathbb{1}$.

The Kraus operators already appeared in the form of measurement operators in sec. 3.3.1. This should not be a surprise since transformations from pre- to post-measurement states are described by quantum operations. Writing $\mathcal{I}(k)(\hat{\rho}) = \hat{E}_k \hat{\rho} \hat{E}_k^\dagger$ and normalizing by $\mathrm{Tr}(\mathcal{I}(k)(\hat{\rho}))$, the system state acquires the form

$$\hat{\rho}_k = \frac{\mathcal{I}(k)(\hat{\rho})}{\mathrm{Tr}(\mathcal{I}(k)(\hat{\rho}))} = \frac{\hat{E}_k \hat{\rho} \hat{E}_k^\dagger}{\mathrm{Tr}(\hat{E}_k \hat{\rho} \hat{E}_k^\dagger)} \,. \tag{3.46}$$

This is nothing but the density operator emerging after a measurement as given in Eq. (3.37). In that form $\hat{\Pi}_k = \hat{E}_k^\dagger \hat{E}_k$ corresponds exactly to elements of a POVM and $p(k) = \mathrm{Tr}(\hat{\Pi}_k \hat{\rho})$ is nothing but the probability for the outcome k. This equivalence permits an extremely useful interpretation of the operator-sum representation and vice-versa: the measurement operators previously encountered can now be interpreted as the Kraus operators $\hat{E}_k \to \hat{M}_k = \langle e_k | \hat{U} | e_0 \rangle$ and Eq. (3.43)

$$\mathcal{E}_{ns}\left(\hat{\rho}\right) = \sum_k \hat{M}_k \hat{\rho} \hat{M}_k^\dagger = \sum_k p(k) \hat{\rho}_k \,, \tag{3.47}$$

can be interpreted as the ensemble after measurement. Notice, that Eq. (3.46) describes an accessible ancilla system that acts as a *probe* at the measurement, while Eq. (3.47) takes the average over all outcomes making the individual results inaccessible. Therefore $\mathcal{E}_{ns}\left(\hat{\rho}\right)$ is a non-selective CPTP map, while $\mathcal{I}\left(\hat{\rho}\right)$ is a selective, trace decreasing quantum operation that discards some terms in the operator sum representation.

3.5 Distance of states

In sec. 2.2.4 some metrics in classical probability theory were introduced. For quantum theory, it is also important to have a kind of distance measure which indicates the similarity or closeness of states. Especially interesting are the ones that have a classical version and a quantum

analogue, for example the *trace distance* and *quantum state fidelity*.

The trace distance for two states $\hat{\rho}$ and $\hat{\sigma}$ is a true metric satisfying the axioms given in sec. 2.2.4 and is defined as half the trace norm $D_{tr}(\hat{\rho}, \hat{\sigma}) = \frac{1}{2}\|\hat{\rho} - \hat{\sigma}\|_{HS}$. When both states commute $[\hat{\rho}, \hat{\sigma}] = 0$ they have a common eigenbasis and the quantum physical trace distance simply becomes the classical total variation distance.

The fidelity of states $\hat{\rho}$ and $\hat{\sigma}$ is another comparability measure defined as

$$F(\hat{\rho}, \hat{\sigma}) = \mathrm{Tr}\left(\sqrt{\sqrt{\hat{\rho}}\,\hat{\sigma}\sqrt{\hat{\rho}}}\right) . \tag{3.48}$$

The fidelity is not a metric in the axiomatic sense, although it has the qualities of a good state discriminator. In case the two states commute, the fidelity $F(\hat{\rho}, \hat{\sigma}) = \mathrm{BC}(r_i, s_i)$ is given by the Bhattacharyya coefficient of the eigenvalue distribution (see sec. 2.2.4). Assume that $\hat{\rho} = |r\rangle\langle r|$ is a pure state, then $F(|r\rangle\langle r|, \hat{\sigma}) = \mathrm{Tr}\left(\sqrt{|r\rangle\langle r|\hat{\sigma}|r\rangle\langle r|}\right) = \sqrt{\langle r|\hat{\sigma}|r\rangle}$, the fidelity is simply the square root of the expectation value. If both states are pure then one receives the overlap between two states

$$F(|r\rangle\langle r|, |s\rangle\langle s|) = \sqrt{\langle r|s\rangle\langle s|r\rangle} = |\langle s|r\rangle| . \tag{3.49}$$

The fidelity ranges from 0 to 1 and is symmetric, invariant under unitary transformation, monotone under CPTP maps $F(\mathcal{E}(\hat{\rho}), \mathcal{E}(\hat{\sigma})) \geq F(\hat{\rho}, \hat{\sigma})$ and concave in its input $F(\sum_i \lambda_i \hat{\rho}_i, \hat{\sigma}) \geq \sum_i \lambda_i F(\hat{\rho}, \hat{\sigma})$.

For qubits, the fidelity can be associated with an angle between states $F(\hat{\rho}, \hat{\sigma}) = \cos(\theta_{\rho,\sigma})$. The state fidelity for two-dimensional states $\hat{\rho}, \hat{\sigma}$ can be simplified to

$$\begin{aligned} F(\hat{\rho}, \hat{\sigma}) &= \sqrt{\mathrm{Tr}(\hat{\rho}\hat{\sigma}) + 2\sqrt{\det(\hat{\rho})\det(\hat{\sigma})}} \\ &= \sqrt{\frac{1}{2}(\mathbf{r}\cdot\mathbf{s}) + \sqrt{(1 - \|\mathbf{r}\|^2)(1 - \|\mathbf{s}\|^2)}} . \end{aligned} \tag{3.50}$$

3.6 Quantum instruments

A crucial object for the mathematical modeling of quantum measurements is the so-called quantum instrument, constructed such that it has a quantum state $\hat{\rho}$ as input and an output consisting of both a quantum state and a classical value. The two outputs correspond to the

measurement outcome and the conditional posteriori state. The quantum instrument related to a proper measurement model has to consider how a microscopic quantum system interacts with another quantum object or a macroscopic device from which out of a multitude of possible values a particular outcome is obtained by a form of registration. After the interaction of the quantum state with the measurement apparatus, the measurement model must also describe the posteriori state. It should be noted that the apparatus must not literally be a large physical device, but is often a quantum system that functions as a probe.

Consider an observable $\hat{\Pi}(X)$ of the object system on Hilbert space $\mathcal{H}$ and classical outcomes represented by the measurable space $(\Omega, \mathcal{F})$. The purpose of the measurement is to determine the 'value' of observable $\hat{\Pi}(X) \in \mathcal{O}(\Omega, \mathcal{F}, \mathcal{H})$ (see sec. 3.3). For the description of the measurement apparatus the following model is used: there is an associated Hilbert space $\mathcal{K}$, usually called the *probe system*, a *meter observable* $\hat{M}(X)$ (respectively *pointer observable*) on the probe system, which is correlated to the observable $\hat{\Pi}(X)$ intended to be measured, and an apparatus outcome space $(\Omega_{\mathcal{K}}, \mathcal{F}_{\mathcal{K}})$. While the measurable spaces may differ, it is legitimate to assume that both have the same outcome spectrum, and therefore corresponding subscripts will simply be omitted from now on. Both systems have initial states $\hat{\rho} \in \mathcal{S}(\mathcal{H})$ and $\hat{\xi} \in \mathcal{S}(\mathcal{K})$, which are physically and stochastically uncorrelated to each other prior to measurement. Eventually, a measurement interaction has to couple the object and probe system, described by a unitary quantum channel $\mathcal{E}_U$, that maps the initial compound state $\hat{\rho} \otimes \hat{\xi}$ to a final transformed state $\hat{U}(\hat{\rho} \otimes \hat{\xi})\hat{U}^\dagger$. A scheme of the described situation is illustrated in Fig. 3.6.

The probe system's elements in this scheme are conveniently collected into a quadruple $\mathcal{M} = \langle \mathcal{K}, \hat{\xi}, \hat{U}, \hat{M} \rangle$ which is called an *indirect measurement model* [21, 48]. In sec. 3.3.1 the direct measurement model was already introduced, where the probability is given by the squared amplitude, respectively $\mathrm{Tr}(\hat{\Pi}(X)\hat{\rho})$. Because the indirect measurement must yield the same results as the direct measurement formula, the measurement model must satisfy the *probability reproducibility condition*

$$\mathrm{Tr}(\hat{\Pi}(X)\hat{\rho}) = \mathrm{Tr}\left(\mathcal{E}_U(\hat{\rho} \otimes \hat{\xi})(\mathbb{1} \otimes \hat{M}(X))\right) \ , \tag{3.51}$$

where $X \in \mathcal{F}$ is a subset of all possible outcome values, indicated by the gauge in Fig. 3.6. The

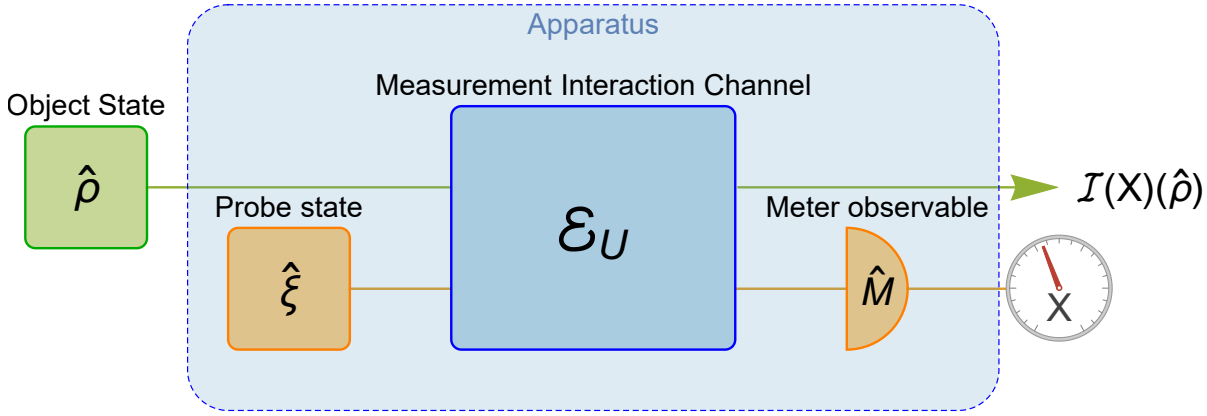


Fig. 3.6: Illustration of an indirect measurement model. The output state, conditioned on the outcome X, is given by a trace non-increasing quantum operation $\mathcal{I}(X)(\rho)$ transforming the initial object state. The result of the measurement is attained by interaction of the apparatus with the object state (a.k.a system state) and pointer read-out of the meter observable $\hat{M}(X)$.

interpretation of Eq. (3.51) is simple; the outcome probability when measuring $\hat{\Pi}(X)$ directly must coincide with the result of the probe system's observable $\hat{M}(X)$.

Now, consider a joint measurement of two observables $\hat{\Pi}^A(X), \hat{\Pi}^B(Y)$, where the first observable $\hat{\Pi}^A(X)$ is measured by means of the measurement model $\langle \mathcal{K}, \hat{\xi}, \hat{U}, \hat{M} \rangle$, while a direct measurement is assumed for observable $\hat{\Pi}^B(Y)$. In order to attain the joint probability $p(X, Y)$ with outcome sets X and Y, a mathematical formalism is needed that describes the state transformation after the object state becomes disentangled from the measurement apparatus. For that purpose the partial trace over the probe state in Eq. (3.51) is applied, which defines the quantum operation

$$\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho}) := \mathrm{Tr}_{\mathcal{K}} \left(\mathcal{E}_U(\hat{\rho} \otimes \hat{\xi})(\mathbb{1} \otimes \hat{M}(X)) \right) . \tag{3.52}$$

This operation valued measure $\mathcal{I}_{\mathcal{M}} : \mathcal{F} \to \mathcal{L}(\mathcal{K})$ from outcome space $(\Omega, \mathcal{F})$ to the non-normalized final states, conditioned on the measurement which has led to a result in X, is called an *instrument*. In Fig. 3.6 the output state $\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})$ is indicated by the green arrow coming out of the apparatus. Consequently, the probability of a joint measurement of

observables $\hat{\Pi}^A, \hat{\Pi}^B$ can now be written in the form

$$\begin{aligned} p(X,Y) &= \mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\hat{\Pi}^B(Y)\right) \\ &= \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{K}}\left(\mathcal{E}_U(\hat{\rho}\otimes\hat{\xi})(\mathbb{1}\otimes\hat{M}(X))\right)\hat{\Pi}^B(Y)\right) \\ &= \mathrm{Tr}\left(\mathcal{E}_U(\hat{\rho}\otimes\hat{\xi})\left(\hat{\Pi}^B(Y)\otimes\hat{M}(X)\right)\right) \ . \end{aligned} \tag{3.53}$$

The properties of instruments are deduced from three defining requirements [48],

1. $\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})$ is a quantum operation.
2. The entire event space Ω or the empty set yield traces $\mathrm{Tr}(\mathcal{I}_{\mathcal{M}}(\Omega)(\hat{\rho})) = 1$, respectively $\mathrm{Tr}(\mathcal{I}_{\mathcal{M}}(\emptyset)(\hat{\rho})) = 0$.
3. Moreover, for a countable mutually disjoint sequence $\{X_i\}_{i\in I} \in \mathcal{F}$ the union of sets satisfies $\mathrm{Tr}(\mathcal{I}_{\mathcal{M}}(\bigcup_i X_i)(\hat{\rho})) = \sum_i \mathrm{Tr}(\mathcal{I}_{\mathcal{M}}(X_i)(\hat{\rho}))$.

Furthermore, consider Ω_B to be the entire outcome space of observable $\hat{\Pi}^B$, then the joint probability Eq. (3.53) reduces to

$$p(X,\Omega_B) = p(X) = \mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\right) \ . \tag{3.54}$$

The normalized posteriori state associated with the pointer observation from the values set X is (assuming $\mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\right) \neq 0$)

$$\hat{\rho}'_X = \frac{\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})}{\mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\right)} \ . \tag{3.55}$$

This conditional output state emerges naturally when calculating the conditional probability using Bayes theorem, i.e. the quotient of the joint probability Eq. (3.53) to the marginal Eq. (3.54) is given by

$$p(Y|X) = \frac{\mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\hat{\Pi}^B(Y)\right)}{\mathrm{Tr}\left(\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho})\right)} = \mathrm{Tr}\left(\hat{\Pi}^B(Y)\hat{\rho}'_X\right) \ . \tag{3.56}$$

This operational approach to quantum mechanical measurement by instruments was among others investigated in the works of Davies and Lewis [51] and Ozawa [56]. An important theorem correlated to the topic: for any apparatus described by a measurement model $\mathcal{M} = \langle \mathcal{K}, \hat{\xi}, \hat{U}, \hat{M}\rangle$ exists a unique CP instrument $\mathcal{I}_{\mathcal{M}}$, where the output probability distribution is given by

Eq. (3.54) and the normalized quantum state reduction by Eq. (3.55). Additionally, without loss of generality the measurement model can be chosen such that $\hat{\xi}$ is initially a pure state, the channel $\mathcal{E}_U$ is a unitary quantum channel $\hat{U}$ and $\hat{M}(X) \to \hat{P}^M(X)$ a sharp observable $(\hat{P}^M(X))^2 = \hat{P}^M(X)$.

The relationship between measurement models and instruments is not one-to-one [48, 57]. That means that while $\mathcal{M} \to \mathcal{I}_\mathcal{M}$ is unique, the inversion of the mapping is not, i.e. different apparatuses can have a similar measurement effect or various possibilities for carrying out the pointer reading can be provided for an apparatus. Consider that the Kraus decomposition of the instrument $\mathcal{I}(x)(\hat{\rho}) = \sum_k \hat{E}_k(\hat{\rho})\hat{E}_k^\dagger$ is not a unique map. Therefore, to each instrument there is an equivalence class of measurement models $\mathcal{I}_\mathcal{M} \to [\mathcal{M}]$ that have the same state transformations for the object. Alternatively, two schemes $\mathcal{M}_1$ and $\mathcal{M}_2$ are equivalent if the transformed states are equal $\mathcal{I}_{\mathcal{M}_1} = \mathcal{I}_{\mathcal{M}_2}$. The annotation of the instrument with $\mathcal{M}$ shall be omitted if the reference is irrelevant.

To connect the instrument and the measured observable the duality relation as defined in Eq. (3.40) must be considered for some arbitrary $\hat{A} \in \mathcal{L}(\mathcal{H})$, that is

$$\begin{aligned} \mathrm{Tr}\left(\hat{A}\ \mathcal{I}(X)(\hat{\rho})\right) &= \mathrm{Tr}\left(\hat{A}\ \mathrm{Tr}_\mathcal{K}\left(\hat{U}(\hat{\rho}\otimes\hat{\xi})\hat{U}^\dagger(\mathbb{1}\otimes\hat{P}^M(X))\right)\right) \\ &= \mathrm{Tr}\left(\hat{\rho}\ \mathrm{Tr}_\mathcal{K}\left(\hat{U}^\dagger(\hat{A}\otimes\hat{P}^M(X))\hat{U}(\mathbb{1}\otimes\hat{\xi})\right)\right) \\ &= \mathrm{Tr}\left(\hat{\rho}\ \mathcal{I}^*(X)(\hat{A})\right)\ . \end{aligned} \tag{3.57}$$

Since the last line has to hold for all states the duality relation yields

$$\mathcal{I}^*(X)(\hat{A}) = \mathrm{Tr}_\mathcal{K}\left(\hat{U}^\dagger\left(\hat{A}\otimes\hat{P}^M(X)\right)\hat{U}(\mathbb{1}\otimes\hat{\xi})\right)\ . \tag{3.58}$$

By comparison to $\mathrm{Tr}\left(\hat{\rho}\ \mathcal{I}^*(X)(\hat{A})\right)$ the POVM of the apparatus $\hat{\Pi}_\mathcal{I}(X)$ emerges from Eq. (3.58) as

$$\hat{\Pi}_\mathcal{I}(X) = \mathcal{I}^*_\mathcal{M}(X)(\mathbb{1})\ , \tag{3.59}$$

which is called the *associated observable* of instrument $\mathcal{I}$. It is particularly convenient to append the formula to the probability in Eq. (3.54) as

$$p(X) = \mathrm{Tr}\left(\mathcal{I}(X)(\hat{\rho})\right) = \mathrm{Tr}\left(\hat{\Pi}_{\mathcal{I}}(X)\hat{\rho}\right) \ . \tag{3.60}$$

The effect observable induced by the instrument is unique $\mathcal{I} \to \hat{\Pi}_{\mathcal{I}}$, while the reverse is once more an equivalence to a collection of instruments $\hat{\Pi} \to [\mathcal{I}]$, because different measurements of the same observable $\hat{\Pi}$ may have quite different effects on the measured object. In conclusion, every measurement scheme fixes an instrument and transitively a unique POVM. Reversely, there are infinitely many CP instruments to a POVM which admit infinitely many models $\mathcal{M}$, ergo

$$\hat{\Pi}_{\mathcal{I}} \Rightarrow [\mathcal{I}], \qquad \mathcal{I} \Rightarrow [\mathcal{M}] \ . \tag{3.61}$$

A fundamental physical principle can be formulated now: "There is no information gain without disturbance of quantum states" (for a proof see [58]). There are instruments that do not change the states and can only perform scaling by a nonnegative probability measure λ:

$$\hat{\Pi}_{\mathcal{I}}(X) = \lambda_X \mathbb{1} = p\,(X)\,\mathbb{1} \ . \tag{3.62}$$

Effects of the form Eq. (3.62) which are proportional to the identity operator are called *trivial observables*. For an instrument of this kind all states of the object system remain unaltered and the particular outcome X does not distinguish any states, which can be simply recognized when inserting $\mathcal{I}(X)(\hat{\rho}) = \lambda_X \hat{\rho}$ into Eq. (3.55). Measurements that do not alter the states of the system give no information at all.

3.6.1 Lüders instrument

In the following section the standard von Neumann measurement model is presented and will be related to the *Lüders instruments*, the most common type of instruments for which the state is updated according to the square root dynamics. For this purpose the corresponding formalism is considered and thereby the ensuing consequences of a von Neumann measurement is revealed. The description of the dynamics of von Neumann's measuring process requires a quantum theoretical description of both the measuring object and the measuring apparatus. Therefore, let $\{|\phi_n\rangle\}$ be an orthonormal basis in $\mathcal{H}$ and the projector $\hat{P}_n = |\phi_n\rangle\langle\phi_n| \in \mathcal{P}(\mathcal{H})$ be the associated sharp observable. Next, $\hat{P}^M_m = |\xi_m\rangle\langle\xi_m|$ is the effect of the apparatus with m being

the pointer value brought forth by the orthonormal probe states $\{|\xi_m\rangle\} \in \mathcal{K}$. The measurement interaction $\hat{U}$ must be designed such that a clear result is obtained when the probe state changes from the initial state $|\xi_1\rangle$ to

$$\hat{U}\left(|\phi_n\rangle \otimes |\xi_1\rangle\right) = |\phi_n\rangle \otimes |\xi_n\rangle \tag{3.63}$$

for all states $|\phi_n\rangle$. The instrument corresponding to the model $\langle \mathcal{K}, |\xi_1\rangle, \hat{U}, \hat{P}_m^M \rangle$ can be calculated with Eq. (3.52). For a general system state $|\psi\rangle = \sum_i c_i |\phi_i\rangle$ calculation yields

$$\begin{aligned}
\mathcal{I}(m)(|\psi\rangle\langle\psi|) &= \mathrm{Tr}_{\mathcal{K}}\left(\left(\hat{U}\left(|\psi\rangle \otimes |\xi_1\rangle\right)\left(\langle\psi| \otimes \langle\xi_1|\right)\hat{U}^\dagger\right)\left(\mathbb{1} \otimes \hat{P}_m^M\right)\right) \\
&= \mathrm{Tr}_{\mathcal{K}}\left(\hat{U}\left(\sum_{i,j} c_i^* c_j \, |\phi_i\rangle\langle\phi_j| \otimes |\xi_1\rangle\langle\xi_1|\right)\hat{U}^\dagger\left(\mathbb{1} \otimes \hat{P}_m^M\right)\right) \\
&= \mathrm{Tr}_{\mathcal{K}}\left(\left(\sum_{i,j} c_i^* c_j \, |\phi_i\rangle\langle\phi_j| \otimes |\xi_i\rangle\langle\xi_j|\right)\left(\mathbb{1} \otimes |\xi_m\rangle\langle\xi_m|\right)\right) \\
&= \mathrm{Tr}_{\mathcal{K}}\left(\sum_i c_i^* c_m \, |\phi_i\rangle\langle\phi_m| \otimes |\xi_i\rangle\langle\xi_m|\right) \\
&= |c_m|^2 \, |\phi_m\rangle\langle\phi_m| = \mathrm{Tr}\left(|\psi\rangle\langle\psi| \, \hat{P}_m^M\right)\hat{P}_m^M = \hat{P}_m^M \, |\psi\rangle\langle\psi| \, \hat{P}_m^M \ .
\end{aligned} \tag{3.64}$$

In case the observable is a projection, the von Neumann type measurements is given by the Lüders instrument as

$$\mathcal{I}_L(m)(\hat{\rho}) = \hat{P}_m \hat{\rho} \hat{P}_m \ , \tag{3.65}$$

which has already been introduced in sec. 3.3.1, but now with more details and clarity. A quick evaluation on the probability confirms the plausibility of the result:

$$p(m) = \mathrm{Tr}\left(\mathcal{I}_L(m)(\hat{\rho})\right) = \mathrm{Tr}\left(\hat{P}_m^2 \hat{\rho}\right) = \mathrm{Tr}\left(\hat{P}_m \hat{\rho}\right) = |c_m|^2 \ . \tag{3.66}$$

Apparatuses of this sort have the property of *ideality*, which means that if outcome m occurs with $p(m) = 1$, then the initial state is an eigenstate of the observable $\hat{P}_m$ and there is no disturbance of the initial state $\hat{P}_m \hat{\rho} \hat{P}_m = \hat{\rho}$. But, as was already pointed out previously, if the measured observable is unsharp, i.e. given by complete set of effects on a finite or countable set $m \in \Omega$, then for a discrete observable

$$\mathcal{I}'_L(m)(\hat{\rho}) = \sqrt{\hat{\Pi}(m)}\hat{\rho}\sqrt{\hat{\Pi}(m)} \ , \tag{3.67}$$

is the correct form and $\mathcal{I}'_L$ is called *general Lüders instrument*. A simple way to see that Eq. (3.65) cannot be generally correct is to look at the total probability:

$$1 = \sum_m p(m) = \sum_m \mathrm{Tr}\left(\hat{\Pi}(m)^2\hat{\rho}\right) \neq \sum_m \mathrm{Tr}\left(\hat{\Pi}(m)\hat{\rho}\right) = 1 \ . \tag{3.68}$$

The statement only holds for projective observables $\hat{\Pi}(m)^2 = \hat{\Pi}(m)$. In close connection to the Lüders instrument is the *Lüders theorem* which states that a sharp discrete observable $\hat{P}_m$ commutes with observable $\hat{A} \in \mathcal{L}(\mathcal{H})$, if measuring $\hat{P}_m$ does not disturb the measurement of $\hat{A}$. Mathematically, this formulates to

$$\mathrm{Tr}\left(\mathcal{I}_L(\Omega)(\hat{\rho})\hat{A}\right) = \mathrm{Tr}\left(\hat{\rho}\hat{A}\right) \quad \Longleftrightarrow \quad [\hat{P}_m, \hat{A}] = 0 \ . \tag{3.69}$$

This conjecture is easily shown, since

$$\mathrm{Tr}\left(\mathcal{I}_L(\Omega)(\hat{\rho})\hat{A}\right) = \sum_m \mathrm{Tr}\left(\hat{P}_m\hat{\rho}\hat{P}_m\hat{A}\right) = \mathrm{Tr}\left(\left(\sum_m \hat{P}_m\right)\hat{\rho}\hat{A}\right) = \mathrm{Tr}\left(\hat{\rho}\hat{A}\right) \ . \tag{3.70}$$

The extension of the proposition to unsharp observables is known as generalized Lüders theorem [59].

An important characteristic of measurements is repeatability. In terms of instruments this can be expressed as the reproducibility of an outcome in a sequential measurement. A common way to phrase this idea mathematically is that the conditional quantum state upon repetition must be identical $\mathrm{Tr}\left(\mathcal{I}(X)(\mathcal{I}(X)(\hat{\rho}))\right) = \mathrm{Tr}\left(\mathcal{I}(X)(\hat{\rho})\right)$. The projective Lüders measurement $\hat{P}_m^2 = \hat{P}_m$ is repeatable as can be quickly proven

$$\mathcal{I}_L(m)(\mathcal{I}_L(m)(\hat{\rho})) = \hat{P}_m\mathcal{I}(m)(\hat{\rho})\hat{P}_m = \hat{P}_m^2\hat{\rho}\hat{P}_m^2 = \hat{P}_m\hat{\rho}\hat{P}_m = \mathcal{I}_L(m)(\hat{\rho}) \ . \tag{3.71}$$

CHAPTER 4

Elements of neutrons optics

4.1 Properties of the neutron

Neutrons belong to the building blocks of the atomic nuclei and are effected by all four fundamental interactions. The neutron has a mass of $m_n = 939.57\,\mathrm{MeV/c^2}$ and can therefore be described by the Schrödinger equation for non-relativist energies, it is not stable outside the nucleus and decays according to the reaction

$$n \to p + e^- + \bar{\nu}_e$$

with an average life-time of $\tau_n \simeq 887\,\mathrm{s}$, it strongly interacts with nucleons and finally is affected by electromagnetic interactions because of the magnetic moment caused by the neutron's $\frac{1}{2}$-spin

$$\boldsymbol{\mu} = \gamma_n \mathbf{S} = \frac{\gamma_n \hbar}{2} \hat{\boldsymbol{\sigma}}\,, \tag{4.1}$$

where $\gamma_n \simeq -1.832\,10^{-8}\ \mathrm{rad\,s^{-1}\,T^{-1}}$ is the neutron's *gyromagnetic ratio*. The energy of a neutron in quantum mechanics is related to *de Broglie wavelength*

$$\lambda = \frac{h}{m_n \cdot v}\,. \tag{4.2}$$

Depending on the value of the wavelength one distinguishes between cold, thermal, epithermal and fast neutrons. Thermal neutrons for example have energies around 25 meV, which is equiv-

alent to a wave length of 1.8 Å ($v \cong 2200\text{m/s}$) and corresponds, according to $E = k_B T$, to a temperature of $\simeq 300\,\text{K}$.

Quantum physically, the density operator of a free neutron wave function is given by a superposition of plane waves

$$\hat{\rho}(\mathbf{r}, t) = \int d^3k \hat{\rho}'(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \,. \tag{4.3}$$

The lack of an electric charge or the absence of the coulomb field allows unimpeded penetration of the neutrons into materials, since the nucleons are neither influenced or hindered by the negatively charged electron orbits nor by the positive nuclear charges. Neutron interactions with atomic nuclei can only take place when the incoming neutron has entered the interaction range of the strong nuclear forces. The neutron is then either scattered or absorbed by a capture reaction with or without excitation. *Neutron optics* can be regarded as the subject that mainly deals with *coherent elastic neutron scattering.*

4.2 Production and detection of thermal neutrons

Research reactors that release neutrons by fission of uranium are the most common source. Only a small fraction, which happens to fly in the direction of the jet pipes leading out of the reactor, can be used. After moderation, i.e. deceleration of neutrons from their high initial energies to thermal energies, the wavelength spread is given according to a Maxwell-Boltzmann distribution. Uncertainty experiments presented in this thesis are all carried out at the TRIGA Mark-II research reactor in Vienna, whose characteristics can be looked up at [60, 61]. Key metrics are: the maximum continuous thermal power of the reactor is 250 kW, which provides a neutron flux of 10^{13} $\text{cm}^{-2}\text{s}^{-1}$. A pyrolytic graphite monochromator installed at the polarimeter beamline of the TRIGA Mark-II delivers a neutron beam with the 1st order wavelength of $\lambda_{(\text{n}=1)} = 1.99\text{Å}$ and a monochromaticity of $(\Delta\lambda)/(\lambda_{(\text{n}=1)}) \cong 0.02$, verified by time of flight measurements.

Due to their electrical neutrality, the neutron is not directly ionizing and can therefore only be detected via a nuclear conversion. The two most commonly used nuclear reactions utilize boron and helium, i.e.

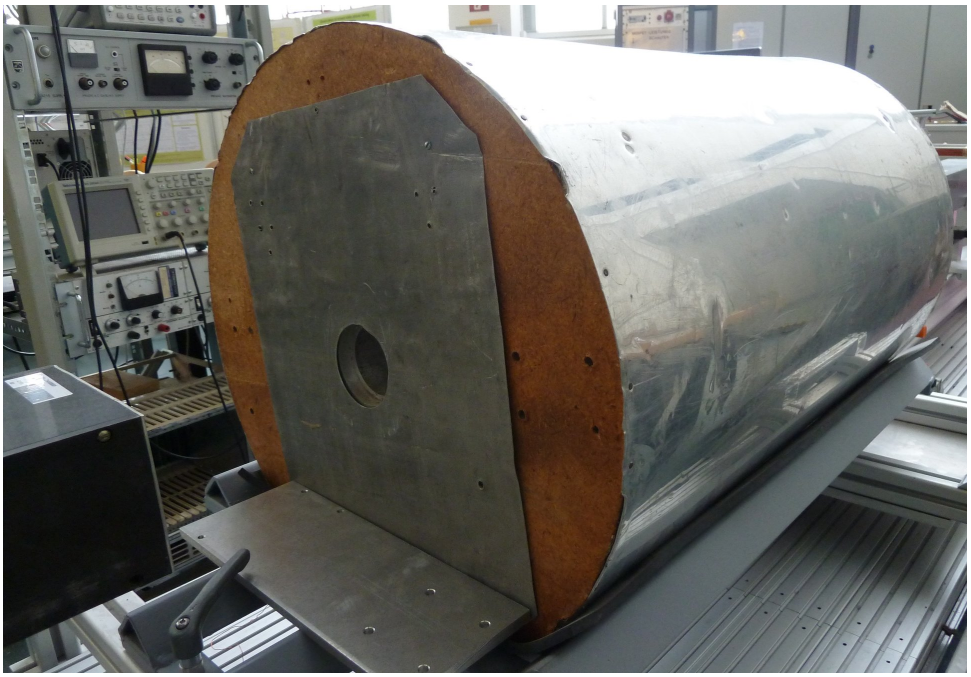


Fig. 4.1: Neutron particle detector containing a BF_3 filled proportional counter tube. The boron is enriched with ${}^{10}_{5}\mathrm{B}$ that produces lithium and alpha particles through nuclear reactions, ionizing the gas in the tube, which is why high voltages are applied between the electrodes. From the size of the round 6cm wide opening of the counter tube it can be recognized that a large part of the detector is made of the protective radiation shielding.

$$ {}^{10}_{5}\mathrm{B} + {}^{1}_{0}\mathrm{n} \rightarrow \begin{cases} {}^{4}_{3}\mathrm{Li} + {}^{4}_{2}\alpha + 2.8 \text{ MeV} \\ {}^{4}_{3}\mathrm{Li}^{*} + {}^{4}_{2}\alpha + 2.3 \text{ MeV} \end{cases} $$

and

$$ {}^{3}_{2}\mathrm{He} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{3}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{p} + 0.764 \text{ MeV} \ . $$

These detectors serve to detect slow or thermal neutrons. In fact, these nuclear reactions are exploited in proportional gas counter tubes, which are filled with either boron trifluoride BF_3 or helium-3 gas at up to about one bar positive pressure. The charged α-particle or Li, respectively the tritium and proton, which are released by secondary ionization are collected at a wire-shaped anode under high voltage. Due to the high electric field in the vicinity of the thin anode, the electrons are accelerated so much that they in turn can ionize further gas atoms. This way a gas

amplification factor of about 10^3 can be achieved. Finally, it is important to mention that the absorption of the neutron in the case of boron is spin-independent, while for ${}^3_2\mathrm{He}$ it is technically possible to have a spin-selective detection. The BF_3 - detector in Vienna, used throughout all experiments presented here, is shielded by moderating and absorbing materials made of paraffin together with boron powder to minimize background. A picture of the cylindric neutron detector is given in Fig. 4.1. The inherent detection efficiency is $> 99\%$ and the statistics for detection is in agreement with the Poisson-distribution given by Eq. (2.15).

4.3 Essentials of neutron optics

The theory of scattering of thermal neutrons is based on the *Born approximation* and the *Fermi pseudo-potential* [62]. The situation to have in mind is a neutron beam that scatters at a certain localized potential $\hat{V}(\mathbf{r})$, constituted by a another free particle or a nucleus in a solid target. For an initial plane wave

$$\langle \mathbf{r}|\mathbf{k}\rangle = \phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{i\mathbf{kr}} \tag{4.4}$$

the wave function of the scattering problem is given by the *Lippmann-Schwinger equation*

$$\psi(\mathbf{r},\mathbf{k}) \propto \phi_{\mathbf{k}}(\mathbf{r}) - \frac{1}{4\pi}\frac{2m_n}{\hbar^2}\int d^3r' \frac{e^{i\mathbf{k}|\mathbf{r}-\mathbf{r'}|}}{|\mathbf{r}-\mathbf{r'}|}\hat{V}(\mathbf{r'})\psi(\mathbf{r'},\mathbf{k})\,. \tag{4.5}$$

Here, $\mathbf{k}$ is the initial momentum of the neutron located at $\mathbf{r}$ and $\hat{V}(\mathbf{r'})$ is the complex optical potential of the scatterer. Since the scattering potential is spatially confined by the extent of the target, the distance of the detector from target is usually considered much larger ($\|\mathbf{r}\| \gg \|\mathbf{r'}\|$). This justifies the evaluation of Eq. (4.5) for an asymptotic behavior

$$k|\mathbf{r}-\mathbf{r'}| = \sqrt{1 - \frac{2\mathbf{r}\cdot\mathbf{r'}}{r^2} + \frac{r'^2}{r^2}} = kr - \mathbf{k'}\cdot\mathbf{r'} + O(r^{-1})\,, \tag{4.6}$$

where the wave vector on the right side is the final momentum $\mathbf{k'} = k\mathbf{e}_r$ of the neutron after the elastic scattering directed at the detector. Inserting the asymptotic formula Eq. (4.6) into the previous wave function gives

$$\psi(\mathbf{r},\mathbf{k}) \propto \phi_{\mathbf{k}}(\mathbf{r}) - \frac{2m_n}{4\pi\hbar^2}\frac{e^{ikr}}{r}\int d^3r' e^{-i\mathbf{k'}\cdot\mathbf{r'}}\hat{V}(\mathbf{r'})\psi(\mathbf{r'},\mathbf{k})\,. \tag{4.7}$$

This solution corresponds to a spherical wave of the type e^{ikr}/r that emanates after the impact as a scattering wave. The amplitude is spatially modulated as a function of the momenta and all information about the inner structure and potential of the impact target is contained in the so-called scattering amplitude $f(\mathbf{k}, \mathbf{k'})$ given by

$$\begin{aligned} f(\mathbf{k}, \mathbf{k'}) &= \frac{2m_n}{4\pi\hbar^2} \int d^3r' e^{-i\mathbf{k'}\cdot\mathbf{r'}} \hat{V}(\mathbf{r'}) \psi(\mathbf{r'}, \mathbf{k}) \\ &\approx \frac{2m_n}{4\pi\hbar^2} \int d^3r' \hat{V}(\mathbf{r'}) e^{-i(\mathbf{k'}-\mathbf{k})\cdot\mathbf{r'}} + \cdots \ . \end{aligned} \tag{4.8}$$

In the second line of Eq. (4.8) the Born approximation was applied which has a simple explanation. The incident primary wave with wave vector $\mathbf{k}$ stimulates a scattered wave $e^{-i\mathbf{k'}\cdot\mathbf{r}}$ by interaction with the respective potential $\hat{V}(\mathbf{r'})$ at the location $\mathbf{r'}$. The total scattering wave of the target results as a superposition of all partial waves from the volume element d^3r. The Born approximation corresponds to the first part in the iterative solution to $f(\mathbf{k}, \mathbf{k'})$ and only takes single scattering events into consideration, while higher terms in the series, concerning multiple scattering, are neglected.

For thermal neutrons, the scattering amplitude can in many situations be described by a single parameter, namely the *scattering length*. It can be derived by the Fermi potential $\hat{V}(\mathbf{r})$, which describes an isotropic scattering on a singular point-like nucleus:

$$\hat{V}(\mathbf{r}) = \frac{2\pi\hbar^2}{m_n} b\delta(\mathbf{r}) \ . \tag{4.9}$$

The parameter b is called the *bound* scattering length which, when inserted into Eq. (4.9), solely determines the scattering amplitude $f(\mathbf{k}, \mathbf{k'}) = b$. Of course, neutrons interact not only with one but with a lattice of bound atoms at position $\mathbf{r}_i$ with scattering lengths b_i building up a mean macroscopic potential as

$$\begin{aligned} V_F(\mathbf{r}) &= \langle \sum_i \frac{2\pi\hbar^2}{m_n} b_i \delta(\mathbf{r} - \mathbf{r}_i) \rangle \\ &= \frac{2\pi\hbar^2}{m_n} N b_c \, . \end{aligned} \tag{4.10}$$

The *coherent scattering length* $b_c = \langle b_i \rangle$ is defined as average over the individual b_i and $N = \langle \sum_i \delta(\mathbf{r} - \mathbf{r}_i) \rangle$ as the average particle number density which provides the means to describe

neutron optical phenomena to a larger extend. With a scattering length of $b_c = 10.3$ fm for nickel, for example, the Fermi potential would be roughly $V_F \cong 245$ neV.

Monochromatization and diffraction

A monochromator is used to select neutrons with certain wavelengths. Perfect single-crystals are often used as monochromators for neutrons, where waves are reflected by *Bragg's law*. In this case, there is a constructive interference only for the wavelengths which occur at the Bragg angle and thus fulfill the Bragg condition

$$n\lambda = 2d_{hkl} \sin\theta \,. \tag{4.11}$$

The integer n determines the order of the reflection of the incident beam at the n-th atomic lattice plane, d_{hkl} is the spacing between the planes in the atomic lattice with h,k,l indicating the Miller indices and θ is the angle of incidence. *Mosaic crystals* can also be used, which consist of small individual spatially displaced perfect crystal areas (mosaic blocks), which allow a broadening of the intensity reflexes due to a wider acceptance angle. The larger the mosaic width and the worse the neutron beam collimation, the larger the wavelength uncertainty $\Delta\lambda$ of the reflected beam becomes. Typically, the wavelength resolution $\Delta\lambda/\lambda$ of mosaic crystal monochromators is about 1%.

Reflection and refraction of neutrons

Refraction refers to the change in the **k**-vector of a wave due to a neutron's transmission through a medium described by a refractive index n. The time independent Schrödinger equation with Fermi potential $(\hat{H}_{kin} + \hat{V}_F)\psi(\mathbf{r}) = E\psi(\mathbf{r})$ has a plain wave solution and therefore an energy relation of the form

$$E = \frac{(\hbar\mathbf{K})^2}{2m_n} + V_F = \frac{(\hbar\mathbf{k})^2}{2m_n} \,, \tag{4.12}$$

where **K** is within the potential region and **k** outside of the target. The refractive index for the neutron wave in this case is defined as

$$\begin{aligned} n = \frac{K}{k} &= \sqrt{1 - \frac{V_F}{E}} = \sqrt{1 - \frac{4\pi}{k^2} N b_c} \\ &\approx 1 - \frac{V_F}{2E} = 1 - \frac{2\pi}{k^2} N b_c \,. \end{aligned} \tag{4.13}$$

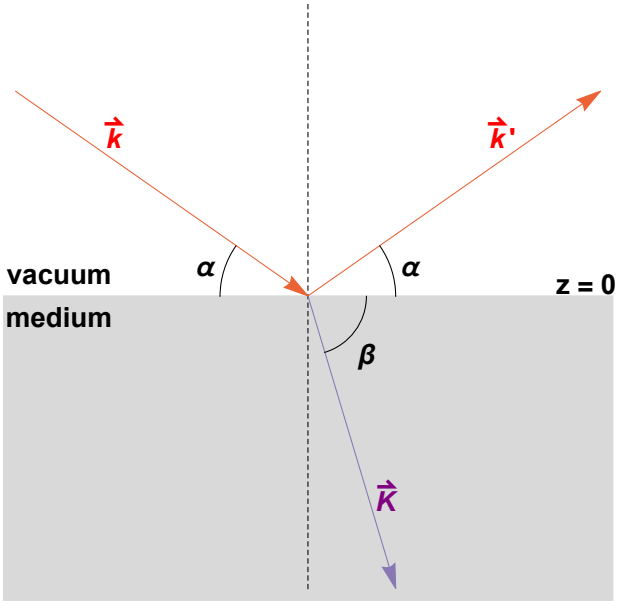


Fig. 4.2: Reflection and refraction of a neutron beam similar to that in light optics. When the neutron meets the surface at $z = 0$ with a different refractive index, the well-known phenomena occur.

The last line approximates the square root by the first term in the Taylor series since $\frac{V_F}{E} \ll 1$, which also implies that the index of refraction is close to $n \lesssim 1$. If the transmission at the interface is between medium A and a different medium B then refraction is given by a relative index $n = \frac{n_A}{n_B}$. Assume, a neutron matter wave with wave vector $\mathbf{k}$ comes from vacuum and enters another medium at level $z = 0$. Then the incident wave at the start (s) will generally have a partially reflected (r) and partially transmitted part (t), that is

$$\psi(\mathbf{r}) = \begin{cases} te^{i\mathbf{K}\cdot\mathbf{r}} & z < 0 \ , \\ se^{i\mathbf{k}\cdot\mathbf{r}} + re^{i\mathbf{k'}\cdot\mathbf{r}} & z > 0 \ . \end{cases} \tag{4.14}$$

The geometry of this problem is illustrated in Fig. 4.2. The reflected wave vector is $\mathbf{k'} = (k_x, -k_z)^T$. The wave function must satisfy the continuity conditions (for $\psi(\mathbf{r})$ and $\hat{\nabla}_{\mathbf{r}}\psi(\mathbf{r})$) at the boundary $z = 0$:

$$\begin{aligned} se^{ik_x x} + re^{ik_x x} &= te^{iK_x x} \ , \\ ik_x(se^{ik_x x} + re^{ik_x x}) &= iK_x(te^{iK_x x}) \ , \\ ik_z(se^{ik_x x} - re^{ik_x x}) &= iK_z(te^{iK_x x}) \ . \end{aligned} \tag{4.15}$$

From the first equation and the partial derivative $\partial_x \psi(\mathbf{r})$ in the second line the equality $k_x = K_x$ can be deduced and subsequently from the last line $k_z(s-r) = K_z(s+r)$, which relates the angles and refractive indices with the amplitudes of the wave function

$$n_z = \frac{K_z}{k_z} = \frac{K \sin\beta}{k \sin\alpha} = n\frac{\sin\beta}{\sin\alpha} = \frac{s+r}{s-r} \ . \tag{4.16}$$

This Eq. (4.16) is Snell's law and provides the expressions

$$r = \frac{1-n_z}{1+n_z} s \ , \qquad t = \frac{2}{1+n_z} s \ , \tag{4.17}$$

which are related to the *reflectivity* $R = |r|^2$ and the *transmissivity* $T = |t|^2$. Mathematically, the reflection and refraction of neutrons is similar to that of electromagnetic radiation and can be described by equivalent relationships known from light optics. Similar to X-ray radiation, total reflection also occurs for neutrons for the so-called grazing incidence. To show this fact the momentum conservation $k_x = K_x$ can be utilized which yields $n = \frac{\cos\alpha}{\cos\beta}$ and then eliminating the angle β from Snell's law Eq. (4.16) delivers

$$R = \left| \frac{\sin\alpha - \sqrt{n^2 - \cos^2\alpha}}{\sin\alpha + \sqrt{n^2 - \cos^2\alpha}} \right|^2 \ . \tag{4.18}$$

The result shows that for $n = 1$ everything is transmitted, while for $\alpha \to 0$ reflectivity is approximately $R \approx 1$. For thermal neutrons the *critical angle for total reflection*

$$\alpha_c = \sqrt{1-n^2} = \sqrt{\frac{\lambda^2}{2\pi} N b_c} \tag{4.19}$$

is typically $\alpha_c \approx 0.2$ degree in which case the series expansion of Eq. (4.18) satisfies $R \propto \alpha_c^2 / \sin^4\alpha$. The results just obtained correspond to the laws of refraction on a single layer. The range of the critical reflection angle can be increased by an order of 5 by incorporation of multilayer structures. This method can be used with non-magnetic materials for neutron guides, but also with magnetic materials in *polarizing supermirrors*.

