Quantum Probability and Quantum Information Theory

H. Maassen

1 Introduction

From its very birth in the 1920s, quantum theory has been characterized by a certain strangeness: It seems to run counter to the intuitions that we humans have about the world we live in.

According to these "realistic" intuitions all things have their definite place and sharply determined qualities, such as speed, color, and weight. Quantum theory, however, refuses to precisely pinpoint them. With respect to this apparent shortcoming of the theory different points of view can be taken. It could be suspected that quantum theory is incomplete, in that it gives a coarse description of a reality that is actually more refined. This is the viewpoint once taken by Einstein, and it still has adherents today. It calls for a search for finer mathematical models of physical reality, based on classical probability, often referred to as "hidden variable models" (see chapter "Photonic Realization of Quantum Information Protocols"). One such attempt is Bohm's theory of non-relativistic quantum mechanics.

However, the work of John Bell in the 1960s and of Alain Aspect in the 1970s and 1980s strongly favors the opposite point of view: Their work has made clear that such models with a classical probabilistic structure are necessarily afflicted with a certain weakness; they must at least allow *action at a distance*. This we regard as a bad property for a theory which aims to describe a physical world where no signals have been observed to travel faster than light. Apart from that, the hidden variable theories which have been found so far are highly artificial and cannot be tested against quantum mechanics since they do not predict any new phenomena.

It is for these reasons that we decide to accept quantum theory with its inherent strangeness and are prepared to modify probability theory accordingly.

H. Maassen (✉)
Radboud University, Nijmegen, The Netherlands, maassen@math.ru.nl

Maassen, H.: *Quantum Probability and Quantum Information Theory*. Lect. Notes Phys. **808**, 65–108 (2010)
DOI 10.1007/978-3-642-11914-9_3

1.1 Quantum Probability

So quantum mechanics does not predict the results of physical experiments with certainty, but calculates probabilities for their possible outcomes.

Now, the classical mathematical theory of probability obtained a unified formulation in the 1930s, when Kolmogorov introduced his axioms, defining the universal structure $(\Omega, \Sigma, \mathbf{P})$ of a probability space. For a long time this theory of probability (dealing with probability distributions, stochastic processes, Markov chains, martingales, etc.) remained completely separate from the mathematical development of quantum mechanics (involving vectors in a Hilbert space, Hermitian operators, unitary transformations, and such like).

In the 1970s and 1980s people around Accardi, Lewis, Davies, Kümmerer, building on ideas of von Neumann's and Segal's, developed a unified framework, a generalized, "non-commutative," probability theory, in which classical probability theory and quantum mechanics can be discussed in unison. It consists of ordinary Hilbert space quantum theory, with the emphasis moved toward operators on Hilbert space, and the algebras which they generate. The main objective of this chapter is to sketch the outlines of this framework and show its usefulness for information theory.

1.2 Quantum Information

In Shannon's (classical) information theory (see chapter "Classical Information Theory"), a single unit, the *bit*, serves to quantify all forms of information, be it in print, computer memory, CD-ROM, or strings of DNA. Such a single unit suffices, because different forms of information can be converted into each other by copying, according to fixed "exchange rates." The physical states of quantum systems, however, cannot be copied into such "classical" information, but *can* be converted into each other. This leads to a new unit of information: the *qubit*.

Quantum information theory studies the handling of this new form of information by information-carrying channels. We shall treat the basic properties of these channels and some impossibilities as well as new possibilities connected with quantum information. The impossibility of copying makes quantum information an ideal means to establish secrecy (see chapter "Quantum Cryptography").

1.3 Quantum Computing

It was Richard Feynman who first thought of *employing* the strangeness of quantum mechanics to do things that would be impossible in a classical world.

The idea was developed in the 1980s and 1990s by David Deutsch, Peter Shor, and many others into a flourishing branch of science called "quantum computing": How to make quantum mechanical systems perform calculations more efficiently than ordinary computers can do. This research is still in a predominantly theoretical stage: The quantum computers actually built are as yet extremely primitive and can by no means compete with even the simplest pocket calculator, but expectations are high (see chapter “Quantum Algorithms”).

1.4 This Chapter

We start with an introduction to quantum probability. In Sect. 2 we demonstrate the “strangeness” of quantum phenomena by very simple polarization experiments, culminating in Bell’s famous inequality, tested in Aspect’s experiment. Bell’s inequality is a statement in classical probability that is violated in quantum probability and in reality.

Taking polarizers as our starting point, in Sects. 3 and 4 we build up the new probability theory in terms of algebras of operators on a Hilbert space. In Sect. 5 operations on these algebras will be characterized, and some aspects will be discussed in which they differ from classical physical operations. They are subject to certain strange limitations: The impossibility of copying, of coding information into bits, of jointly measuring incompatible observables, of observation without perturbing the object (cf. Sect. 6). But they also open up surprising possibilities: entangling remote systems, teleportation of this entanglement, sending two bits in a single qubit (cf. Sect. 7). Further luring perspectives as highly efficient algorithms for sorting, Fourier transformation, and factoring very large numbers will be treated in chapter “Quantum Algorithms”.