Supermirror

The neutron is a spin-1/2 particle with a magnetic moment $\boldsymbol{\mu}$, which for an exterior magnetic field $\mathbf{B}$ results in a potential $V_m = \boldsymbol{\mu} \cdot \mathbf{B}$. The phenomenon behind neutron polarization is spin-

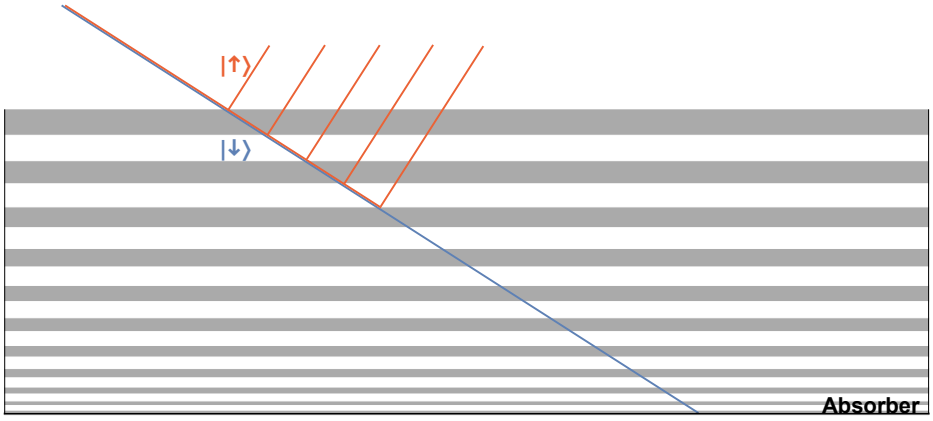


Fig. 4.3: Schematic illustration of the alternating multilayer structure of magnetic and non-magnetic materials in a supermirror polarizer. Bragg reflection on certain magnetic single crystals is a standard method for producing polarized monochromatic thermal neutron beams.

dependent interaction of the particles with magnetic materials. Among the most common tools are magnetic supermirrors made out of a ferromagnetic material (e.g. Fe, Co, Ni) which induce *magnetic scattering* (b_m) at the interface. Effectively, this scattering behavior is described by adapting the scattering length; Eq. (4.10) is modified to

$$V_F(\mathbf{r}) = \frac{2\pi\hbar^2}{m_n} N(b_c \pm b_m) \ . \tag{4.20}$$

The sign indicates whether the neutrons spins are oriented parallel or anti-parallel to the magnetization of the nuclei in the sample. The polarizing supermirror is constructed of several layers of ferro- and non-ferromagnetic materials (usually two) applied to a neutron absorbing substrate with decreasing layer thickness towards the substrate. The magnitude of nuclear and magnetic scattering is determined by their respective scattering lengths. The reflectivity for the spin magnetic moment in the multi-bilayer with material A and B in the guide tube has a proportionality of the form

$$R_\pm \propto [N_A b_c^A - N_B(b_c^B \pm b_m^B)] \ . \tag{4.21}$$

For a suitable sequence of bilayers ($b_c^B = \mp b_m^B$) the neutron reflectivity of one spin component can be immensely reduced. Each of these bilayers in these systems produces a Bragg peak and

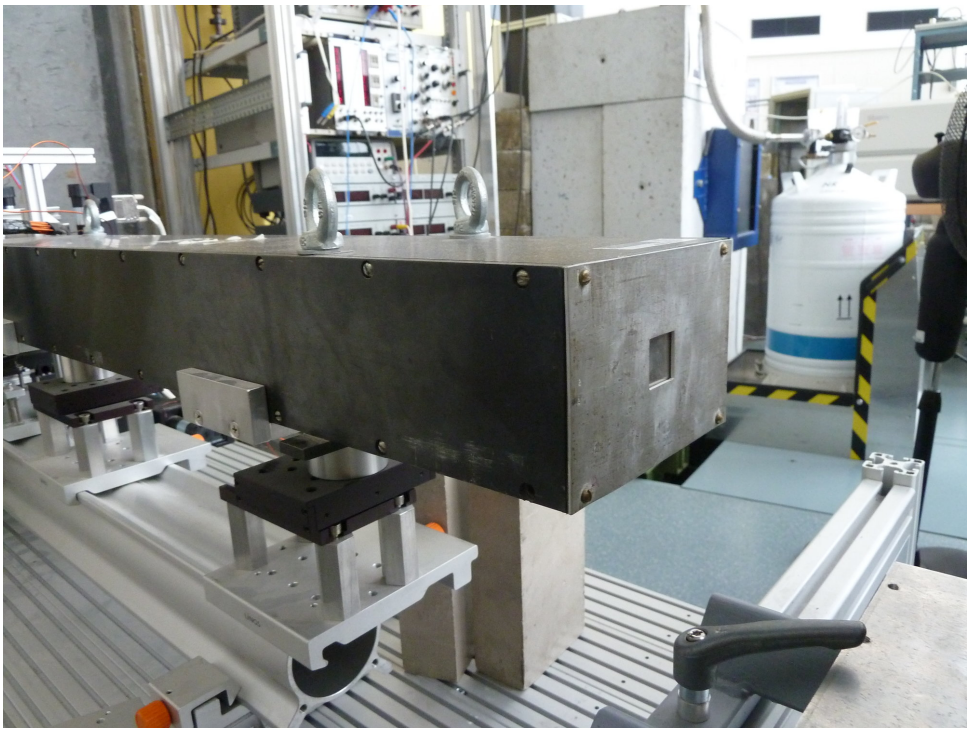


Fig. 4.4: A polarizing Ni-Ti supermirror is a rectangular cuboid with a typical size of (HWD) 10cm x 25cm x 80cm depending on the design, materials or manufacturer. Small windows of dimensions up to 5cm x 5cm at both ends of the supermirror serve as input and output for the neutrons. The curvature of the multilayer structure can not be seen from the outside, but can be noticed when adjusting with a subsequent detector. Because of the elements making up the inside, the object is rather heavy.

the sequence of multi-layers extends the range of reflection (see Fig. (4.3)). A slight curvature of the total, typically 0.5 - 1 m long arrangement prevents direct transmission of neutrons that would otherwise not touch the walls of the mirror. However, the inside of a supermirror is not visible from the outside, as Fig. 4.4 shows. The typically achievable degrees of polarization P are in the range of 0.9 - 0.99, where the efficiency of polarization

$$P = \left| \frac{N_+ - N_-}{N_+ + N_-} \right| \tag{4.22}$$

is defined as the difference of particles polarized up (+) or down (-) divided by the total particle number. The polarization varies from 0 for a totally unpolarized beam to 1 for a fully polarized beam.

Phase shift and absorption

If the neutron is not reflected by the multilayer but enters the medium, the neutron in the material is shifted in phase and can also be absorbed depending on the specific cross section. With a few exceptions, the probability of neutron absorption in the atomic nucleus for most isotopes is much smaller than the scattering probability. For a material with thickness D, there is a weakening of the neutron transmittance according to the *Beer-Lambert law* $T = e^{-\mu D}$, where μ is the absorption coefficient. Should the neutron not be absorbed, then the Fermi potential V_F causes nonetheless a phase shift. For a neutron propagating perpendicular to a material and exiting after a length D a phase shift of $\chi = -\lambda N b_c D$ will occur. For example, copper with 1 mm thickness shifts the phase of the neutron with 2Å wavelength by $\chi \approx -140$ rad assuming a neutron incidence normal to the surface. Such a relative phase shift can be measured with neutron interferometers [31].

4.4 Interaction of free neutrons with a classical magnetic field

The appropriate quantum mechanical description of the dynamics of a neutron with non-relativistic energies in an external magnetic field is given by the Schrödinger (- Pauli) equation

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t) = \hat{H}(\mathbf{r},t)\psi(\mathbf{r},t) = \left(-\frac{\hbar^2}{2m_n}\hat{\nabla}^2 - \mu\,\hat{\boldsymbol{\sigma}}\cdot\mathbf{B}(\mathbf{r},t)\right)\psi(\mathbf{r},t)\ . \tag{4.23}$$

The Schrödinger equation Eq. (4.23) contains the time derivative of the wave function as well as the kinetic and potential operator. The magnetic interaction term contains the Pauli spin operators $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)^T$ indicating that the wave function is constituted by a two dimensional spinor function $|\chi(\mathbf{r},t)\rangle = (\chi_+(\mathbf{r},t), \chi_-(\mathbf{r},t))^T$. The evolution and dynamics of the wave function depends on the form and type of the magnetic field. The behavior of neutrons in a homogeneous static magnetic field is explained in the following.

In case the field vector is constant and uniform, i.e. $\mathbf{B}(\mathbf{r},t) = B_0\mathbf{e}_n$, Eq. (4.23) can be reduced to the time-independent Schrödinger equation solved by

$$|\psi(x,t)\rangle = \frac{1}{\sqrt{2\pi}}\int \phi(k)e^{i\left(kx-\frac{\hbar k^2}{2m_n}t\right)}\mathrm{d}k\ |\chi(t)\rangle\,, \tag{4.24}$$

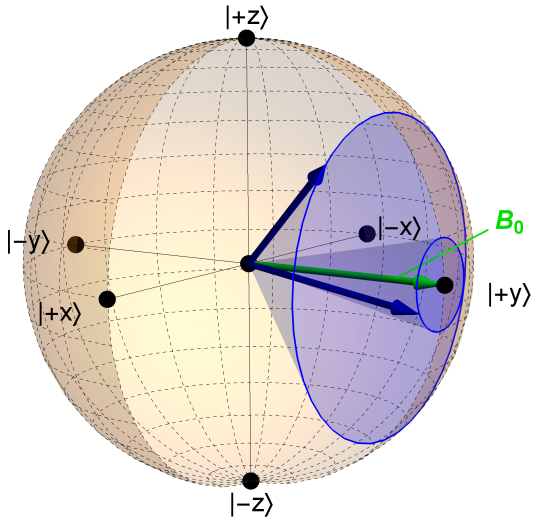


Fig. 4.5: Depiction of Larmor precession. The external field B_0 points into the y-direction. Depending on the initial angle of the spin, it will undergo a gyroscopic precession about the magnetic field. For states orthogonal to the magnetic field, the cone degenerates into a circle allowing to flip the spin from up to down.

where the spin evolution $|\chi(t)\rangle$ is given by the unitary operator $\hat{U}_{DC}$

$$\begin{aligned} |\chi(t)\rangle &= \hat{U}_{DC}(t)\,|\chi(0)\rangle = e^{-i\frac{\mu B_0}{\hbar}\hat{\boldsymbol{\sigma}}\mathbf{e}_n t}\,|\chi(0)\rangle \\ &= \left(\cos\left(\frac{\mu B_0}{\hbar}t\right)\mathbb{1}_2 - i\hat{\boldsymbol{\sigma}}\cdot\mathbf{e}_n \sin\left(\frac{\mu B_0}{\hbar}t\right)\right)|\chi(0)\rangle \ . \end{aligned} \tag{4.25}$$

Two interesting phenomena can be deduced from the previous formulae, *Zeeman energy splitting* and *Larmor precession*. Zeeman splitting of the neutron kinetic energy occurs due to interaction of the spin parallel to the external magnetic field. The initial energy $E_0 = \hbar^2 k_0^2/2m_n$ of the free neutron changes for up- and down spin states to

$$\langle \hat{H} \rangle = \frac{\hbar^2 k_0^2}{2m_n} \pm \mu B_0 \,. \tag{4.26}$$

The Larmor precession describes the evolution of the neutron spin in the constant magnetic field. Introducing the Larmor frequency $\omega_L = 2\mu B_0/\hbar$ the polarization of the neutron, calculated by the expectation value of the Pauli spin operators, yields

$$\mathbf{P} = \langle \hat{\boldsymbol{\sigma}} \rangle = \langle \chi(0)| \left(\hat{\boldsymbol{\sigma}}\cos(\omega_L t) + \mathbf{e}_n \times \hat{\boldsymbol{\sigma}}\sin(\omega_L t)\right) + \mathbf{e}_n(\mathbf{e}_n\cdot\hat{\boldsymbol{\sigma}})\left(1-\cos(\omega_L t)\right)|\chi(0)\rangle \ . \tag{4.27}$$

The Eq. (4.27) makes use of Rodrigues' rotation formula and a short mathematical exercise reveals that it fulfills

$$\dot{\boldsymbol{P}} = \frac{d}{dt}\langle\hat{\boldsymbol{\sigma}}\rangle = \boldsymbol{\omega} \times \mathbf{P}\,. \tag{4.28}$$

This equation is nothing but the Bloch equation without relaxation describing how a magnetic field gives a torque on the system. The cone-like movement of spin in the field, illustrated in Fig. 4.5, is called Larmor precession.

Homogeneous, constant magnetic fields can be applied in various ways. The most common methods include, on the one hand, permanent magnets on the other hand coils with a *solenoid* or a *Helmholtz* design. An ideal solenoid is geometrically a helical course of wire which has a length l relatively large to the diameter and thus a very high number of windings N. The field strength of such a magnetic field can be regulated by the current I and is approximately [63]

$$B(I) \approx \frac{\mu_0 N}{l} I\,. \tag{4.29}$$

The photo in Fig. 4.6 shows a commonly used coil. This consists of two insulated wires wound in different directions. Observing the course of the turns, it can be seen that the coils each generates a homogeneous magnetic field in the horizontal and vertical directions, while there is no field along the short side (orthogonal to the image plane).

The Helmholtz coil is a special coil arrangement in which two short coils with a radius R are arranged in parallel at the distance equal to the radius R fed with the same current. At the center of the coils the field is given by

$$B(I) \approx \frac{8\mu_0}{\sqrt{125}R} I\,. \tag{4.30}$$

Unlike neutrons in spatially homogeneous fields, a different effect occurs for the spin if the field is inhomogeneous. A neutron beam in a magnetic field, where $\mathbf{F} = \boldsymbol{\nabla}(\boldsymbol{\mu} \cdot \mathbf{B})$ is non-zero will be deflected according to Stern-Gerlach experiment which will always occur as soon as a non-homogeneous magnetic field transition is present. However, strong fields in the Tesla range are necessary to realize a splitting of a thermal neutron beam of a few arc-seconds [64]. A supermirror can work as a spin-separating *analyzer* much like a Stern-Gerlach apparatus.

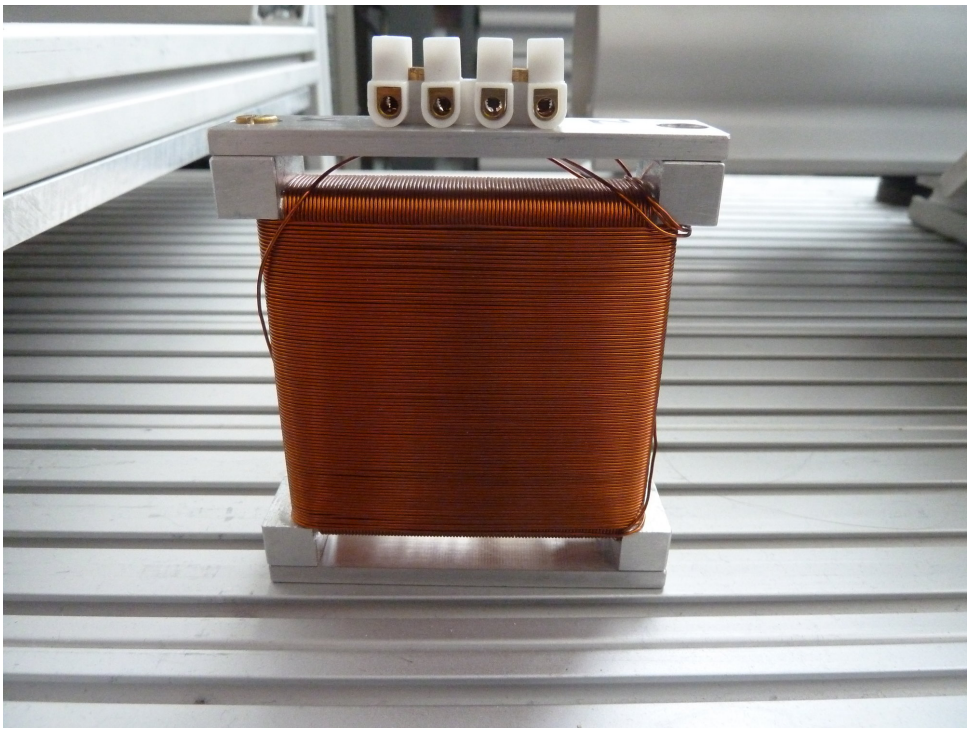


Fig. 4.6: The DC spin rotating coil. Basically, it consists of two helically wound wires (2mm diameter), that lie on top of each other and produce local magnetic fields. These spin rotators have a size of (HWD) 6cm x 6cm x 2cm and are typically standing on small goniometer stages to adjust them with respect to a guide field and the neutron beam propagating along the geometric center through the copper. The wires are connected to a current source using screw terminals sitting on top of the coil.

CHAPTER 5

Preparation uncertainty relations

State preparation uncertainties refer to uncorrelated measurements of observables $\hat{A}$ and $\hat{B}$ on distinct but identically prepared systems of an ensemble $\hat{\rho}$. The term preparation uncertainty was introduced as such to make a distinction between conceptually different types of uncertainty relations. The byword preparation emphasizes the reciprocal inaccuracy in the production or processing of those quantum states which have no common eigenbasis. Consequently, there is no device or mechanism that would allow to prepare arbitrarily sharp eigenstates of two non-commutating observables jointly. However, this type of uncertainties idealize the actual measurement and do not take into account the quantum mechanical nature of the measurement procedure, which was not possible at Heisenberg's time as the necessary mathematical tools had not been sufficiently developed. Originally, the uncertainty principle had also been understood to set a limitation on measurements, i.e. each measurement necessarily has a back-action on the measured object and consequently limits the achievable accuracy of incompatible measurements. The impossibility to infer the individual results of the measurements from the jointly measured statistics is called the measurement uncertainty principle.

There are two categories that give three distinguished types of uncertainties [47, 65]

1. State preparation uncertainty describes the impossibility of preparing a condition in which conjugated quantities simultaneously have an exact value. This comprises the standard

textbook presentation in terms of standard deviation.

2. a) The uncertainty of a simultaneous joint measurement. It is not possible to realize a measurement device that accurately estimates the outcomes of two physically incompatible observables $\hat{A}$ and $\hat{B}$ such that their joint measurement distribution is perfectly decomposable.

   b) The uncertainty of a successive joint measurement. Because a measurement ensues a disturbance on the quantum state a subsequent measurement of an incompatible observable will be compromised.

Point 1 differs from point 2, because it involves correlated measurements. Point 2 a) and 2 b) in turn are different, because the successive type accounts for both the inaccuracy and the imparted disturbance in one detection, while the joint type involves two measurements. It should be mentioned that e.g. according to Hall [22] a 4th kind exists still, which might be called overlap-uncertainty:

3. A measurement apparatus cannot be calibrated such that it unambiguously discriminates different physically incompatible quantum states.

In this kind of uncertainty principle, various systematically selected states are sent in sequence to a measurement apparatus. The term overlap then refers to the fact that the measurements cannot unambiguously tell apart all prepared particles unless they are all orthogonal to each other. Somehow, this statement can be regarded as the reversal of the principle in point 1, in which state and observables exchange roles. In this chapter, a small overview of important intermediate results are summarized, which are then partly needed in the subsequent chapters.

5.1 Variational preparation uncertainty relations

Heisenberg did not use any rigorous formulation in his initial work [2] about the uncertainty. The first form of uncertainty relation had therefore a more qualitative character. However, a short time after Heisenberg's vague presentation an inequality holding for arbitrary states was proven by Kennard [5], later by Weyl [6], who established the relation which is now simply known as Heisenberg's uncertainty relation

$$\sigma(\hat{Q})\sigma(\hat{P}) \geq \frac{\hbar}{2}\,. \tag{5.1}$$

Here, $\sigma(\hat{Q})$ and $\sigma(\hat{P})$ denote the standard deviations of position and momentum, whose product has to always be greater equal to half the reduced Planck constant $\hbar$. Soon thereafter, a generalization of the inequality to arbitrary observables $\hat{A}$, $\hat{B}$ was reported by Robertson [8]

$$\sigma(\hat{A})\sigma(\hat{B}) \geq \left| \frac{1}{2i} \langle\psi|[\hat{A},\hat{B}]|\psi\rangle \right| := C_{AB} \ , \tag{5.2}$$

and circa one year after, Schrödinger [9] derived a slightly improved version which contains Eq. (5.2) as special case

$$\sigma(\hat{A})^2\sigma(\hat{B})^2 \geq \left| \frac{1}{2i} \langle\{\hat{A} - \langle\hat{A}\rangle\,\mathbb{1}, \hat{B} - \langle\hat{B}\rangle\,\mathbb{1}\}\rangle \right|^2 + \left| \frac{1}{2i} \langle[\hat{A},\hat{B}]\rangle \right|^2 \ . \tag{5.3}$$

The proof of the relation Ineq. (5.3) relies on the Cauchy - Schwarz inequality: $|\,\langle\hat{A}'^2\rangle\,|\,|\,\langle\hat{B}'^2\rangle\,| \geq |\,\langle\hat{A}'\hat{B}'\rangle\,|^2$. With the identification $\hat{A}' = \hat{A} - \langle\hat{A}\rangle\,\mathbb{1}$, likewise for $\hat{B}'$, the left side yields the standard deviations $\sigma(\hat{A}) = \langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle$ for both operators. The right side is attained by making the decomposition $\hat{A}'\hat{B}' = \frac{1}{2}\left([\hat{A}',\hat{B}'] + \{\hat{A}',\hat{B}'\}\right)$ and substituting the primed observables with the shifted ones. When $\langle\{\hat{A}',\hat{B}'\}\rangle = 0$ then Robertson's relation is attained. The terms added by Schrödinger correspond to additional statistical correlations and, in fact, can be understood as quantum mechanical covariance [66] given as $\mathrm{cov}(\hat{A},\hat{B}) = \frac{1}{2}\langle\{\hat{A},\hat{B}\}\rangle - \langle\hat{A}\rangle\langle\hat{B}\rangle$. When the covariance is taken into account, then the *Robertson-Schrödinger equation* can be written particularly elegantly [67, 68] by means of the covariance matrix Eq. (2.18)

$$\det\left(\Sigma_{ij}(\hat{A},\hat{B})\right) \geq C_{AB} \ . \tag{5.4}$$

A close look at these inequalities is mandatory at this point. The standard deviation in Ineq. (5.2) is $\sigma(\hat{A}) = \sqrt{\langle\psi|\hat{A}^2|\psi\rangle - \langle\psi|\hat{A}|\psi\rangle^2}$ and likewise for observable $\hat{B}$. The standard deviations characterize the dispersion of a distribution and are a measure for the spread of a set of values around the mean. This means that the argued inequalities are concerned with the statistical nature of the outcome distribution and are therefore not related to the systematic imperfections of a quantum measurement. The theoretical standard deviation is in reality determined by numerous repetitions of the measurement on exactly identical states. Physically, one can associate preparation uncertainty relations with a reservoir of particles which contains a large number of copies of one $|\psi\rangle$, as illustrated in Fig 5.1, so that an error-free measuring apparatus determines the distribution of $\sigma(\hat{A})$ on a sub-ensemble of the states and with the

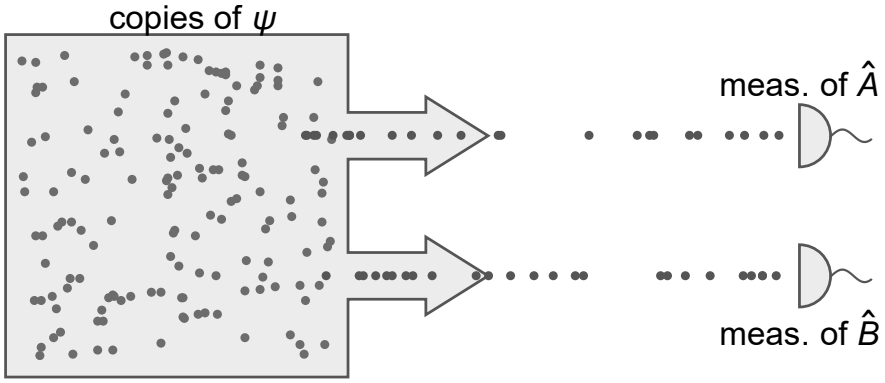


Fig. 5.1: Measurements on different parts of a sample of equally prepared systems.

second sub-ensemble the second distribution width $\sigma(\hat{B})$; consequently, there will be trade-off between the spreads of those distributions.

The position-momentum uncertainty is in many ways extraordinary in the sense that the involved operators are defined on an infinite-dimensional Hilbert space, related to each other by a Fourier transformation. The Fourier transformation provides for instance the link between the time domain and the frequency domain or a location and the spatial frequencies (wave numbers). Denoting Fourier transformation of the function $f(x)$ by $F(k)$, the bijective mapping and its inverse are given by

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2i\pi xk} dx, \qquad F(k) = \int_{-\infty}^{\infty} f(x) e^{-2i\pi xk} dx\,. \tag{5.5}$$

Suppose $f \in L^2(\mathbb{R})$ with unit norm $\|f\|_2 = 1$, then the variance can be written as $\sigma(f)^2 = \int_{-\infty}^{\infty}(x - \langle x\rangle)^2 |f(x)|^2 dx$ and for the Fourier transformed function likewise $\sigma(F)^2 = \int_{-\infty}^{\infty}(k - \langle k\rangle)^2 |f(k)|^2 dk$. So, mathematically for probability distributions with well defined Fourier transforms, an inequality of the form

$$\sigma(f)^2 \sigma(F)^2 \geq \frac{1}{(4\pi)^2} \tag{5.6}$$

can be obtained [6, 7, 69], saturated by Gaussian functions. The reciprocal nature of the spreads in the Fourier transform of, say, a plane wave $\psi = e^{\frac{i}{\hbar}(\mathbf{p}\cdot\mathbf{x} - Et)}$ also suggests an energy-time uncertainty. However, there is no time operator in quantum physics and hence no uncertainty relation can literally involve the time variable t displayed by a clock in a laboratory. Nevertheless, uncertainty relation of this kind do exist [70, 71] and classical Fourier transform seems to

be one of the few options to deal with such observables at all.

Over the time it has been questioned, whether the standard deviation is an adequate measure of uncertainty or not. On one hand, criticisms arose concerning the appropriateness of the standard deviation as a measure of distribution concentration, on the other hand, there are some downsides of the variance in finite dimensional Hilbert spaces. It seems, the less a probability function resembles a normal distribution, the less useful the standard deviations is. Here is a short list of potential problems:

- Multimodal distributions, i.e. distributions with multiple peaks. Consider the convex combination of two normal distributions $f(x) = \frac{1}{2}\mathcal{N}(\mu, \sigma_1) + \frac{1}{2}\mathcal{N}(-\mu, \sigma_2)$. The total standard deviation for two narrow bumps $\sigma_1 \approx \sigma_2 \to 0$ is given by $\sigma(f) \simeq \mu$. So the standard deviation of the function and its Fourier transform does not depend on the widths of the independent Gaussians, but on the distance of the peaks to the coordinate center.

- There is nothing special about the second moment of a distribution. One could as well take the kurtosis, the fourth standardized moment, or any higher $2n$-th moment (see 2.2.2).

- A big issue for continuous distributions are diverging standard deviations $\sigma \to \infty$ or distributions without higher moments. The Cauchy distribution given in Tab. 2.1, which is used in physics to describes resonances, has no mean value and ergo no higher stochastic moments.

- No invariance under relabeling $\{p_1, p_2\} \leftrightarrow \{p_2, p_1\}$. Given a probability vector $\mathbf{p} = (\mathrm{Tr}(\hat{\Pi}_1\rho), \mathrm{Tr}(\hat{\Pi}_2\rho), ...\mathrm{Tr}(\hat{\Pi}_n\rho))^T$, where is $\hat{\Pi}_i$ are elements of a POVM, a permutation π of the of the measurement outcomes $\pi(\mathbf{p})$ changes the standard deviation.

- In finite dimensions the standard deviation vanishes for the eigenstates of $\hat{A}$ or $\hat{B}$. Particularly for inequalities containing only products, like Robertson's relation, this leads to trivial statements.

The last point above can be circumvented by deriving a formula that contains the sum instead of the product of standard deviations. Recently, such inequalities were derived, among the more renown is the Maccone-Pati [72] uncertainty relation $\sigma(\hat{A})^2 + \sigma(\hat{B})^2 \geq \max\{L_1, L_2\}$ with

$L_1 = \pm i \langle [\hat{A}, \hat{B}] \rangle + \left| \langle \psi | \hat{A} \pm i\hat{B} | \psi^{\perp} \rangle \right|^2$ and $L_2 = \frac{1}{2} \left| \langle \psi^{\perp}_{\hat{A}+B} | \hat{A} + \hat{B} | \psi \rangle \right|^2$. This stronger uncertainty relations can be related to Robertson's equation by observing that from $(\sigma(\hat{A}) - \sigma(\hat{B}))^2 \geq 0$ it follows that $\sigma(A)^2 + \sigma(B) \geq 2\sigma(A)\sigma(B)$. Erasing the second summand of L_1 then yields the well-known form Ineq. (5.2).

5.2 Entropic preparation uncertainty relations

Unlike the standard deviation, which quantifies the weighted dispersion of the random variable along the abscissa x, the information entropy determines the uncertainty of a distribution only by the quantum physically relevant values on the ordinate axis, namely the probabilities $p(x)$. In order words, entropy measures the uncertainty only by the 'heights' in a histogram. While working on the many-worlds interpretation of quantum mechanics Everett discovered already in 1957 an inequality, which he would publish later in an expanded version of his doctoral thesis [29, 73]. Likewise Hirschmann [12] studied the same entropic representation of the uncertainty relation for position and momentum and proposed in the same year for the sum of differential entropies

$$h(|f|^2) = -\int_{-\infty}^{\infty} |f(x)|^2 \log(|f(x)|^2) dx \tag{5.7}$$

the uncertainty relation $h(|f|^2) + h(|F|^2) \geq 0$, where $\|f(x)\|_2 = \|F(k)\|_2 = 1$ are normalized and related to each other by Fourier transformation. The relation was tightened by Beckner [13] and independently by Białynicki-Birula and Mycielski in 1975 [14] to the renown form

$$h(|f|^2) + h(|F|^2) \geq \log\left(\frac{e}{2}\right) \ . \tag{5.8}$$

The Ineq. (5.8) faithfully satisfies the maximum property of the differential entropy for a Gaussian Ineq. (2.43) . For position and momentum the following inequality can be derived

$$\begin{aligned} \log(e\pi\hbar) &\leq \log(2\pi e\sigma(Q)\sigma(P)) \\ &= \log\left(\sqrt{2\pi e\sigma(\hat{Q})^2}\right) + \log\left(\sqrt{2\pi e\sigma(\hat{P})^2}\right) \leq h(Q) + h(P) \ . \end{aligned} \tag{5.9}$$

Because the entropic uncertainty in Ineq. (5.9) implies the standard deviation relation it is considered as stronger.

The change of the representation form of the uncertainty led to a re-consideration of existing examples and highlighted some issues with the standard deviation. The usefulness of entropy motived Deutsch [15] to find an information-theoretic inequality for finite dimensional Hilbert spaces which would not only be valid for continuous Fourier transformed pairs of variables. Among the problems already given for the standard deviation as inappropriate measures of uncertainty, Deutsch criticized that unlike in the position momentum case the right hand side of Robertson's relation is state dependent. The commutator cannot be proportional to the identity in finite dimensions, hence would inherently not constitute a constant bound. It is easy to realize that by looking at the trace of the commutator $[\hat{A}, \hat{B}] = \alpha\hat{C}$, $\alpha \in \mathbb{C}$ which is

$$\mathrm{Tr}([\hat{A}, \hat{B}]) = \mathrm{Tr}(\hat{A}\hat{B}) - \mathrm{Tr}(\hat{B}\hat{A}) = 0 = \alpha\mathrm{Tr}(\hat{C}) \ . \tag{5.10}$$

So $\hat{C}$, if the trace is well defined, must be traceless, which is clearly not the case for the identity operator.

For finite dimensions entropic uncertainty relations first utilized the Shannon entropy $H(A)$, where A is the random variable related to the physical observable $\hat{A}$, defined by

$$H(A) = H(\hat{A}|\hat{\rho}) = -\sum_i \mathrm{Tr}(|a_i\rangle\langle a_i|\,\hat{\rho}) \log\left(\mathrm{Tr}(|a_i\rangle\langle a_i|\,\hat{\rho})\right) \ . \tag{5.11}$$

This is evidently the same as Eq. (2.32), but the formula receives now a physical interpretation as the deficiency in the information which a prepared state $\hat{\rho}$ gives about measurements made on $\hat{A}$. For two non-commuting observables $\hat{A}$ and $\hat{B}$, having complete sets of non-degenerate, normalized eigenstates $|a_i\rangle$ and $|b_i\rangle$, Deutsch derived the relation

$$H(A) + H(B) \geq -2\log\left(\frac{1+c}{2}\right) \ , \qquad c = \max_{i,j} |\langle a_i|b_j\rangle|^2 \ , \tag{5.12}$$

where the uncertainty bound c is defined as the maximal overlap of the quantum states and can also be written as the maximum of the squared fidelity $c = \max_{i,j} F(|a_i\rangle\langle a_i|\,, |b_j\rangle\langle b_j|)^2$. The right-hand side has now generally all features of a reasonable state independent limit: it does not explicitly dependent on the input state and equals zero if $\hat{\rho}$ is a simultaneously eigenstate of $\hat{A}$ and $\hat{B}$, i.e. when both observables posses definite values. The maximal overlap ranges between $1/d$ for two mutually unbiased bases, d is the dimension of the Hilbert space, to 1, when the observables have a common eigenvector; so $\frac{1}{d} \leq c \leq 1$.

After Deutsch's work this subject developed quickly. Partovi [74] first studied the inequality for degenerate spectra, and afterwards Maassen and Uffink [16] derived, based on a conjecture of Kraus [75], a tighter bound for the uncertainty measurements given by

$$H(A) + H(B) \geq -2\log(c) \ . \tag{5.13}$$

Amazingly, Maassen and Uffink proved that Ineq. (5.13) is just one member of a general class of inequalities given by the sum of Rényi entropies with indices $\{\alpha, \beta\} \geq \frac{1}{2}$

$$H_\alpha(A) + H_\beta(B) \geq -2\log(c) \ , \qquad \frac{1}{\alpha} + \frac{1}{\beta} = 2 \, . \tag{5.14}$$

While the validity of such entropic inequalities is undoubted, it is a great challenge to find out when saturation can be achieved depending on certain parameters like the degree of state mixture, the type and number of observables and the dimension of the system. The Maassen-Uffink uncertainty relation serves as useful guidepost since it is optimal for mutually unbiased quantum states in all dimensions. For other cases, finding the optimum in the inequality remains a highly active and intensively studied field of research.

5.3 Optimal preparation uncertainty relations for qubits

Because qubits belong to the simplest quantum system, entropic uncertainty relations in two dimensions were studied in depth and are one of the few cases for which an exact lower bound on the sum of the information entropies exists, even if only numerically. First Sanchez [76, 77] and then Ghirardi et. al. [78] contributed significantly to the analysis. Here is how the optimization works: the most general qubit observables are of the form Eq. (3.33). First, notice that if $|a\rangle$ is an eigenvector of $\mathbf{a} \cdot \hat{\boldsymbol{\sigma}}$, $\|\mathbf{a}\| = 1$, it is also an eigenvector of $\beta\mathbb{1} + \alpha\mathbf{a} \cdot \hat{\boldsymbol{\sigma}}$. Because the Shannon entropy depends only on the eigenstates and not on the eigenvalues, without loss of generality one can fix the pair of observables to $\hat{A} = \mathbf{a} \cdot \hat{\boldsymbol{\sigma}}$ and $\hat{B} = \mathbf{b} \cdot \hat{\boldsymbol{\sigma}}$, where both Bloch vectors have unit length. Accordingly, the problem simplifies to finding the Bloch vector $\mathbf{r}$ corresponding to the initial state $|\psi\rangle$ that minimizes the sum of the binary entropies $H_b(\mathbf{a} \cdot \mathbf{r}) + H_b(\mathbf{b} \cdot \mathbf{r})$ (see Eq. (2.33)). It is clear that $\mathbf{r}$ must be coplanar to the vectors $\mathbf{a}$, $\mathbf{b}$ to yield minimal values. By

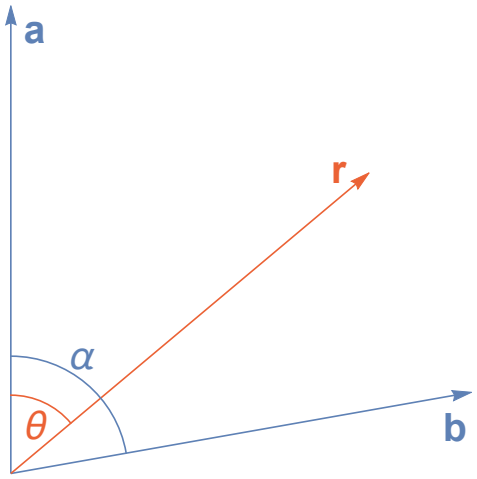


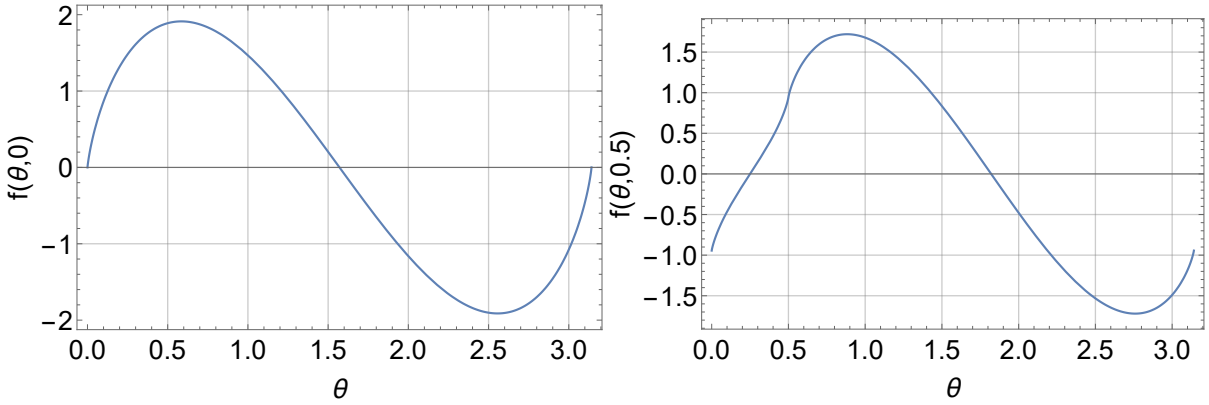
Fig. 5.2: The vectors and angles visualize the situation in the deduction of the optimal lowest bound for in two-dimensional Hilbert space. Here, **a** and **b** belong the respective observables $\hat{A}$ and $\hat{B}$ with an included angle α, while the others θ and $\alpha - \theta$ are the angles between the Bloch vectors and the system state.

identification of $\mathbf{a} \cdot \mathbf{r} = \cos\theta$ and expressing the observables' Bloch vector angle $\mathbf{a} \cdot \mathbf{b}$ by the maximal overlap $c = \cos\frac{\alpha}{2}$, the optimal bound is obtained by the minimization

$$\begin{aligned} f(\theta,\alpha) &= \frac{\partial}{\partial\theta}\left(H_b(\cos\theta) + H_b(\cos(\alpha-\theta))\right) \\ &= -\sin\theta\log_2\left(\frac{1+\cos\theta}{1-\cos\theta}\right) + \sin(\alpha-\theta)\log_2\left(\frac{1+\cos(\alpha-\theta)}{1-\cos(\alpha-\theta)}\right) = 0\ , \end{aligned} \tag{5.15}$$

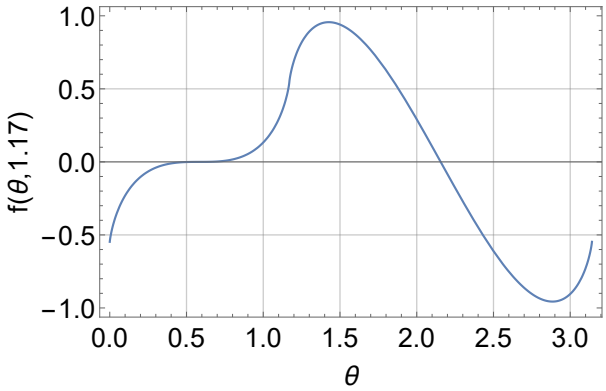
where θ runs over the interval $[0, 2\pi)$ and α over $[0, \pi)$. The situation is illustrated in Fig. 5.2 and the derivative of Eq. (5.15) is plotted for four different α in Fig. 5.3. These plots indicate that the form of the derivative depends on α and that for some critical value α_c the number of zeros of the function changes from 2 to 4. This critical value apparently occurs at the point where the inflection of the curve is stationary. In order to determine that particular angle the second derivative has to be calculated. From the first graph Fig. 5.3(a) one sees that the function has a peak when

$$\left.\frac{\partial f(\theta,\alpha)}{\partial\theta}\right|_{\alpha=0} = 0 \quad \rightarrow \quad \cos\theta\log_2\left(\frac{1+\theta}{1-\theta}\right) = 0\ . \tag{5.16}$$

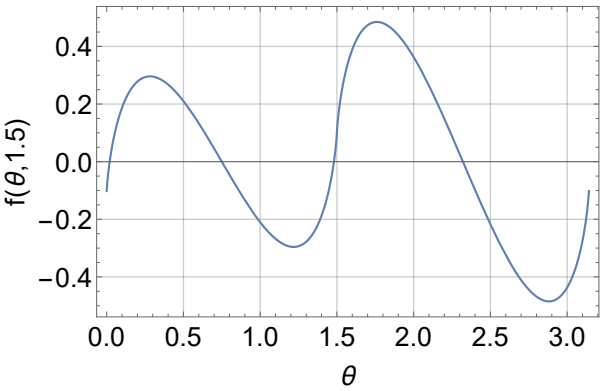


(a) $\alpha = 0$.

(b) $\alpha < \alpha_c$.



(c) $\alpha \cong \alpha_c$.



(d) $\alpha > \alpha_c$.

Fig. 5.3: Plots of the derivative as given in Eq. (5.15). The function $f(\theta, \alpha)$ is shown for four different values of α, starting with $\alpha = 0$ and then $\alpha = 0.5$ in Fig. 5.3(b), $\alpha = 1.17$ in Fig. 5.3(c) and $\alpha = 1.5$ in Fig. 5.3(d). Depending on this angle, a bifurcation of the solution from two to four intersection with the abscissa occurs, which determines the critical angle.

The solution of the transcendental equation Eq. (5.16) shall be denoted as θ_c and has for $0 \leq \theta \leq \frac{\pi}{2}$ a value of $\theta_c \simeq 0.585$. Inserting the attained solution yields at $f(\theta_c, \alpha) = 0$ the critical angle $\alpha_c \simeq 1.169$. Summarizing the result, the optimal entropic preparation uncertainty relation in two dimensional Hilbert space can be divided into three regions, first [78]

$$H_b(A) + H_b(B) \geq 2H_b(\cos\alpha)\,, \qquad \alpha \in [0, \alpha_c]\,, \tag{5.17}$$

second, the range $\alpha \in (\alpha_c, \frac{\pi}{2})$ larger then the critical angle for which one has to resort to nu-

merical calculations and third $\alpha = \frac{\pi}{2}$ which is the limit as given by the Maassen-Uffink bound.

In general, the analysis of uncertainty relations depend on many circumstances and therefore generalizations of system specific results are only partially possible. General optimization for other dimensions is fairly difficult. Recently, Abbott et. al [79] provided a full characterization of a tight state-independent uncertainty relations for qubits, which will be briefly excerpted in the following paragraphs.

Consider a matrix $\hat{M}$ which has the Bloch vectors from $\hat{A} = \mathbf{a} \cdot \boldsymbol{\sigma}$ and $\hat{B} = \mathbf{b} \cdot \boldsymbol{\sigma}$ as rows

$$\hat{M} = \begin{pmatrix} \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} , \tag{5.18}$$

then it is possible to write the components of the inner products as

$$\mathbf{u} := \hat{M}\mathbf{r} = \begin{pmatrix} \mathbf{a} \cdot \mathbf{r} \\ \mathbf{b} \cdot \mathbf{r} \end{pmatrix} = \begin{pmatrix} \langle \hat{A} \rangle \\ \langle \hat{B} \rangle \end{pmatrix} . \tag{5.19}$$

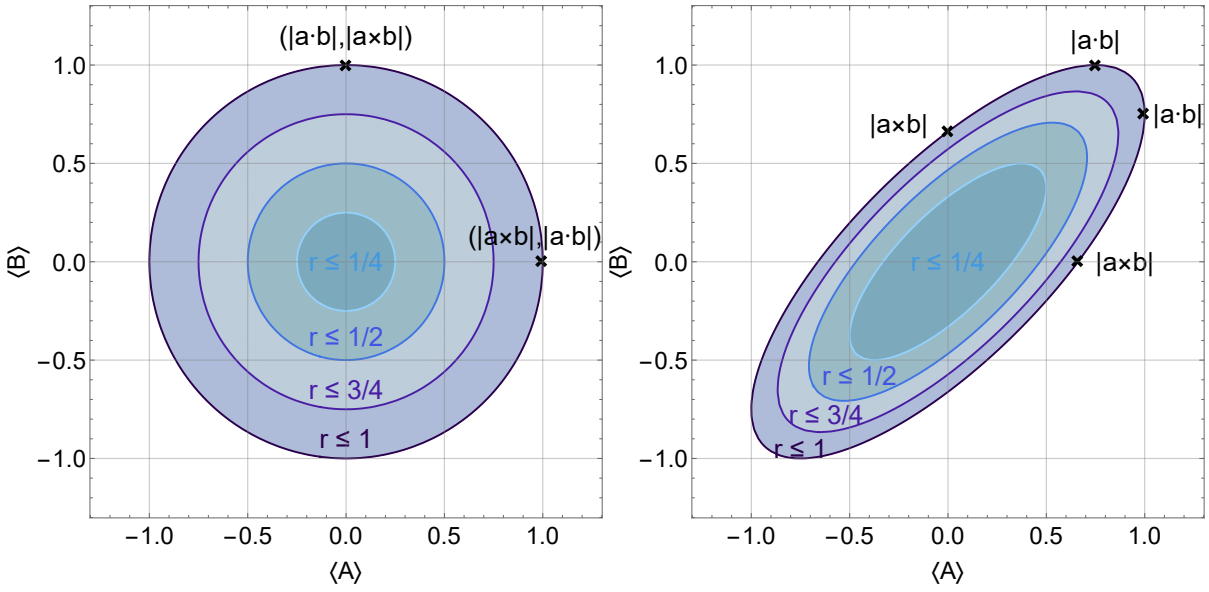
Because $\hat{M}$ is not a quadratic matrix it is not invertible, but as mentioned in sec. 3.1.1 it is possible to calculate a pseudoinverse. Geometrically $\hat{M}^+\hat{M}$ is an orthogonal projection onto the subspace spanned by $\{\mathbf{a}, \mathbf{b}\}$ (i.e. range of $\hat{M}^T$). Here is an important lemma required for further proceeding: if $\mathbf{r}$ is the Bloch vector of the quantum state $\hat{\rho}$ then the following inequality holds

$$\left|\hat{M}^+\mathbf{u}\right| = \left|\hat{M}^+\hat{M}\mathbf{r}\right| \leq |\mathbf{r}| \tag{5.20}$$

with equality given when $\mathbf{r} \in \text{span}\{\mathbf{a}, \mathbf{b}\}$. Choosing a particular basis for the Bloch vectors $\mathbf{a} = (\sqrt{\frac{1}{2}(1 + \mathbf{a} \cdot \mathbf{b})}, \sqrt{\frac{1}{2}(1 - \mathbf{a} \cdot \mathbf{b}}, 0)^T$ and $\mathbf{b} = (\sqrt{\frac{1}{2}(1 + \mathbf{a} \cdot \mathbf{b})}, -\sqrt{\frac{1}{2}(1 - \mathbf{a} \cdot \mathbf{b}}, 0)^T$ allows the determination of the pseudoinverse

$$\hat{M}^+ = \frac{1}{\sqrt{1 - (\mathbf{a} \cdot \mathbf{b})}} \begin{pmatrix} \sqrt{\frac{1+\mathbf{a}\cdot\mathbf{b}}{2}} & \sqrt{\frac{1-\mathbf{a}\cdot\mathbf{b}}{2}} \\ \sqrt{\frac{1-\mathbf{a}\cdot\mathbf{b}}{2}} & -\sqrt{\frac{1+\mathbf{a}\cdot\mathbf{b}}{2}} \\ 0 & 0 \end{pmatrix} . \tag{5.21}$$

Inserting this result into Eq. (5.20) provides a tight uncertainty relation for the expectation



(a) $\mathbf{a} \cdot \mathbf{b} = 0$. (b) $\mathbf{a} \cdot \mathbf{b} = \frac{3}{4}$.

Fig. 5.4: The uncertainty in Eq. (5.23) changes its form depending on the inner product $\mathbf{a} \cdot \mathbf{b}$ and its radius determined by $\|\mathbf{r}\|$. In the case of mutually unbiased states the expectation values are confined within a circle, whose radius decreases for mixed states. When the inner product becomes larger a deformation to an ellipse occurs.

values of the observables

$$\begin{aligned} |\langle A\rangle \mathbf{a} - \langle B\rangle \mathbf{b}|^2 &= \langle \hat{A}\rangle + \langle \hat{B}\rangle - 2(\mathbf{a}\cdot\mathbf{b})\,\langle \hat{A}\rangle\,\langle \hat{B}\rangle \\ &\leq \left(1-(\mathbf{a}\cdot\mathbf{b})^2\right)\|\mathbf{r}\|^2 \leq 1-(\mathbf{a}\cdot\mathbf{b})^2 = |\mathbf{a}\times\mathbf{b}|^2 \ . \end{aligned} \tag{5.22}$$

The quadric in Eq. (5.22) is of the form $x^2 - Axy + y^2 + B = 0$ and therefore describes some conic section. To find out which, the following transformations $\langle \hat{A}\rangle = \frac{1}{\sqrt{2}}(\langle \hat{A}'\rangle - \langle \hat{B}\rangle')$ and $\langle \hat{A}\rangle = \frac{1}{\sqrt{2}}(\langle \hat{A}'\rangle + \langle \hat{B}'\rangle)$ are employed, which gives

$$\frac{\langle \hat{A}'\rangle^2}{(1+\mathbf{a}\cdot\mathbf{b})} + \frac{\langle \hat{B}'\rangle^2}{(1-\mathbf{a}\cdot\mathbf{b})} \leq \|\mathbf{r}\|^2 \ . \tag{5.23}$$

So the uncertainty relation in Eq. (5.22) describes an ellipse where the radius is given by the degree of mixture $\|\mathbf{r}\|$ and the inner product $\mathbf{a}\cdot\mathbf{b}$ controls the lengths of major and minor axes. Because the ellipse is only canonical in the primed coordinates, it will be skewed for the original expectations values. A plot of the situation is given in Fig. 5.4 for two different choices of scalar

products.

Because, the square of the Pauli matrices is given by the identity operator, it is possible to reformulate Eq. (5.22) in terms of standard deviations. So, for or all observables with $\hat{A}^2 = \mathbb{1}$ the expectation value is $\langle \hat{A} \rangle = \pm\sqrt{1 - \sigma(\hat{A})^2}$ and consequently the following tight preparation uncertainty relation is obtained

$$\sigma(A)^2 + \sigma(B)^2 + 2\,|\mathbf{a}\cdot\mathbf{b}|\,\sqrt{1-\sigma(A)^2}\sqrt{1-\sigma(B)^2} \geq 2 - \left(1 - (\mathbf{a}\cdot\mathbf{b})^2\right)\|\mathbf{r}\|\,, \tag{5.24}$$

which is only saturated when $\mathbf{r}$ is in the span of $\mathbf{a}$-$\mathbf{b}$ and $(\mathbf{a}\cdot\mathbf{b})\,\langle \hat{A} \rangle\,\langle \hat{B} \rangle \geq 0$.

The article elaborates further how to derive an inequality for more than 2 observables, e.g. the well-known relation $\langle \hat{A} \rangle + \langle \hat{B} \rangle + \langle \hat{C} \rangle \leq \|\mathbf{r}\|$. However, for the continuation of this thesis all required relationships are summed up so that the topic of preparation uncertainty relations can be completed. Of course, there are many more and significant developments on this subject and a plethora of other quantities or viewpoints that can be exploited to quantify the respective trade-offs. Overviews on entropic uncertainty relations and corresponding developments are provided in [29, 80].

CHAPTER 6

Operational error-disturbance

The measurement uncertainty is about the inevitable intrusion through the act of measurement and the concomitant limitation of the precision of conjugate observables. Decisive in this category is that a joint probability is determined in one shot, unlike in the determination of statistical dispersions consisting of uncorrelated single measurements. The main focus in this segment will be on Ozawa's theory based on operational rms-errors.

6.1 Theory of quantum operational error and disturbance

The noise in a measurement describes the difference between the output of a measurement apparatus and the 'true value', respectively the intended quantity to be measured. Denote the intended observable in a given state $\hat{\rho}$ as $\hat{A}$ and regard the actually made precise measurement of another observable $\hat{X}$, the meter observable in the same state, the difference between them is represented by the *noise operator*

$$\hat{N}(\hat{A}) = \hat{X} - \hat{A} \ . \tag{6.1}$$

The *root mean square* (*rms*) *noise* or *error* is the average noise defined as

$$\epsilon(\hat{A}) := \langle \hat{N}(\hat{A})^2 \rangle^{\frac{1}{2}} \ . \tag{6.2}$$

The question arises how the error $\epsilon(\hat{A})$ can be utilized to quantify the accuracy of a measurement. In this case the observables must be related to the measurement model of the apparatus as shown in Fig. 3.6. Let the quadruple $\langle \mathcal{K}, \hat{\xi}, \hat{U}, \hat{M} \rangle$ constitute an indirect measurement model, which describes the measurement process of apparatus $\mathbf{M}$, where $\mathcal{K}$ is the apparatus state space, $\hat{\xi}$ is the initial apparatus state, $\hat{U}$ is a unitary operator providing the interaction between object and apparatus system and $\hat{M}$ is the meter observable of the apparatus. Defining $\hat{\rho}$ as the initial object state, the error $\epsilon(\hat{A})$ of apparatus $\mathbf{M}$ for a general observable $\hat{A}$ is defined [21] as

$$\epsilon(\hat{A}) := \langle \hat{N}(\hat{A})^2 \rangle^{\frac{1}{2}} = \mathrm{Tr}\left(\left(\hat{U}(\mathbb{1} \otimes \hat{M})\hat{U}^\dagger - \hat{A} \otimes \mathbb{1} \right)^2 \hat{\rho} \otimes \hat{\xi} \right)^{\frac{1}{2}} . \tag{6.3}$$

This way the noise operator is related to the difference $\hat{N}(\hat{A}) = \hat{M}^{out} - \hat{A}^{in}$, which describes how accurate the value of the observable $\hat{A}^{in} = \hat{A} \otimes \mathbb{1}$ on the product state $\hat{\rho} \otimes \hat{\xi}$ before the measurement is transferred to the apparatus's actual observation $\hat{M}^{out} = \hat{U}^\dagger(\mathbb{1} \otimes \hat{M})U$. The definition can also be rewritten as a distance with the Hilbert-Schmidt norm

$$\epsilon(\hat{A}) := \|(\hat{M}^{out} - \hat{A}^{in})\sqrt{\hat{\rho} \otimes \hat{\xi}}\|_{HS} . \tag{6.4}$$

In this form the error in Eq. (6.3) has apparently the property to vanish when the experimental output equals the eigenvalues of $\hat{A} = \sum_i a_i |a_i\rangle \langle a_i|$ (neglecting degeneracy). Assume the probe state to be $\hat{\xi} = |\xi\rangle\langle\xi|$ then from $\hat{N}(\hat{A}) |a_i\rangle |\xi\rangle = (\hat{M}^{out} - a_i) |a_i\rangle |\xi\rangle$, it follows a precise measurement of $\hat{A}$ by the apparatus is made when

$$\epsilon(\hat{A}) = 0 \quad \Rightarrow \quad \hat{M}^{out} |a_i\rangle |\xi\rangle = a_i |a_i\rangle |\xi\rangle . \tag{6.5}$$

Expanding the square of expression Eq. (6.3) gives

$$\epsilon(\hat{A}) = \mathrm{Tr}\left(\left((\hat{M}^{out})^2 - \{\hat{M}^{out}, \hat{A}^{in}\} + (\hat{A}^{in})^2 \right) (\hat{\rho} \otimes \hat{\xi}) \right) . \tag{6.6}$$

Then, defining the partial trace of an arbitrary system observable $\hat{\mathcal{O}} \in \mathcal{H} \otimes \mathcal{K}$ over the apparatus state $\hat{\xi}$, given by

$$\hat{O}^{(n)} = \mathrm{Tr}_{\mathcal{K}}\left(\hat{\mathcal{O}}^{(n)}(\mathbb{1} \otimes \hat{\xi}) \right) , \tag{6.7}$$

shows that the root mean square noise of the measurement is determined by the first $\hat{O}$ and second moment $\hat{O}^{(2)}$ of operators (see sec. 2.2.2), because

$$\begin{aligned}\mathrm{Tr}(\hat{O}_A^{(2)}\hat{\rho}) &= \mathrm{Tr}\left((\hat{M}^{out})^2(\hat{\rho}\otimes\hat{\xi})\right) \ , \\ \mathrm{Tr}(\hat{O}_A\hat{A}\hat{\rho}) &= \mathrm{Tr}\left(\hat{M}^{out}\hat{A}^{in}(\hat{\rho}\otimes\hat{\xi})\right) \ , \\ \mathrm{Tr}(\hat{A}\hat{O}_A\hat{\rho}) &= \mathrm{Tr}\left(\hat{A}^{in}\hat{M}^{out}(\hat{\rho}\otimes\hat{\xi})\right)\end{aligned} \tag{6.8}$$

and the squared error of observable $\hat{A}$ in terms of the system state only with the dissected detector state can be denoted as

$$\epsilon(\hat{A})^2 = \mathrm{Tr}\left(\left(\hat{O}_A^{(2)} - \{\hat{O}_A, \hat{A}\} + \hat{A}^2\right)\hat{\rho}\right) \ . \tag{6.9}$$

It was elucidated that the initial state $\hat{\rho}$ of a system will in general change under the influence of a measurement to $\mathcal{I}(m)(\hat{\rho})$. For any bounded Hermitian operator $\hat{B}$ on $\mathcal{H}$ the probability after the first measurement are given by $\mathrm{tr}(\mathcal{I}(m)(\hat{\rho})\hat{B}) = \mathrm{tr}(\mathcal{I}^*(m)(\hat{B})\hat{\rho})$, where the dual transform $\mathcal{I}^*(\hat{B})$ is employed. The disturbance of a measurement observable for a given state is determined by the CPTP map that describes the non-selective operation of the type Eq. (3.47) of the apparatus. This non-selective state change and its dual transformation in the system plus apparatus picture are determined by

$$\begin{aligned}\mathcal{E}_{ns}(\hat{\rho}) &= \mathrm{Tr}_{\mathcal{K}}\left(\hat{U}(\hat{\rho}\otimes\hat{\xi})\hat{U}^\dagger\right) \ , \\ \mathcal{E}^*_{ns}(\hat{A}) &= \mathrm{Tr}_{\mathcal{K}}\left(\hat{U}(\hat{A}\otimes\mathbb{1})\hat{U}^\dagger(\mathbb{1}\otimes\hat{\xi})\right) \ .\end{aligned} \tag{6.10}$$

The average over all measurement outcomes of the non-selective map discards all information about the correlations between meter value and the eigenvalue of $\hat{B}$. A lot of the mathematical technicalities of the disturbance are analogous to the treatment of the noise operator. The root mean square (rms) disturbance $\eta(\hat{B})$ is defined by the disturbance operator $\hat{D} = \hat{B}^{out} - \hat{B}^{in}$ in the following way

$$\begin{aligned}\eta(\hat{B}) :&= \langle\hat{D}(\hat{B})^2\rangle^{\frac{1}{2}} = \langle\hat{B}^{out} - \hat{B}^{in}\rangle^{\frac{1}{2}} \ . \\ &= \mathrm{Tr}\left(\left(\hat{U}(\hat{B}\otimes\mathbb{1})\hat{U}^\dagger - \hat{B}\otimes\mathbb{1}\right)^2\hat{\rho}\otimes\hat{\xi}\right)^{\frac{1}{2}} \ .\end{aligned} \tag{6.11}$$

Here, the observables $\hat{B}^{in} = \hat{B}\otimes\mathbb{1}$ and $\hat{B}^{out} = \hat{U}^\dagger(\hat{B}\otimes\mathbb{1})\hat{U}$ are linked to the instrument of the noise measurement by $\mathrm{Tr}\left(\hat{B}^{out}(\hat{\rho}\otimes\hat{\xi})\right) = \mathrm{Tr}\left(\mathcal{I}^*(m)(\hat{B})\hat{\rho}\right)$. It can be demonstrated that $\eta(\hat{B})$

vanishes if and only if the estimation of $\hat{A}$ does not affect the subsequent measurement of $\hat{B}$. By the same methods all formulas can be used again

$$\begin{aligned}\eta(\hat{B})^2 &= \|(\hat{B}^{out} - \hat{B}^{in})\sqrt{\hat{\rho}\otimes\hat{\xi}}\|_{HS} \\ &= \mathrm{Tr}((\hat{B}^{out})^2(\hat{\rho}\otimes\hat{\xi})) - \mathrm{Tr}(\hat{B}^{out}(\hat{B}\otimes\mathbb{1})(\hat{\rho}\otimes\hat{\xi})) - \mathrm{Tr}((\hat{B}\otimes\mathbb{1})\hat{B}^{out}(\hat{\rho}\otimes\hat{\xi})) \\ &+ \mathrm{Tr}((\hat{B}^2\otimes\mathbb{1})(\hat{\rho}\otimes\hat{\xi})) \\ &= \mathrm{Tr}\left((\mathcal{E}^*_{ns}(\hat{B}^2) - \{\mathcal{E}^*_{ns}(\hat{B}),\hat{B}\} + \hat{B}^2)\hat{\rho}\right) \ .\end{aligned} \tag{6.12}$$

Notice, the dual map from the Schrödinger to Heisenberg picture is used to transfer the disturbance of the state system to the perturbed observable, which in terms of Kraus operators has a close analogy to Eq. (3.47) given by

$$\mathcal{E}^*_{ns}(\hat{O}^{(n)}) = \sum_k \hat{M}_k\hat{O}^{(n)}\hat{M}_k^\dagger \ . \tag{6.13}$$

Instead of the given $\epsilon(\hat{A})$ and $\eta(\hat{B})$ definitions, a different decomposition can be done that separates both error and disturbance into two terms, which are called *operator bias* and *fuzziness*, respectively [81, 82]. For example, in the case of error, Eq. (6.9) is equivalent to

$$\epsilon(\hat{A})^2 = \mathrm{Tr}\left(\left(\hat{O}^{(2)}_A - \hat{O}^2_A\right)\hat{\rho}\right) + \mathrm{Tr}\left(\left(\hat{O}_A - \hat{A}\right)^2\hat{\rho}\right) \ . \tag{6.14}$$

The first term is called fuzziness which characterizes the sharpness of the measurement observable. The fuzziness vanishes when observable $\hat{O}_A$ is a sharp projector for which the second moment is identical to the squared first moment, while it becomes maximal for trivial effects proportional to the identity [96]. The second quantity is the bias, which quantifies how well the operator $\hat{O}_A$ approximates $\hat{A}$ in a measurement. This form of separation of the error is more reminiscent of classical estimation theory, where *bias estimators* and *sample variances* appear as analogies to their quantum version.

6.1.1 Error - Disturbance uncertainty relations

In the process of quantifying Heisenbergs original formulation in terms of measurement errors, the use of a noise operator as introduced in sec. 6.1 appeared first for the special case of position and momentum by Arthurs and Kelly [17], which would later be generalized to arbitrary

observables by Arthurs and Goodman [18] who provided the error-error trade-off

$$\epsilon(\hat{A})\epsilon(\hat{B}) \geq C_{AB} \ . \tag{6.15}$$

The issue with Ineq. (6.15) is that it does not generally hold as it was deduced under the assumption of unbiased approximation. After about two decades, Ozawa achieved to come up with a universally valid formulation of the measurement uncertainty relation, which includes two additional terms. A closer look at the issue follows for the case of error and disturbance.

The idea behind Ozawa's derivation of the error-disturbance uncertainty relation is as follows; first, because the meter observable $\hat{M}$ and $\hat{B}$ are in distinct Hilbert spaces the commutator $[\hat{M}^{out}, \hat{B}^{out}] = 0$ commutes and accordingly

$$\begin{aligned} 0 &= [\hat{M}^{out}, \hat{B}^{out}] \\ &= [\hat{D}(\hat{B}) + \hat{B}^{in}, \hat{N}(\hat{A}) + \hat{A}^{in}] \\ &= [\hat{D}(\hat{B}), \hat{N}(\hat{A})] + [\hat{D}(\hat{B}), \hat{A}^{in}] + [\hat{B}^{in}, \hat{N}(\hat{A})] + [\hat{B}^{in}, \hat{A}^{in}] \, . \end{aligned} \tag{6.16}$$

This commutativity is a key idea of the measurability: while the observables $\hat{A}, \hat{B}$ do not commute, their approximations do. Building the expectation values, taking the modulus on the entire term and splitting the expression into subterms by applying the triangular inequality leads to

$$\begin{aligned} -[\hat{B}^{in}, \hat{A}^{in}] &= [\hat{D}(\hat{B}), \hat{N}(\hat{A})] + [\hat{D}(\hat{B}), \hat{A}^{in}] + [\hat{B}^{in}, \hat{N}(\hat{A})] \ , \\ \left|\langle [\hat{B}^{in}, \hat{A}^{in}] \rangle\right| &= \left|\langle [\hat{D}(\hat{B}), \hat{N}(\hat{A})] \rangle + \langle [\hat{D}(\hat{B}), \hat{A}^{in}] \rangle + \langle [\hat{B}^{in}, \hat{N}(\hat{A})] \rangle\right| \ , \\ \left|\langle [\hat{B}, \hat{A}] \rangle\right| &\leq \left|\langle [\hat{D}(\hat{B}), \hat{N}(\hat{A})] \rangle\right| + \left|\langle [\hat{D}(\hat{B}), \hat{A}^{in}] \rangle\right| + \left|\langle [\hat{B}^{in}, \hat{N}(\hat{A})] \rangle\right| \ . \end{aligned} \tag{6.17}$$

For further processing an estimation of the terms must be done. The rms value is always greater than or equal to the standard deviation

$$\epsilon(\hat{A}) \geq \sigma(\hat{N}(\hat{A})) \ , \qquad \eta(\hat{B}) \geq \sigma(\hat{D}(\hat{B})) \tag{6.18}$$

and accordingly Robertson's relation Ineq. (5.2) delivers the inequality

$$\epsilon(\hat{A})\eta(\hat{B}) \geq \sigma(\hat{N}(\hat{A}))\sigma(\hat{D}(\hat{B})) \geq \left|\frac{1}{2} \langle [\hat{D}(\hat{B}), \hat{N}(\hat{A})] \rangle\right| \ . \tag{6.19}$$

Thus, a very useful intermediate result is obtained, when the relation Eq. (6.17) is updated by the newest result

$$\epsilon(\hat{A})\eta(\hat{B}) + \frac{1}{2}\left(\left|\langle[\hat{D}(\hat{B}), \hat{A}^{in}]\rangle\right| + \left|\langle[\hat{B}^{in}, \hat{N}(\hat{A})]\rangle\right|\right) \geq \frac{1}{2}\left|\langle[\hat{B}, \hat{A}]\rangle\right| = C_{AB} \ . \tag{6.20}$$

The joint measurement of two observables whose uncertainty is $\epsilon(\hat{A})\eta(\hat{B}) \geq C_{AB}$[4] only holds when the correlation terms vanish. Making the same trick with the other terms on the left hand side

$$\begin{aligned} \epsilon(\hat{A})\sigma(\hat{B}) \geq \sigma(\hat{N}(\hat{A}))\sigma(\hat{B}) \geq \left|\frac{1}{2}\langle[\hat{B}^{in}, \hat{N}(\hat{A})]\rangle\right| \ , \\ \eta(\hat{B})\sigma(\hat{A}) \geq \sigma(\hat{D}(\hat{B}))\sigma(\hat{A}) \geq \left|\frac{1}{2}\langle[\hat{D}(\hat{B}), \hat{A}^{in}]\rangle\right| \ , \end{aligned} \tag{6.21}$$

the error-disturbance uncertainty relation (EDUR) by Ozawa, universally valid for arbitrary pairs of observables and for arbitrary generalized measurements, is derived [20] as

$$\epsilon(\hat{A})\eta(\hat{B}) + \epsilon(\hat{A})\sigma(\hat{B}) + \eta(\hat{B})\sigma(\hat{A}) \geq C_{AB} \ . \tag{6.22}$$

First Improvement

Although the inequality Ineq. (6.22) is valid in its form, the question arose under which circumstances does equality hold. The weak point in the previous derivation is the triple application of Robertson's inequality, since it only happens for limited situations that all three terms are simultaneously tight. Branciard [83] in 2013 came up with more stringent inequalities derived from two geometric lemmas.

By adopting the following notations

$$|a\rangle = \frac{\hat{A}^{in} - \langle\hat{A}\rangle}{\sigma(\hat{A})}\,|\psi, \xi\rangle \ , \qquad |b\rangle = \frac{\hat{B}^{in} - \langle\hat{B}\rangle}{\sigma(\hat{B})}\,|\psi, \xi\rangle \ , \tag{6.23}$$

$$|x\rangle = \frac{\hat{M}^{out} - \langle\hat{A}\rangle}{\sigma(\hat{A})}\,|\psi, \xi\rangle \ , \qquad |y\rangle = \frac{\hat{B}^{out} - \langle\hat{B}\rangle}{\sigma(\hat{B})}\,|\psi, \xi\rangle \ , \tag{6.24}$$

[4] A great dispute in the scientific community arose regarding this inequality as it is not clear whom it can be attributed to [21, 23]. According to Ozawas, this inequality is to be blamed on Heisenberg.

$$\hat{\mathbf{a}} = \begin{pmatrix} \Re(|a\rangle) \\ \Im(|a\rangle) \end{pmatrix}, \qquad \hat{\mathbf{b}} = \begin{pmatrix} \Im(|b\rangle) \\ -\Re(|b\rangle) \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} \Re(|x\rangle) \\ \Im(|x\rangle) \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} \Im(|y\rangle) \\ -\Re(|y\rangle) \end{pmatrix}, \tag{6.25}$$

it is possible to employ the following

1. Lemma: For any two orthogonal vectors $\mathbf{x}$ and $\mathbf{y}$ of the Euclidean space, the following inequality is satisfied

$$\|\hat{\mathbf{a}} - \mathbf{x}\|^2 + \|\hat{\mathbf{b}} - \mathbf{y}\|^2 + 2\|\hat{\mathbf{a}} - \mathbf{x}\|\|\hat{\mathbf{b}} - \mathbf{y}\|\sqrt{1 - (\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})^2} \geq (\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})^2 \,. \tag{6.26}$$

A thorough calculation

$$\begin{aligned} \|\hat{\mathbf{a}} - \mathbf{x}\|^2 &= (\langle x| - \langle a|)(|x\rangle - |a\rangle) = \left(\frac{\epsilon(\hat{A})}{\sigma(\hat{A})}\right)^2 , \\ \|\hat{\mathbf{b}} - \mathbf{y}\|^2 &= (\langle y| - \langle b|)(|y\rangle - |b\rangle) = \left(\frac{\eta(\hat{B})}{\sigma(\hat{B})}\right)^2 \end{aligned} \tag{6.27}$$

allows a strengthening of Ozawa's relation for the case of pure states to

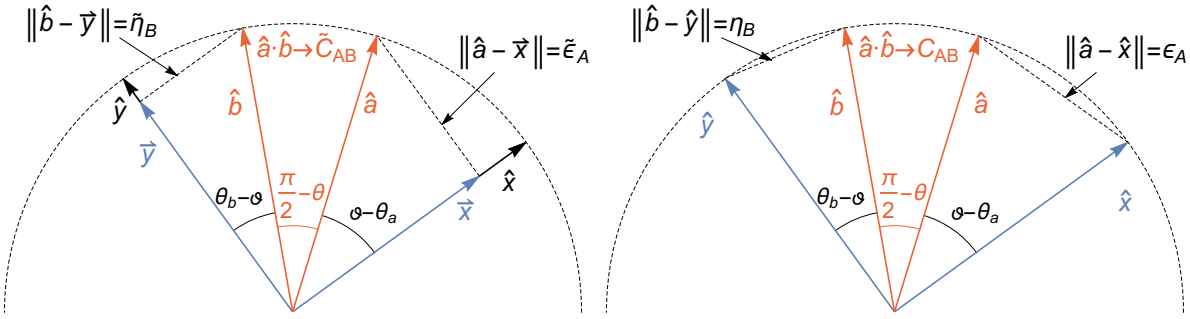
$$\epsilon(\hat{A})^2\sigma(\hat{B})^2 + \eta(\hat{B})^2\sigma(\hat{A})^2 + 2\epsilon(\hat{A})\eta(\hat{B})\sqrt{\sigma(\hat{A})^2\sigma(\hat{B})^2 - C_{\hat{A}\hat{B}}^2} \geq C_{\hat{A}\hat{B}}^2 \,. \tag{6.28}$$

A geometrical meaning of the lemma is supplied in Fig. 6.1(a) and as mentioned the inequality is in fact tight for any observables $\hat{A}, \hat{B}$ and system state $\hat{\rho} = |\psi\rangle\langle\psi|$ in the sense that equality is obtainable by measurements.

A special, nevertheless important sub-case concerns observables with $\hat{A}^2 = \hat{B}^2 = \mathbb{1}$, to which the Pauli observables belong to. Assuming that the meter observable has the same spectrum $\hat{M}^2 = \mathbb{1}$, it follows that also the approximate observables are $\hat{O}_A^2 = \hat{O}_B^2 = \mathbb{1}$. For states satisfying $\sigma(\hat{A}) = \sigma(\hat{B}) = 1$ it is possible to utilize

2. Lemma: For any two orthonormal vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ of the Euclidean space, the following inequality is satisfied for $\hat{\mathbf{a}}_\perp = \sqrt{1 - (\hat{\mathbf{a}} \cdot \hat{\mathbf{x}})^2}$ and $\hat{\mathbf{b}}_\perp = \sqrt{1 - (\hat{\mathbf{b}} \cdot \hat{\mathbf{y}})^2}$

$$(\hat{\mathbf{a}}_\perp)^2 + (\hat{\mathbf{b}}_\perp)^2 + 2\hat{\mathbf{a}}_\perp 2\hat{\mathbf{b}}_\perp\sqrt{1 - (\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})^2} \geq (\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})^2 \,. \tag{6.29}$$



(a) Lemma 1. (b) Lemma2.

Fig. 6.1: Geometric depiction of the employed lemmas. The vectors $\mathbf{x}$ and $\mathbf{y}$ correspond to the observables that are jointly measurable and which are supposed to approximate the observables being represented by $\mathbf{a}$ and $\mathbf{b}$. These pictures of error $\epsilon(\hat{A})$ and disturbance $\eta(\hat{B})$ make the quantities as operator difference very intelligible. Note that, in 2. Lemma it is necessary for all vectors to have unit length in order to satisfy the restriction of $\hat{A}^2 = \hat{B}^2 = \hat{O}_A^2 = \hat{O}_B^2 = \mathbb{1}$.

The geometric picture in Fig. 6.1(b) shows the arrangement forces all vectors to have unit length and relate to error and disturbance by

$$\begin{aligned}(\hat{\mathbf{a}}_\perp)^2 &= \|\hat{\mathbf{a}}-\mathbf{x}\|^2\left(1-\frac{\|\hat{\mathbf{a}}-\mathbf{x}\|^2}{4}\right) = \epsilon(\hat{A})^2\left(1-\frac{\epsilon(\hat{A})^2}{4}\right) := \tilde{\epsilon}(\hat{A})^2\,,\\ (\hat{\mathbf{b}}_\perp)^2 &= \|\hat{\mathbf{b}}-\mathbf{y}\|^2\left(1-\frac{\|\hat{\mathbf{b}}-\mathbf{y}\|^2}{4}\right) = \eta(\hat{B})^2\left(1-\frac{\eta(\hat{B})^2}{4}\right) := \tilde{\eta}(\hat{B})^2\,.\end{aligned} \tag{6.30}$$

Summing up: For $\hat{A}^2 = \hat{B}^2 = \hat{O}_A^2 = \hat{O}_B^2 = \mathbb{1}$ and $\sigma(\hat{A}) = \sigma(\hat{B}) = 1$ the improved inequality in Ineq. (6.28) simplifies further to

$$\tilde{\epsilon}(\hat{A})^2 + \tilde{\eta}(\hat{B})^2 + 2\tilde{\epsilon}(\hat{A})\tilde{\eta}(\hat{B})\sqrt{1-\tilde{C}_{AB}^2} \geq \tilde{C}_{AB}^2\,. \tag{6.31}$$

Second Improvement

In uncertainty relations, the case of mixed states is rarely treated. If uncertainty relations are designed to be state-independent the inaccuracies can only be a consequence of incompatible operators. The bound C_{AB} for a totally mixed state $\hat{\rho} = \mathbb{1}/d$ would lead to Eq. (5.10) and therefore vanish. To avoid the disappearance of the bound for error and disturbance Ozawa

suggested a modification [84].

The central idea relies on two modifications, first a purification similar to the Schmidt decomposition Eq. (3.26): in the *canonical purification* the ancilla Hilbert space is given by the dual space $\mathcal{H}_{anc} = \mathcal{H}^*$, which contains all *bra*-states. The purification in this form changes to

$$|\chi\rangle = \sum_i \sqrt{p_i}\, |\phi_i\rangle \otimes \langle \phi_i| \ , \tag{6.32}$$

which gives the original state back accordingly $\hat{\rho} = \mathrm{Tr}_{\mathcal{H}^*}(|\chi\rangle\langle\chi|)$. The second ingredient is changing the operators in the extended Hilbert space. First, define

$$\sqrt{\hat{\rho}}(-i[\hat{A},\hat{B}])\sqrt{\hat{\rho}} = \hat{U}_W^\dagger \left|\sqrt{\hat{\rho}}(-i[\hat{A},\hat{B}])\sqrt{\hat{\rho}}\right| \tag{6.33}$$

to be the polar decomposition as shown in Eq. (3.7), where $\hat{U}_W$ can be chosen as a self-adjoint unitary operator on $\mathcal{H}$ (likewise for its dual $\hat{U}_W \in \mathcal{H}^*$). The observables on the extended Hilbert space $\mathcal{H} \otimes \mathcal{H}^*$ now change to

$$\begin{aligned} \hat{A}' &= \hat{A} \otimes I_{\mathcal{H}^*} \ , \\ \hat{U}' &= \hat{U} \otimes I_{\mathcal{H}^*} \ , \\ \hat{B}'_W &= \left(\hat{B} - \langle \hat{B} \rangle\, I_{\mathcal{H}}\right) \otimes \hat{U}_W^* \ . \end{aligned} \tag{6.34}$$

Physically, there is no change in the observable $\hat{A}'$ and the unitary operator in the measurement model $\langle \mathcal{K}, \hat{\xi}, \hat{U}', \hat{M}\rangle$. The disturbance of the new observable $\hat{B}'_W$ is however different as it now interacts with the extended dual state $\langle\phi_i|$. The shift of the operator $\hat{A} \to \hat{A}'$ does not effect the standard deviation $\sigma(\hat{A}') = \sigma(\hat{A})$ nor the error $\epsilon(\hat{A}') = \epsilon(\hat{A})$. For $\hat{B} \to \hat{B}'_W$ the standard deviation becomes smaller $\sigma(\hat{B}'_W) \leq \sigma(\hat{B})$, while disturbance remains as it is $\eta(\hat{B}'_W) = \eta(\hat{B})$. Finally, a closer look at the bound $C_{\hat{A}'\hat{B}'_W}$ unveils the following

$$\begin{aligned} 2iC_{\hat{A}'\hat{B}'_W} &= \langle\chi|[\hat{A}', \hat{B}'_W]|\chi\rangle = \langle\chi|[\hat{A}\otimes I_{\mathcal{H}^*}, \hat{B}\otimes \hat{U}_W^*]|\chi\rangle \\ &= \langle\chi|[\hat{A},\hat{B}]|\chi\rangle \otimes \hat{U}_W^* \\ &= \sum_{i,j} \sqrt{p_i}\sqrt{p_j}\, \langle\phi_i|[\hat{A},\hat{B}]|\phi_j\rangle\, \langle\phi_j|\hat{U}_W|\phi_i\rangle \\ &= \mathrm{Tr}\left(\hat{U}_W\sqrt{\hat{\rho}}[\hat{A},\hat{B}]\sqrt{\hat{\rho}}\right) = i\mathrm{Tr}\left(\left|\sqrt{\hat{\rho}}[\hat{A},\hat{B}]\sqrt{\hat{\rho}}\right|\right) \\ &:= 2iD_{AB} \ . \end{aligned} \tag{6.35}$$

That is, the bound can be replaced with a new quantity given by the expectation value of the modulus of the operator in Eq. (6.33). Accordingly, the results can be used to reformulate the previous relations Ineq. (6.28) and Ineq. (6.31) to

$$\epsilon(\hat{A})^2\sigma(\hat{B})^2 + \eta(\hat{B})^2\sigma(\hat{A})^2 + 2\epsilon(\hat{A})\eta(\hat{B})\sqrt{\sigma(\hat{A})^2\sigma(\hat{B})^2 - D_{AB}^2} \geq D_{AB}^2 \ , \tag{6.36}$$

and

$$\tilde{\epsilon}(\hat{A})^2 + \tilde{\epsilon}(\hat{B})^2 + 2\tilde{\epsilon}(\hat{A})\tilde{\epsilon}(\hat{B})\sqrt{1 - D_{AB}^2} \geq D_{AB}^2 \ . \tag{6.37}$$

In comparison to C_{AB} the new bound

$$D_{AB} = \frac{1}{2}\mathrm{Tr}\left(\left|\sqrt{\hat{\rho}}[\hat{A},\hat{B}]\sqrt{\hat{\rho}}\right|\right) \tag{6.38}$$

has the property to become 1 for totally mixed states when $[\hat{A},\hat{B}]$ is traceless. This quantity can also be used to replace the boundary on the right-hand side of the Robertson relation [85]. Also, if $\hat{\sigma} = \frac{1}{|\alpha|^2}[\hat{A},\hat{B}]^\dagger\hat{\rho}[\hat{A},\hat{B}]$ is a quantum state, where the constant α is determined by the commutator $[\hat{A},\hat{B}] = \alpha\hat{C}$, then there a nice interpretation for D_{AB}, namely

$$\begin{aligned} D_{AB} &= \frac{1}{2}\mathrm{Tr}\left(\sqrt{\sqrt{\hat{\rho}}[\hat{A},\hat{B}]^\dagger\hat{\rho}[\hat{A},\hat{B}]\sqrt{\hat{\rho}}}\right) \\ &= \frac{|\alpha|}{2}\mathrm{Tr}\left(\sqrt{\sqrt{\hat{\rho}}\,\hat{\sigma}\sqrt{\hat{\rho}}}\right) \\ &= \frac{|\alpha|}{2}F(\hat{\rho},\hat{\sigma}) \ . \end{aligned} \tag{6.39}$$

Thus, if the action of the commutator is proportional to some quantum operation, then D_{AB} has the meaning of quantum state fidelity as in Eq. (3.48).

6.2 Measurement scheme of error and disturbance

An issue appeared at first regarding the accessibility of errors and disturbances in practical experiments, since physically the expectation value of $\{\hat{A},\hat{O}_A\}$ in Eq. (6.9) implies a joint measurement of two non-commuting observables of the form $\mathrm{Tr}(\hat{A}\hat{O}_A\hat{\rho})$. In order to overcome this obstacle the *weak measurement method* and the *three-state method* were proposed.

6.2.1 Weak measurement method

In recent years, the concept of *weak measurements* has gained much attention in the quantum physics community. In such weak measurements, the mean value of the quantum mechanical probability distribution is measured without causing excessive disturbances on the quantum state, so that a (partial) characterization of the wave function is possible after completion of the measurements.

In 1988 Aharonov et al. [86] introduced the *weak value* $\langle \hat{A} \rangle_w$ of observable $\hat{A}$: in its most general form, assuming $\mathrm{Tr}(\hat{\Pi}_f \hat{\rho}_i) \neq 0$, it is written as

$$\langle \hat{A} \rangle_w = \frac{\mathrm{Tr}(\hat{\Pi}_f \hat{A} \hat{\rho}_i)}{\mathrm{Tr}(\hat{\Pi}_f \hat{\rho}_i)} \,. \tag{6.40}$$

Here, $\hat{\rho}_i$ is the initial quantum state of the system and $\hat{\Pi}_f$ is the post-selection observable. If $\hat{\rho}_i = |\psi_i\rangle\langle\psi_i|$ is a pure state and $\hat{\Pi}_f = |\psi_f\rangle\langle\psi_f|$ is a discrete projector, then the originally published form $\langle \hat{A} \rangle_w = \langle \psi_f | \hat{A} | \psi_i \rangle \,/\, \langle \psi_f | \psi_i \rangle$ is obtained. The weak measurement is a special kind of joint measurement, usually extracted by preparing a particular initial state, coupling the quantum system weakly to the apparatus and then making a specific (succeeding) post-selection. Assume the special case where the projectors of the spectral decomposition $\hat{A} = \sum_m a_m \hat{P}_m$ commutes with the post-selection operator $[\hat{\Pi}_f, \hat{P}_m] = 0$ for all outputs m, then from the identification $p(m|f) = \mathrm{Tr}(\hat{\Pi}_f \hat{P}_m \hat{\rho}_i)$ it follows that the weak value is identical with a conditional expectation

$$\langle \hat{A} \rangle_w = \sum_m a_m \frac{p(m|f)}{p(f)} \,. \tag{6.41}$$

However, observable $\hat{A}$ will in general not commute with the post-selection effect, which especially implies that Born's rule Eq. (3.34) does not generally yield a true probability. However, if *quasi-probabilities* p_{wv} with negative outcomes are admissible then the weak-valued probability, conditioned on an initial state and a final measurement, is defined by the real part of the weak-valued projector $\hat{P}_m$ [65]

$$p_{wv}(m|f) = \Re \langle \hat{P}_m \rangle_w = \Re \frac{\mathrm{Tr}(\hat{\Pi}_f \hat{P}_m \hat{\rho}_i)}{\mathrm{Tr}(\hat{\Pi}_f \hat{\rho}_i)} \,, \tag{6.42}$$

which is connected to the joint quasi-probability by a Bayesian kind of rule

$$p_{wv}(m,f) = p_{wv}(m|f)p(f) = \Re\mathrm{Tr}(\hat{\Pi}_f \hat{P}_m \hat{\rho}_i) \ . \tag{6.43}$$

Among a variety of new applications, weak-valued probabilities have also been linked to the formal definitions of error-disturbance Eq. (6.3), Eq. (6.11) and hence to the uncertainty principle. The connection is quickly confirmed for sharp observables, e.g.

$$\begin{aligned}
\epsilon(\hat{A})^2 &= \mathrm{Tr}\left(\left(\hat{U}(\mathbb{1}\otimes\hat{M})\hat{U}^\dagger - \hat{A}\otimes\mathbb{1}\right)^2 \hat{\rho}\otimes\hat{\xi}\right) \\
&= \mathrm{Tr}\left(\left(\hat{U}(\mathbb{1}\otimes\sum_m m\hat{P}_m)\hat{U}^\dagger - \sum_a a\hat{P}_a\otimes\mathbb{1}\right)^2 \hat{\rho}\otimes\hat{\xi}\right) \\
&= \sum_{a,m}(m-a)^2\ \Re\mathrm{Tr}\left(\left(\hat{U}(\mathbb{1}\otimes\hat{P}_m)\hat{U}^\dagger(\hat{P}_a\otimes\mathbb{1})\right)\hat{\rho}\otimes\hat{\xi}\right) \\
&= \sum_{a,m}(m-a)^2 p_{wv}(m,a)
\end{aligned} \tag{6.44}$$

where the resemblance to Eq. (6.43) can be increased by denoting $\hat{\Pi}'_m = \hat{U}(\mathbb{1}\otimes\hat{P}_m)\hat{U}^\dagger$, $\hat{P}'_a = \hat{P}_a\otimes\mathbb{1}$ and $\hat{\rho}'_i = \hat{\rho}\otimes\hat{\xi}$, which returns the intended form $p_{wv}(m,a) = \Re\mathrm{Tr}(\hat{\Pi}'_m\hat{P}'_a\hat{\rho}'_i)$. The error $\epsilon(\hat{A})$ can hence be determined by a weak measurement [87] of $\hat{P}_a$ on the initial quantum state, after which a strong post-selective measurement takes place to read out the final meter observable $\hat{M}$. As before the disturbance can also be interpreted as the weak valued quasi-probability distribution between the initially intended and finally measured observables after disturbance

$$\eta(\hat{B})^2 = \sum_{b,b'}(b-b')^2 p_{wv}(b,b') \ . \tag{6.45}$$

Due to the relationship between error/disturbance and weak measurements, implementation of experiments for testing the weak valued formulation have also been carried out for two-dimensional quantum systems. These set-ups have in common that they are based on a quantum circuit suggested by Lund and Wiseman [88], who proposed to measure the joint quasi-probabilities in a sequential set-up as illustrated in Fig. 6.2. Experimental realizations of the proposal would be difficult with neutron optical set-ups available at the moment, but there are photonic tests of the weakly-measured error-disturbance [89–91]. A thorough calculation of the quantum circuit suggested by Lund is given in [92] that shows a mismatch between the joint distribution and the actual rms-deviation.