2 Why Classical Probability Does Not Suffice

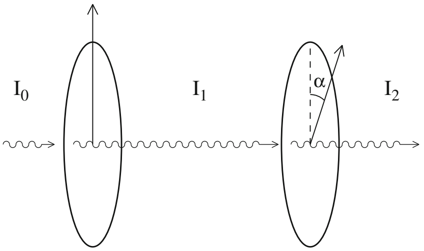
(This section is based on [8].)

2.1 An Experiment with Polarizers

To start with, we consider a simple experiment. In a beam of light of a fixed color we put a pair of polarizing filters, each of which can be rotated around the axis formed by the beam. As is well known, the light falling through both filters changes in intensity when the filters are rotated relative to each other. Starting from the orientation where the resulting intensity is maximal and rotating one of the filters through an angle α, the light intensity decreases with α, vanishing for $\alpha = \frac{1}{2}\pi$. If we call the intensity of the beam before the filters I_0, after the first I_1, and after the second I_2, then $I_1 = \frac{1}{2}I_0$ (we assume the original beam to be unpolarized), and (Fig. 1)

$$I_2 = I_1 \cos^2 \alpha \ . \tag{1}$$

Fig. 1 Two polarizers in conjunction



So far the phenomenon is described well by classical physics. During the last century, however, it has been observed that for very low intensities (monochromatic) light comes in small packages, which were called *photons*, whose energy depends on the color, but not on the total intensity. So the intensity must be proportional to the *number* of these photons, and formula (1) must be given a statistical meaning: A photon passing through the first filter has a probability $\cos^2\alpha$ to pass through the second. Formula (1) then holds only on the average, for large numbers of photons.

Thinking along the lines of classical probability, we may associate with a polarization filter in the direction α a random variable P_α, taking the value $P_\alpha(\omega) = 0$ if the photon ω is absorbed by the filter and $P_\alpha(\omega) = 1$ if it passes through. For two filters in the directions α and β these random variables then should be correlated as follows:

$$\mathbb{E}(P_\alpha P_\beta) \;=\; \mathbf{P}[P_\alpha = 1,\, P_\beta = 1] \;=\; \tfrac{1}{2}\cos^2(\alpha - \beta). \tag{2}$$

Here we hit on a difficulty: The function on the right-hand side is not a possible correlation function! This can be seen as follows. Take three polarizing filters, having polarization directions α_1, α_2, and α_3, respectively. Put them on the optical bench in pairs. They should give rise to random variables P_1, P_2, and P_3 satisfying

$$\mathbb{E}(P_i P_j) = \tfrac{1}{2}\cos^2(\alpha_i - \alpha_j)\;. \tag{3}$$

Proposition 1 (Bell's three-variable inequality) *[2] For any three 0–1-valued random variables P_1, P_2, and P_3 on a probability space $(\Omega, \mathbf{P})$ the following inequality holds:*

$$\mathbf{P}[P_1 = 1,\, P_3 = 0] \;\leq\; \mathbf{P}[P_1 = 1,\, P_2 = 0] + \mathbf{P}[P_2 = 1,\, P_3 = 0]. \tag{4}$$

Proof

$$\begin{aligned}\mathbf{P}[P_1 = 1,\, P_3 = 0] &= \mathbf{P}[P_1 = 1,\, P_2 = 0,\, P_3 = 0] + \mathbf{P}[P_1 = 1,\, P_2 = 1,\, P_3 = 0]\\ &\leq \mathbf{P}[P_1 = 1,\, P_2 = 0] + \mathbf{P}[P_2 = 1,\, P_3 = 0]. \qquad \square\end{aligned} \tag{5}$$

In our example, we have

$$\begin{aligned}\mathbf{P}[P_i = 1, P_j = 0] &= \mathbf{P}[P_i = 1] - \mathbf{P}[P_i = 1, P_j = 1] \\ &= \tfrac{1}{2} - \tfrac{1}{2}\cos^2(\alpha_i - \alpha_j) = \tfrac{1}{2}\sin^2(\alpha_i - \alpha_j).\end{aligned} \tag{6}$$

Bell's inequality thus reads

$$\tfrac{1}{2}\sin^2(\alpha_1 - \alpha_3) \leq \tfrac{1}{2}\sin^2(\alpha_1 - \alpha_2) + \tfrac{1}{2}\sin^2(\alpha_2 - \alpha_3), \tag{7}$$

which is clearly violated for the choices $\alpha_1 = 0$, $\alpha_2 = \frac{1}{6}\pi$, and $\alpha_3 = \frac{1}{3}\pi$, where it says that

$$\frac{3}{8} \leq \frac{1}{8} + \frac{1}{8}\,. \tag{8}$$

This example suggests that classical probability cannot even describe this simple experiment!