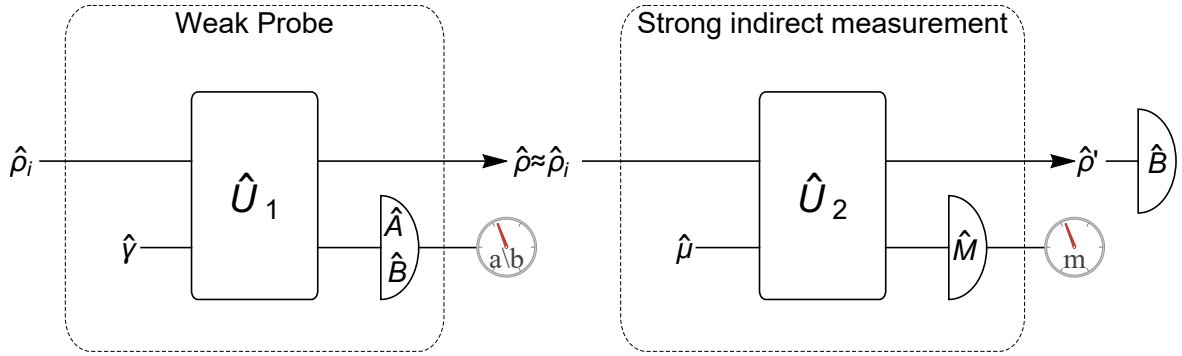


Fig. 6.2: Schematic diagram of sequential set-up to measure the weak-valued probabilities p_{wv}. The idea is to make a weak measurement of $\hat{A}$ to obtain the value a (or $[\hat{B}]$ and b), such that it leaves the initial state $\hat{\rho}_i$ almost undisturbed and then make an indirect strong measurement with meter output m, immediately followed by a strong $\hat{B}$ determination on the perturbed state $\hat{\rho}'$ with outcome b'. The read-out of the weak probe and the meter or final third measurement give joint probabilities which approximately reconstruct the quasi-probability distributions.

6.2.2 Tomographic three-state method

Here, experiments at the research reactor facility TRIGA Mark II of the Vienna University of Technology that have been made over the years in the field of error-disturbance tests in the case of two-dimensional quantum systems are described. In the experiments [36, 93, 94] the set-up is of the form illustrated in Fig. 6.3. A thermal monochromatic neutron beam with a mean wavelength of 1,99Å and a monochromaticity of $\Delta\lambda/\lambda \cong 0.02$ propagates in the y-direction. The neutrons that emerge from the reactor are mixed spin states which are polarized after crossing a bent Co-Ti supermirror resulting all states to enter the set-up in spin state $|+z\rangle$. The entire configuration from the first polarizer to analyzer-2 is enclosed by a Helmholtz coil, which generate two magnetic guide fields in $+z$ - direction of the magnitude of roughly $B_g \cong 13G$. This external magnetic field causes a Zeemann split of the potential energy where the population in the respective levels can be changed by applying magnetic field perpendicular to the y-z plane.

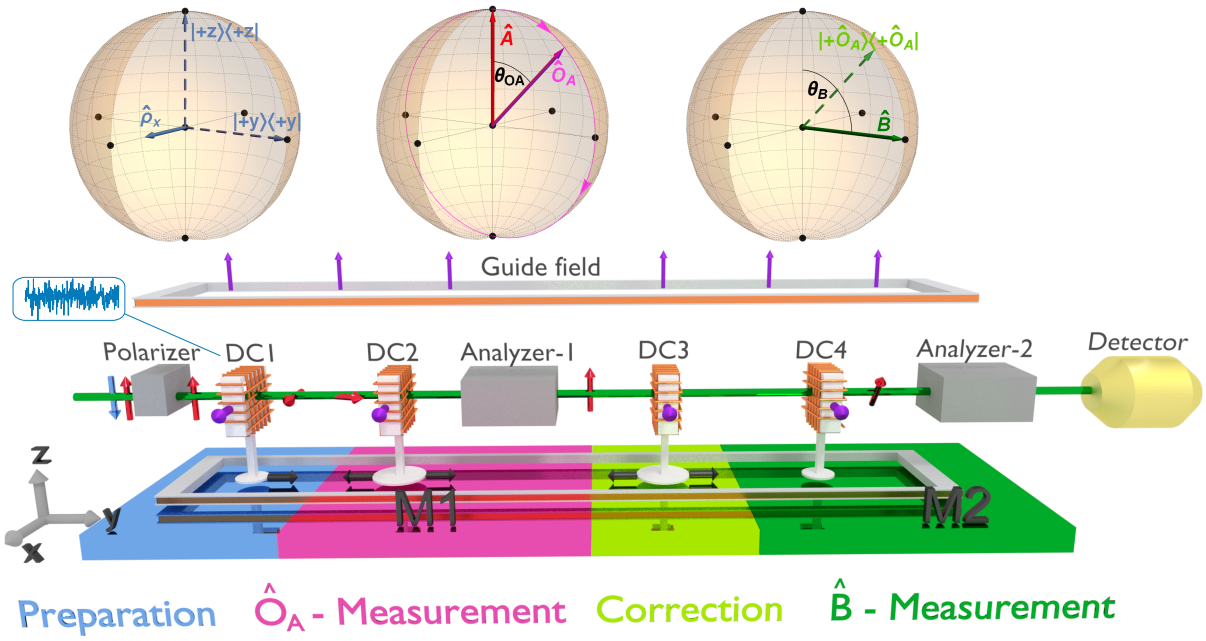


Fig. 6.3: Illustration of the experimental set-up together with Bloch sphere description. The neutron-polarimeter set-up consists of three stages. (1) Preparation (blue region): the incoming mono-energetic neutron beam is polarized in $+z$-direction by passing through a polarizing supermirror. In coil DC1 the required directions of the input states are generated. In case the coil is fed with a noise current an ensemble mixed in time is generated controlled via a tunable noisy magnetic field. (2) Apparatus M1, consisting of a projective $\hat{O}_A$ measurement (pink region) and a correction operation (light green region): analyzer-1 together with the coils DC2/3 carry out the first measurement and allow for a unitary rotation of the output state of the $\hat{O}_A$ measurement. (3) Apparatus M2, measuring $\hat{B}$ (dark green region): the second measurement fixes $\hat{B}$ and is carried out by the the coil DC4 and the analyzer-2. Three Bloch-spheres are depicted: indicated on the left is the input states to be sent to the apparatus M1 and M2, shown in the middle are the directions of the $\hat{A}$ (red) and $\hat{O}_A$ (pink) measurements and exhibited on the right are the directions of the $\hat{B}$ measurement in the apparatus M2 (dark green) as well as the state after the correction (light green).

As shown in Fig. 6.3, the experimental set-up is divided into three larger stages: 1. state preparation (blue), 2. $\hat{O}_A$ measurement (pink), and 3. $\hat{B}$ measurement (light and dark green).

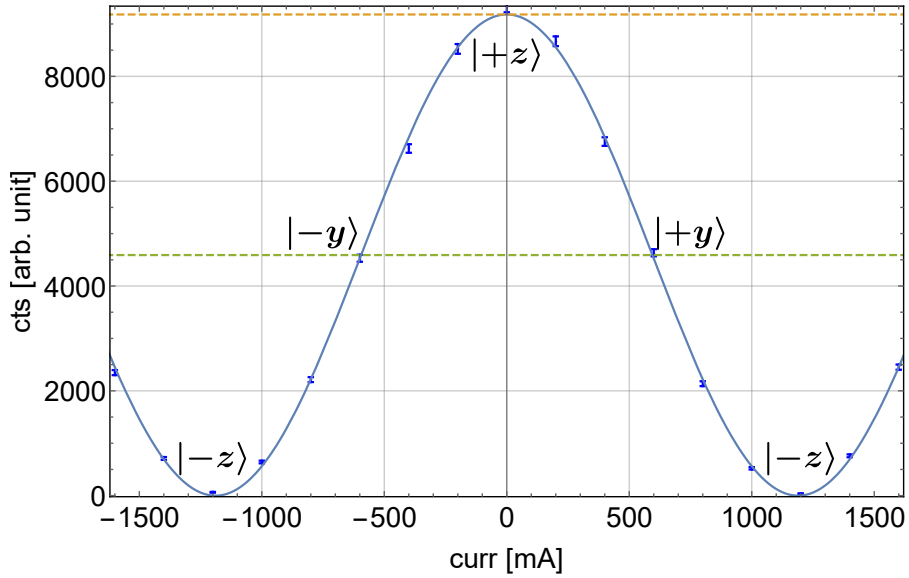


Fig. 6.4: Illustration of a well aligned DC-coil which shows the oscillation of the neutron count rate between the up and downs spin states. At 0 current all states are in the $|+z\rangle$ state, while at $\cong \pm 595$ mA at half the intensity the states correspond to $|\mp y\rangle$, Finally for currents twice as strong $\cong \pm 1190$ mA the spin is flipped to $|-z\rangle$. The contrast of such an oscillation can be up to 0.99 and the phase misalignment below 1°.

1. State preparation: In the first stage the initial $|+z\rangle$ up spin of the neutron is changed by static magnetic field in x-direction, whose unitary evolution $\hat{U}_{DC1}$ is given in sec. 4.4 and typical diagram of a polarization oscillation is shown in Fig. 6.4. The coils are solenoids fed with a direct current in two axes, one that produces a field in the negative z-direction B_z which compensates the guide field $B_g \mathbf{e}_z = -B_z \mathbf{e}_z$ and a second field B_x that induces a the unitary transformation. The current in the coil changes the magnetic field strength and hence the polarization of the neutron's magnetic moment. By checking the transmission probability through a supermirror analyzer $p = |\langle +z|\hat{U}_{DC1}|+z\rangle|^2$ the rotation angle can be controlled by the coil current. It should be noted that all four coils involved are adjusted in exactly the same way. Their performances, i.e. contrast and phase-stability of oscillations are virtually identical.

From the Bloch sphere representation it is clear that coils change the polar angle of the qubit state. When a neutron spin immediately after the coil re-enters the guide field it will Larmor precess about the vertical z-direction by some azimuthal angle. Depending on the time of flight

t_{12} between DC1 and DC2 for instance, it is possible to tune the incident angle of the spin to second coil: if $|+y\rangle$ is the state after DC1 then from $\exp(i\phi\hat{\sigma}_z)\,|+y\rangle$ the azimuthal angle $\phi = \mu|B_g|t_{12}/\hbar$ is determined by the neutron's propagation time between the coils. Since the neutron velocity is constant, the angle is controlled by the relative coil distance. Once more, this concept works also for DC2, DC3 and DC4.

2. $\hat{O}_A$ **measurement:** The second stage consists of the apparatus $M1$ defined as DC2 + analyzer-1, and the unitary operation $\hat{U}$ carried out by DC3. The apparatus $M1$ makes a measurement of

$$\hat{O}_A = \cos\theta_{OA}\hat{\sigma}_z + \sin\theta_{OA}\hat{\sigma}_y = \sum_{m=\pm1} m\hat{P}_{OA}(m)\ . \tag{6.46}$$

The (Kraus) measurement operators of $M1$ are

$$\{\hat{M}_m\} = \{|\hat{\sigma}_z = m\rangle\,\langle\hat{O}_A = m|\} \tag{6.47}$$

and thus, the output state after the measurement in apparatus $M1$ can be written as

$$\hat{\rho}' = \sum_m |\hat{\sigma}_z = m\rangle\langle\hat{O}_A = m|\hat{\rho}|\hat{O}_A = m\rangle\langle\hat{\sigma}_z = m|\ . \tag{6.48}$$

The whole process consisting of DC2 + analyzer-1 + DC3 can be considered as a measuring apparatus for measuring $\hat{O}_A$ with the measurement operators

$$\{\hat{M}_m\} = \{\hat{U}_{DC3}\,|\hat{\sigma}_z = m\rangle\langle\hat{O}_A = m|\} \tag{6.49}$$

and the POVM

$$\hat{P}_{OA}(m) = \hat{M}_m^\dagger\hat{M}_m\,. \tag{6.50}$$

Let $\hat{A} = \hat{\sigma}_z$, then evaluation of the error Eq. (6.9) for *dichotomic spin observables* (i.e. they have only two outcomes) reveals

$$\{\hat{A}, \hat{O}_A\} = 2\cos\theta_{OA}\mathbb{1} \tag{6.51}$$

and remembering that the square of the Pauli matrices are $\hat{A}^2 = \hat{O}_A^2 = \mathbb{1}$ the error in Eq. (6.9) is expected to become

$$\epsilon(\hat{A}) = \sqrt{2 - 2\cos\theta_{OA}} = 2\left|\sin\left(\frac{\theta_{OA}}{2}\right)\right|\ . \tag{6.52}$$

3. $\hat{B}$ **measurement:** In a successive measurement the determination of $\hat{O}_A$ causes the disturbance on the measurement of the following observable $\hat{B}$. Most of the results are derived analogously to the error case. The whole process consisting of DC4 + analyzer-2 can be considered as a measuring apparatus for measuring $\hat{B}$ with outcome $b = \pm 1$ and measurement operators

$$\{\hat{M}_{m2}\} = \{|\hat{\sigma}_y = b\rangle \langle \hat{\sigma}_z = m| \hat{U}^\dagger_{DC3}\} . \tag{6.53}$$

Recall, that the measurement is projective measurement of $\hat{O}_A$ if and only if

$$\hat{U}_{DC3} |\hat{\sigma}_z = m\rangle = |\hat{O}_A = m\rangle \ . \tag{6.54}$$

In the first publications [36, 93], the coil DC3 was used in order to make DC2+analyzer-1 + DC3 a projective measurement of $\hat{O}_A$. Let $\hat{B} = \hat{\sigma}_y$, thus the disturbance is given by

$$\begin{aligned}
\hat{O}_B &= \sum_m \hat{M}^\dagger_m \hat{B} \hat{M}_m \\
&= \sum_{m=\pm 1} |\hat{O}_A = m\rangle \langle \hat{O}_A = m|\hat{B}|\hat{O}_A = m\rangle \langle \hat{O}_A = m| \\
&= \sin\theta_{OA} \hat{O}_A \\
\hat{B}\hat{O}_B &= \sin\theta_{OA} \hat{B}\hat{O}_A = \sin\theta_{OA} \left(\hat{\sigma}_y \left(\cos\theta_{OA}\hat{\sigma}_z + \sin\theta_{OA}\hat{\sigma}_y\right)\right) \\
&= \sin^2\theta_{OA} \mathbb{1} + i \sin\theta_{OA}\cos\theta_{OA}\hat{\sigma}_x \\
\eta(\hat{B})^2 &= 2 - 2\Re(\hat{B}\hat{O}_B) = 2 - 2\sin^2\theta_{OA} = 2\cos^2\theta_{OA} \ .
\end{aligned} \tag{6.55}$$

The theoretical prediction of disturbance $\eta(\hat{B})$ for a projective measurement of $\hat{O}_A$ is therefore

$$\eta(\hat{B}) = \sqrt{2}\,|\cos\theta_{OA}| \ . \tag{6.56}$$

Extraction of error and disturbance from data

The three state method is based on the decomposition of the anti-commutator into three parts

$$\{\hat{A}, \hat{O}_A\} = 4\hat{P}^+_A \hat{O}_A \hat{P}^+_A - \hat{A}\hat{O}_A\hat{A} - \hat{O}_A \ , \tag{6.57}$$

where $\hat{P}^+_A = \frac{(\mathbb{1}+\hat{A})}{2}$ is the two-dimensional projector belonging to the positive eigenvalue of $\hat{A}$ [5]. Utilizing this formula gives the algebraically equivalent expression

[5] There also exists a two state method which uses a combination of $\hat{P}^+_A$ and $\hat{P}^-_A$ for a split into two parts.

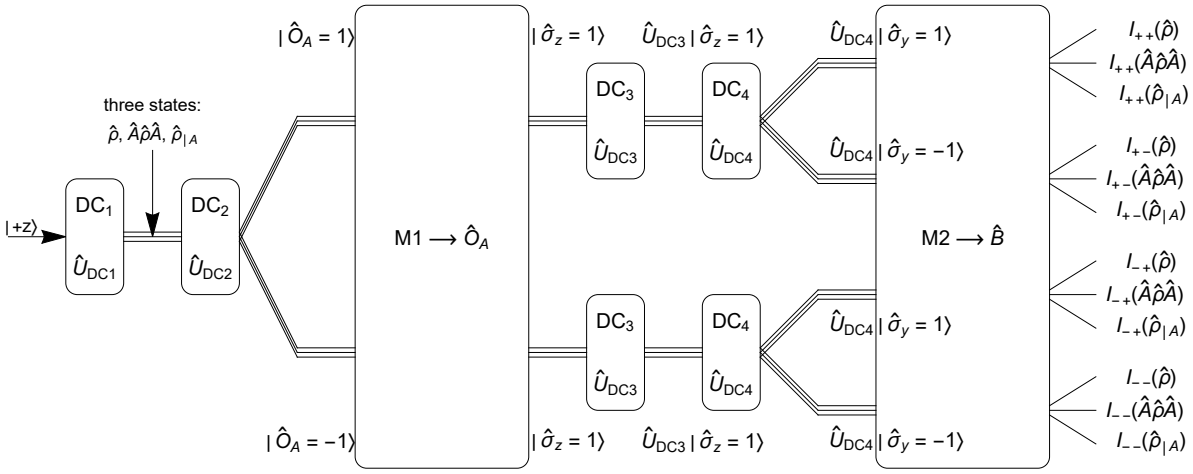


Fig. 6.5: Diagram of measurement procedure. In the first operation $\hat{U}_{DC1}$ generates from the initial state $|+z\rangle$ in the case $\hat{A} = \sigma_z$ the three states $\{|+x\rangle, |-x\rangle$ and $|+z\rangle\}$. The second operation is the transformation to $|\hat{O}_A = \pm 1\rangle$ by $\hat{U}_{DC2}$ which pass the first analyzer with probability $p = |\langle \hat{O}_A = \pm 1 | \hat{\sigma}_z = 1\rangle|^2$. Then $|\hat{\sigma}_z = 1\rangle$ is transformed to $\hat{U}_{DC3}|\sigma_z = 1\rangle = \hat{U}_{DC3}|\sigma_z = m\rangle$ depending on the state from the $\hat{O}_A$ measurement. The last coil DC4 has the same purpose as the second coil and controls the transition probability $p = \left|\langle \sigma_z = b | \hat{U}_{DC4}\hat{U}_{DC3} | \sigma_z = m\rangle\right|^2$ through the second analyzer. At the end for every of the three input states 4 different count-rates are obtained, denoted by $I_{m,b}$ where $\{m, b\} = \{-1, 1\}$.

$$\epsilon(\hat{A})^2 = 2 + \mathrm{Tr}(\hat{O}_A\hat{\rho}) + \mathrm{Tr}(\hat{O}_A\,\hat{A}\hat{\rho}\hat{A}) - 4\,\mathrm{Tr}(P_A^+\hat{\rho})\,\mathrm{Tr}(\hat{O}_A\,\hat{\rho}_{|A}) \ . \tag{6.58}$$

which allows to access the error statistically by measurements of $\hat{O}_A$ with three different states $\{\hat{\rho}, \hat{A}\hat{\rho}\hat{A}, \hat{\rho}_{|A} = \hat{P}_A^+\hat{\rho}\hat{P}_A^+/\mathrm{Tr}(\hat{P}_A^+\hat{\rho})\}$. Furthermore, this implies that in the state preparation step three different states must be generated with $\hat{U}_{DC1}$. Physically, the disturbance can be measured with the three-state method as well, i.e.

$$\eta(B)^2 = 2 + \mathrm{Tr}(\hat{O}_B\hat{\rho}) + \mathrm{Tr}(\hat{O}_B\,\hat{B}\rho\hat{B}) - 4\,\mathrm{Tr}(P_B^+\rho)\,\mathrm{Tr}(\rho_{|B}\hat{O}_B) \ . \tag{6.59}$$

The diagram in Fig. 6.5 helps to understand how the data and the statistics of the measurement are obtained. In the first part the three input states are generated sequentially by three different

transformations $\hat{U}_{DC1}\,|+z\rangle$. The measurement operator in the experiment is realized by a supermirror analyzer which takes $|\hat{O}_A = \pm 1\rangle = \hat{U}_{DC2}\left(\hat{U}_{DC1}\,|+z\rangle\right)$ as input states and will always produce $|\hat{\sigma}_z = 1\rangle$ as output state. In order to realize both outcomes $m = \pm 1$, it is necessary to incorporate the third coil DC3. By appropriate choice of the unitary operations the input and output states can be properly correlated $\hat{U}_{DC3}\,|\sigma_z = \pm 1\rangle$, which rotate in each case on two antipodal points on the Bloch sphere. The purpose of the second apparatus is to measure $\hat{B} = \hat{\sigma}_y$ so that the transition probabilities through the analyzer would be $p = \left|\langle \sigma_y = b|\hat{U}_{DC3}(\sigma_z = m)\rangle\right|^2$. The fourth DC4 is there to change the axis of filtering by two different rotations: $\hat{U}_{DC4}$ is exactly chose such that

$$\begin{aligned} p &= \left|\langle \sigma_y = b|\hat{U}^{\dagger}_{DC4}\hat{U}_{DC4}|\hat{U}_{DC3}(\sigma_z = m)\rangle\right|^2 \\ &= \left|\langle \sigma_z = b|\hat{U}_{DC4}\hat{U}_{DC3}|\sigma_z = m\rangle\right|^2 , \end{aligned} \tag{6.60}$$

which implies a state transformation $\hat{U}_{DC4}\,|\sigma_y = b\rangle = |\sigma_z = b\rangle$ without changing the total probability.

To every combination of $m = \pm 1$ and $b = \pm 1$ and state $\hat{\rho}$, a count rate is obtained, denoted by $I_{m,b}(\hat{\rho})$. Probability is calculated from the relative frequency with which one of the outcomes occurs. The expectation values in Eq. (6.58) or Eq. (6.59) are then determined by the ratio of the count rates for each state as illustrated in Fig. 6.5, i.e.

$$\mathrm{Tr}(\hat{O}_A\hat{\rho}) = \frac{\sum_{(m,b)} m I_{m,b}(\hat{\rho})}{\sum_{(m,b)} I_{m,b}(\hat{\rho})} , \qquad \mathrm{Tr}(\hat{O}_B\hat{\rho}) = \frac{\sum_{(m,b)} b I_{m,b}(\hat{\rho})}{\sum_{(m,b)} I_{m,b}(\hat{\rho})} , \tag{6.61}$$

where $\hat{\rho}$ must be alternated according to the three-state-method.

6.3 Adjustments

The adjustment of the system, shown in Fig. 6.6, can be done as follows. Immediately at the beam port of the reactor wall, a supermirror is placed, which is adjusted in height and angle so that a maximum number of neutrons are detected. The second, subsequent supermirror is set at a distance of approximately 0.5m to maximum transmission of the neutrons. Subsequently, a slight misalignment is performed, whereby higher orders of the Brag-reflected neutrons can

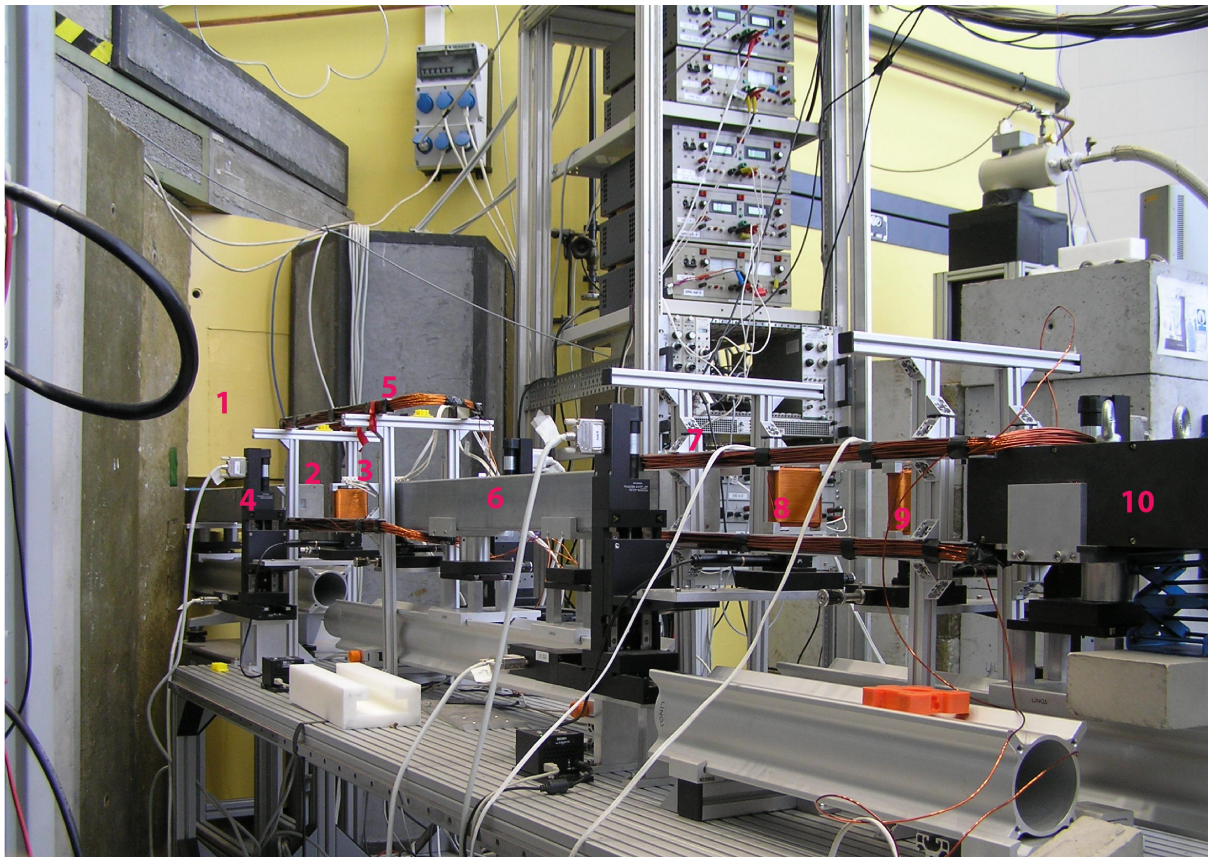


Fig. 6.6: Photo of the polarimeter beamline at the TRIGA Mark-II research reactor. The beam port is located on the yellow wall of the reactor (1) from which a gray, polarizing supermirror (2) protrudes. A solenoid coil (3), standing on a translation stage (4), is aligned centrally to the window of the first supermirror and is surrounded by a Helmholtz coil (5), which has been raised for better visibility in the photo. The second DC coil is covered by the middle supermirror (6). The whole scenario is repeated in the back segment: a Helmholtz coil (7) in which there are two solenoid coils (8,9), followed by another supermirror (10), which finally ends at the detector.

be removed. Although this reduces the count rate, it does increase the degree of polarization, which is typically of the order of 98%. For the second supermirror, which is followed by a detector, the same procedure is carried out.

The DC coils are used, as explained in sec. 4.4, to rotate the neutron spin. They stand on small goniometers, which are not recognizable on the image in Fig. 6.6, controlling the alignment to

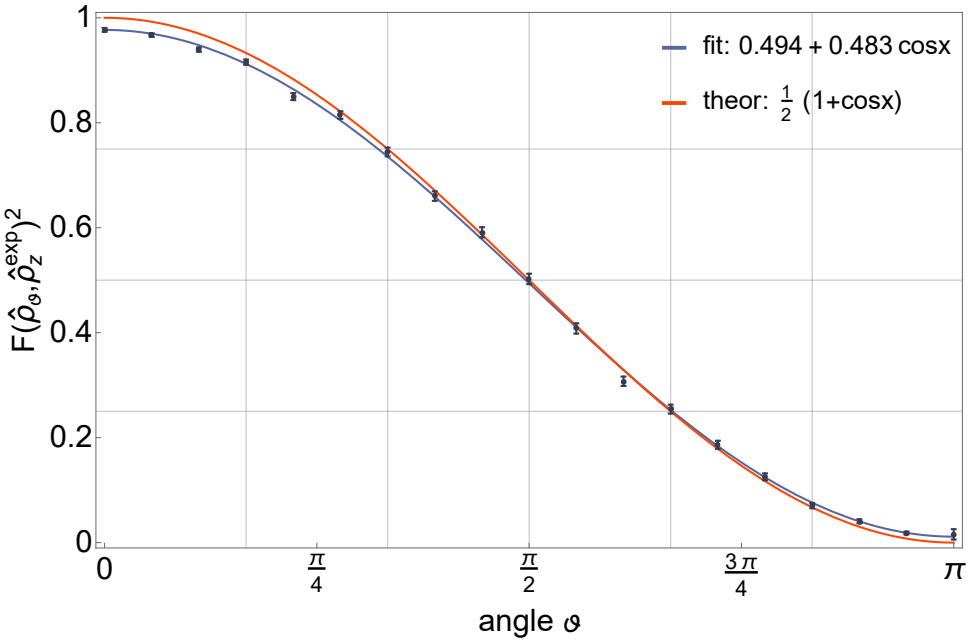
the neutron beam and the external magnetic field. The neutrons sense a non-adiabatic change in the magnetic field during transition through the coil and precesses according to Eq. (4.25). A well-established coil is shown in Fig 6.4. In addition, each of these small solenoids are placed on motorized linear stages, which can change the relative distance between the coils and to the respective supermirrors continuously. To avoid problems in connection with the Earth magnetic field or stray fields Helmholtz coils producing circa 1.3 mT cancel possible influences to a large extend. An unfortunate influencing factor of the copper coils is that it causes a loss rate of about 25 neutrons/sec due to scattering and low absorption.

The imperfect degree of polarization, coming from suboptimal spin purification of the supermirrors, is a systematic error that degrades the data, which can be crucial for testing optimal uncertainty trade-offs. The finite efficiency of the supermirror is shown in the plots in Fig. 6.7. The continuous change of the incident spin relative to the static projection of the supermirror in the z-direction would ideally have to result in $|\langle\chi(\vartheta,0)|+z\rangle| = \cos^2\frac{\vartheta}{2}$. In Fig. 6.7(a), the fit of the data is shown by the black curve, where the parameters of the fit function $f(x) = a + b\cos x$ resulted to $a \cong 0.494$ and $b \cong 0.483$. A potential phase error is neglected because it is much smaller and only statistical in nature. The deviation from the red ideal curve is strongest at the extreme values, which is shown in Fig. 6.7(b). This shows that the static fluctuations are smaller than the systematic insufficiency of the polarization measurement.

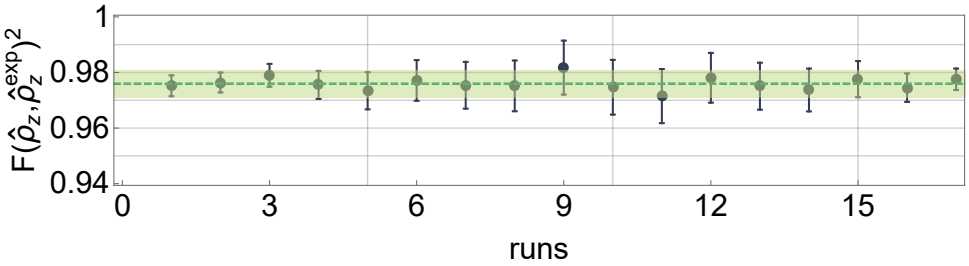
To undo this deficiency from the data, the values must be slightly scaled. For this purpose, one makes a singular measurement at the beginning of the adjustment and then determines the fit, with which the fit parameters are obtained. These can be used to offset the new ordinate values according to

$$f_{new}(x) = \frac{1}{2}\left(1 + \frac{f_{meas}(x) - a}{b}\right) \ . \tag{6.62}$$

The adaptation of the finite degree of polarization must be carried out for each supermirror involved.



(a) Discrepancy between the black-colored function obtained by fitting the data and the theoretical prediction represented by the red curve. The fidelity $F(\hat{\rho}_\vartheta, \hat{\rho}_z^{\text{exp}})^2$ quantifies the degree of overlap, which experimentally is strictly smaller than 1 and larger than 0.



(b) Imperfection of polarization. Each point corresponds to packages of neutrons in 60 seconds. The green line and confidence band demonstrate that the maximal overlap has a mean of 0.976 ± 0.005 and cannot reach the maximal probability.

Fig. 6.7: Very high but imperfect efficiency in polarization measurements. The upper panel shows data points acquired by preparing the state $\hat{\rho}_\vartheta = |\chi(\vartheta,0)\rangle\langle\chi(\vartheta,0)|$ (the vector $|\chi(\vartheta,\varphi)\rangle$ is defined in Eq. (3.15)) and making a projection onto $\hat{\rho}_Z^{\text{exp}}$. The impurity of $\hat{\rho}_z^{\text{exp}}$ is a systematic error that can be taken into account by scaling the data with a correction factor which for example lift the mean value shown in the lower panel.

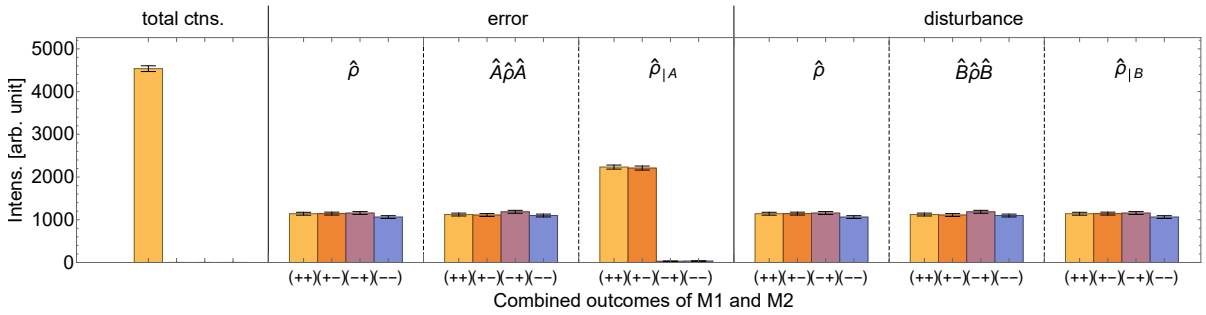
6.4 Results

6.4.1 Standard configuration

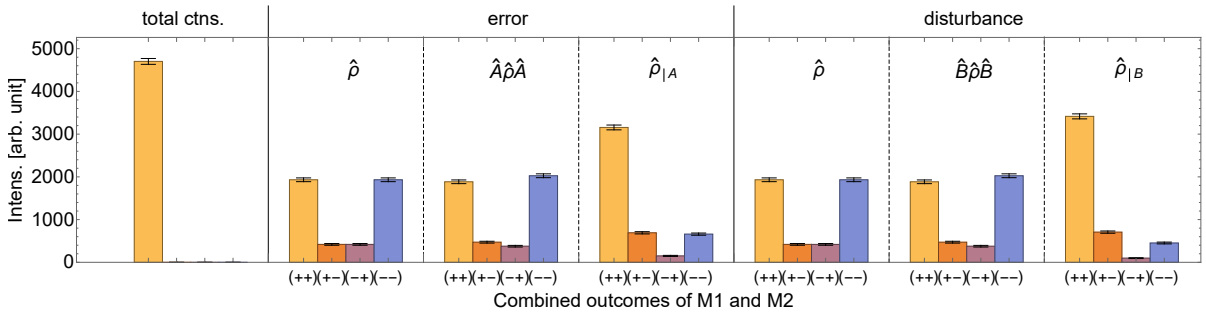
The error and disturbance are obtained by measuring the frequency of a particular outcome relative to all possible sets. So, when the first apparatus measures for example $m = 1$ and the second apparatus measures $b = -1$, this means that the neutrons have a certain probability of being transmitted through the supermirrors in that configuration. This gives a count rate $I_{(m=+1),(b=-1)}$ in a predefined time-scale, usually 60 secs. or 90secs., which is always kept constant throughout the entire experiment. The intensities, according to the corresponding outcomes are depicted in Fig. 6.8 for the so-called *standard configuration* $\hat{A} = \hat{\sigma}_z$, $\hat{B} = \hat{\sigma}_y$, hence their eigenstates are mutually unbiased and the states of $\hat{O}_A$ are coplanar with $|\pm a\rangle, |\pm b\rangle$. The initial input state is $\hat{\rho}_x = |+x\rangle\langle+x|$.

In the upper panel Fig. 6.8(a) the angle is $\theta_{OA} = 0$ which means that an exact measurement of $\hat{A}$ is performed. The first four outputs occur all with $\frac{1}{4}$ of the total intensity, because each $|\langle+x|\hat{\sigma}_z = m\rangle|^2$ and $|\langle+z|\hat{\sigma}_y = b\rangle|^2$ ($|+z\rangle$ is the state after the $\hat{A}$ measurement) occur each with the same probability of $\frac{1}{2}$. Because $\hat{A}\hat{\rho}_x\hat{A}$ just reflects the state to $\hat{\rho}_{-x} = |-x\rangle\langle-x|$ there is nothing changing in the next intensity quadruple. The error measurement is completed with the count rates for $\hat{\rho}_{|A} = |+z\rangle\langle+z|$: the up-spin state is completely transmitted in the first apparatus $m = 1$ and then measured with half the probability in apparatus M2, while the neutrons are not transmitted when $m = -1$ is analyzed . For the disturbance measurement, the first two histograms do not change, because the input states are once more $\hat{\rho}_x, \hat{\rho}_{-x}$ and the two expectation values Eq. (6.61) are obtained by different evaluation only. The last count rate distribution is also uniform, because $\hat{\rho}_{|B} = |+y\rangle\langle+y|$ behaves the same as $\hat{\rho}_x$ in the first histogram.

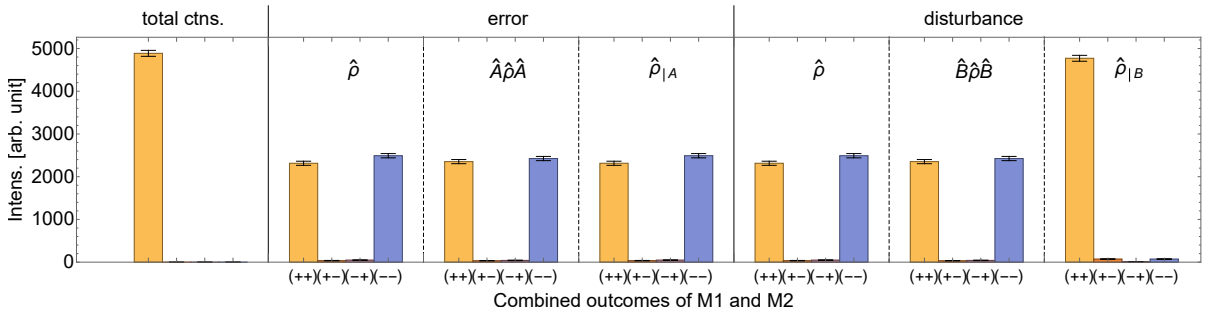
The histograms for $\hat{\rho}$ and $\hat{A}\hat{\rho}\hat{A} = \hat{B}\hat{\rho}\hat{B}$ in Figs. 6.8(b) and 6.8(c) are identical due to the symmetry of the observable configuration. The error and disturbance are entirely dominated by the two eigenstates of the non-commuting observables $\hat{\rho}_{|A}$ and $\hat{\rho}_{|B}$ in the measurement. The last bar chart in Fig. 6.8(c) indicates that $\hat{\rho}_{|B}$ has been measured accurately and therefore the disturbance vanishes.



(a) Histogram for $\theta_{OA} = 0$ resulting in $\hat{O}_A = \hat{A}$.



(b) Histogram for $\theta_{OA} = \frac{2\pi}{9}$ close to the angle bisector of $\hat{A}$ and $\hat{B}$.



(c) Histogram for $\theta_{OA} = 0$ resulting in $\hat{O}_A = \hat{B}$.

Fig. 6.8: Count rate histograms. Determination of the expectation values Eq. (6.61) by count rates $I_{m,b}$ for three different angles θ_{OA}, namely 0 in Fig. 6.8(a), $\frac{2\pi}{9}$ in Fig. 6.8(b) and $\frac{\pi}{2}$ in Fig. 6.8(c). The input state is $\hat{\rho} = \hat{\rho}_x$. The panel on the left always shows the total count rate, which is approximately 4800, the three bar charts to the right of the first column are the required intensities to measure the error with all three states and finally the three leftmost histograms are required for the disturbance. The label below the charts correspond to the subscripts of the intensities, e.g. for the error $(+-) = I_{(m=+1),(b=-1)}$.

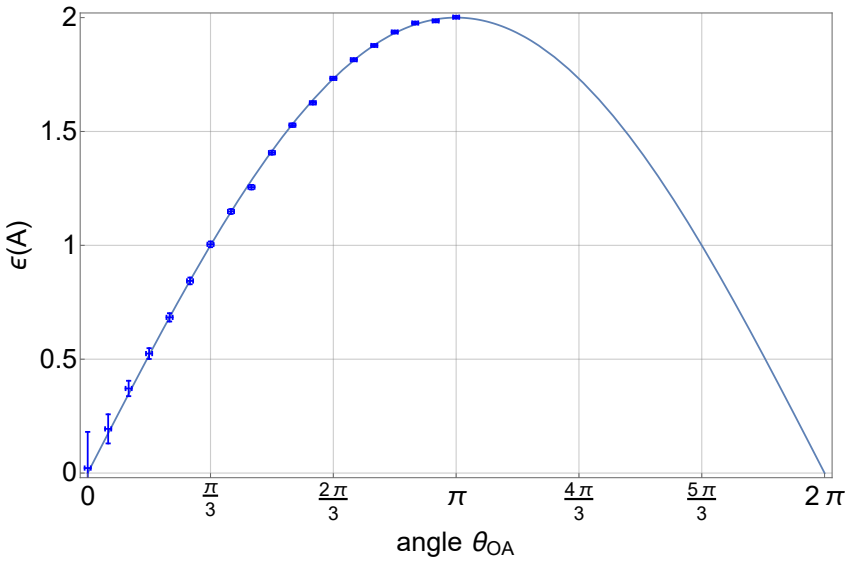


Fig. 6.9: Measurement results of error as function of the polar angle of the observable $\hat{O}_A = \cos\theta_{OA}\hat{\sigma}_z + \sin\theta_{OA}\hat{\sigma}_y$. The blue line corresponds to the theoretical prediction given by the function $\epsilon(\hat{A}) = 2\left|\sin\frac{\theta_{OA}}{2}\right|$. The data points represent the measured quantities from $\theta_{OA} = 0$ to π. The statistical error-bars are larger close to 0 since they obtained at low count rates.

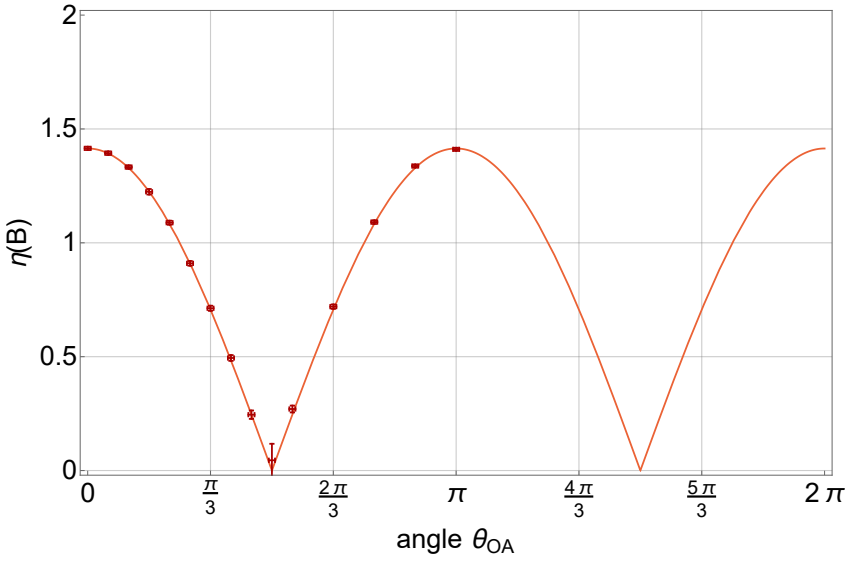


Fig. 6.10: Measurement results of $\eta(\hat{B}) = \sqrt{2}\,|\cos\theta_{OA}|$ for projective measurement of $\hat{O}_A$ (for the uncorrected case). The red line correspond once more to the given function and points correspond to the results. The statistical error-bars are larger close to 0 since they obtained at low count rates.

Error-disturbance results obtained by projective measurements with neutrons are given in Fig. 6.9 and Fig. 6.10. They were first confirmed in the article [36]. The error at $\theta_{OA} = 0$ (or multiples of 2π) is zero, because at this angle the approximate observable $\hat{O}_A = \hat{A}$ is identical to the intended observable and there is no difference between the two. In other words if the observables are perfectly correlated then perfect accuracy measurements are possible. A change of the angle θ_{OA} misaligns the directions of the observables and therefore alters the distance between them. Accordingly the error increases and reaches a maximum at $\hat{O}_A = -\hat{A}$. A rotation of the observable beyond this point reverts the error back symmetrically (and is thus not measured). The disturbance has an opposite behavior. When $\hat{O}_A = \hat{A}$ then the measurement imparts the maximal disturbance on the state that measures $\hat{B}$. The disturbance $\eta(\hat{B})$ decreases however as $\hat{O}_A$ approaches $\hat{B}$ and has a minimum at equality for $\theta_{OA} = \frac{\pi}{2}$.

Both error and disturbance depend on a the same angle θ_{OA}, therefore it is appropriate to plot them parametrically $(\epsilon(\hat{A}), \eta(\hat{B}))$. Four different error-disturbance trade-offs are given in Fig. 6.11 for the standard configuration. The results of the measurement are plotted again in the range $\theta_{OA} \in [0, \frac{\pi}{2}]$ with their theoretical prediction. They obviously violate a naive Arhturs-Kelly-like version of the uncertainty relation $\epsilon(\hat{A})\eta(\hat{B}) \geq \frac{1}{2}$. On the other hand the blue line given by Ozawa's first prediction Ineq. (6.22) is far-off the results and there seems no way for the outcomes to touch that curve. This is a strong evidence that the inequality is too loose. The green curve corresponding to Ineq. (6.31) for pure states $C_{AB} = 1$ on the other hand can actually be approached. It was suggested by Branciard that instead of recreating the eigenstate of the projector $\hat{P}_{OA}(m)$ it is better to transform the state after analyzer-1 to the eigenstates of $\hat{B}$ [83]. This can be realized by a correction procedure.

Correction and anti-correction: The data in Fig. 6.10 satisfy the error-disturbance relation under consideration but do not lead to saturation. From relation Ineq. (6.31), if $\epsilon(\hat{A}) = 2\left|\sin\frac{\theta_{OA}}{2}\right|$ then $\eta(\hat{B})$ should satisfy

$$2\left|\sin\left(\frac{\pi}{4} - \frac{\theta_{OA}}{2}\right)\right| \leq \eta(\hat{B}) \leq 2\left|\sin\left(\frac{\pi}{4} + \frac{\theta_{OA}}{2}\right)\right| . \tag{6.63}$$

This means that that the region of attainable values is bounded from below and also from above. The lower limits turn out to be accessible by 'correcting' the state after the $\hat{O}_A$ measurement

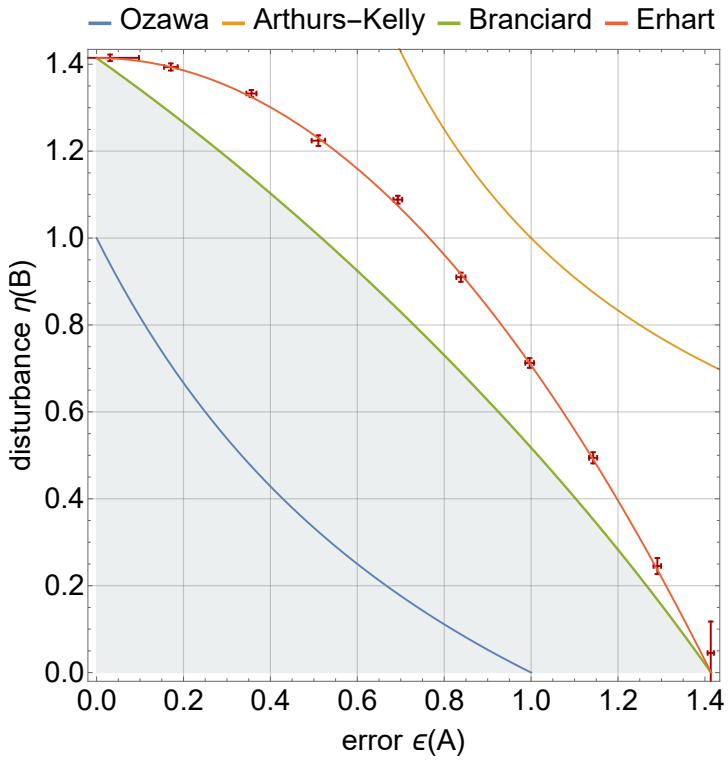


Fig. 6.11: Plot of $\epsilon(\hat{A})$-$\eta(\hat{B})$ with four different predictions. The blue curve corresponds to the first prediction of Ozawa as given in Ineq. (6.22) when $\sigma(\hat{A}) = \sigma(\hat{B}) = 1$ and the red curve corresponds to the first result obtained by Erhart et. al. [36]. Apparently experimental outcomes are below the Arthurs-Kelly curve, i.e. the naive reinterpretation of the Robertson-like inequality $\epsilon(\hat{A})\eta(\hat{B}) \geq \frac{1}{2}$, given by the yellow line, and therefore falsify its general validity. The ominous gap between the results and the Ozawa prediction encouraged the search for an improvement, obtained by Branciard, whose solution is plotted in green.

while the upper limit is obtained by 'anti-correcting'. To obtain an optimal operator $\hat{U}$, let $\hat{U}_{DC3}$ be such that $\hat{U}_{DC3}\,|\hat{\sigma}_z = m\rangle = |\hat{B} = m\rangle$ for $m = \pm 1$ e.g.,

$$\hat{U}_{DC3} = \frac{1}{\sqrt{2}}\,(\hat{\sigma}_z + \hat{\sigma}_y) = \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} \tag{6.64}$$

independent of the value of θ_{OA}. Then, the measurement operators are given by $\{\hat{M}_m\} =$

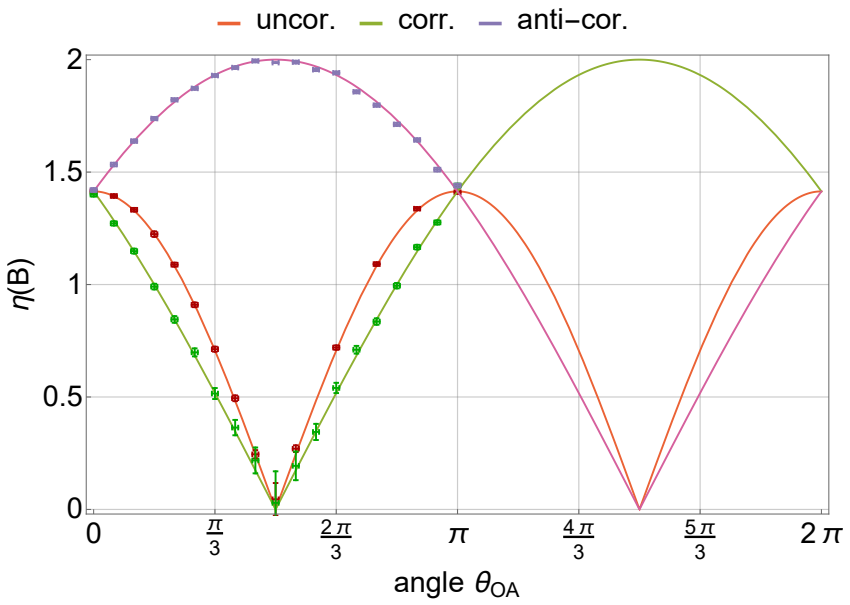


Fig. 6.12: Comparison of disturbances. The red line and points show once more the result as given in Fig. 6.10, while the green curve $\eta(\hat{B}) = 2\left|\sin\left(\frac{\pi}{4} - \frac{\theta_{OA}}{2}\right)\right|$ and the measurement points demonstrate the decrease of disturbance between $0 \leq \theta_{OA} \leq \pi$. The disturbance can be increased to find the worst case as given by the purple curve $\eta(\hat{B}) = 2\left|\sin\left(\frac{\pi}{4} + \frac{\theta_{OA}}{2}\right)\right|$ with the corresponding measurement results.

$\{|\hat{B} = m\rangle \langle \hat{O}_A = m|\}$ and eventually

$$\begin{aligned}
\hat{O}_B &= \sum_m \hat{M}_m^\dagger \hat{B} \hat{M}_m \\
&= \sum_{\pm 1} |\hat{O}_A = m\rangle \langle \hat{B} = m|\hat{B}|\hat{B} = m\rangle \langle \hat{O}_A = m| = \hat{O}_A \\
\hat{B}\hat{O}_B &= \hat{B}\hat{O}_A \\
&= \hat{\sigma}_y \left(\cos\theta_{OA}\hat{\sigma}_z + \sin\theta_{OA}\hat{\sigma}_y\right) \\
&= \sin\theta_{OA}\mathbb{1} + i\cos\theta_{OA}\hat{\sigma}_x \\
\eta(\hat{B})^2 &= 2 - 2\Re(\hat{B}\hat{O}_B) = 2 - 2\sin\theta_{OA} = 4\sin\left(\frac{\pi}{4} - \frac{\theta_{OA}}{2}\right)^2 .
\end{aligned} \tag{6.65}$$

To obtain an anti-optimal $\hat{U}$, let $\hat{U}_{DC3}$ be such that $\hat{U}_{DC3}|\hat{\sigma}_z = m\rangle = |\hat{B} = -m\rangle$ for $m = \pm 1$ e.g.,

$$\hat{U}_{DC3} = \frac{1}{\sqrt{2}}(\hat{\sigma}_z - \hat{\sigma}_y) = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \tag{6.66}$$

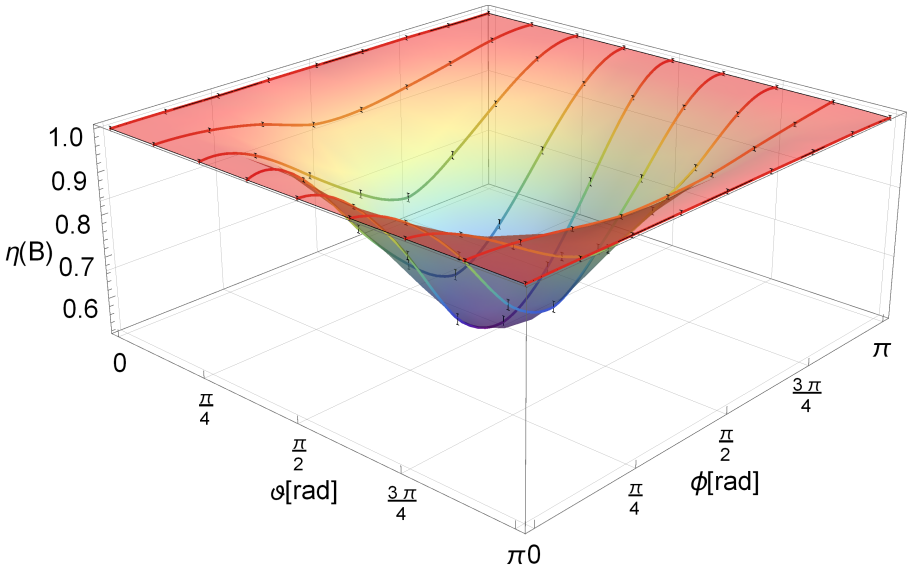


Fig. 6.13: Experimental search of minimal disturbance. After the projective measurement of $\hat{O}_A$ at an angle of $\theta_{OA} = 50°$ the unitary $\hat{U}_{DC3}$ is used to create different states $|\chi(\vartheta, \varphi)\rangle$. The minimum at $\vartheta = \phi = \frac{\pi}{2}$ corresponds exactly to $|+y\rangle$.

independent of the value of θ_{OA}. Then, the measurement operators are given by $\{\hat{M}_m\} = \{|\hat{B} = -m\rangle \langle \hat{O}_A = m|\}$ and maximal disturbance becomes

$$\begin{aligned} \hat{O}_B &= \sum_{\pm 1} |\hat{O}_A = m\rangle \langle \hat{B} = -m|\hat{B}|\hat{B} = -m\rangle \langle \hat{O}_A = m| = -\hat{O}_A \ , \\ \eta(\hat{B})^2 &= 2 - 2\Re(\hat{B}\hat{O}_B) = 2 + 2\sin\theta_{OA} = 4\sin\left(\frac{\pi}{4} + \frac{\theta_{OA}}{2}\right)^2 \ . \end{aligned} \tag{6.67}$$

The results with correcting and anti-correcting unitary transformations are plotted in Fig. 6.12. After the measurement of $\hat{O}_A$ and before the measurement of $\hat{B}$ a unitary quantum channel $\hat{U}_{DC3}$ allows to generate different states. When the state after the apparatus M1 yields the outcome $m = 1$, the state afterwards is $|+z\rangle$ and can appropriately be rotated to $|+y\rangle$ for every angle of θ_{OA} to reduce the disturbance by a correlated transformation. At zero error the disturbance is a fixed point and at $\theta_{OA} = \frac{\pi}{2}$ the uncorrected states already matches the eigenstate of $\hat{B}$, so there is no need for improvements there. For intermediate values the effect of the optimization is clearly visible. If instead for $m = 1$ the unitarily transformed state is always $|-y\rangle$ then physically the set-up adds disturbance to the quantum state before it is finally detected. Corrected and anti-corrected disturbance are inversely phased (phase-shifted by π).

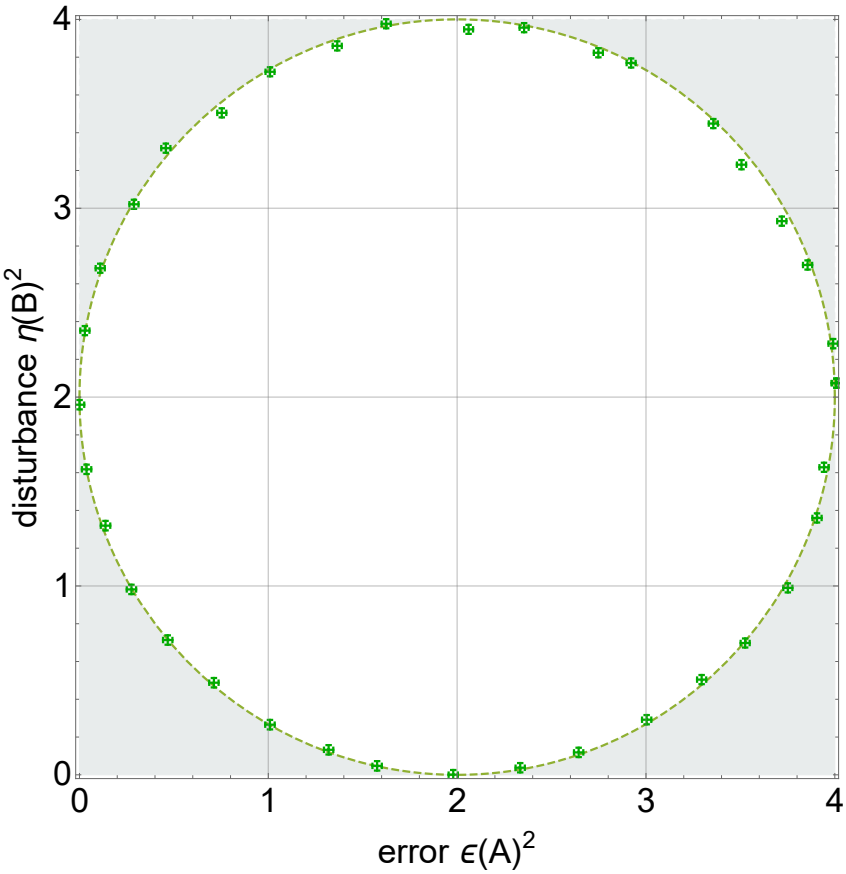


Fig. 6.14: Squared error-disturbance $\epsilon(\hat{A})^2$ vs. $\eta(\hat{B})^2$ uncertainty relation as indicated by Ineq. (6.68). The region in gray is inaccessible, the green boundary is the saturated limit of the inequality and the white area the entire accessible region.

The optimization can be investigated with experimental methods as well. For each angle θ_{OA}, a stepwise variation of the unitary transformations has to be done to find the position of the minimum. The dependence of the disturbance on the configuration of the post-measurement state is plotted in Fig. 6.13. This plot shows the result when the unitary operator $\hat{U}_{DC3}$ generates all possible qubit states $|\chi(\vartheta, \varphi)\rangle$ as defined in Eq. (3.15) in the range from $0 < \vartheta < \pi$ and $0 < \phi < \pi$. The lowest disturbance is obtained at a the point that corresponds to the positive eigenstate of $\hat{B}$.

The parametric plot of the result Eq. (6.52) vs. Eq. (6.63) must in its squared form yield a

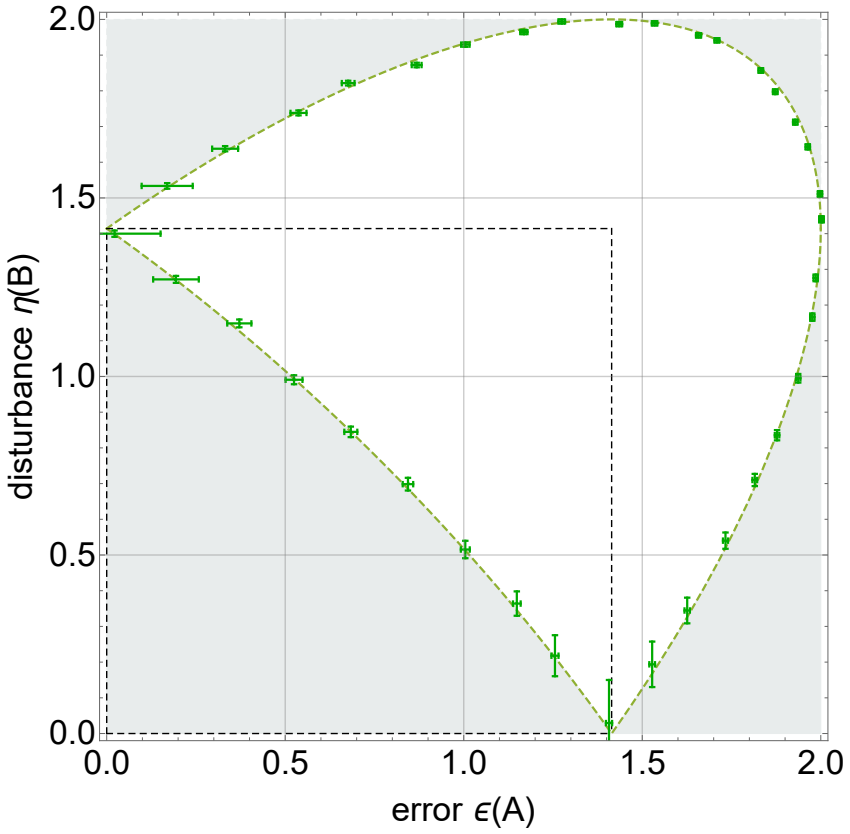


Fig. 6.15: Plots of error-disturbance obtained by building the square root in Fig. 6.14. The region in gray is inaccessible, while the dashed green line represent the smallest and largest obtainable uncertainty outcomes, which have been obtained properly. The framed section matches the picture Fig. 6.11 and is usually the most interesting part, because it contains the lowest achievable pairs of values.

circle as indicated by Ineq. (6.31)

$$\tilde{\epsilon}(\hat{A})^2 + \tilde{\eta}(\hat{B})^2 \geq 1 \qquad \Leftrightarrow \qquad (\epsilon(\hat{A})^2 - 2)^2 + (\eta(\hat{B})^2 - 2)^2 \leq 4 \; . \tag{6.68}$$

This is shown in Fig. 6.14, where the full range of obtainable $(\epsilon(\hat{A})^2, \eta(\hat{B})^2)$ pairs and the 'forbidden' region are colored in white and gray. The measured points at the boundary below the value $\eta(\hat{B})^2 \leq 2$ correspond to the optimized uncertainty relation, when the inequality is saturated, while the points above that value can be obtained by anti- correction. In Fig. 6.15 the quadrant $(0, 0)$ to $(\sqrt{2}, \sqrt{2})$ is identical to the prediction of the tight bound shown in Fig. 6.11, first measured in [91].

6.4.2 Non-standard configuration

The research of error and disturbance is extended now by changing the polar angle of $\hat{B}$, thereby altering the observable's relative orientation. That is, the choices for the operators are $\hat{A} = \hat{\sigma}_z$ and

$$\hat{B}' = \cos\theta_B \hat{\sigma}_z + \sin\theta_B \hat{\sigma}_y \tag{6.69}$$

for two angles $\theta_B = \{\frac{\pi}{3}, \frac{\pi}{6}\}$ and for the input state $\hat{\rho} = |+x\rangle\langle+x|$. Overall, the uncertainty is expected to decrease, since the involved observables $\hat{A}, \hat{B}'$ are no longer maximally non-commuting, or more precisely their eigenstates are not mutually unbiased. The choice of the operator changes the optimal disturbance to

$$\eta(\hat{B}') = 2\left|\sin\left(\frac{\theta_{OA} - \theta_B}{2}\right)\right| \tag{6.70}$$

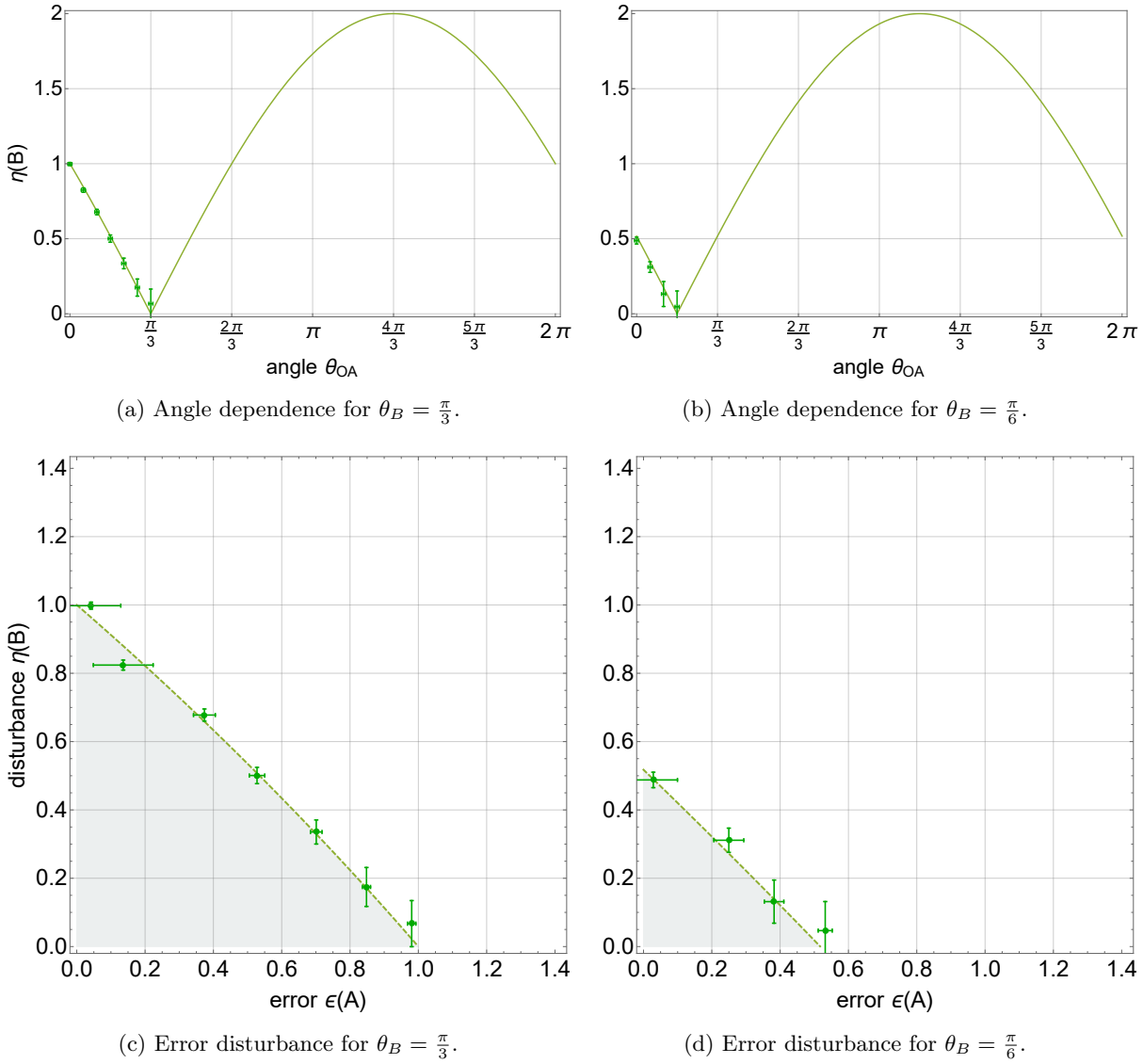
and also the bound $C_{AB} = |\sin\theta_B|$. Since all observables are still in the y-z plane orthogonal to the system state Ineq. (6.31) still holds. Analogous to the standard configuration after apparatus M1, a correction of the state after the $\hat{O}_A$ measurement the closest eigenstate of $\hat{B}'$ must be performed to reach the lowest disturbance. The experimental results are plotted for both angles θ_B in Fig. 6.16. In the angle dependent plots it is apparent that the disturbance disappears once more when the observables coincide $\hat{B}' = \hat{O}_A$ and the closer these two operators are, the more the minimum shifts to the left. In the error-disturbance trade-offs it is obviously the case that the uncertainty shrinks the smaller the angle θ_B.

6.4.3 Error-disturbance for mixed states

As was pointed out, the limit C_{AB} in the uncertainty relation given by Ineq. (6.31) predicts a disappearance of the error-disturbance trade-off for mixed states, while other new bound D_{AB} as introduced by Ineq. (6.37) is unaffected for changes of mixture in the standard configuration, i.e. for $\hat{A} = \hat{\sigma}_z, \hat{B} = \hat{\sigma}_y$ and states

$$\hat{\rho}_x = \frac{\mathbb{1} + \|\mathbf{r}_x\|\sigma_x}{2} \ , \tag{6.71}$$

which are mixed if $\|\mathbf{r}_x\|$ has not unit length.



(a) Angle dependence for $\theta_B = \frac{\pi}{3}$.

(b) Angle dependence for $\theta_B = \frac{\pi}{6}$.

(c) Error disturbance for $\theta_B = \frac{\pi}{3}$.

(d) Error disturbance for $\theta_B = \frac{\pi}{6}$.

Fig. 6.16: The upper panel of the figure contains the variations of the measurement operator angle θ_{OA} for a fixed $\theta_B = \frac{\pi}{3}$ in Fig. 6.16(a) and $\theta_B = \frac{\pi}{6}$ in Fig. 6.16(b). In the lower panel the error-disturbance is plotted which displays the property that the closer the observables in context are the more the limit C_{AB} decreases and hence the more the uncertainty region disappears.

The set-up for this measurement is equivalent to the experiments that have been used previously. The only difference is located in the first coil DC1 which is also fed with a random noise-signal and thereby produces a 'noisy' magnetic field. This means the neutrons with different arriving

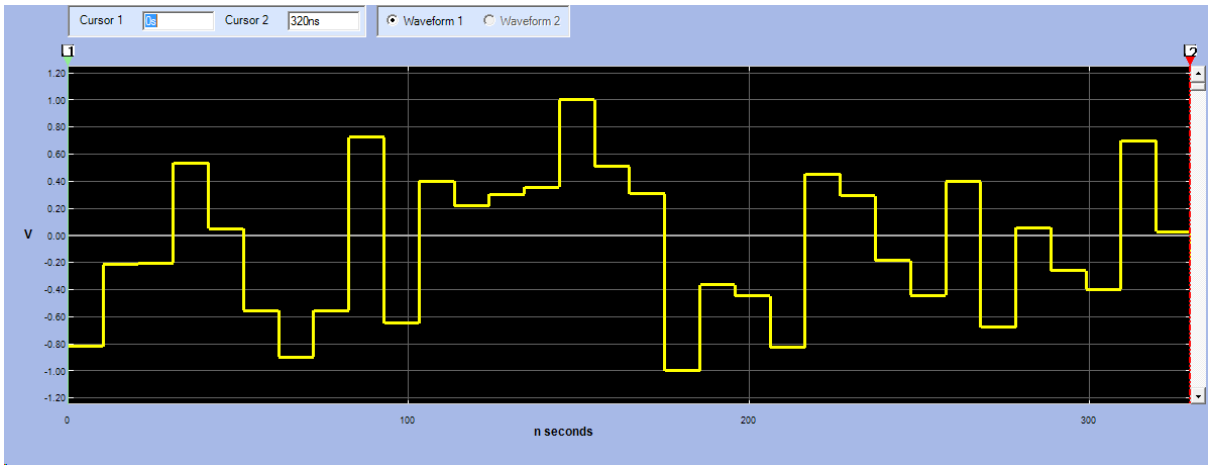


Fig. 6.17: Noisy voltage as plotted by the ArbExpress® signal generator software from Tektronix. Such a random wave form can either be generated by the software and then uploaded onto the signal generator (AFG 3022B) or directly created at the device itself. If one chooses the rate of change of the voltage as long as the neutrons' time of transmission through the coil, each particle will see a different magnetic field. Therefore, feeding such a signal into the solenoid producing the B_x field mixes the total spin ensemble.

times at the coil DC1 are unitarily rotated individually by some $\hat{U}_{noise}(\frac{\pi}{2} + \Delta\xi_{noise}(t))$, so that the whole state given by the time average is a mixed spin ensemble [95]. Experimentally, the conversion to mixed states is carried out using an AFG 3022B signal generator, which is connected to the windings of the x-coil, fed with a random voltage amplitudes looking something like in Fig. 6.17. The degree of mixture and alignment of the states was monitored by measuring the polarization vector $\mathbf{P} = \mathrm{Tr}(\boldsymbol{\sigma}\hat{\rho})$. Tuning of the amplitude of the noise field will accordingly change $\mathbf{P}$ and allow to determine the desired degrees of mixtures. The checking of the efficiency of mixing is listed in Tab. 6.1.

The results of error-disturbance for mixed input states as given by Eq. (6.71) for five different mixtures are depicted in Fig. 6.18. There are two limits in each plot colored in green and blue if the results differ and the according measurement points. Starting with the pure state obviously nothing changes, then continuing with mixed states $\|\mathbf{r}_x\| \cong \{\frac{3}{4}, \frac{1}{2}, \frac{3}{4}, 0\}$ the measurement outcome reveals that the form of the familiar error-disturbance trade-off remains the same. Hence Ineq. (6.31) is indeed only tight for pure states and the second improved version Eq. (6.37) is a legitimate replacement to explain the invariance of the $\epsilon(\hat{A}), \eta(\hat{B})$ - uncertainty relation for

$P_x = \mathrm{Tr}(\hat{\sigma}_x\hat{\rho})$	$P_y = \mathrm{Tr}(\hat{\sigma}_y\hat{\rho})$	$P_z = \mathrm{Tr}(\hat{\sigma}_z\hat{\rho})$
0.998 ± 0.003	-0.004 ± 0.008	-0.008 ± 0.008
0.748 ± 0.011	0.016 ± 0.008	-0.025 ± 0.008
0.499 ± 0.014	0.003 ± 0.008	-0.005 ± 0.008
0.241 ± 0.015	0.038 ± 0.006	-0.047 ± 0.006
-0.030 ± 0.017	0.022 ± 0.009	-0.036 ± 0.009

Tab. 6.1: Result of the polarization vector $\mathbf{P} = \mathrm{Tr}(\hat{\boldsymbol{\sigma}}\hat{\rho})$ measurement indicating the degree of mixture.

mixed states. This statement remains true as long as the input state is 'orthogonal' to the regarded observables as indicated by Fig. 6.18(a).

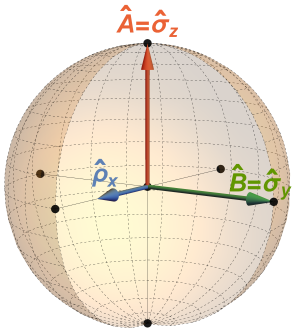
6.5 Measurement uncertainty relation based on statistical distances

Whether the definitions of error and disturbance as given by Ozawa are appropriate measures to quantify measurement uncertainty lead to intensive controversies. As competitor to the construct of rms operator deviations, an alternative approach to the topic was presented by Busch, Lahti and Werner (BLW) [23–25] whose theory is based on the statistical distance, more accurately the Wasserstein-2 metric (see. Eq. (2.28)) between their outcome values. The distance in the error-disturbance relations is obtained by a comparison of two reference outcome distributions of observables $\hat{A}$, $\hat{B}$ to a measurement of $\hat{C}$, approximating $\hat{A}$ and likewise $\hat{D}$ measuring $\hat{B}$ with different, but identically prepared ensembles $\hat{\rho}$. A scheme of the idea is shown in Fig. 6.19.

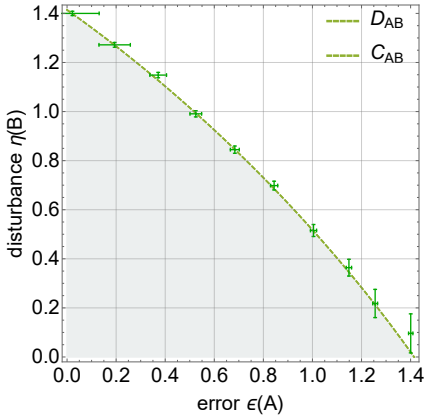
The theory starts with a state-specific error $\Delta(\hat{A}, \hat{C}, \hat{\rho})$ which assesses a degree of closeness of $\hat{A}$ to $\hat{C}$ for a specific quantum state $\hat{\rho}$. By the same reasoning for the disturbance the distance of the distribution shall be defined as

$$\Delta(\hat{A}, \hat{C}, \hat{\rho}) = W_2(\hat{A}_\rho, \hat{C}_\rho) \ , \qquad \Delta(\hat{B}, \hat{D}, \hat{\rho}) = W_2(\hat{B}_\rho, \hat{D}_\rho) \ , \tag{6.72}$$

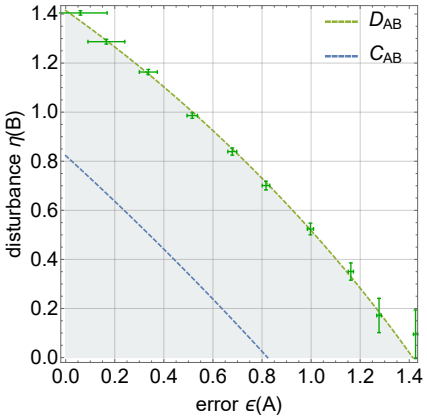
where subscript to the observables in the Wasserstein-2 distance $W_2(\hat{A}_\rho, \hat{C}_\rho)$ have been appended to indicate state dependency. However, the difference in a particular state is not a good indicator



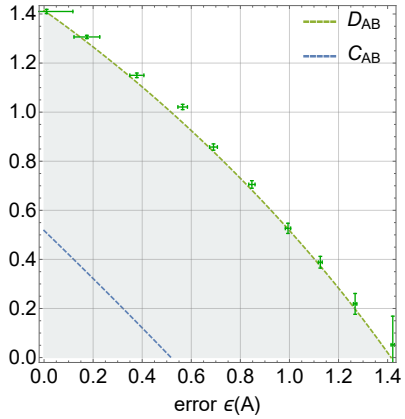
(a) Geometric situation.



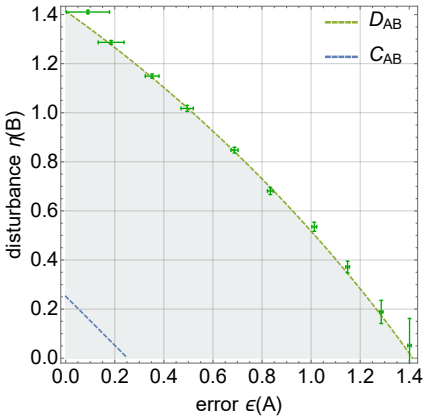
(b) Result for pure state $\|\mathbf{r}_x\| = 1$.



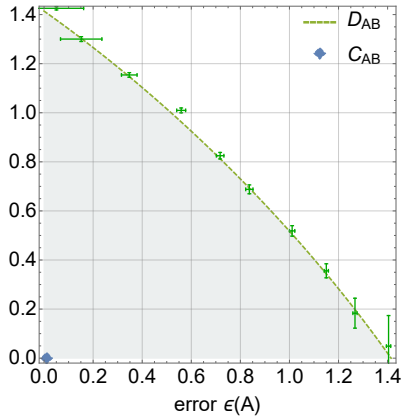
(c) Result for mixed state $\|\mathbf{r}_x\| \cong 0.75$.



(d) Result for mixed state $\|\mathbf{r}_x\| \cong 0.5$.



(e) Result for mixed state $\|\mathbf{r}_x\| \cong 0.25$.



(f) Result for mixed state $\|\mathbf{r}_x\| \cong 0$.

Fig. 6.18: Plots of $\epsilon(\hat{A})$-$\eta(\hat{B})$ showing for five different mixtures $\|\mathbf{r}_x\|$: (b) 1, (c) 0.75, (d) 0.5, (e) 0.25 and (f) 0 no effect on the uncertainty as predicted by D_{AB}, while the other bound C_{AB} given by the blue dashed line suggest a disappearance of uncertainty the more mixed the quantum states become [94].

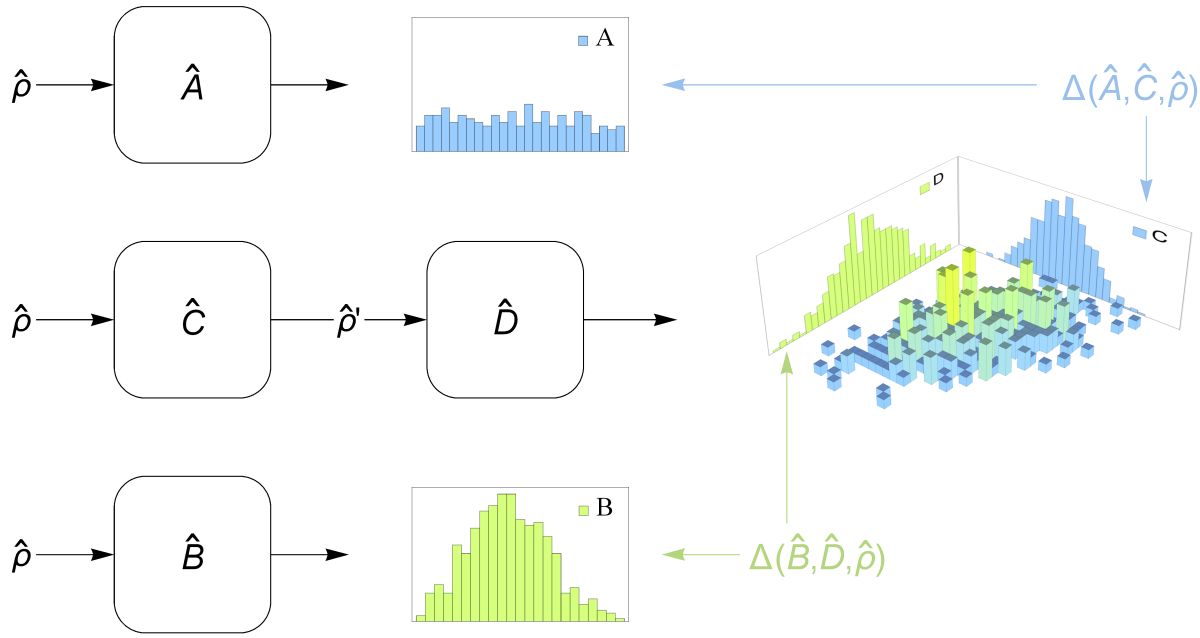


Fig. 6.19: Schematic implementation of a distance-type error-disturbance experiment. Three separate measurements are performed to record the three different distributions: two single measurements that provide the reference values (top and bottom histograms) and a bivariate distribution from the joint measurement, whose marginals are then compared with the reference outcome distributions. The difference in the output statistics is quantified by the state dependent value $\Delta(\hat{A},\hat{C},\hat{\rho}) = W_2(\hat{A}_\rho,\hat{C}_\rho)$ for the error and $\Delta(\hat{B},\hat{D},\hat{\rho}) = W_2(\hat{B}_\rho,\hat{D}_\rho)$ for disturbance. These Δ's in turn are used to obtain state independent measures of imprecision.

for the overall performance in the sense that its smallness implies high measurement quality of the device. Therefore, BLW propose to consider two types of total error/disturbance:

1. The 'worst case' estimate: a *figure of merit (FoM)* of the measurement apparatus given by the maximal difference between the measurements with respect to all quantum state

$$\Delta(\hat{A},\hat{C})_{FoM} = \sup_\rho W_2(\hat{A}_\rho,\hat{C}_\rho)\ , \qquad \Delta(\hat{B},\hat{D})_{FoM} = \sup_\rho W_2(\hat{B}_\rho,\hat{D}_\rho)\ . \tag{6.73}$$

2. Alternatively, since it might be difficult to retrieve the supremum over all states, a *calibration error (Cal)* is proposed in which the prepared states are eigenstates of the observable to be determined, respectively for which $\hat{C}_\rho$ is nearly a point (Dirac-) measure δ_x [96]

$$\Delta(\hat{A},\hat{C})_{Cal} = \lim_{\varepsilon\to 0}\ \sup_{\rho,x}\Big(W_2(\hat{C}_\rho,\delta_x)\ \big|\ W_2(\hat{A}_\rho,\delta_x)\le\varepsilon\Big)\ . \tag{6.74}$$

The analysis of the trade-off relation for errors $\Delta(\hat{A},\hat{C})$ and $\Delta(\hat{B},\hat{D})$ for qubits measurements is made in [25]. Calculations yield the following relationships. Let the joint probability distribution of $\hat{A}_\rho,\hat{C}_\rho$ be collectively written into the matrix

$$\gamma_{ij}=\begin{pmatrix}\gamma & \langle P_A^+\rangle-\gamma\\ \langle P_C^+\rangle-\gamma & \langle P_C^-\rangle\end{pmatrix} \tag{6.75}$$

with the respective projectors of the observables under consideration, then the state-dependent Wasserstein-2-metrics yield [97]

$$\begin{aligned} W_2(\hat{A}_\rho,\hat{C}_\rho) &= \sqrt{\inf_{\gamma\in\Gamma(\hat{A}_\rho,\hat{C}_\rho)}\sum_{(i,j)=\pm1}(i-j)^2\gamma_{ij}}\\ &= \sqrt{\inf_{\gamma}4(\gamma_{1,-1}+\gamma_{-1,1})}=\sqrt{4\left|\langle P_A^+\rangle-\langle P_C^-\rangle\right|}\\ &= \sqrt{2\left|\mathbf{r}\cdot(\mathbf{a}-\mathbf{c})\right|}\\ W_2(\hat{B}_\rho,\hat{D}_\rho) &= \sqrt{2\left|\mathbf{r}\cdot(\mathbf{b}-\mathbf{d})\right|}\ . \end{aligned} \tag{6.76}$$

These state-dependent quantities are important from an experimental point of view because they show some differences between Ozawa and BLW's results. In contrast, in the qubit case for sharp observables the state-independent expression $\Delta(\hat{A},\hat{C})_{FoM}$ and $\Delta(\hat{A},\hat{C})_{Cal}$ are de-facto identical with Ozawa' error definition $\epsilon(\hat{A})$, at least for pure states. Indeed, as proven in BLW's works an indisputable similarity is given

$$\begin{aligned} \epsilon(\hat{A}) &= \frac{1}{2}\Delta(\hat{A},\hat{C})^2_{FoM}=\Delta(\hat{A},\hat{C})_{Cal}\ ,\\ \eta(\hat{B}) &= \frac{1}{\sqrt{2}}\Delta(\hat{B},\hat{D})^2_{FoM}=\Delta(\hat{B},\hat{D})_{Cal}\ . \end{aligned} \tag{6.77}$$

Hence, the BLW error-disturbance relation $(\Delta(\hat{A},\hat{C})_{Cal},\Delta(\hat{B},\hat{D})_{Cal})$ in two dimensional quantum systems with projective measurements is given as illustrated in Figs. 6.9, 6.10 and 6.15 and the figure of merit errors are obtained by the square roots of the calibration definitions up to some constant factor.

There exist a few tests of BLW uncertainty relation [98–100]. Here, only the neutron optical method [97] is summarized. The experimental setup has once more the design as in Fig. 6.3. The target observables are the maximally non-commuting Pauli operators $\hat{A}=\mathbf{a}\cdot\hat{\boldsymbol{\sigma}}=\hat{\sigma}_z$ and

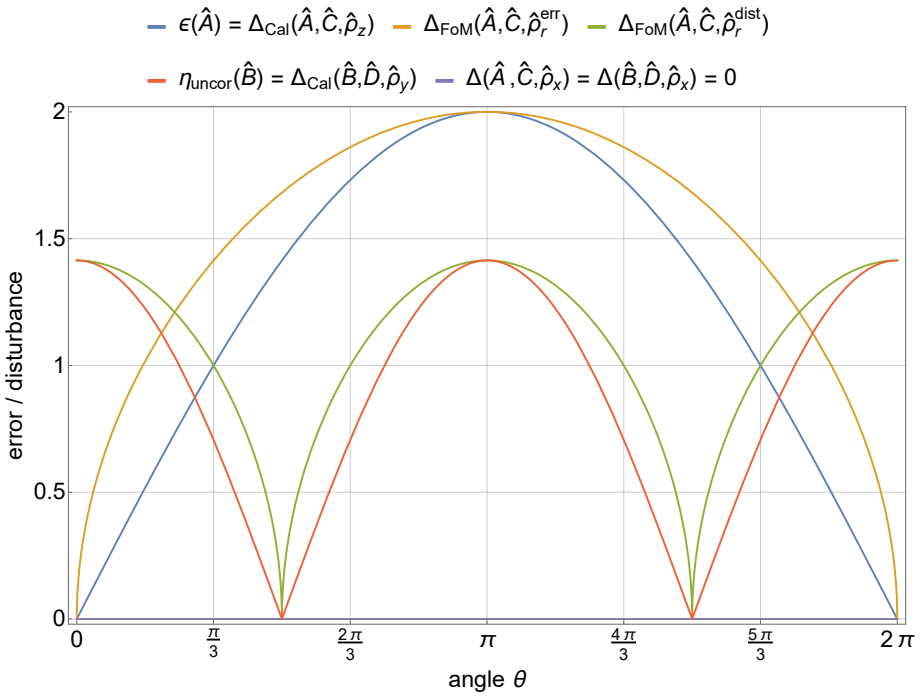


Fig. 6.20: A collection of the various state-dependent errors for projective Pauli observables. The calibration errors in blue and red for the eigenstates of $\hat{A}$ and $\hat{B}$ exactly match Ozawa's error-disturbance definition. The figure of merit errors colored yellow and green are obtained by $\hat{\rho}^{err}_{\mathbf{r}}$ and $\hat{\rho}^{dist}_{\mathbf{r}}$ according to Eq. (6.77) and share the same extremity and monotonicity as the calibration errors. Finally, for a state 'perpendicular' to all involved observables, errors vanish $\Delta(\hat{A},\hat{C},\hat{\rho}_x) = \Delta(\hat{B},\hat{D},\hat{\rho}_x) = 0$ shown by the violet line on the x-axis, since all probabilities in that case are uniform and there is no statistical difference between the outcomes.

$\hat{B} = \mathbf{b}\cdot\hat{\boldsymbol{\sigma}} = \hat{\sigma}_y$, while the approximating observables are chosen as $\hat{C} = \mathbf{c}\cdot\hat{\boldsymbol{\sigma}} = \sin\theta\,\hat{\sigma}_y + \cos\theta\,\hat{\sigma}_z$ and the modified output operator $\hat{D} = \mathbf{d}\cdot\hat{\boldsymbol{\sigma}} = \mathcal{E}^*_{ns}(\hat{B}) = \sum_k \hat{C}_k\hat{B}\hat{C}_k^\dagger = \sin^2\theta\,\hat{\sigma}_y + \frac{1}{2}\sin(2\theta)\,\hat{\sigma}_z$ (see. Eq. (6.13)). To obtain the apparatus' figure of merit it is necessary to prepare the input state $\hat{\rho}^{err}_{\mathbf{r}}$ with Bloch vector $\mathbf{r} = (\mathbf{a}-\mathbf{c})/\|\mathbf{a}-\mathbf{c}\|$, or in the case of disturbance $\mathbf{r} = (\mathbf{b}-\mathbf{d})/\|\mathbf{b}-\mathbf{d}\|$ associated with $\hat{\rho}^{dist}_{\mathbf{r}}$. If the measurement operators are known beforehand, then calibration errors are obtained by the eigenstates $\hat{\rho}_{\pm\mathbf{a}} = |\pm a\rangle\langle\pm a|$ and $\hat{\rho}_{\pm\mathbf{b}} = |\pm b\rangle\langle\pm b|$. The various cases are grouped and plotted in Fig. 6.20. One finds that there is no significant difference to Ozawa's theory in most cases. The calibration error and disturbance coincide with Eq. (6.52) and Eq. (6.56), the figures of merit are simple rescaled versions of these results. However,

Ozawa's definitions for projective measurements of the Pauli observables are state-independent. Hence, there is a discrepancy in the case for the states for which the expectation value vanish $\langle\hat{A}\rangle = \langle\hat{B}\rangle = \langle\hat{B}\rangle = \langle\hat{D}\rangle = 0$, because all outcome distributions are equiprobable. This is indicated in the plot Fig. 6.20 by the pruple line congruent with the abscissa.

6.6 Discussion

What remains to complete the topic is a discussion about the presented theories and to make certain criticisms and comparisons. The issue with the error-disturbance uncertainty relation Ineq. (6.37) are rooted in the error and disturbance definitions

$$\begin{aligned} \epsilon(\hat{A})^2 &= \mathrm{Tr}\left(\left(\hat{O}_A^{(2)} - \{\hat{O}_A, \hat{A}\} + \hat{A}^2\right)\hat{\rho}\right) \ , \\ \eta(\hat{B})^{(2)} &= \mathrm{Tr}\left(\left(\hat{O}_B^2 - \{\hat{O}_B, \hat{B}\} + \hat{B}^2\right)\hat{\rho}\right) \ . \end{aligned} \tag{6.78}$$

The anti-commutator necessitates a measurement of a product of two generally non-compatible observables $\mathrm{Tr}(\hat{O}_A\hat{A}\hat{\rho})$ which cannot be done in one shot. The two existing workarounds are the weak measurement method and the three-state method.

Although weak measurements are an elegant technique, it seems as if the principle of the simultaneous indeterminacy of the involved observables has been lead to absurdity. One of the principles of quantum physics, as already mentioned in sec. 3.6, is that there can be no information without disturbance. Looking again at Fig. 6.2 one recognizes that in this form the pointer value is tried to be determined without causing a disturbance, which is not possible. But if one then accepts the finite effectiveness of the weakly interacting apparatus, the objects of consideration changes to something like $\epsilon(\hat{A}, \hat{A}_w)$ and $\eta(\hat{A}_w, \hat{M})$, i.e the error of the weak observable and the disturbance thereby. That means instead of the intended error and the disturbance one measures a combination of the weak error before the apparatus and the disturbance of the strong measuring apparatus. This charade to define the value of $\hat{A}$ results in inaccurate approximations of the error.

In the case of the three state method, there is again the criticism that no 'direct' measurement

is made. This means that three different states have to be prepared with which separate, multiple measurements are performed. Now one could argue that the same is the case with the entropic definitions, with which the overlap is determined and the fourth type of uncertainty (see section heading of chapter 5) is being regarded. Here, however, the difference with Ozawa's definition is that the states are not eigenvectors of one operator. Furthermore, the number of states to be prepared does not increase with increasing dimension of the Hilbert state, but remains the same. The second aspect to consider is that the algebraic transformation Eq. (6.57) inevitably contains a term $(\hat{A} + \mathbb{1}) > 1$ and thus has a spectrum greater than 1. In the spin case, the operator is proportional to the projector and a factor of 4 has been introduced into the transformations in Eq. (6.58) or Eq. (6.59). In general, however, the super-normalized state $(\hat{A} + \mathbb{1})\hat{\rho}(\hat{A} + \mathbb{1})$ will have to be normed by the term $\mathrm{Tr}((\hat{A} + \mathbb{1})^2\hat{\rho})$, which still is not a valid effect $(\hat{A} + \mathbb{1})^2 \notin \mathcal{E}(\mathcal{H})$. If certain algebraic transformations and the multitude of necessary single measurements are accepted and error-disturbance can be formally reconstructed there is still the question whether the interpretation as mean square error is still meaningful. At least, the advantage over the weak valued method is that high accuracies in the measurements are achievable.

When $\hat{A}^2 = \hat{B}^2 = \hat{O}_A^2 = \hat{O}_B^2 = \mathbb{1}$ and the initial system state $\hat{\rho}$ is orthogonal to these, hence $\sigma(\hat{A}) = \sigma(\hat{B}) = 1$, then the optimal uncertainty relation is given in Ineq. (6.37). For the Pauli matrices the commutator $[\hat{A} = \hat{\sigma}_z, \hat{B} = \hat{\sigma}_y] = 2i\hat{\sigma}_x$ inserted to Eq. (6.39) and with the help of Eq. (3.50) then yields

$$D_{AB} = \frac{|2i|}{2} F(\hat{\rho}_x, \hat{\sigma}_x\hat{\rho}_x\hat{\sigma}_x) = F(\hat{\rho}_x, \hat{\rho}_x) = 1\,. \tag{6.79}$$

From this the uncertainty relation Ineq. (6.68) is obtained. For the Pauli matrices, not only is there square equal to $\hat{\sigma}_i^2 = \mathbb{1}$, but also the anti-commutator $\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbb{1}$ is proportional to the identity. Hence the generally state dependent error $\epsilon(\hat{A})$ and disturbance $\eta(\hat{B})$ are state independent in this particular case. The following problem occurs: let the input state be $\hat{\rho}_y$, then $\sigma(\hat{B}) = 0$ and therefore the requirements of the second geometric lemma Eq. (6.29) are not met and the proper description should be Ineq. (6.36). The bound C_{AB} is zero and likewise

$$D_{AB} = \frac{|2i|}{2} F(\hat{\rho}_y, \hat{\sigma}_x\hat{\rho}_y\hat{\sigma}_x) = F(\hat{\rho}_y, \hat{\rho}_{-y}) = 0\,. \tag{6.80}$$

Both Branciard's and Ozawa's uncertainty relation Ineq. (6.28) and Ineq. (6.36) just suggest

$\eta(\hat{B}) \geq 0$. But for projective qubit measurements Ineq. (6.68) holds under all circumstances since it is state independent and the inequality derived for a particular ensemble is valid generally. Under the current status, the inequality that holds for all states follows only for a few selective cases.

The theory of Busch-Lahti-Werner has merely been touched here, hence for a recent criticism consider [101]. In this thesis, it was only shown that BLW and Ozawa in the two-dimensional projective case are actually identical and therefore no quantitative statement can be made about which theory is preferable. In order for theoretical differences to crystallize, it would be possible to resort to the case of mixed states. For the totally mixed state $\Delta(\hat{A}, \hat{C}, \frac{\mathbb{1}}{d})$ would always yield a uniform probability and hence vanish similar to the case of $\Delta(\hat{A}, \hat{C}, \hat{\rho}_x) = 0$ for $\mathrm{Tr}(\hat{A}\hat{\rho}_x) = \mathrm{Tr}(\hat{C}\hat{\rho}_x) = 0$.

CHAPTER 7

Entropic noise-disturbance

A measurement uncertainty relation can also be put in terms of accessible information, i.e. how much information can be obtained by a given measurement and how much correlation is lost after the invasive observation. Buscemi et al. [26] introduced definitions of noise and disturbance, which quantify the performance of the measurement apparatuses by conditional Shannon entropies. Unlike in Ozawa's error-disturbance definition which indicate the statistical distance between the approximate observable and the observable actually measured, the entropic definitions measure the ability to infer the system observable from the measurement outcome. Other notable differences are that this entropic uncertainty relation only relies on the probabilities of the measurement outcomes and that it is conceptually state-independent.

7.1 Theory of entropic measurement uncertainty relations

The proposed information-theoretic uncertainty describes the trade-off between *entropic noise* $N(\mathcal{M}, \hat{A})$ and *entropic disturbance* $D(\mathcal{M}, \hat{B})$, where $\hat{A}$ and $\hat{B}$ are two generally non - commuting observables with eigenstates $|a_i\rangle$ and $|b_j\rangle$ respectively and $\{\hat{M}(m) = \hat{M}_m\}_m$ is a POVM associated with the apparatus' instrument $\mathcal{M}$. To understand the whole framework a closer look on the definitions of $N(\mathcal{M}, \hat{A})$ and $D(\mathcal{M}, \hat{B})$ is due. Consider the scheme in Fig. 7.1 to

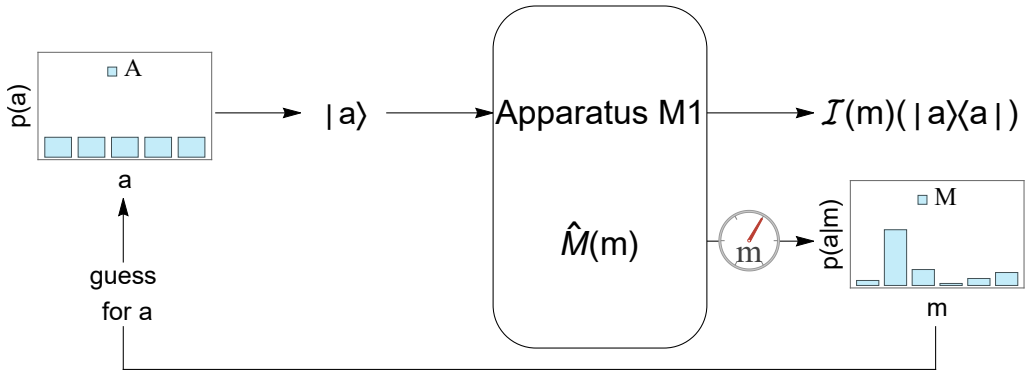


Fig. 7.1: The probabilistic histogram on the left indicates that a classical uniform distribution $p(a) = \frac{1}{d}$ of the eigenstates $|a\rangle$ each labeled by the eigenvalues a is prepared. The unknown states $|a\rangle$ are measured by apparatus M1 which ejects a conditional output state $\mathcal{I}(m)(|a\rangle\langle a|)$ together with a pointer value m from the measurement operator $\hat{M}_m$. The outcome provides a probability $p(a|m)$ conditioned on the result m from which the initial input state can be guessed. The ability to guess or discriminate all input states by the measurement is given by taking an average over all results indicated in the figure by the right side histographic distribution.

get a grasp on the idea of entropic noise. Initially, the eigenstates of $\hat{A}$ are prepared with equal probability and measured by the apparatus M1. In a d-dimensional Hilbert space this implies that the the joint probability distribution for the a-th input state and outcome m is given by

$$p(m,a) = p(a)p(m|a) = \frac{1}{d}p(m|a)\,, \tag{7.1}$$

where the conditional probability is

$$p(m|a) = \mathrm{Tr}\left(\mathcal{I}(m)(|a\rangle\langle a|)\right) = \mathrm{Tr}(\hat{M}_m\,|a\rangle\langle a|)\,. \tag{7.2}$$

The information-theoretic noise of such a measurement is defined to be the conditional Shannon entropy Eq. (2.36), so

$$\begin{aligned} N(\mathcal{M},\hat{A}) &:= H(A|M) = H(A,M) - H(M) \\ &= \sum_m p(m)H(A|M=m) \\ &= -\sum_{m,a} p(m)p(a|m)\log(p(a|m))\ , \end{aligned} \tag{7.3}$$

where $H(A, M)$ denotes the entropy of the joint distribution $p(m, a)$ associated with the random variables $A \ni a$ and $M \ni m$, and $H(M)$ denotes the entropy of $p(m)$. Note that the logarithm base is arbitrary, but is chosen here to be 2, corresponding to bits. This noise is the information-theoretic uncertainty of the input value of A, conditioned on the obtained measurement outcome. It vanishes only if the input value can be obtained without error from the outcome. The dependency in the conditional probabilities can also be swapped with Bayes' theorem Eq. (2.8)

$$p(a|m) = \frac{p(m)}{p(a)} p(m|a) = \frac{\mathrm{Tr}(\hat{M}_m \left|a\right\rangle \left\langle a\right|)}{\mathrm{Tr}(\hat{M}_m)} \,. \tag{7.4}$$

This inversion of the conditional probability also has the consequence that the converse physical situation is realizable in which the input state $\hat{\rho}_m = \frac{\hat{M}_m}{\mathrm{Tr}(\hat{M}_m)}$ with initial distribution $p(m) = \frac{1}{d}\mathrm{Tr}(\hat{M}_m)$ determines the observable $\hat{A}$. The entropic noise operator can be defined equivalently with the notation of Eq. (5.11) as

$$N(\mathcal{M}, \hat{A}) = \sum_m p(m) H(\hat{A}|\hat{\rho}_m) \,. \tag{7.5}$$

The definition of entropic disturbance bears strong resemblance to the noise case. The unidentified eigenstate $|b\rangle$ in the d-dimensional Hilbert space with uniform distribution $p(b) = \frac{1}{d}$ is measured by apparatus M1, but this time a subsequent projective measurement of observable $\hat{B}$ is made in the second apparatus M2. The experimental set-up needs to be modified, to allow for possible error correction as illustrated in Fig. 7.2. The error correction operation $\mathcal{E}$ is a CPTP map that can reduce the loss of information and depend the state and the meter outcome following the measurement of $\hat{M}$. Denote B' as the random variable with respect to the result of the subsequent measurement of $\hat{B}$ labeled by b', then one can calculate the disturbance of observable $\hat{B}$, for a given error correction operation $\mathcal{E}_m$, as the conditional entropy

$$\begin{aligned} D_{\mathcal{E}}(\mathcal{M}, \hat{B}) &:= H(B|B') = H(B, B') - H(B') \\ &= -\sum_{b,b'} p(b')p(b|b') \log(p(b|b')) \,. \end{aligned} \tag{7.6}$$

which can be evaluated with $p(b', b) = p(b)p(b'|b) = \frac{1}{d}p(b'|b)$ and

$$p(b'|b) = \sum_m \mathcal{E}_m \left(\mathrm{Tr}\left(\mathcal{I}(m)(|b\rangle\langle b|)\right)\right) \,. \tag{7.7}$$

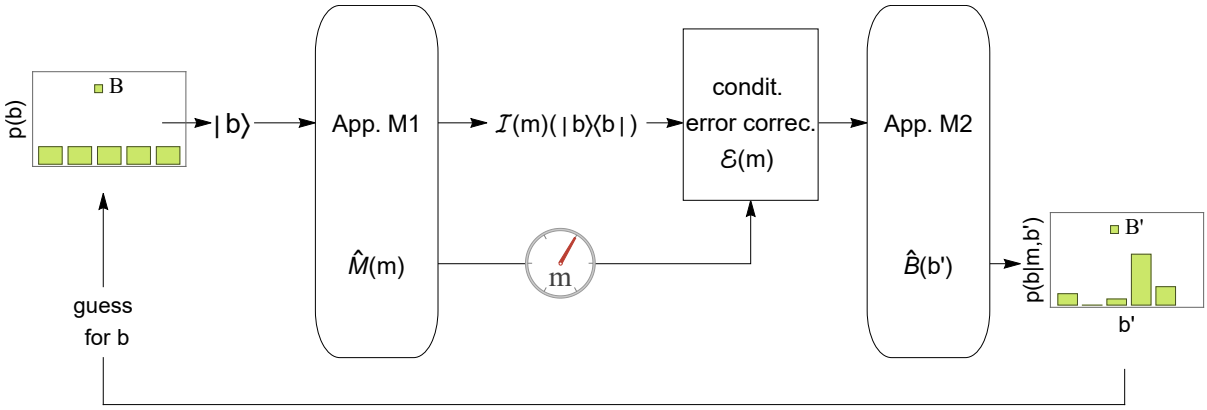


Fig. 7.2: Scheme for the entropic disturbance measurement. The classical uniform distributed $p(b) = \frac{1}{d}$ eigenstates $|b\rangle$ are send into the set-up. After a measurement in apparatus M1 a result m and an output state is obtained. Depending on the outcome m an error correction procedure conditioned on the result m on $\mathcal{I}(m)(|b\rangle\langle b|)$ can be applied in order to reduce the loss of correlation due to M1. The second apparatus then measures observable $\hat{B}$ and generates outcome b', which can be used to infer the initial input state.

The information-theoretic disturbance then corresponds to the minimum value of the disturbance optimized over all possible error-correction operations

$$D(\mathcal{M}, \hat{B}) := \inf_{\mathcal{E}} D_{\mathcal{E}}(\mathcal{M}, \hat{B}). \tag{7.8}$$

The noise-noise uncertainty relation measures the mutual exclusiveness to which an apparatus can simultaneously perform as an $\hat{A}$-measuring device and a $\hat{B}$-measuring device and is given here by

$$N(\mathcal{M}, \hat{A}) + N(\mathcal{M}, \hat{B}) \geq -\log(c) = -\log(\max_{i,j} |\langle a_i|b_j\rangle|^2) \ , \tag{7.9}$$

while the entropic noise-disturbance uncertainty reads

$$N(\mathcal{M}, \hat{A}) + D(\mathcal{M}, \hat{B}) \geq \log(c) := -\log(\max_{i,j} |\langle a_i|b_j\rangle|^2) \tag{7.10}$$

quantifying the degree of joint measurability in a successive set-up. These inequalities are conceptually similar to the Maassen-Uffink relation Ineq. (5.13). The accessible region of the preparation uncertainty [27] for the state $\hat{\rho}$ is defined by

$$E(\hat{A},\hat{B}) = \{(H(\hat{A}|\hat{\rho}), H(\hat{B}|\hat{\rho})) \mid \hat{\rho} \text{ is a density matrix}\} \ . \tag{7.11}$$

As implied by Eq. (7.5) and in contrast to region $E(\hat{A},\hat{B})$ the measurement uncertainty region for noise-noise measurements is given by

$$\begin{aligned} R_{NN}(\hat{A},\hat{B}) &= \{(N(\mathcal{M},\hat{A}), N(\mathcal{M},\hat{B})) \mid \mathcal{M} \text{ is an instrument}\} \\ &= \left\{ \sum_m p(m) \left(H(A|\hat{\rho}_m), H(B|\hat{\rho}_m)\right) \middle| \{p(m),\hat{\rho}_m\}_m \right\} \end{aligned} \tag{7.12}$$

respectively for noise-disturbance by

$$R_{ND}(\hat{A},\hat{B}) = \{(N(\mathcal{M},\hat{A}), D(\mathcal{M},\hat{B})) \mid \mathcal{M} \text{ is a quantum instrument}\} . \tag{7.13}$$

The weighted ensemble $\{p(m),\hat{\rho}_m\}_m$ is a particular convex combination which satisfies the summation rule $\sum_m p(m)\hat{\rho}_m = \frac{\mathbb{1}}{d}$. In [27] theoretical studies are conducted about the uncertainty regions to link preparation and measurement uncertainties and find out how their limits are correlated:

1. Proposition: The entropic noise-noise region $R_{NN}(\hat{A},\hat{B}) \subseteq \text{conv}(E(\hat{A},\hat{B}))$ is in the convex hull of the preparation region $E(\hat{A},\hat{B})$.

2. Proposition: The entropic noise-disturbance region can be related to the entropic noise-noise region via $R_{ND}(\hat{A},\hat{B}) \subseteq \text{cl}(R_{NN}(\hat{A},\hat{B})) \not\subseteq R_{NN}(\hat{A},\hat{B})$, where cl denotes the monotone closure, i.e. the closure under increasing either coordinate.

7.2 Experiment 1 - Entropic noise-disturbance measurement uncertainty

7.2.1 Prediction for projective qubit measurements

If a projective measurement of $\hat{M} = \boldsymbol{\sigma}\cdot\mathbf{m}$ is performed on such an eigenstate (see Fig. 7.1), then the measurement outcome will be $m = \pm 1$, with conditional probability distribution $p(m|a) = \frac{1}{2}(1 + a\,m\,\mathbf{a}\cdot\mathbf{m})$. Hence, if the two eigenstates are sent with equal probabilities, the

joint probability of input a and output m is

$$p(m,\alpha) = p(a)\,p(m|a) = \frac{1}{4}(1 + a\,m\,\mathbf{a}\cdot\mathbf{m}) \ . \tag{7.14}$$

Note it follows immediately that the average input and output distributions are given by the marginals

$$p(a) = \frac{1}{2} = p(m) \ . \tag{7.15}$$

From Eq. (7.3) the noise may be calculated as

$$N(\mathcal{M},\hat{A}) = H_b\left(\mathbf{a}\cdot\mathbf{m}\right) \ , \tag{7.16}$$

where H_b denotes the binary entropy function as introduced by Eq. (2.33).

To calculate the information-theoretic disturbance of $\hat{B}$, the eigenstates of $\hat{B}$ must have equal prior probabilities, and the correlation between b and b' must be determined. In particular, from the measured joint probability distribution $p(b,b')$ one can calculate the disturbance of $\hat{B}$, for a given error correction operation $\mathcal{E}$, as the conditional entropy. The first relation containing expression Eq. (7.6) can be experimentally tested for any particular error correction operation, while inequality Ineq. (7.10) can only be tested in practice if the optimal error correction operation is known. Note that, assuming a von Neumann-like collapse of the initial state in a projective Lüders instrument, the theoretical joint probability distribution of b, m and b' may be calculated as

$$p(b,m,b') = p(b,m)\,p(b'|b,m) = \frac{1}{4}\left(1 + bm\,\mathbf{b}\cdot\mathbf{m}\right)\,\langle b'\mathbf{b}|\mathcal{E}(|m\mathbf{m}\rangle\langle m\mathbf{m}|)|b'\mathbf{b}\rangle \ . \tag{7.17}$$

The joint distribution $p(b,b')$ can then be found by summing over m. Two particular error correction operations are of interest, which are considered in turn below.

No error correction: If no error correction is carried out, i.e., $\mathcal{E} \equiv \mathbb{1}$, then Eq. (7.17) yields the joint probability distribution

$$p(b,b') = \sum_{m=\pm 1}\frac{1}{8}\left(1 + bm\,\mathbf{b}\cdot\mathbf{m}\right)\left(1 + b'm\,\mathbf{b}\cdot\mathbf{m}\right) = \frac{1 + bb'(\mathbf{b}\cdot\mathbf{m})^2}{4} \ . \tag{7.18}$$

Hence, the corresponding disturbance is given by

$$D_{\mathbb{1}}(\mathcal{M}, \hat{B}) = H_b\left((\mathbf{b}\cdot\mathbf{m})^2\right) \ . \tag{7.19}$$

Note that the disturbance is zero for $\mathbf{m} = \pm\mathbf{b}$, as one would expect, also that $D_{\mathbb{1}}(M, B) > N(M, B)$ (where the right hand side is calculated analogously to Eq. (7.16)). Thus, by performing no error correction at all, the disturbance to $\hat{B}$ is greater than the noise added to any estimate of $\hat{B}$ from the measurement of $\hat{M}$.

Optimal error correction in projective measurements: Define the error correction operation $\mathcal{E}_{\text{opt}}$ by

$$\mathcal{E}_{\text{opt}}(|m\mathbf{m}\rangle\langle m\mathbf{m}|) := \begin{cases} |m\mathbf{b}\rangle, & \mathbf{b}\cdot\mathbf{m} \geq 0 \\ |-m\mathbf{b}\rangle, & \mathbf{b}\cdot\mathbf{m} < 0. \end{cases} \tag{7.20}$$

Hence, the output state is rotated from $|m\mathbf{m}\rangle$ to $|m\mathbf{b}\rangle$ if the angle between directions $\mathbf{m}$ and $\mathbf{b}$ is no more than π radians, and to $|-m\mathbf{b}\rangle$ otherwise. Note that, rather than implementing $\mathcal{E}_{\text{opt}}$ via rotations, one can simply discard the state $|m\mathbf{m}\rangle$ and prepare the desired state directly. Recalling b' is the outcome of a subsequent measurement of $\hat{B}$, as depicted in Fig. 7.2, it immediately follows that

$$b' := \begin{cases} m, & \mathbf{b}\cdot\mathbf{m} \geq 0 \\ -m, & \mathbf{b}\cdot\mathbf{m} < 0. \end{cases} \tag{7.21}$$

Thus, the error correction acts to ensure that b' is determined solely from knowledge of the result m. Further, this formula immediately implies that $p(b'|m) = \frac{1}{2}(1 + b'm\cdot\text{sign}\{\mathbf{b}\cdot\mathbf{m}\})$. It follows, using Eq. (7.17), that the corresponding joint probability distribution of b and b' is given by

$$p(b, b') = \sum_{m=\pm1} p(b, m)\, p(b'|b, m) = \sum_{m=\pm1} p(b, m)\, p(b'|m) = \frac{1 + b\, b'|\mathbf{b}\cdot\mathbf{m}|}{4} \ . \tag{7.22}$$

The corresponding disturbance is therefore given by

$$D_{\mathcal{E}_{\text{opt}}}(\mathcal{M}, \hat{B}) = H_b\left(\mathbf{b}\cdot\mathbf{m}\right) = N(\mathcal{M}, \hat{B}) \leq D_{\mathbb{1}}(\mathcal{M}, \hat{B}) \ . \tag{7.23}$$

where $H_b(x) = H_b(1 - x)$ has been used. Note that the equality with $N(\mathcal{M}, \hat{B})$ means that the error correction and measurement of $\hat{B}$ add no further noise. This is the minimum possible disturbance, i.e. $D_{\mathcal{E}_{\text{opt}}}(\mathcal{M}, \hat{B}) = D(\mathcal{M}, \hat{B})$. This is easily shown by noticing that, if the

measurement along m is projective (as is assumed here), then the state of the quantum system coming out of the measurement apparatus is a function of the outcome m. In other words, the output quantum system is in fact a function of the output random variable M.

An experiment testing the predictions presented above was successfully performed with neutrons [102] with the exact same construction as shown in Fig. 6.3. The first coil DC1 makes the state preparation $|a\mathbf{a}\rangle$ and $|b\mathbf{b}\rangle$, DC2 plus analyzer-1 perform the measurement $\hat{M} = \cos\theta\hat{\sigma}_z + \sin\theta\hat{\sigma}_y$, DC3 is once more responsible for error correction and DC4 plus analyzer-2 makes the measurement of $\hat{B} = \hat{\sigma}_y$. Then for every combination of the triple $\{a, m, b'\}$, $\{b, m, b'\}$ a neutron intensity $I_{\{a,m,b'\}}$, $I_{\{b,m,b'\}}$ is recorded, whose ratio provides a probability through

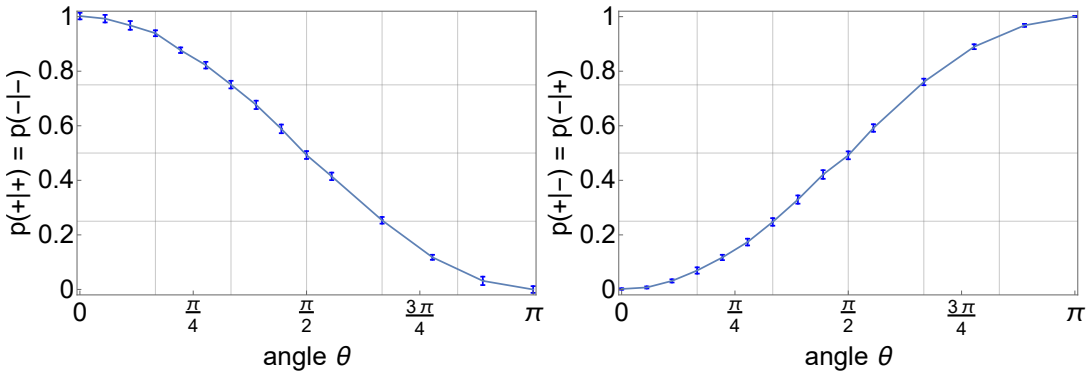
$$\begin{aligned} p(a) &= \frac{\sum_{m,b'} I_{\{a,m,b'\}}}{\sum_{a,m,b'} I_{\{a,m,b'\}}} , \qquad p(m|a) = \frac{\sum_{b'} I_{\{a,m,b'\}}}{\sum_{m,b'} I_{\{a,m,b'\}}} . \\ p(b) &= \frac{\sum_{m,b'} I_{\{b,m,b'\}}}{\sum_{b,m,b'} I_{\{b,m,b'\}}} , \qquad p(m|b) = \frac{\sum_{b'} I_{\{b,m,b'\}}}{\sum_{m,b'} I_{\{b,m,b'\}}} . \end{aligned} \tag{7.24}$$

A profound difference of the conducted experiment and the presented theory is that the conditioning of the states and measurement outcomes $m \leftrightarrow a$ have been swapped. Converting Eq. (7.24) according to Bayes' rule Eq. (2.8) it is possible to evaluate the information theoretic definitions of noise and disturbance nevertheless.

7.2.2 Results

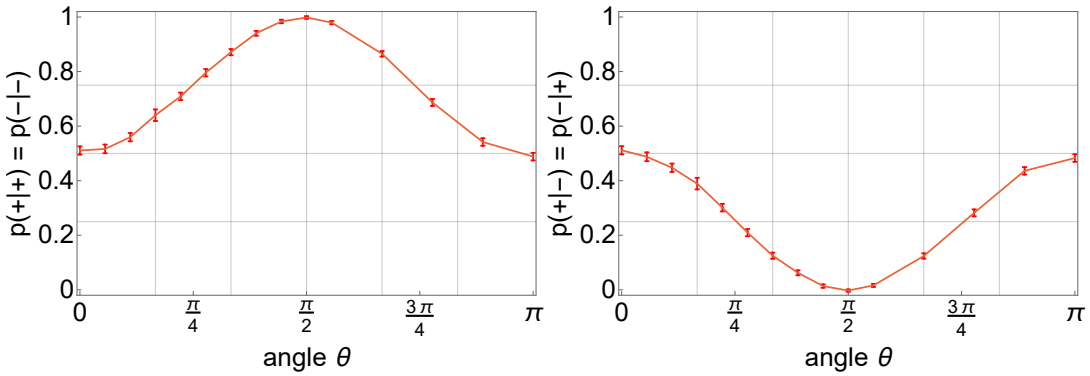
With entropies everything is determined by measuring the conditional probabilities which are plotted in Fig. 7.3. To determine the noise, the probability $p(m|a)$ is recorded, which is represented by the blue curves. The angle $\theta = 0$ means that the projectors $(1 \pm \mathbf{m}\cdot\hat{\boldsymbol{\sigma}})/2$ let the neutrons corresponding to eigenstates $a = \pm 1$ of the same sign pass completely, while spins with opposite values do not reach the detector at this angle θ. Of course, turning the projectors by $\theta = \pi$ causes a reversal of the situation.

For the disturbance, the prepared eigenstates are $|\pm b\rangle$, so the probability at first is $\frac{1}{2}$. The extreme values at $\frac{\pi}{2}$ come about for a similar reason as in the noise case. The first apparatus



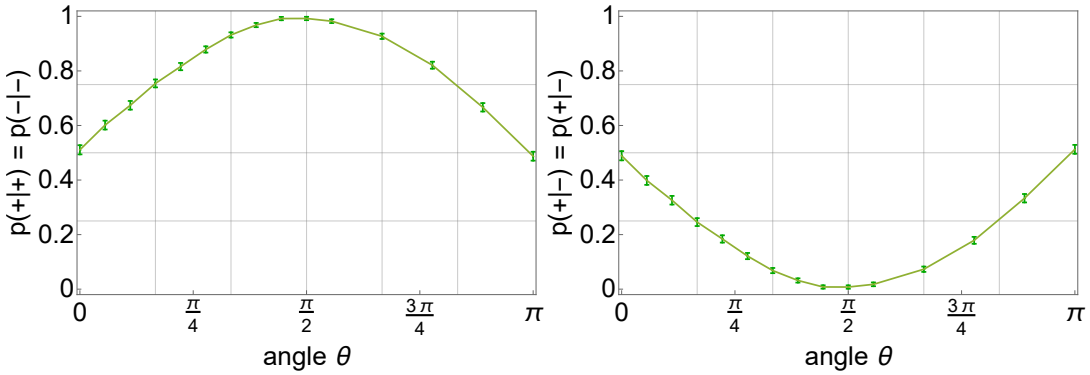
(a) Probability $p(m|a)$ with $m = a$.

(b) Probability $p(m|a)$ with $m = -a$



(c) [Probability $p_{uncor}(b'|b)$ with $b' = b$.

(d) Probability $p_{uncor}(b'|b)$ with $b' = -b$.



(e) Probability $p_{cor}(b'|b)$ with $b' = b$.

(f) Probability $p_{cor}(b'|b)$ with $b' = -b$.

Fig. 7.3: Conditional probabilities obtained by Eq. (7.24), from which the entropic noise and disturbance are determined. In the upper panels the conditional probabilities $p(m|a)$ are plotted versus the angle of the projective measurement with Bloch vector $\mathbf{m} = \cos\theta\,\hat{\mathbf{e}}_z + \sin\theta\,\hat{\mathbf{e}}_y$. In the middle panel are the red curves of the probability measurement $p_{uncor}(b'|b)$ required for the uncorrected disturbance, while the lowest panel show the corrected conditional probabilities $p_{cor}(b'|b)$ in green.

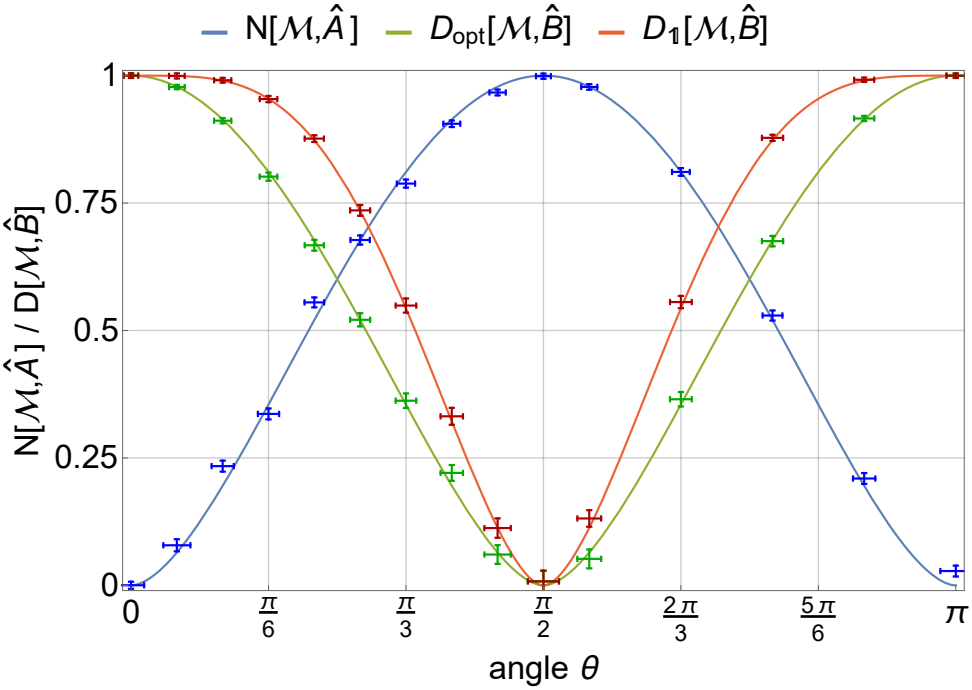


Fig. 7.4: Results of entropic noise $N(\mathcal{M}, \hat{A})$ (blue) as predicted by Eq. (7.16) for $\mathbf{a} = \hat{\mathbf{e}}_z$, corrected disturbance Eq. (7.23) (green) and uncorrected disturbance Eq. (7.19) (red) for $\mathbf{b} = \hat{\mathbf{e}}_y$ in a projective measurement with $\mathbf{m} = \cos\theta\,\hat{\mathbf{e}}_z + \sin\theta\,\hat{\mathbf{e}}_y$.

M1 does not disturb the prepared eigenstates $|\pm b\rangle$ at this point, but in the second apparatus M2 the eigenstate can either be in the range or in the kernel of the projector that is making the disturbance measurement. One recognizes very well how the unitary correction procedure, performed after apparatus M1, increases the curvature of the probability function from the uncorrected curves in Figs. 7.3(c) and 7.3(d) to the corrected ones in Figs. 7.3(e) and 7.3(f).

Varying the angle θ of the projector $\hat{M}$ with increment $\Delta\theta = 10°$ in the interval $[0°, 90°]$ and with $\Delta\theta = 20°$ in $[100°, 180°]$ the oscillation of noise and disturbance have been recorded which are plotted in Fig. 7.4. The information theoretic aspect of the results can be understood as preparing a message in form of the system's state while the apparatus has to acquire knowledge about the system through the PVM with outcomes m. At $\theta = 0$ the measurement observable is $\hat{M} = \hat{\sigma}_z$ so that for the two eigenstates $|a\mathbf{a}\rangle$ the probability distribution has maximal peaks at $\{p(a)\} = \{1, 0\}$ or $\{p(a)\} = \{0, 1\}$, hence the Shannon entropy is zero, or in other words there is no information obtained by the measurement. An alteration of the polar angle reduces the

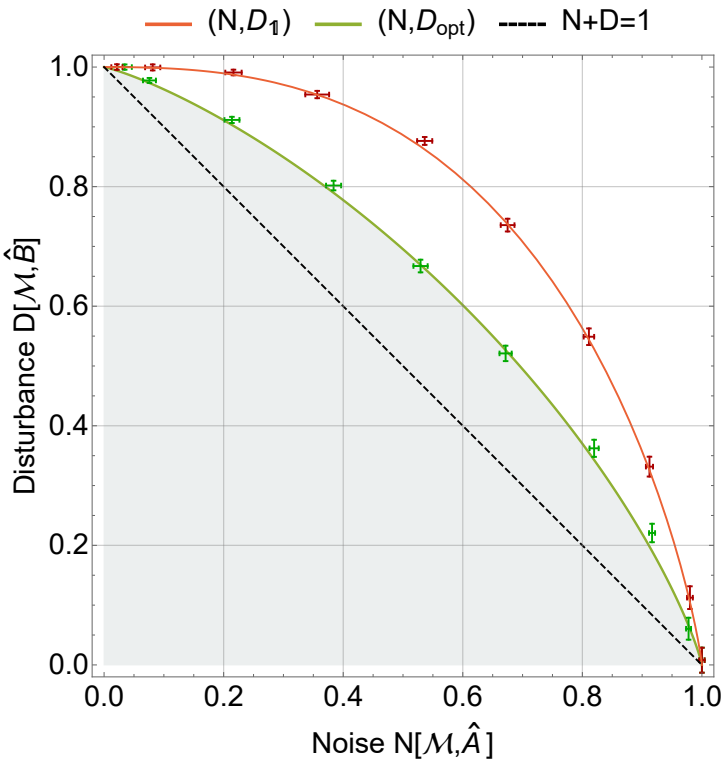


Fig. 7.5: Parametric plot of information noise $N(\mathcal{M}, \hat{A})$ versus disturbance $D(\mathcal{M}, \hat{B})$. The measured points in red corresponding to an uncorrected uncertainty trade-off are generally equal or above the optimized green points. Furthermore, the black dashed line $N(\mathcal{M}, \hat{A}) + D(\mathcal{M}, \hat{B}) = 1$ as suggested by Eq. (7.10) cannot be saturated, except for the top-right and bottom-left points in the corner.

capability to foretell the measurement result until at at $\theta = \frac{\pi}{2}$ each outcome will be equiprobable $\{p(a)\} = \{\frac{1}{2}, \frac{1}{2}\}$, the uncertainty with which one of the two results is obtained and the information gained by the meter value is maximal. Unlike in Ozawa's definition of error, the information-theoretic noise is the largest in the case of mutually unbiased states, since both operators $\pm\hat{A}$ can perfectly discriminate all $|a_i\rangle$.

The disturbance is reciprocal to the noise, when sending the two eigenstates $|b\mathbf{b}\rangle$ the conditional probabilities $p(b'|b)$ would be perfectly correlated without the first measurement apparatus. For $\hat{M} = \hat{\sigma}_z$ the correlation after M1 disappears maximally and information gain by the second measurement has an extreme as indicated in Fig. 7.4. Rotating $\hat{M}$ towards $\hat{\sigma}_y$ reduced the loss of

correlation in device M1 and allows for a perfect inference of the input states again at $\theta = \frac{\pi}{2}$.

The decisive uncertainty behavior is seen in Fig. 7.5 which displays three different curves. The black dashed line corresponds to Ineq. (7.10) with the boundary $c = 1$; obviously the measured points are above the general prediction, however the points in the case of Pauli observables seem to not get close to the boundary except for the extremes $(N, D) = (1, 0)$ and $(N, D) = (0, 1)$. The red theoretical line corresponding to $(H_b(\cos\theta), H_b(\sin^2\theta))$ lies significantly above the green line $(H_b(\cos\theta), H_b(\sin\theta))$, whose implicit inequality describing the uncertainty relation is described by

$$g\left(N(\mathcal{M}, \hat{\sigma}_z)\right) + g\left(D(\mathcal{M}, \hat{\sigma}_y)\right) \leq 1 \ . \tag{7.25}$$

The function $g(x) = H_b^{-1}(x)$ is the inverse function[6] of the binary entropy Eq. (2.33) in the domain $x \in [0, 1]$. The uncertainty relation in terms of the inverse function therefore corresponds to the saturated limit when the measurement is performed by a Lüders instrument [27].

7.3 Experiment 2 - Entropic noise-noise measurement uncertainty

7.3.1 Prediction for general qubit measurements

After the first experiment [102], a revision of the theory suggested the use of general POVM measurements rather than a projective measurement. It turned out [27] that for such a change of observables the inequality Ineq. (7.25) is not always satisfied and it is possible find pairs of entropic noise and disturbance values that are below the boundary. A proper analysis of noise and disturbance for unsharp observables (see sec. 3.3) is very difficult and has not been entirely solved so far. Therefore, only entropic noise-noise uncertainty principle [103] are considered here for the time being, afterward the latest findings in the other case will be ventured briefly.

In chapter 5 it was stated that the Maassen-Uffink inequality is not generally tight and a lower bound for arbitrary $\mathbf{a}$ and $\mathbf{b}$ is found by minimizing the left-hand side in Eq. (5.15); this region will be identified with $E(\hat{A}, \hat{B})$ as defined in Eq. (7.11). Combining previous results an optimal entropic noise-noise uncertainty relation can be constituted in case of a two dimensional quantum system. First, consider the tight preparation uncertainty relation for qubits given

[6] This function has no analytic inverse, making many calculations with it rather cumbersome.

by Eq. (5.22), then substitute the expectation value with the inverse entropy, since $H(A) = H_b(\mathbf{a}\cdot\mathbf{m}) = H_b(\langle\hat{A}\rangle)$, i.e.

$$g\left(H(A)\right)^2 + g\left(H(B)\right)^2 - 2\left|\mathbf{a}\cdot\mathbf{b}\right| g\left(H(A)\right) g\left(H(B)\right) \leq \left(1-(\mathbf{a}\cdot\mathbf{b})^2\right)\|\mathbf{r}\|^2 \ . \tag{7.26}$$

In 1. proposition it is suggested that the measurement uncertainty region $R_{NN}(\hat{A},\hat{B})$ is given by the convex hull of $E(\hat{A},\hat{B})$, therefore

$$R_{NN}(\hat{A},\hat{B}) = \mathrm{conv}\left\{(s,t)\,\middle|\, g(s)^2 + g(t)^2 - 2\left|\mathbf{a}\cdot\mathbf{b}\right| g(s)g(t) \leq \left(1-(\mathbf{a}\cdot\mathbf{b})^2\right)\|\mathbf{r}\|^2\right\} \ . \tag{7.27}$$

The region shows different behavior depending on the critical angle $\alpha_c \cong 0.691$ obtained by solving Eq. (5.15). In Fig. 5.3 it was visualized that at $|\mathbf{a}\cdot\mathbf{b}| \cong \cos(0.691) \cong 0.391$ a bifurcation of the values occurs, which in this instance implies that $E(\hat{A},\hat{B})$ spans a convex area when $|\mathbf{a}\cdot\mathbf{b}| \geq 0.391$. Consequently, the uncertainty relation satisfying this conditions is given by

$$g\left(N(\mathcal{M},\hat{A})\right)^2 + g\left(N(\mathcal{M},\hat{B})\right)^2 - 2\left|\mathbf{a}\cdot\mathbf{b}\right| g\left(N(\mathcal{M},\hat{A})\right) g\left(N(\mathcal{M},\hat{B})\right) \leq \left(1-(\mathbf{a}\cdot\mathbf{b})^2\right)\|\mathbf{r}\|^2 \tag{7.28}$$

and the preparation and measurement uncertainty region are congruent $R_{NN}(\hat{A},\hat{B}) = E(\hat{A},\hat{B})$. On the other hand, when $|\mathbf{a}\cdot\mathbf{b}| \leq 0.391$ then the region $E(\hat{A},\hat{B})$ is not convex and has no implicit analytic expression, except in the case $|\mathbf{a}\cdot\mathbf{b}| = 0$ for which the Maassen-Uffink like relation Ineq. (7.9)

$$N(\mathcal{M},\hat{A}) + N(\mathcal{M},\hat{B}) \geq 1 \tag{7.29}$$

is tight. To grasp the procedure of measuring this uncertainty relation projective measurements must be considered first. With $p(+1) = p(-1) = \frac{1}{2}$ and $H(\hat{A}|\hat{\rho}_+) = H(\hat{A}|\hat{\rho}_-)$ the noise in Eq. (7.5) evaluates to $N(\mathcal{M},\hat{A}) = H(\hat{A}|\hat{\rho}_+)$. This means that for projective two-outcome measurements the reachable entropic noises span the area $E(\hat{A},\hat{B}) = \{(H(\hat{A}|\hat{\rho}), H(\hat{B}|\hat{\rho}))\} = \{(N(\mathcal{M},\hat{A}), N(\mathcal{M},\hat{B}))\}$ when the measurement apparatus $\mathcal{M}$ implements the PVM $\{\hat{\rho}_+,\hat{\rho}_-\}$. Now, concerning points contained in the complement $R_{NN}(\hat{A},\hat{B}) \setminus E(\hat{A},\hat{B})$, it is clear that these points are obtained by the combination that makes the region convex, so if $\{\hat{\rho}_{1+},\hat{\rho}_{1-}\}$ corresponds to one point at which the area changes its convexity and $\{\hat{\rho}_{2+},\hat{\rho}_{2-}\}$ to the other point; then joining these by the line $\lambda\{\hat{\rho}_{1+},\hat{\rho}_{1-}\} + (1-\lambda)\{\hat{\rho}_{2+},\hat{\rho}_{2-}\}$ gives a POVM that allows

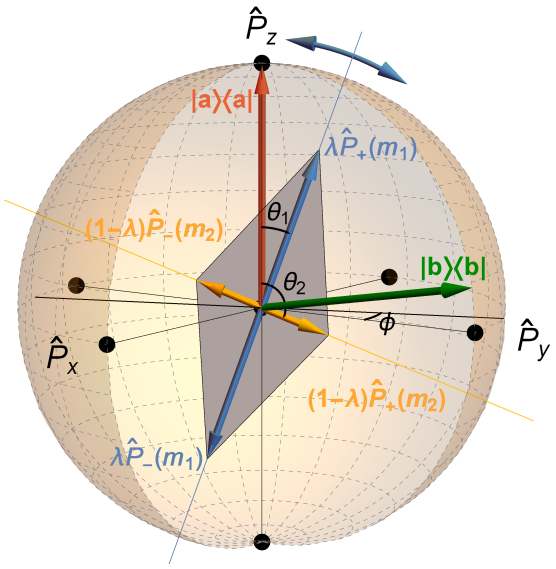


Fig. 7.6: Graphical illustration of the POVM in Eq. (7.30) in the operator representation. The blue vectors correspond to $\hat{P}_{\pm}(\mathbf{m}_1)$ with length $\|\hat{M}_1\| = \|\hat{M}_2\| = \lambda$ of the POVM operators $\hat{M}_m$, while the yellow arrows correspond to $\hat{P}_{\pm}(\mathbf{m}_2)$ with length $1-\lambda = \|\hat{M}_3\| = \|\hat{M}_4\|$. Because each axes-vectors face in opposite directions, the spanned area is always a parallelogram, albeit it can be very deformed. The red arrow represents the pure state $\hat{\rho}_a = |a\rangle\langle a|$ and the green one $\hat{\rho}_b = |b\rangle\langle b|$, which are drawn coplanar with the parallelogram of the measurement operators. Overall, the parallelogram POVM has 4 degrees of freedom: angles θ_1, θ_2, ϕ and the lengths denoted by $\{\lambda, 1-\lambda\}$.

to obtain values outside of $E(\hat{A}, \hat{B})$. Swapping the role in definition Eq. (7.5) $\hat{A} \leftrightarrow \hat{\rho}_m$, one concludes that with the four-outcome POVM

$$\hat{M}_m = \Big\{\lambda\hat{P}_+(\mathbf{m}_1), \lambda\hat{P}_-(\mathbf{m}_1), (1-\lambda)\hat{P}_+(\mathbf{m}_2), (1-\lambda)\hat{P}_-(\mathbf{m}_2)\Big\} \tag{7.30}$$

it is possible to reach the optimal information theoretic noise-noise trade-off, which quantifies how well the unknown eigenstates of observables $\hat{A}, \hat{B}$ can both be inferred by the measurement apparatus. A graphical interpretation of expression Eq. (7.30) is illustrated in Fig. 7.6.

So far only the uncertainty region, i.e parametric relationship between the noises has been specified. It remains as a task to find out the explicit behavior $N(\mathcal{M}, \hat{A})$ as a function of the

relative angle between the observables' Bloch vectors. The marginal probabilities are given by

$$p(m) = \frac{1}{d}\mathrm{Tr}\left(\hat{M}_m\right) = \begin{cases} \frac{\lambda}{2}\,, & m = \{1,2\}\,, \\ \frac{1-\lambda}{2}\,, & m = \{3,4\}\,. \end{cases} \tag{7.31}$$

and with $\hat{\rho}_m = \frac{\hat{M}_m}{\mathrm{Tr}(\hat{M}_m)} = \{\hat{P}_+(\mathbf{m}_1), \hat{P}_-(\mathbf{m}_1), \hat{P}_+(\mathbf{m}_2), \hat{P}_-(\mathbf{m}_2)\}$ it follows that

$$p(a|m) = \mathrm{Tr}(\hat{\rho}_m\hat{\rho}_a) = \begin{cases} \frac{1}{2}\left(1 + a\tilde{m}\,\mathbf{m}_1\cdot\mathbf{a}\right)\,, & m = \{1,2\}\,, \\ \frac{1}{2}\left(1 + a\tilde{m}\,\mathbf{m}_2\cdot\mathbf{a}\right)\,, & m = \{3,4\}\,, \end{cases} \tag{7.32}$$

where a and $\tilde{m}$ both range over ± 1. Therefore inserting both results into Eq. (7.3) the noise of operator $\hat{A}$ (likewise for $\hat{B}$) is the convex sum of the binary Shannon entropies with the respective scalar products $\mathbf{m}_i \cdot \mathbf{a}$

$$N(\mathcal{M}, \hat{A}) = \lambda H_b(\mathbf{m}_1\cdot\mathbf{a}) + (1-\lambda)H_b(\mathbf{m}_2\cdot\mathbf{a})\,. \tag{7.33}$$

7.3.2 Experimental procedures and data acquisition

A generalized measurement is typically realized by restricting observations to a subspace of a larger Hilbert space, which are part of an orthogonal measurement but cannot in general be described as projections on the subspace. Since Eq. (7.30) (see also Fig. 7.6) is a statistical combination of two projections and the post-measurement states are irrelevant, one can adapt a successive measurement scheme that can realize the POVM by performing a sequential projective measurement. This reproduces the probability distribution of the outcomes for an arbitrary input state. A succession of two Lüders instrument does the trick, where additionally two unitary channels are also introduced to account for possible transformations

$$p(a|m) = \mathrm{Tr}\left(\mathcal{I}_L(\tilde{m})\left(\mathcal{E}_U^a\left(\mathcal{I}_L(\lambda_i)(\hat{\rho}_\theta)\right)\right)\right)\,. \tag{7.34}$$

The realization of output m will be identified with the four result of two separate outcomes $p(a|m) = p(a|\lambda_i, \tilde{m})$.

In Fig. 7.7 a schematic example of the entire procedure for the determination of noise $\mathcal{N}(\mathcal{M}, \hat{A})$ is illustrated. The transmissivity of the first analyzer determines the probability weighting and is controlled in the coil DC1, which rotates the incoming spin $\hat{\rho} = |{+z}\rangle\langle{+z}|$ by applying a magnetic field B_x in x-direction. The rotation angle in the Larmor precession is given by

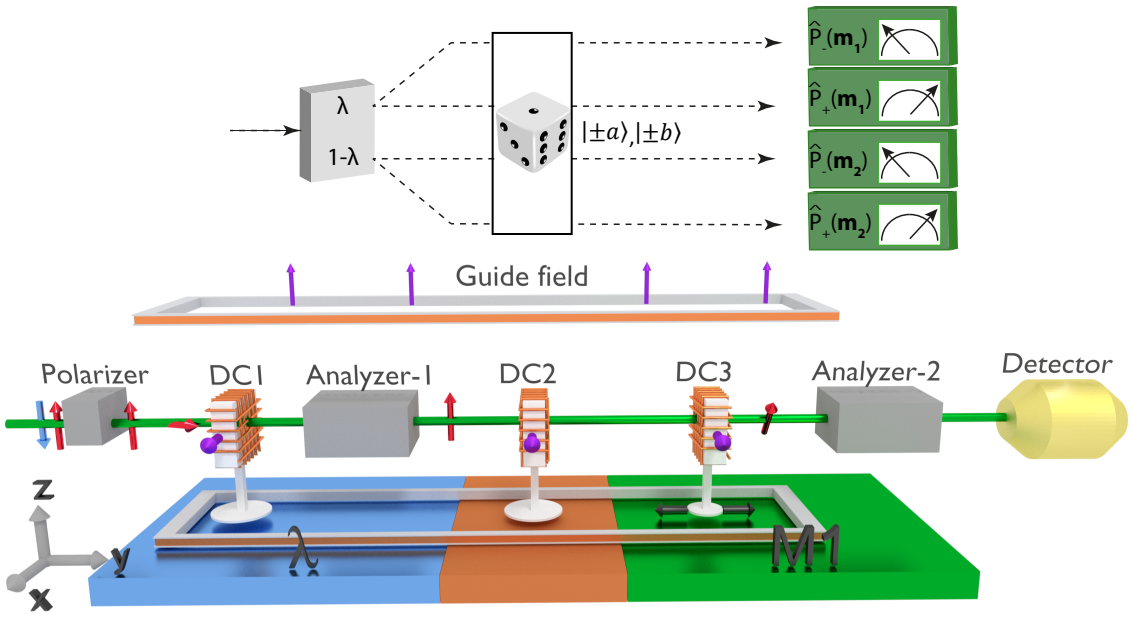


Fig. 7.7: Neutron polarimeter set-up of the measurement. The initial state $|+z\rangle\langle+z|$, indicated by the red arrow after the polarizer, is transformed to $\hat{\rho}_\theta$ with two different angles in DC1 and passes analyzer-1 with probabilities $p(\lambda_1) = |\langle+z|\hat{\rho}_\theta|+z\rangle|^2$ or $p(\lambda_2) = |\langle+z|\hat{-\rho}_\theta|+z\rangle|^2$. A random number generator makes a virtual die roll by computational methods and equiprobably generates one of the two possible eigenstates of the observables $\hat{A}$ and $\hat{B}$ by two magnetic fields in DC2. Finally, DC3 and analyzer-2 make a projection into the direction $\mathbf{m}_1$ on the neutrons transmitted with probability λ and a projection to $\mathbf{m}_2$ on the ensemble transmitted with $1-\lambda$.

$\theta = \gamma B_x t$, where γ is the neutron's gyromagnetic ratio and t the time of flight through the coil. The corresponding unitary operator $\hat{U}_1(\theta) = \exp(i\frac{\theta}{2}\sigma_x)$ transforms the initial states to $\hat{\rho}_\theta = \hat{U}_1(\theta)\hat{\rho}\hat{U}_1^\dagger(\theta)$ which consequently yields

$$\begin{aligned}\mathcal{I}_L(\lambda_i)(\hat{\rho}_\theta) &= \hat{P}_+(\mathbf{z})\hat{U}_1(\theta)\,|+z\rangle\langle+z|\,\hat{U}_1^\dagger(\theta)\hat{P}_+(\mathbf{z}) \\ &= \left|\langle+\mathbf{z}|\hat{U}_1(\theta)|+\mathbf{z}\rangle\right|^2 |+z\rangle\langle+z| = p(\lambda_i)\,|+z\rangle\langle+z| \ .\end{aligned} \tag{7.35}$$

Because two complementary probabilities are required, $p(\lambda_i) = \{\lambda, 1-\lambda\}$ is defined such that

$$\begin{aligned}p(\lambda_1) = \lambda &= \left|\langle+\mathbf{z}|\hat{U}_1(\theta)|+\mathbf{z}\rangle\right|^2 = \cos^2\left(\frac{\theta}{2}\right) \ , \\ p(\lambda_2) = 1-\lambda &= \left|\langle+\mathbf{z}|\hat{U}_1(\theta+\pi)|+\mathbf{z}\rangle\right|^2 = \sin^2\left(\frac{\theta}{2}\right) \ .\end{aligned} \tag{7.36}$$

The magnetic field in DC2 turns the spin by the same effect as in the first coil. For every point that is measured a total of four different currents are fed into DC2 which will prepare the eigenstates $|+\mathbf{a}\rangle$ and $|-\mathbf{a}\rangle$ of observable $\hat{A}$ and likewise the eigenstates $|+\mathbf{b}\rangle$ and $|-\mathbf{b}\rangle$ of observable $\hat{B}$. The probability with which $\mathcal{E}_U^a(p(\lambda_i)\,|+z\rangle\langle+z|)$ creates one of the eigenstates is controlled by a random number generator specifying the probability distribution $p(a) = p(b) = \frac{1}{2}$ as required. Experimentally, this means that a random number generating software in the range $\{0, 1\}$ creates one outcome and then applies one of two currents in DC1 (see Fig. 6.4) to prepare $|\pm z\rangle$ or likewise $|\pm y\rangle$. Be aware, that the randomization happens only once for every particle ensemble and not for individual neutrons. The eigenstates are pure with a uniform probability to obtain one of the possible choices. From the entire ensemble contributing to the statistics projections $\hat{P}_{\tilde{m}}(\mathbf{m}_1) = \hat{P}_{\pm}(\mathbf{m}_1)$ are made on the fraction λ and projection $\hat{P}_{\tilde{m}}(\mathbf{m}_2) = \hat{P}_{\pm}(\mathbf{m}_2)$ likewise on the fraction $1-\lambda$. Re-evaluating Eq. (7.34) fully, delivers

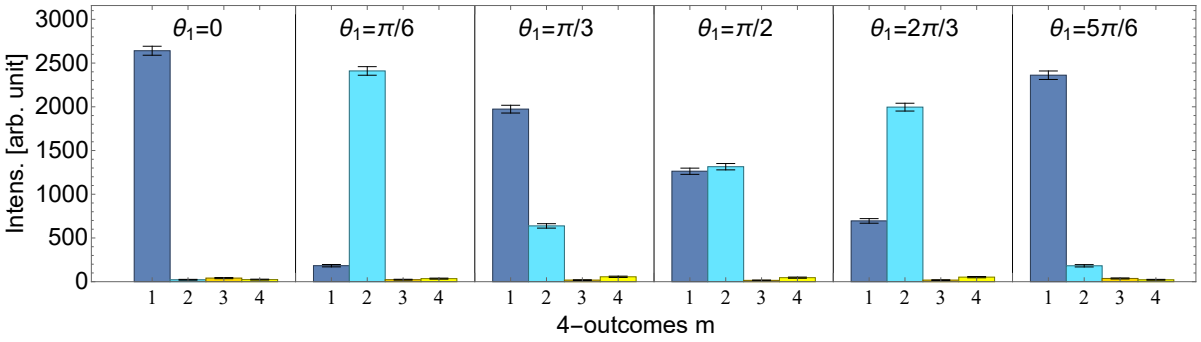
$$\begin{aligned} p(a|m) = p(a|\lambda_i, \tilde{m}) &= \mathrm{Tr}\left(\mathcal{I}_L(\tilde{m})\left(\mathcal{E}_U^a(\mathcal{I}_L(\lambda_i)(\hat{\rho}_\theta))\right)\right) \\ &= \mathrm{Tr}\left(\mathcal{I}_L(\tilde{m})\left(\mathcal{E}_U^a(\hat{P}_\lambda\hat{\rho}_\theta\hat{P}_\lambda)\right)\right) \\ &= \mathrm{Tr}\left(\mathcal{I}_L(\tilde{m})(p(\lambda_i)\hat{\rho}_a)\right) = \mathrm{Tr}\left(p(\lambda_i)\mathcal{I}_L(\tilde{m})(\hat{\rho}_a)\right) \\ &= \mathrm{Tr}\left(p(\lambda_i)\hat{P}_{\tilde{m}}(\mathbf{m}_i)\,|a\mathbf{a}\rangle\langle a\mathbf{a}|\right) = \mathrm{Tr}\left(\hat{M}_m\,|a\mathbf{a}\rangle\langle a\mathbf{a}|\right)\ . \end{aligned} \tag{7.37}$$

as desired. This gives a total of 8 different intensities for the 2 times 4 combinations of $a = \{-, +\}$ and $m = \lambda_i \times \tilde{m} = \{1, 2, 3, 4\}$, and 8 count rates for b respectively. The output intensities get labeled with lower indices $I_{\{a,m\}}$, where $m = \{1, 2, 3, 4\}$ is partitioned into parts

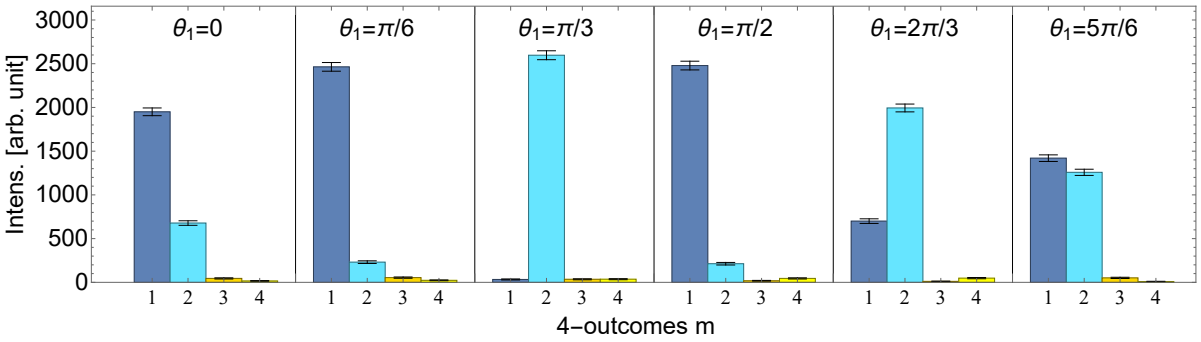
$$p(a|m) \to \begin{cases} \lambda & \begin{cases} |+\mathbf{a}\rangle \vee |-\mathbf{a}\rangle & \begin{cases} \hat{P}_+(\mathbf{m}_1) & \to I_{a,m=1}\ , \\ \hat{P}_-(\mathbf{m}_1) & \to I_{a,m=2}\ , \end{cases} \\ |+\mathbf{a}\rangle \vee |-\mathbf{a}\rangle & \begin{cases} \hat{P}_+(\mathbf{m}_2) & \to I_{a,m=3}\ , \\ \hat{P}_-(\mathbf{m}_2) & \to I_{a,m=4}\ . \end{cases} \end{cases} \\ 1-\lambda & \end{cases} \tag{7.38}$$

Analogous to formulas in Eq. (7.24) the ratio of intensities provides the probabilities

$$p(\lambda_i) = \frac{\sum_{m=2i-1}^{2i} I_{\{a,m\}}}{\sum_{m=1}^{4} I_{\{a,m\}}}\ , \qquad p(a|m) = \begin{cases} \lambda\frac{I_{\{a,m\}}}{\sum_{m=1}^{2} I_{\{a,m\}}}\ , & m = \{1, 2\}\ , \\ (1-\lambda)\frac{I_{\{a,m\}}}{\sum_{m=3}^{4} I_{\{a,m\}}}\ , & m = \{3, 4\}\,. \end{cases} \tag{7.39}$$



(a) Histogram of the projective measurement outcomes with Bloch vector $\mathbf{a} = \hat{\mathbf{e}}_z$ and $\mathbf{m}_1 = \cos\theta_1 \mathbf{e}_z + \sin\theta_1 \mathbf{e}_y$ required to determine $N(\mathcal{M}, \hat{A})$ at $\lambda = 1$. The stepsize of the rotation angle in the plot is $\Delta\theta_1 = \pi/6$. The intensity on the left is maximally peaked and therefore allows a clear distinction of the input states, while at $\theta_1 = \frac{\pi}{2}$ outcome 1 and 2 registered approximately the same number of neutrons: the noise at that point will be minimal. When the intensity in dark blue is higher it is more likely that the input state is $|a\rangle$, else if the bar in light blue is taller then the outcome suggests that $|-a\rangle$ is the measured eigenstate.



(b) Histogram of the projective measurement outcomes with Bloch vector $\mathbf{m}_1$ same as in Fig. 7.8(a) and $\mathbf{b} = \cos(\frac{\pi}{3})\mathbf{e}_z + \sin(\frac{\pi}{3})\mathbf{e}_y$ required to determine $N(\mathcal{M}, \hat{B})$ at $\lambda = 1$. When the intensity in dark blue is higher it is more likely that the input state is $|b\rangle$, else for light blue $|-b\rangle$ is more probable. Because the states $|a\rangle, |b\rangle$ are not mutually unbiased in this example, the intensity on the left is distributed unevenly and peaks when $\theta_1 = \frac{\pi}{3}$.

Fig. 7.8: Intensity distributions in the projective case where the 4-outcome POVM is simply $\hat{M}_m = \left\{\hat{P}_+(\mathbf{m}_1), \hat{P}_-(\mathbf{m}_1), \emptyset, \emptyset\right\}$. The Bloch vectors are note perfectly orthogonal, instead the inner product is $\mathbf{a}\cdot\mathbf{b} \cong 0.5$, which corresponds to an angle of $\frac{\pi}{3}$. The histograms indicate how the count rates in Eq. (7.39) provide the probabilities. The value of λ corresponds to the frequency of occurrence of outcomes 1+2 and 3+4 to the total count rate (1+2+3+4). The ratio of $m = 1$ to $m = 2$, in turn, determines the probability associated with the first projector and $m = 3$ to $m = 4$, which do not occur in this figure, the probabilities from $\hat{P}_+(\mathbf{m}_2)$ and $\hat{P}_-(\mathbf{m}_2)$.

7.3.3 Results

The regarded observables of the spin-1/2 system are the Pauli matrices $\hat{A} = \mathbf{a}\cdot\hat{\boldsymbol{\sigma}}$ and $\hat{B} = \mathbf{b}\cdot\hat{\boldsymbol{\sigma}}$. To keep the experiment simple $\mathbf{a} = \mathbf{e}_z$ is chosen for the first Bloch vector and only $\mathbf{b}$ is varied over the y-z plane. In the case $\lambda = 1$, the value at which the parallelogram degenerates into antipodal projectors, the results are known to be determined by rotating the operators $\hat{P}(\mathbf{m}_1)$, where $\mathbf{m}_1 = \cos\theta_1\mathbf{e}_z + \sin\theta_1\mathbf{e}_y$. It is no surprise that the noise in this case is given by Eq. (7.16) and Eq. (7.23), i.e. $H_b(\mathbf{m}_1\cdot\mathbf{a})$ and $H_b(\mathbf{m}_1\cdot\mathbf{b})$.

The raw data of the projective noise measurement before evaluation is demonstrated in Fig. 7.8. The upper half Fig. 7.8(a) shows the count rates from which the noise $N(\mathcal{M}, \hat{A})$ can be extracted. It should be clear by now that the maximally peaked distribution provides no new information and the uniform distribution at $\theta = \frac{\pi}{2}$ yields the maximal entropy. The inference on the unknown input state is color-coded in dark ($|a\rangle$) and light blue ($|-a\rangle$) meaning that the bar that is taller is more likely to be the true initial state. The tradef-off is visually recognizable, since the ability to draw a concise conclusion about the initial state is not peaked simultaneously. Due to the choice of the relative angle of $\mathbf{b}$ to be at $\frac{\pi}{3}$, there is no maximum offset of the distribution in the lower Fig. 7.8(b). Note that event 3 and 4 never take place, which means that a projective measurement $\lambda = 1$ is made.

A plot of one set of the results is given in Fig. 7.9 which shows the variation of noise depending on the angle of $\mathbf{m}_1(\theta_1)$ when $\mathbf{a} = \mathbf{e}_z$ and $\mathbf{b} \cong \cos(\frac{4\pi}{9})\mathbf{e}_z + \sin(\frac{4\pi}{9})\mathbf{e}_y$. The measured points in purple display how the noise $N(\mathcal{M}, \hat{B})$ increases when $\mathbf{m}_1' = \cos\theta_1\mathbf{e}_z + \sin\phi\sin\theta_1\mathbf{e}_y + \cos\phi\sin\theta_1\mathbf{e}_x$ is also changed by some azimuthal angle; this does of course not affect the noise $H_b(\mathbf{m}_1\cdot\mathbf{a}) = H_b(\mathbf{m}_1\cdot\mathbf{e}_z)$ of observable $\hat{A}$. When the observables' Bloch vector $\mathbf{a}$, $\mathbf{b}$ span a plane, then points of disturbance are lifted the more the measurement direction lies outside that plane $\mathbf{m}_1'\cdot(\mathbf{a}\times\mathbf{b}) \neq 0$. Understandably, this happens, because entropy increases for more uniform distributions.

As was mentioned before, for $|\mathbf{a}\cdot\mathbf{b}| \leq 0.391$ the projective measurement does not provide the lowest boundary. How the measurement points for the minimal uncertainty relation are obtained is illustrated in Fig. 7.10. For this purpose, the neutron spin is measured not only in the direction $\mathbf{m}_1$, but also in $\mathbf{m}_2$, so that all four slots of the bar chart contribute to the

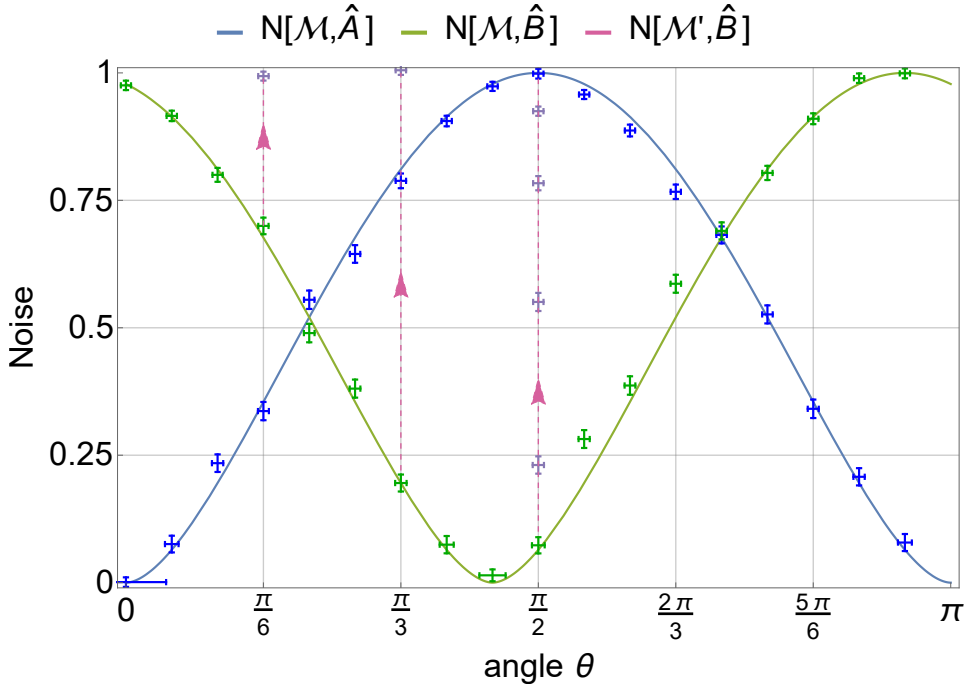
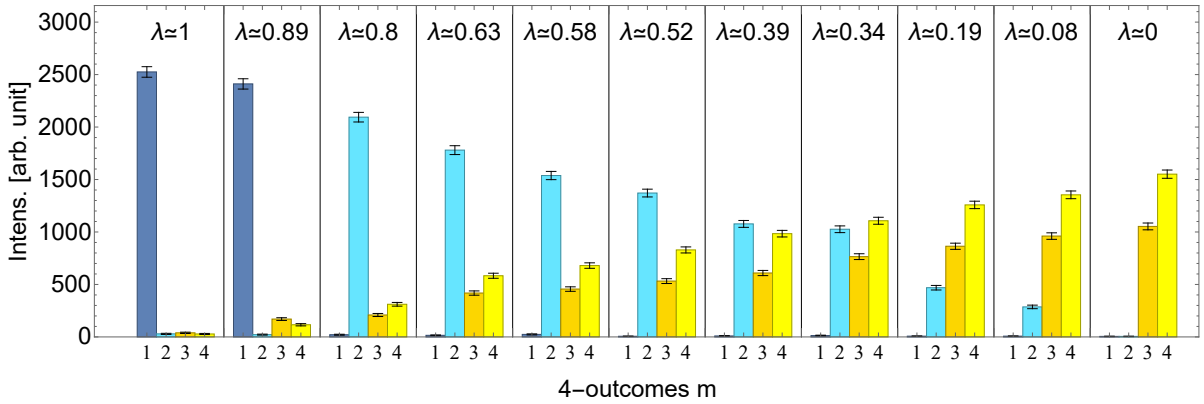


Fig. 7.9: Results of entropic noises $N(\mathcal{M},\hat{A}) = H_b(\cos\theta)$ in blue and $N(\mathcal{M},\hat{A}) = H_b(\mathbf{m}_1\cdot\mathbf{b})$ in green for a relative angle $\mathbf{a}\cdot\mathbf{b} = b_z \cong \cos(\frac{4\pi}{9})$. The minimal uncertainty occurs when the Bloch vector of the measurement operator is in the plane spanned by **a**-**b**, while the uncertainty increases when the azimuthal angle in $\mathbf{m}_1' = \cos\theta_1\mathbf{e}_z + \sin\phi\sin\theta_1\mathbf{e}_y + \cos\phi\sin\theta_1\mathbf{e}_x$ is not an integer multiple of π. The purple colored data points indicate the increase of noise, when the projector $\hat{P}(\mathbf{m}_1')$ is rotated out of the y-z plane at the positions $\theta = \frac{\pi}{6}, \frac{\pi}{3}$ and $\frac{\pi}{2}$.

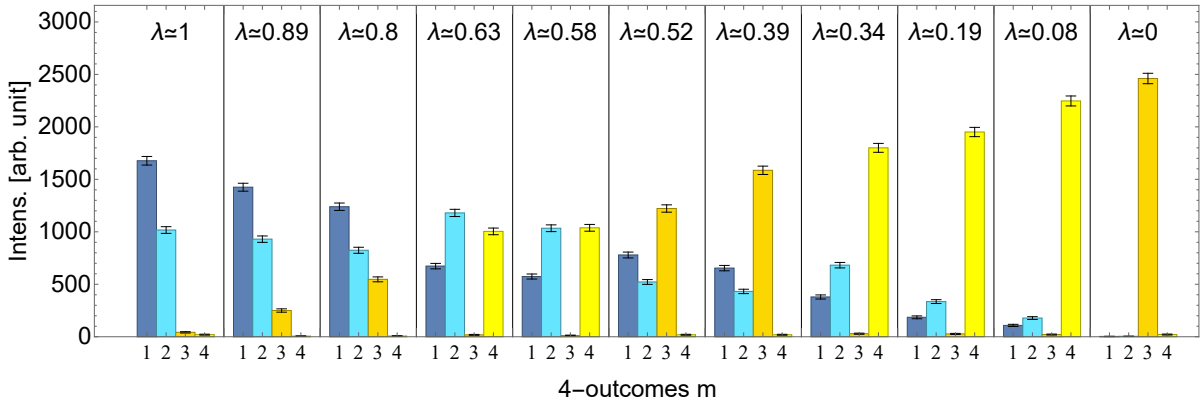
conditional entropy. In a projective measurement, only the ratio of the count rates of two bars of the chart determine the entire probability distribution. However, the interplay of 4 intensities can increase success rate for the inference of the unknown state. Instead of acquiring information with the total number of neutrons, it can be more informative if one makes an inquiry with a sub-ensemble only and perform a second inquiry with the remainder of the particles. The blue and turquoise colored bar of the histogram build one block and the orange/yellow the other, where both correspond to the projector as shown in Fig. 7.6. From the minimization

$$\min_{\theta}\left(H_b(\mathbf{m}_1(\theta)\cdot\mathbf{a}) + H_b(\mathbf{m}_1(\theta)\cdot\mathbf{b})\right)\ , \tag{7.40}$$

according to Eq. (5.17), the entropy minimizing projection directions are $\mathbf{m}_1 \cong \cos(4°)\mathbf{e}_z + \sin(4°)\mathbf{e}_y$ and likewise $\mathbf{m}_2 = \cos(76°)\mathbf{e}_z + \sin(76°)\mathbf{e}_y$. The following statement can be made



(a) Histogram of the measurement outcomes with Bloch vector $\mathbf{a} = \hat{\mathbf{e}}_z$, $\mathbf{m}_1 \cong \cos(4°)\mathbf{e}_z + \sin(4°)\mathbf{e}_y$ and $\mathbf{m}_2 \cong \cos(76°)\mathbf{e}_z + \sin(76°)\mathbf{e}_y$ required to determine $N(\mathcal{M}, \hat{A})$. The angles of the measurment directions are a consequence of the minimization problem introduced in sec. 5.3 and further analyzed in Fig. 7.11. For dark blue and orange the probability of the input state suggests an initial eigenstate of $|a\rangle$, while the lighter counterparts favor $|-a\rangle$. The advantage of the POVM measurement is that more information about the original state can be obtained from the bar charts in intermediate areas.



(b) Histogram required to determine $N(\mathcal{M}, \hat{B})$ with Bloch vector $\mathbf{b} \cong \cos(\frac{4\pi}{9})\mathbf{e}_z + \sin(\frac{4\pi}{9})\mathbf{e}_y$ and identical directions for the projectors $\mathbf{m}_1 \cong \cos(4°)\mathbf{e}_z + \sin(4°)\mathbf{e}_y$ and $\mathbf{m}_2 \cong \cos(76°)\mathbf{e}_z + \sin(76°)\mathbf{e}_y$. Like in Fig. 7.10(a) darker colors suggest $|+b\rangle$ while light blue and light yellow indicate $|-b\rangle$ to be the true initial state.

Fig. 7.10: Intensity distribution of the 4-outcome POVM of Eq. (7.30) for $\mathbf{a}\cdot\mathbf{b} \cong 0.17$. In contrast to Fig. 7.8 the outputs $m = 3$ and $m = 4$ also appear for $\lambda \neq 1$ and become maximal for $\lambda = 0$ at which the projective measurement of $\mathbf{m}_2$ is performed. The heights of bar 1+2 and 3+4 divided by the total intensity determines once more the statistical mixture of the measurements. From left to right the decrease of the weight parameter is circa $\Delta\lambda = -0.1$. The probabilities $\mathrm{Tr}(\hat{P}_\pm(\mathbf{m}_1)\hat{\rho})$ are given by the cool, bluish colored bars and $\mathrm{Tr}(\hat{P}_\pm(\mathbf{m}_2)\hat{\rho})$ by the warm, bright colored bars.

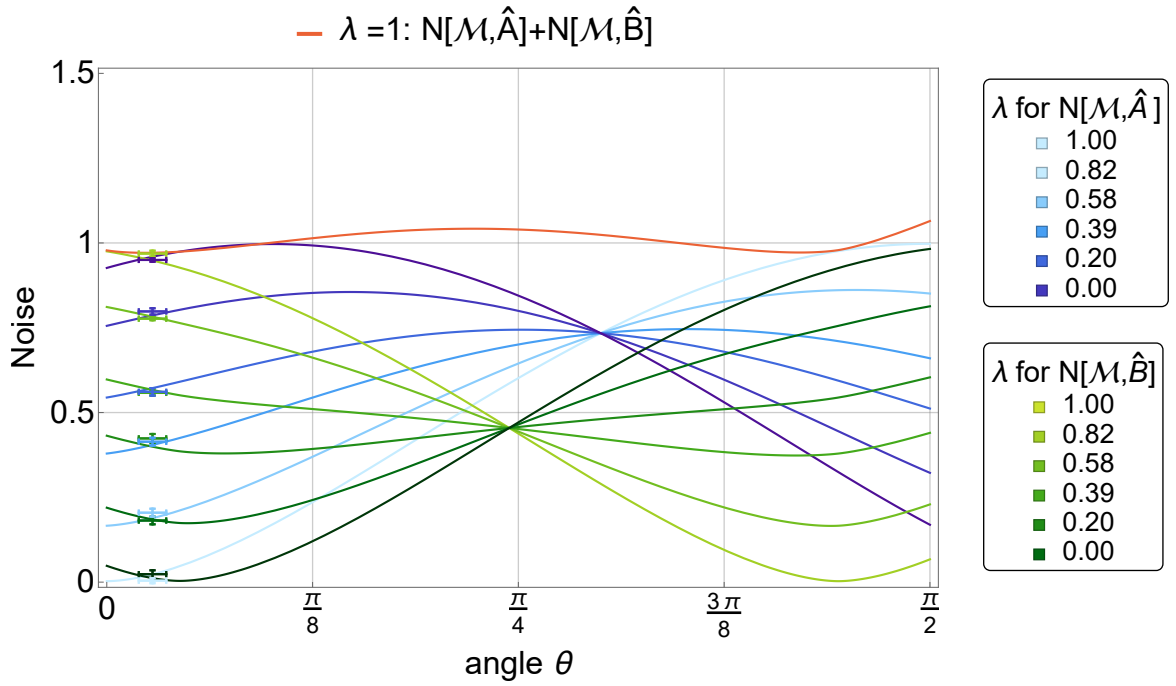


Fig. 7.11: Determination of noises of the four-outcome POVM. The red curve indicates the sum $N(\mathcal{M},\hat{A})+N(\mathcal{M},\hat{B})$ at $\lambda=1$ for $\mathbf{a}\cdot\mathbf{b}\cong\cos(\frac{4\pi}{9})$, which has two local minima, namely one at $\theta_1\approx 4°$ and $\theta_2\cong 76°$. The noises for different values of mixing λ are colored bluish for $N(\mathcal{M},\hat{A})=H_b(\mathbf{m}_1\cdot\mathbf{a})$ and greenish for $N(\mathcal{M},\hat{B})=H_b(\mathbf{m}_1\cdot\mathbf{b})$. Unlike in Fig. 7.9 the angle is no longer varied, instead the measurement results are obtained for different mixing of the POVM at the two directions $\mathbf{m}_1(\theta_1)$ and $\mathbf{m}_2(\theta_2)$. The color legend on the right side indicates that lighter colors correspond to larger and darker colors to smaller values of λ. The data points show that for $\lambda=1$ the projective measurement at $\mathbf{m}_1(4°)$ close to $\mathbf{a}$ has almost vanishing noise $N(\mathcal{M},\hat{A})$ while the noise of $N(\mathcal{M},\hat{B})$ is large. The effect of decreasing the weighting λ means that projective measurement at $\mathbf{m}_1(76°)$ contributes also to the entropy and hence reduces the noise of $\hat{B}$ at the expanse of the noise in observable $\hat{A}$.

now: since the positive operators $\hat{P}_+(\mathbf{m}_1),\hat{P}_+(\mathbf{m}_2)$ are in the northern hemisphere, output 1 and 3 tend to favor state $|+a\rangle$, while 2 and 3 will indicate higher probabilities at $|-a\rangle$. At the same time, 4° and 76° are also in the right hemisphere, so that outcome 1 and 3 also imply $|+b\rangle$ and 2+4 again $|-b\rangle$.

In Fig. 7.11 the respective noises for different weightings λ, as indicated by Eq. (7.33), are plot-

ted. The graph show that entropic noise for the non-projective cases is obtained as a function of λ, but not as a continuous function of the polar angle θ_1. From the sum $N(\mathcal{M}, \hat{A}) + N(\mathcal{M}, \hat{B})$ given by the red curve for $\lambda = 1$, the position of the two local minima $\{\theta_1, \theta_2\}$ can be deduced which fix both angles and depend on the inner product $|\mathbf{a} \cdot \mathbf{b}|$.

The result of the noise-noise uncertainty relation for an eigenstate overlap of $|\mathbf{a} \cdot \mathbf{b}| \cong 0.5$ is plotted in Fig. 7.12. The projective noise measurement starts at $\theta_1 = 0$, which yields $N(\mathcal{M}, \hat{A}) = 0$ and makes contact with the coordinate axes at approximately $H_b(\cos(\frac{\pi}{3})) \approx 0.81$. Constant variation of the angle θ_1 by an increment of $\Delta\theta_1 = 10°$ decreases the noise of observable $\hat{B}$ at the expense of the noise of observable $\hat{A}$, vanishing again at 0.81. At $\theta_1 = \frac{\pi}{2}$ the noise of $\hat{A}$ has its maximum $N(\mathcal{M}, \hat{A}) = 1$ and reverts back along the blue curve. The noise $N(\mathcal{M}, \hat{B}) = 1$ is given at $\theta_1 = \frac{5\pi}{6}$. The measurement points in purple are obtained for a fixed noise of $\hat{A}$ by rotating $\mathbf{m}_1(\phi)$ out of the **a**-**b** plane. In Fig. 7.12 the three points with the lowest noise $N(\mathcal{M}, \hat{B})$ and the rightmost point at $N(\mathcal{M}, \hat{A}) = 1$ point are vertically elevated to the upper edge: ergo this area belongs to the accessible uncertainty region $E(\hat{A}, \hat{B})$ as well.

The case depicted in Fig. 7.12 is a generic example that the optimal uncertainty relations are already obtainable by a projective measurement. A different situation occurs when the inner product is smaller than the critical angle, which is plotted for 4 different values of $|\mathbf{a} \cdot \mathbf{b}|$ in Fig. 7.13. The white area coincides with the uncertainty $E(\hat{A}, \hat{B})$ and the blue curve and points corresponds to theoretical and experimental results obtained by projective measurements. The blue shaded area represents the region $R_{NN}(\hat{A}, \hat{B}) \setminus E(\hat{A}, \hat{B})$, which contains the elements that are in the convex hull of $E(\hat{A}, \hat{B})$ constituting the green line, the lowest achievable uncertainty bound. The gray areas is the set of points inaccessible by any kind of POVM.

The sequence of graphs starts with the mutually unbiased case plotted in Fig. 7.13(a); the Bloch vectors are perpendicular, the boundary is $c = 1$ and the uncertainty relation is given by Ineq. (7.9). Indeed, choosing $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{2}$ for the 4-outcome POVM Eq. (7.30) and varying the probability weighting λ improves the noise-noise trade-off noticeably. Each measured point on the bottom green line has a different value of λ marked by slightly different colors. As the legend on the top of the plots implicates the largest probability weighting is colored the darkest and decreasing by $\Delta\lambda \approx -0.1$ the experimental results wander from the bottom-right

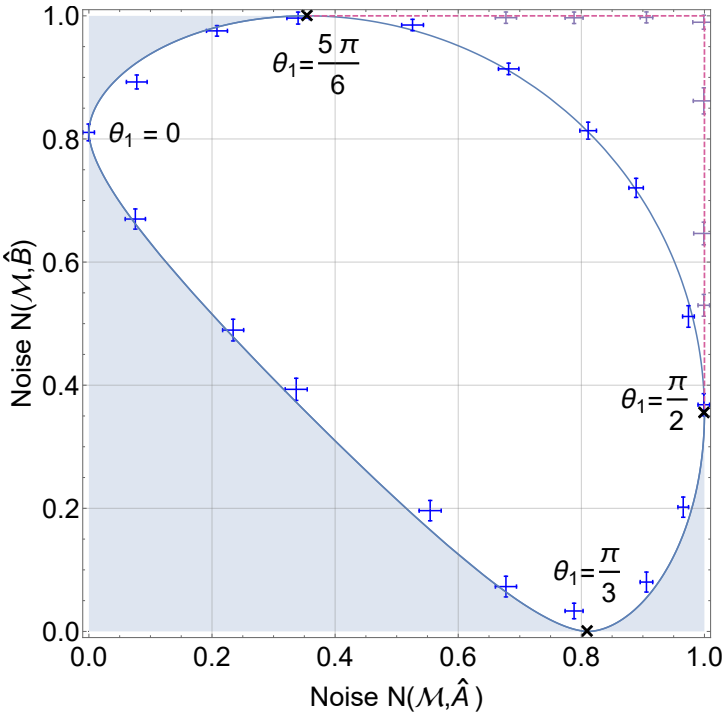
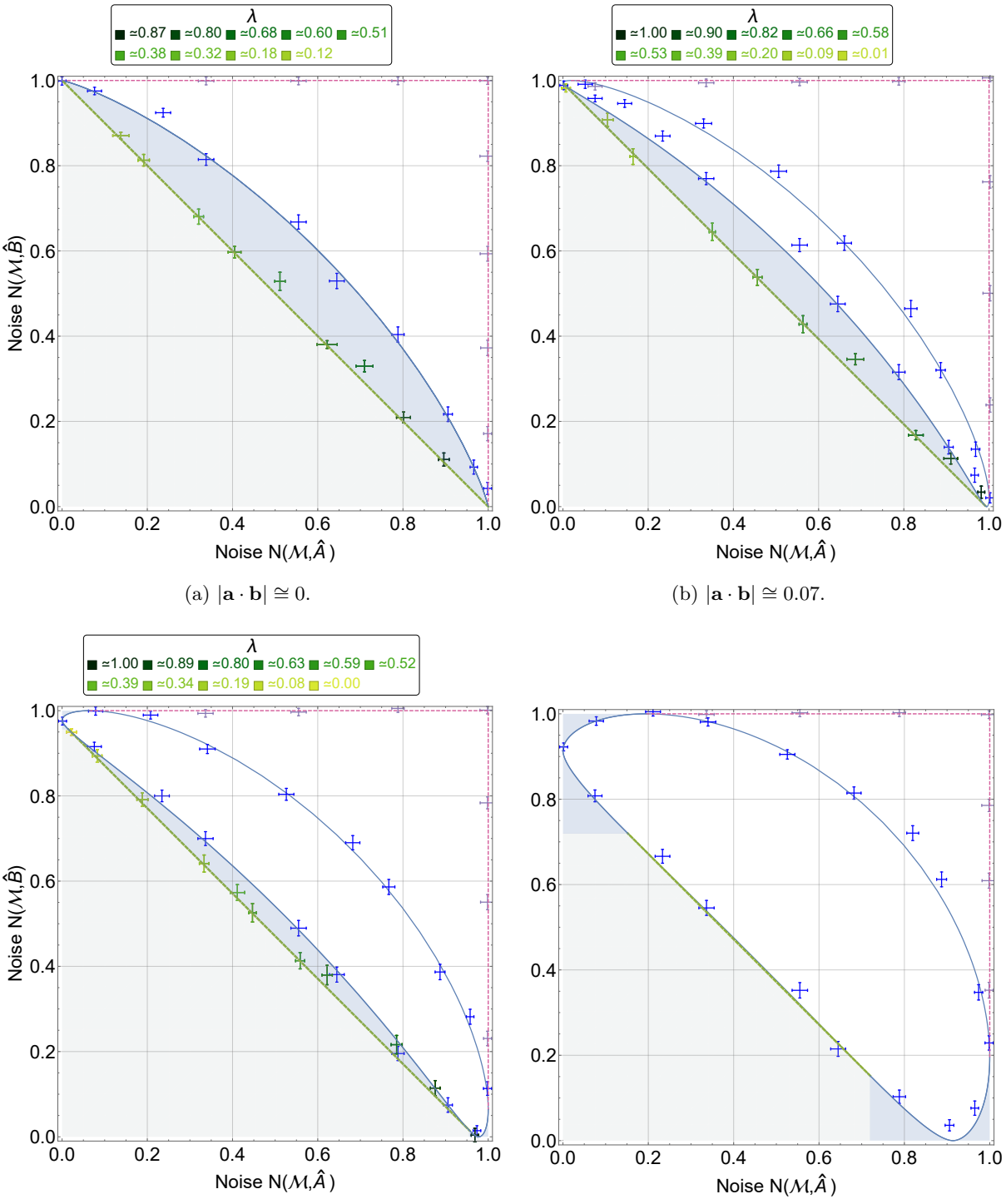


Fig. 7.12: Noise $N(\mathcal{M}, A)$ - $N(\mathcal{M}, B)$ trade-off between the two non-commuting observables at $\mathbf{a} \cdot \mathbf{b} \cong 0.5$. The blue boundary is obtained by projective measurements in the **a**-**b** (i.e. y-z) plane, while the purple corners are results of non-coplanar projective measurements, obtained by lifting the points as illustrated by Fig. 7.9. The white, convex area $E(\hat{A}, \hat{B})$ is the domain of values obtainable by projectors $\hat{P}_{\pm}(\mathbf{m}_1)$, while the blue region is excluded by projective measurements.

to the top-left corner with colors getting brighter. Physically the employment of the POVM increases the ability to correctly infer the unknown eigenstates except for the fixed points in the corners.

When the eigenstates of $\hat{B}$ come closer to $\hat{A}$ as is the case in Fig. 7.13(b) and Fig. 7.13(c), $E(\hat{A}, \hat{B})$ becomes less concave, the region $R_{NN}(\hat{A}, \hat{B}) \setminus E(\hat{A}, \hat{B})$ starts shrinking and the improvements made by the POVM become closer to the PVM uncertainty domain. Notice that the angles in $\mathbf{m}_1(\theta_1)$ and $\mathbf{m}_2(\theta_2)$ are determined by analyzing the projective noise-noise region which has two points at which the area becomes non-convex, i.e. the start and end point of the green line. These can, as already indicated in Fig. 7.11, be numerically calculated by finding the minima of $\min_\theta \left(H_b(\mathbf{m}_1(\theta) \cdot \mathbf{a}) + H_b(\mathbf{m}_1(\theta) \cdot \mathbf{b})\right)$. Eventually, as is exhibited in



(a) $|\mathbf{a}\cdot\mathbf{b}| \cong 0$.

(b) $|\mathbf{a}\cdot\mathbf{b}| \cong 0.07$.

(c) $|\mathbf{a}\cdot\mathbf{b}| \cong 0.17$.

(d) $|\mathbf{a}\cdot\mathbf{b}| \cong 0.35$.

Fig. 7.13: Plots of noise $N(\mathcal{M},\hat{A})$ versus $N(\mathcal{M},\hat{B})$ for four different $|\mathbf{a}\cdot\mathbf{b}| \leq 0.391$. The gray area demarcates a new uncertainty region with a new domain of outcomes only attainable by the POVM. The green line constructs the convex hull of the white area $E(\hat{A},\hat{B})$.

Fig. 7.13(d) for a scalar product close to the critical value, the bean shaped curve overlaps with the green line. This means the measurement uncertainty region becomes almost congruent with the preparation uncertainty domain $R_{NN}(\hat{A},\hat{B}) \approx E(\hat{A},\hat{B})$. Although it is no longer visually recognizable, the gray area along with the green line marks the generally inaccessible area as in Figs. 7.13(a)–7.13(c), while the blue areas are already excluded by projective measurements. Attempts to access the non-convex area from $(N(\mathcal{M},A),N(\mathcal{M},B)) \approx (0.17,0.70)$ to $(N(\mathcal{M},A),N(\mathcal{M},B)) \approx (0.70,0.17)$ is of no use, since the bottom projective measurement result is practically collinear with the green line obtainable by the POVM.

The above experiment confirms the validity of the 1. Proposition for qubit measurements. Indeed the limit in the uncertainty relation can be significantly improved for intermediate values between some outer, fixed points (see Fig. 7.13). It is particularly useful that both preparation and measurement uncertainty principles can be connected with each other geometrically. In this case convexity poses a simple criterion on the set of allowed noise-noise values.

7.4 Entropic noise-disturbance measurement uncertainty for general qubit measurements

Finding optimal entropic noise-disturbance uncertainty relations is difficult and analytical solutions have not been generally derived yet. It is worth having a glimpse at the issue and present some developments. To recognize the problem consider Ineq. (7.25) not for sharp observables, but instead for the qubit POVM of the form Eq. (3.32), then calculations show: if $\hat{\Pi}_k = \frac{1}{n}(1+\mathbf{m}_k\cdot\hat{\boldsymbol{\sigma}})$, with the vector $\mathbf{m}_k = (0,\sin(\phi+\frac{2\pi k}{n}),\cos(\phi+\frac{2\pi k}{n}))^T$, $[k>1\in\mathbb{N}]$, the noise would simply be $N(\mathcal{M},\hat{A}) = \frac{1}{n}\sum_{k=1}^{n} H_b(\mathbf{m}_k\cdot\mathbf{a})$. Assuming that there is a correction that makes the disturbance identical to the noise case[7], it would likewise mean that $D_\epsilon(\mathcal{M},\hat{B}) = N(\mathcal{M},\hat{B}) = \frac{1}{n}\sum_{k=1}^{n} H_b(\mathbf{m}_k\cdot\mathbf{b})$. Accordingly, for $n=2$ the projective measurement result Ineq. (7.25) is obtained, while for the trine POVM $n=3$ the uncertainty relation

$$g(N(\mathcal{M},\hat{A}))^2 + g(D_{\mathcal{E}}(\mathcal{M},\hat{B}))^2 \approx 1.1 \not\leq 1 \tag{7.41}$$

does not hold [27], for $n=4$ the violation is even 1.22.

[7] This correction is based on a maximum-likelihood procedure.

Based on some basic considerations and numerical calculations, Abbott et. al. [27] suggest a test for an entropic noise-disturbance uncertainty whose lowest region might be presentable by

$$R_{ND}(\hat{A},\hat{B}) = \{(N(\mathcal{M},\hat{A}), D_{\mathbb{1}}(\mathcal{M},\hat{B})) \mid \mathcal{M} \text{ is purity preserving}\} \ . \tag{7.42}$$

For a *purity preserving quantum instrument* the Kraus decomposition is given by a single summand $\mathcal{M}_m(\hat{\rho}) = \hat{M}_m\hat{\rho}\hat{M}_m^\dagger = \hat{U}_m\sqrt{\hat{\Pi}_m}\hat{\rho}\sqrt{\hat{\Pi}_m}\hat{U}_m^\dagger$. In other words realizing a general Lüders instrument Eq. (3.67) with an additional unitary correction $\hat{U}_m$ originating from the polar decomposition of the measurement operator might suffice to obtain the lowest bound. The conjecture in [27] is that for the two orthogonal Pauli observables, say $\hat{\sigma}_z, \hat{\sigma}_x$, and the three-outcome POVM $\hat{\Pi}_m(x) = p_m(x)(\mathbb{1} + \mathbf{r}_m(x)\cdot\hat{\boldsymbol{\sigma}})$, where $p_0(x) = \frac{\cos x}{(1+\cos x)}$, $p_1(x) = p_{-1}(x) = \frac{1}{2(1+\cos x)}$ and $\mathbf{r}_m = ((-1)^m\cos(mx), 0, \sin(mx))^T$, the noise-disturbance region

$$R_{ND}(\hat{\sigma}_z,\hat{\sigma}_x) = \mathrm{cl}\left\{\frac{\cos x + H_b(\sin x)}{1+\cos x}, \frac{H_b(\cos x)}{1+\cos x} \mid 0 \le x \le \frac{\pi}{2}\right\} \tag{7.43}$$

provides a tight uncertainty region. It remains an open issue to verify this speculative prediction.

7.5 Discussion

Entropic definitions of uncertainty relations are easy to handle, interpretation is plausible, they are state independent, and it should be emphasized that it depends only on probabilities. There is also a bigger concept [104, 105] supporting the fundamental ideas of entropic formulations. The choice of conditional Shannon entropy as noise and disturbance is reasonable, since it is the only Rényi entropy for which the chain rule is well defined. Physically, however, it requires the production of all eigenstates of observables $\hat{A}$ and $\hat{B}$, especially in infinite dimensions of Hilbert spaces interpretation and implementation can not be continued logically.

The fact that one sends unknown, uniformly distributed states on the measurement apparatus appears somehow artificial. Ideally, an appropriate device would have to generate each states sequentially without a physical possibility of recording the randomness of the physical process. Furthermore, rms and entropic disturbances differ in that the former is transformed by a non-selective measurement (in Ozawa's theory, error correction comes from the non-uniqueness of the measurement effects, see Eq. (3.61)) while the latter gives a conditional, selective output.

This drastically changes the approach of error correction, since in the entropic case not only unitary transformations, but any optimization depending on the meter outcome can be performed. Whether correction 'on the fly', i.e. forwarding the meter outcome to the correction device while the set-up is running, is meaningful should also be up for debate. Finally, due to the occurrence of the logarithm, many values and results are often numerical in nature. This can make the subject of optimization a tedious task.

CHAPTER 8

Conclusion and Outlook

The developments and studies at the Atomic Institute in Vienna were described and the necessary methodologies presented. The motivation and the current developments are rooted in the wish to break up the confusion of different ideas about the uncertainty principle presented at the historical beginning of quantum mechanics. The first half of the thesis was devoted to the mathematical and physical basics that are necessary for a deeper understanding of the subject. The focus here was on the modern treatment of observables and quantum operations, witch enable the discussion of the incompatibility of conjugate quantities. Neutron optics is an established and versatile tool for experimental tests. An emphasis was made on the experiments and the obtained results regarding the characterization of measurement uncertainties of the rms error-disturbance type introduced by Ozawa [20] and the information-theoretic noise-disturbance type first published by Buscemi et. al [26]. Both error/noise - disturbance theories have given new impetus to the research of uncertainty relations and already underwent several iterations of improvements recently.

The first experimental test of Ozawa's theory gave new stimulus to this area of research. However, it turned out that the original inequalities were not satisfactory, because for most cases no saturation was achievable. Therefore, revisions were necessary to remedy this situation. The previous chapters 6 and 7 illustrate the importance of error correction procedures for uncer-

tainty relations for sequential measurements (see Figs. 6.12–6.16). It has also been shown that Ozawa's theory for mixed spin states exhibits exactly the same behavior as for pure spin states (see Fig. 6.18); this residual or unaffected behavior for mixtures is indeed amazing, because it is generally not expected that uncertainty relations are entirely insensitive to the form of the prepared state. The partial invalidation of the bound C_{AB} for the error-disturbance uncertainty relation proved that the alternative bound D_{AB} provides the tightness for projective measurements when $\mathrm{Tr}(\hat{A}\hat{\rho}) = \mathrm{Tr}(\hat{B}\hat{\rho}) = 0$.

The entropic version, in turn, is of great importance because of its thematic proximity to information theory. Here, the idea is based on the amount of information obtained by the approximate measurement. As shown in Figs. 7.4 and 7.5, the more knowledge acquired by the measurement, the more the system is disturbed; a subsequent determination has less information to offer because of the loss of correlation. Again, for spin measurements, there is plenty of room for improvement. The results shown for a successive measurement are optimal only for projective Pauli observables, in other cases it is still an open problem to find out how the optimal noise-disturbance bound can be characterized. Because of the difficulties, instead of a sequential uncertainty relation for generalized observables, an entropic noise-noise trade-off was tested. For this purpose, a successive Lüders instrument was realize, which provides the necessary statistics to implement the 4-output POVM of interest. The joint-measurement uncertainty dictates that no device can be calibrated such that a joint determination of unknown eigenstates of two non-commuting observables is possible. Interestingly, with POVMs more information can be retained when the observables under consideration are non-commuting 'enough', i.e. optimality by general measurements depends on the inner product $|\mathbf{a} \cdot \mathbf{b}|$ of the Bloch vectors. This property has been shown in the experimental results in Figs. 7.12 and 7.13.

Several points for innovations are available. Starting from Ozawa's theory, the most obvious challenge is to describe new uncertainty relations with POVMs. In this case, the fuzziness as mentioned in Eq. (6.14) would give a non-zero value and contribute to the average rms noise as well. A corresponding proposal with a quantum circuit is found in the article by Lu et al. [82]. From the experimental point of view uncertainty relations for quantum systems in higher dimensions, e.g. 2x2 = 4 dimensions, should be the next target due to availability of established manipulation and measurement techniques. However, one should avoid choosing observables,

which lead to state-independent behavior of error and disturbance. This requires care with the observables build by tensor products of Pauli-observables that construct the higher Clifford algebras over the real numbers in which the square and anti-commutator can be proportional to the identity: generally, $(\hat{\sigma}_i \otimes \hat{\sigma}_j)^2 = \mathbb{1}_4$ is always true, however some of the combinations like $\{\hat{\sigma}_x \otimes \hat{\sigma}_x, \hat{\sigma}_y \otimes \hat{\sigma}_z\} = \mathrm{diag}(2\hat{\sigma}_y, -2\hat{\sigma}_y)$ are non trivial and imply new behavior. There are definitions similar to Ozawa's measures [19, 22, 65] which can also be tackled next. As already briefly mentioned in sec. 6.5, the BLW theory could be inspected with mixed spin particles in order to highlight discrepancies with error and disturbance in the case of qubits.

In the entropic case, as with Ozawa's error-disturbance, a test of noise-disturbance uncertainty for generalized observables should be studied next. In order to determine the region in Eq. (7.43), a 3-output rank-1 POVM has to be realized. It would be advisable to implement a suitable ancilla system, although it should be possible to perform a one-dimensional system of successive measurements (see [106]). It is important to make sure that the post-measurement state is given by the generalized Lüders instrument given by Eq. (3.67). In addition to the Shannon entropy, an alternative formulation of a measurement uncertainty relation based on relative entropy described by Eq. (2.40) has also been introduced [28, 107]. To what extent this theory is more useful or expedient, has yet to be clarified. Theoretically, it might be preferable because the KL-divergence should not require a specification of a distribution *a priori* and seems more flexible. Furthermore, it does not have the complications of the Shannon entropy for non-discrete spectra and is related to a true distance, the *Fischer information metric*. It would also be interesting to find an extension that allows to formulate a weak-valued (information) entropy. This would however require a formula that can deal with negative quasi-probabilities which are ill-defined for the logarithm.

More than ever, it would again be of great interest to explore position-momentum trade-offs using neutron-optical phenomena like diffraction or scattering. The new measurement uncertainty experiments have so far been limited to 2-dimensional quantum systems. Uncertainty relation is also related to other physical issues. The most obvious example is complementarity, which can be regarded as uncertainty relations in disguise [108]. Similar to the previous analysis, a which way detector could be modeled by an indirect measurement model, which would correlate error and disturbance with *visibility* and *distinguishability*. A new evaluation of Englert's definitions

[109] in this context in a neutron interferometer would be feasible.

At the polarimeter beam line at the TRIGA Mark-II the next task should be testing the preparation uncertainty relation Eq. (5.24). The peculiarity of this inequality is that it can naturally be expanded to trade-offs of 3 or more observables. Working on the theory [107] and comparing it to the existing entropic noise and disturbance could also be interesting in the near future. If coils with oscillating magnetic fields are installed in the beamline, energy-dependent measurements with time-resolved detections could be made. In any case, the experimental research of two-dimensional uncertainty relations still carries on.

Index

Frequently used symbols

$(\Omega, \mathcal{F}, p)$	probability space.
$D(\mathcal{M}, \hat{B})$	optimal entropic disturbance of $\hat{B}$ w.r.t instrument $\mathcal{M}$.
$D_{\mathcal{E}}(\mathcal{M}, \hat{B})$	entropic disturbance of $\hat{B}$ w.r.t instrument $\mathcal{M}$ and error correction $\mathcal{E}$.
$E(\hat{A}, \hat{B})$	preparation uncertainty region of observables $\hat{A}, \hat{B}$.
$F(\hat{\rho}, \hat{\sigma})$	quantum state fidelity of states $\hat{\rho}, \hat{\sigma}$.
$H(X)$	Shannon entropy of random variable X.
$H(X, Y)$	joint Shannon entropy.
$H(X\|Y)$	conditional Shannon entropy.
$H_b(x)$	binary Shannon entropy.
$N(\mathcal{M}, \hat{A})$	entropic noise of $\hat{A}$ w.r.t instrument $\mathcal{M}$.
$R_{ND}(\hat{A}, \hat{B})$	measurement uncertainty region of noise of $\hat{A}$ and disturbance of $\hat{B}$.
$R_{NN}(\hat{A}, \hat{B})$	measurement uncertainty region of noise of $\hat{A}$ and of $\hat{B}$.
$W_p(\mu, \nu)$	Wasserstein p-metric of measures μ and ν.
Ω	sample space.
$\Sigma, \mathcal{F}$	σ - algebras.
$\epsilon(\hat{A})$	error of observable $\hat{A}$.
$\eta(\hat{B})$	disturbance of observable $\hat{B}$.
$\hat{A}^+$	pseudoinverse of $\hat{A}$.
$\hat{D}(\hat{B})$	disturbance operator of observable $\hat{B}$.
$\hat{N}(\hat{A})$	noise operator of observable $\hat{A}$.

$\hat{P}(a_i)$ or $\hat{P}_a$	projector with eigenvalue outcome space $\{a_i\}$.
$\hat{\Pi}(X)$	effect with outcome X.
$\hat{\rho}$	density operator.
$\hat{\sigma}_i$	Pauli matrices.
$\langle \mathcal{K}, \hat{\xi}, \hat{U}, \hat{M} \rangle$	indirect measurement model.
$\mathcal{E}$	quantum operations.
$\mathcal{I}_L(m)(\hat{\rho})$	Lüders instrument.
$\mathcal{I}_{\mathcal{M}}(X)(\hat{\rho}), \mathcal{M}$	quantum instrument.
$\mathcal{L}(\mathcal{H})$	set of all bounded linear maps.
$\mathcal{O}(\Omega, \mathcal{F}, \mathcal{H})$	set of observables.
$\mathcal{S}(\mathcal{H})$	set of states.
$\sigma(X)$	standard deviation of random variable X.
$conv(X)$	convex hull of set X.
$cov(X, Y)$	covariance of random variables X and Y.
$g(x)$	inverse function of the binary Shannon entropy.
$h(x)$	differential entropy.
$p(A, B)$	joint probability.
$p(A\|B)$	conditional probability.

Bibliography

[1] M. Born, W. Heisenberg, and P. Jordan, "Zur Quantenmechanik. II." *Zeitschrift für Physik* **35**, 557 (1926).

[2] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen Kinematik und Mechanik," *Zeitschrift für Physik* **43**, 172 (1927).

[3] W. Heisenberg, C. H. Eckhart, and F. C. Hoyt, *The Physical Principles of the Quantum Theory* (University of Chicago Press, 1930).

[4] V. Jijnasu, "The uncertainty principle - A simplified review of the four versions," *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **55**, 62 (2016).

[5] E. H. Kennard, "Zur Quantenmechanik einfacher Bewegungstypen," *Zeitschrift für Physik* **44**, 326 (1927).

[6] H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, 1928).

[7] G. H. Hardy, "A theorem concerning fourier transforms," *Journal of the London Mathematical Society* **s1-8**, 227 (1933).

[8] H. P. Robertson, "The Uncertainty Principle," *Phys. Rev.* **34**, 163 (1929).

[9] E. Schrödinger, *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (Akademie der Wissenschaften, 1930).

[10] K. Popper and v. Weizscker, "Zur Kritik der Ungenauigkeitsrelationen," *Naturwissenschaften* **22**, 807 (1934).

[11] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty - II," *The Bell System Technical Journal* **40**, 65 (1961).

[12] I. I. Hirschman, "A Note on Entropy," *American Journal of Mathematics* **79**, 152 (1957).

[13] W. Beckner, "Inequalities in Fourier Analysis," *Annals of Mathematics* **102**, 159 (1975).

[14] I. Białynicki-Birula and J. Mycielski, "Uncertainty relations for information entropy in wave mechanics," *Communications in Mathematical Physics* **44**, 129 (1975).

[15] D. Deutsch, "Uncertainty in Quantum Measurements," *Phys. Rev. Lett.* **50**, 631 (1983).

[16] H. Maassen and J. B. M. Uffink, "Generalized entropic uncertainty relations," *Phys. Rev. Lett.* **60**, 1103 (1988).

[17] E. Arthurs and J. L. Kelly, "B.S.T.J. briefs: On the simultaneous measurement of a pair of conjugate observables," *The Bell System Technical Journal* **44**, 725 (1965).

[18] E. Arthurs and M. S. Goodman, "Quantum Correlations: A Generalized Heisenberg Uncertainty Relation," *Phys. Rev. Lett.* **60**, 2447 (1988).

[19] D. M. Appleby, "Concept of Experimental Accuracy and Simultaneous Measurements of Position and Momentum," *International Journal of Theoretical Physics* **37**, 1491 (1998).

[20] M. Ozawa, "Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement," *Phys. Rev. A* **67**, 042105 (2003).

[21] M. Ozawa, "Uncertainty relations for noise and disturbance in generalized quantum measurements," *Annals of Physics* **311**, 350 (2004).

[22] M. J. W. Hall, "Prior information: How to circumvent the standard joint-measurement uncertainty relation," *Phys. Rev. A* **69**, 052113 (2004).

[23] P. Busch, P. Lahti, and R. F. Werner, "Proof of Heisenberg's Error-Disturbance Relation," *Phys. Rev. Lett.* **111**, 160405 (2013).

[24] P. Busch, P. Lahti, and R. F. Werner, "Colloquium: Quantum root-mean-square error and measurement uncertainty relations," *Rev. Mod. Phys.* **86**, 1261 (2014).

[25] P. Busch, P. Lahti, and R. F. Werner, "Heisenberg uncertainty for qubit measurements," *Phys. Rev. A* **89**, 012129 (2014).

[26] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, "Noise and Disturbance in Quantum Measurements: An Information-Theoretic Approach," *Phys. Rev. Lett.* **112**, 050401 (2014).

[27] A. A. Abbott and C. Branciard, "Noise and disturbance of qubit measurements: An information-theoretic characterization," *Phys. Rev. A* **94**, 062110 (2016).

[28] A. Barchielli, M. Gregoratti, and A. Toigo, "Measurement Uncertainty Relations for Position and Momentum: Relative Entropy Formulation," *Entropy* **19**, 301 (2017).

[29] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, "Entropic uncertainty relations and their applications," *Rev. Mod. Phys.* **89**, 015002 (2017).

[30] J. Klepp, S. Sponar, and Y. Hasegawa, "Fundamental phenomena of quantum mechanics explored with neutron interferometers," *Progress of Theoretical and Experimental Physics* **2014**, 082A01 (2014).

[31] H. Rauch and S. Werner, *Neutron Interferometry: Lessons in Experimental Quantum Mechanics, Wave-particle Duality, and Entanglement* (Oxford University Press, 2015).

[32] M. Zawisky, M. Baron, and R. Loidl, "Three-beam interference and which-way information in neutron interferometry," *Phys. Rev. A* **66**, 063608 (2002).

[33] C. G. Shull, "Single-Slit Diffraction of Neutrons," *Phys. Rev.* **179**, 752 (1969).

[34] H. Kaiser, S. A. Werner, and E. A. George, "Direct Measurement of the Longitudinal Coherence Length of a Thermal Neutron Beam," *Phys. Rev. Lett.* **50**, 560 (1983).

[35] A. G. Klein, G. I. Opat, and W. A. Hamilton, "Longitudinal Coherence in Neutron Interferometry," *Phys. Rev. Lett.* **50**, 563 (1983).

[36] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, "Experimental demonstration of a universally valid error-disturbance uncertainty relation in spin measurements," *Nature Physics* **8**, 185 (2012).

[37] E. Desurvire, *Classical and Quantum Information Theory: An Introduction for the Telecom Scientist* (Cambridge University Press, 2009).

[38] R. Mathar, *Informationstheorie: Diskrete Modelle und Verfahren* (Vieweg+Teubner Verlag, 2013).

[39] C. Villani, *Optimal Transport: Old and New* (Springer Berlin Heidelberg, 2008).

[40] A. Khrennikov, *Interpretations of Probability* (De Gruyter, 2009).

[41] A. Kolmogorov, *Foundations of the Theory of Probability* (Martino Fine Books, 2013).

[42] S. Banach and A. Tarski, "Sur la décomposition des ensembles de points en parties respectivement congruentes," *Fund. math* **6**, 924 (1924).

[43] M. M. Deza and E. Deza, *Encyclopedia of Distances* (Springer Berlin Heidelberg, 2014).

[44] R. Adamczak, "Metric and classical fidelity uncertainty relations for random unitary matrices," *Journal of Physics A: Mathematical and Theoretical* **50**, 105302 (2017).

[45] S. Fehr and S. Berens, "On the Conditional Rényi Entropy," *IEEE Transactions on Information Theory* **60**, 6801 (2014).

[46] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, *Handbook of Differential Entropy* (Taylor & Francis, 2013).

[47] D. Greenberger, K. Hentschel, and F. Weinert, *Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy* (Springer Berlin Heidelberg, 2009).

[48] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement* (Cambridge University Press, 2011).

[49] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, 2007).

[50] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[51] E. B. Davies and J. T. Lewis, "An operational approach to quantum probability," *Comm. Math. Phys.* **17**, 239 (1970).

[52] S. Massar and P. Spindel, "Uncertainty Relation for the Discrete Fourier Transform," *Phys. Rev. Lett.* **100**, 190401 (2008).

[53] S. Bagchi and A. K. Pati, "Uncertainty relations for general unitary operators," *Phys. Rev. A* **94**, 042104 (2016).

[54] G. Ludwig, *Foundations of Quantum Mechanics I* (Springer-Verlag Berlin Heidelberg, 1983).

[55] A. Furusawa and P. van Loock, *Quantum Teleportation and Entanglement: A Hybrid Approach to Optical Quantum Information Processing* (Wiley, 2011).

[56] M. Ozawa, "Quantum measuring processes of continuous observables," *Journal of Mathematical Physics* **25**, 79 (1984).

[57] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer Berlin Heidelberg, 2014).

[58] P. Busch, "No Information Without Disturbance: Quantum Limitations of Measurement," in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony* (Springer Netherlands, Dordrecht, 2009) pp. 229–256.

[59] P. Busch and J. Singh, "Lüders theorem for unsharp quantum measurements," *Physics Letters A* **249**, 10 (1998).

[60] H. Böck, Y. Hasegawa, E. Jericha, G. Steinhauser, and M. Villa, "The past and the future of the triga reactor in vienna," *J. Energ. Power. Eng.* **7**, 654 (2013).

[61] "The TRIGA Mark-II Reactor," `http://ati.tuwien.ac.at/reactor/EN/`.

[62] V. Sears, *Neutron optics: an introduction to the theory of neutron optical phenomena and their applications*, Oxford series on neutron scattering in condensed matter (Oxford University Press, 1989).

[63] W. Greiner, *Classical Electrodynamics* (Springer New York, 2012).

[64] G. Badurek, R. J. Buchelt, G. Kroupa, M. Baron, and M. Villa, "Permanent magnetic field-prism polarizer for perfect crystal neutron interferometers," *Physica B: Condensed Matter* **283**, 389 (2000).

[65] J. Dressel and F. Nori, "Certainty in Heisenberg's uncertainty principle: Revisiting definitions for estimation errors and disturbance," *Phys. Rev. A* **89**, 022106 (2014).

[66] P. Gibilisco and T. Isola, "How to distinguish quantum covariances using uncertainty relations," *Journal of Mathematical Analysis and Applications* **384**, 670 (2011).

[67] M. de Gosson and F. Luef, "Quantum States and Hardys Formulation of the Uncertainty Principle: a Symplectic Approach," *Letters in Mathematical Physics* **80**, 69 (2007).

[68] C. Bastos, A. E. Bernardini, O. Bertolami, N. Costa Dias, and J. a. N. Prata, "Robertson-Schrödinger-type formulation of Ozawa's noise-disturbance uncertainty principle," *Phys. Rev. A* **89**, 042112 (2014).

[69] D. Gabor, *Theory of communication* (Institution of Electrical Engineering, 1946).

[70] L. I. Mandelshtam and I. E. Tamm, "The uncertainty relation between energy and time in nonrelativistic quantum mechanics," *J. Phys. (USSR)* **9**, 249 (1945).

[71] P. Busch, "The Time-Energy Uncertainty Relation," in *Time in Quantum Mechanics* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 73–105.

[72] L. Maccone and A. K. Pati, "Stronger Uncertainty Relations for All Incompatible Observables," *Phys. Rev. Lett.* **113**, 260401 (2014).

[73] B. S. Dewitt and N. Graham, *The Many Worlds Interpretation of Quantum Mechanics* (Princeton University Press, 2015).

[74] M. H. Partovi, "Entropic Formulation of Uncertainty for Quantum Measurements," *Phys. Rev. Lett.* **50**, 1883 (1983).

[75] K. Kraus, "Complementary observables and uncertainty relations," *Phys. Rev. D* **35**, 3070 (1987).

[76] J. Sánchez, "Entropic uncertainty and certainty relations for complementary observables," *Physics Letters A* **173**, 233 (1993).

[77] J. Sánches-Ruiz, "Optimal entropic uncertainty relation in two-dimensional hilbert space," *Physics Letters A* **244**, 189 (1998).

[78] G. Ghirardi, L. Marinatto, and R. Romano, "An optimal entropic uncertainty relation in a two-dimensional hilbert space," *Physics Letters A* **317**, 32 (2003).

[79] A. Abbott, P.-L. Alzieu, M. Hall, and C. Branciard, "Tight State-Independent Uncertainty Relations for Qubits," *Mathematics* **4**, 8 (2016).

[80] S. Wehner and A. Winter, "Entropic uncertainty relations–a survey," *New Journal of Physics* **12**, 025009 (2010).

[81] P. Busch, T. Heinonen, and P. Lahti, "Noise and disturbance in quantum measurement," *Physics Letters A* **320**, 261 (2004).

[82] X.-M. Lu, S. Yu, K. Fujikawa, and C. H. Oh, "Improved error-tradeoff and error-disturbance relations in terms of measurement error components," *Phys. Rev. A* **90**, 042113 (2014).

[83] C. Branciard, "Error-tradeoff and error-disturbance relations for incompatible quantum measurements," *Proceedings of the National Academy of Sciences* **110**, 6742 (2013).

[84] M. Ozawa, "Error-disturbance relations in mixed states," *arXiv:1404.3388v1 [quant-ph]* (2014).

[85] M. Hayashi, *Quantum Information: An Introduction* (Springer Berlin Heidelberg, 2006).

[86] Y. Aharonov, D. Z. Albert, and L. Vaidman, "How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100," *Phys. Rev. Lett.* **60**, 1351 (1988).

[87] M. J. W. Hall, A. K. Pati, and J. Wu, "Products of weak values: Uncertainty relations, complementarity, and incompatibility," *Phys. Rev. A* **93**, 052118 (2016).

[88] A. P. Lund and H. M. Wiseman, "Measuring measurement-disturbance relationships with weak values," *New Journal of Physics* **12**, 093011 (2010).

[89] F. Kaneda, S.-Y. Baek, M. Ozawa, and K. Edamatsu, "Experimental Test of Error-Disturbance Uncertainty Relations by Weak Measurement," *Phys. Rev. Lett.* **112**, 020402 (2014).

[90] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, "Violation of Heisenberg's Measurement-Disturbance Relationship by Weak Measurements," *Phys. Rev. Lett.* **109**, 100404 (2012).

[91] M. Ringbauer, D. N. Biggerstaff, M. A. Broome, A. Fedrizzi, C. Branciard, and A. G. White, "Experimental Joint Quantum Measurements with Minimum Uncertainty," *Phys. Rev. Lett.* **112**, 020401 (2014).

[92] P. Busch and N. Stevens, "Direct Tests of Measurement Uncertainty Relations: What It Takes," *Phys. Rev. Lett.* **114**, 070402 (2015).

[93] G. Sulyok, S. Sponar, J. Erhart, G. Badurek, M. Ozawa, and Y. Hasegawa, "Violation of Heisenberg's error-disturbance uncertainty relation in neutron-spin measurements," *Phys. Rev. A* **88**, 022110 (2013).

[94] B. Demirel, S. Sponar, G. Sulyok, M. Ozawa, and Y. Hasegawa, "Experimental Test of Residual Error-Disturbance Uncertainty Relations for Mixed Spin-$\frac{1}{2}$ States," *Phys. Rev. Lett.* **117**, 140402 (2016).

[95] J. Klepp, S. Sponar, S. Filipp, M. Lettner, G. Badurek, and Y. Hasegawa, "Observation of Nonadditive Mixed-State Phases with Polarized Neutrons," *Phys. Rev. Lett.* **101**, 150404 (2008).

[96] P. Busch, P. Lahti, J. P. Pellonp, and K. Ylinen, *Quantum Measurement* (Springer International Publishing, 2016).

[97] G. Sulyok and S. Sponar, "Heisenberg's error-disturbance uncertainty relation: Experimental study of competing approaches," *Phys. Rev. A* **96**, 022137 (2017).

[98] W. Ma, Z. Ma, H. Wang, Z. Chen, Y. Liu, F. Kong, Z. Li, X. Peng, M. Shi, F. Shi, S.-M. Fei, and J. Du, "Experimental Test of Heisenberg's Measurement Uncertainty Relation Based on Statistical Distances," *Phys. Rev. Lett.* **116**, 160405 (2016).

[99] F. Zhou, L. Yan, S. Gong, Z. Ma, J. He, T. Xiong, L. Chen, W. Yang, M. Feng, and V. Vedral, "Verifying Heisenbergs error-disturbance relation using a single trapped ion," *Sci Adv* **2** (2016).

[100] Y.-X. Zhang, Z.-E. Su, X. Zhu, S. Wu, and Z.-B. Chen, "Quantum uncertainty switches on or off the error-disturbance tradeoff," *Scientific Reports* **6**, 26798 (2016).

[101] D. Appleby, "Quantum Errors and Disturbances: Response to Busch, Lahti and Werner," *Entropy* **18**, 174 (2016).

[102] G. Sulyok, S. Sponar, B. Demirel, F. Buscemi, M. J. W. Hall, M. Ozawa, and Y. Hasegawa, "Experimental Test of Entropic Noise-Disturbance Uncertainty Relations for Spin-1/2 Measurements," *Phys. Rev. Lett.* **115**, 030401 (2015).

[103] B. Demirel, S. Sponar, A. A. Abbott, C. Branciard, and Y. Hasegawa, "Experimental test of an entropic measurement uncertainty relation for arbitrary qubit observables," *arXiv:1711.05023v1 [quant-ph]* (2017).

[104] M. H. Partovi, "Majorization formulation of uncertainty in quantum mechanics," *Phys. Rev. A* **84**, 052117 (2011).

[105] S. Friedland, V. Gheorghiu, and G. Gour, "Universal Uncertainty Relations," *Phys. Rev. Lett.* **111**, 230401 (2013).

[106] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, "Simulating Positive-Operator-Valued Measures with Projective Measurements," *Phys. Rev. Lett.* **119**, 190501 (2017).

[107] A. Barchielli, M. Gregoratti, and A. Toigo, "Measurement uncertainty relations for discrete observables: Relative entropy formulation," *arXiv:1608.01986v2 [math-ph]* (2017).

[108] P. J. Coles, J. Kaniewski, and S. Wehner, "Equivalence of wave-particle duality to entropic uncertainty," *Nature Communications* **5**, 5814 (2014).

[109] B.-G. Englert, "Fringe Visibility and Which-Way Information: An Inequality," *Phys. Rev. Lett.* **77**, 2154 (1996).

Curriculum Vitae

Personal information

Name	Bülent Demirel
Address	Liechtensteinstraße 101/4
	1090 Wien
Date of birth	29.03.1989
Nationality	Austrian
E-mail	b.demirel89@hotmail.com

Work experience

2010 - 2011	Community service
	Jugend am Werk, Vienna (Austria)
2011 - present	Voluntary social service
	Jugend am Werk, Vienna (Austria)
2013 - present	Project assistant (FWF)
	Atominstitut - Vienna University of Technology, Vienna (Austria)

Education and training

2000 - 2007	Gymnasium
	AHS Albertus Magnus School, Vienna (Austria)
2007 - 2010	Bachelor of science
	Vienna University of Technology, Vienna (Austria)
2010 - 2013	Master of Science
	Vienna University of Technology, Vienna (Austria)

Language skills

Languages	German (native level)
	Turkish (native level)
	English (good)

Scientific publications

Related to thesis

G. Sulyok, S. Sponar, B. Demirel, F. Buscemi, M. J. W. Hall, M. Ozawa, and Y. Hasegawa, Experimental Test of Entropic Noise-Disturbance Uncertainty Relations for Spin-1/2 Measurements," Phys. Rev. Lett. 115, 030401 (2015).

B. Demirel, S. Sponar, G. Sulyok, M. Ozawa, and Y. Hasegawa, Experimental Test of Residual Error-Disturbance Uncertainty Relations for Mixed Spin-$\frac{1}{2}$ States," Phys. Rev. Lett. 117, 140402 (2016).

B. Demirel, S. Sponar, A. A. Abbott, C. Branciard, and Y. Hasegawa, Experimental test of an entropic measurement uncertainty relation for arbitrary qubit observables," arXiv:1711.05023v1 [quant-ph] (2017).

Unelated to thesis

B. Demirel, S. Sponar, and Y. Hasegawa, Measurement of the spin-rotation coupling in neutron polarimetry," New Journal of Physics 17, 023065 (2015).