Remark 1 The above calculation could be summarized as follows: we are in fact looking for a family of 0–1-valued random variables $(P_\alpha)_{0 \leq \alpha < \pi}$ with $\mathbf{P}[P_\alpha = 1] = \frac{1}{2}$, satisfying the requirement that

$$\mathbf{P}[P_\alpha \neq P_\beta] = \sin^2(\alpha - \beta). \tag{9}$$

Now, on the space of 0–1-valued random variables on a probability space the function $(X, Y) \mapsto \mathbf{P}[X \neq Y]$ equals the L^1-distance of X and Y:

$$\mathbf{P}[X \neq Y] = \int_\Omega |X(\omega) - Y(\omega)|\, \mathbf{P}(\mathrm{d}\omega) = \|\, X - Y \,\|_1\,. \tag{10}$$

On the other hand, the function $(\alpha, \beta) \mapsto \sin^2(\alpha - \beta)$ does not satisfy the triangle inequality for a distance function on the interval $[0, \pi)$. Therefore no family $(P_\alpha)_{0 \leq \alpha < \pi}$ exists which meets requirement (9).

2.2 An Improved Experiment

On closer inspection the above example is not very convincing. Indeed, when two polarizers are arranged on the optical bench, why should not the random variable for the second polarizer depend on the angle of the first? The correlation in (2) would then read

$$\mathbb{E}(P_\alpha P_{\alpha,\beta}) \;=\; \mathbf{P}[P_\alpha = 1, P_{\alpha,\beta} = 1] \;=\; \tfrac{1}{2}\cos^2(\alpha - \beta), \tag{11}$$

which can easily be satisfied, and the whole refutation collapses.

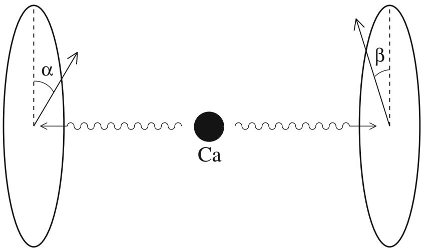


Fig. 2 Photon pair production

So we should do a better experiment. We must let the filters act on the photons without influence on each other. Maybe we can separate them spatially?

Here a clever technique from quantum optics comes to our aid. It is possible to build a device that produces *pairs* of photons, such that the members of each pair move in opposite directions and show opposite behavior toward parallel polarization filters: If one passes the filter, then the other is surely absorbed. The device contains calcium atoms, which are excited by a laser to a state they can only leave under emission of such a pair (Fig. 2).

With these photon pairs, the very same experiment can be performed, but this time the polarizers are far apart, each one acting on its own photon. The same correlations are measured, say first between P_{α_1} on the left and P_{α_2} on the right, then between P_{α_1} on the left and P_{α_3} on the right, and finally between P_{α_2} on the left and P_{α_3} on the right. The same outcomes are found, violating Bell's three-variable inequality, thus strengthening the case against classical probability.

2.3 The Decisive Experiment

Advocates of classical probability could still find serious fault with the argument given so far. Indeed, do we really *have to* assume that we are measuring the same random variable P_{α_2} on the right as later on the left? Is it really true that the polarizations in these pairs are exactly opposite? There could exist a probabilistic explanation of the phenomena without this assumption.

So the argument has to be tightened still further. This brings us to the experiment which was actually performed by A. Aspect in Orsay (near Paris) in 1982 [1]. In this experiment a random choice out of two different polarization measurements was performed on each side of the pair-producing device, say in the direction α_1 or α_2 on the left and in the direction β_1 or β_2 on the right, giving rise to *four* random variables $P_1 := P(\alpha_1)$, $P_2 := P(\alpha_2)$ and $Q_1 := Q(\beta_1)$, $Q_2 := Q(\beta_2)$, two of which are measured and compared at each trial (see chapter "Photonic Realization of Quantum Information Protocols" for more details).

Proposition 2 (Bell's four-variable inequality) *For any quadruple P_1, P_2, Q_1, and Q_2 of 0–1-valued random variables on $(\Omega, \mathbf{P})$ the following inequality holds:*

$$\mathbf{P}[P_1 = Q_1] \;\leq\; \mathbf{P}[P_1 = Q_2] + \mathbf{P}[Q_2 = P_2] + \mathbf{P}[P_2 = Q_1]. \tag{12}$$

(In fact, by symmetry, neither of these four probabilities is larger than the sum of the other three.)
Proof It is easy to see that for all ω

$$P_1(\omega) = Q_1(\omega) \Rightarrow P_1(\omega) = Q_2(\omega) \quad \text{or} \quad Q_2(\omega) = P_2(\omega) \quad \text{or} \quad P_2(\omega) = Q_1(\omega) . \tag{13}$$

□

Bell's four-variable inequality can be viewed as a "quadrangle inequality" with respect to the metric $(X, Y) \mapsto \| X - Y \|_1$ on random variables X, Y.

On the other hand, quantum mechanics predicts (cf. Sect. 3.6), and the experiment of Aspect showed, that one has,

$$\mathbf{P}[P(\alpha) = Q(\beta) = 1] = \tfrac{1}{2} \sin^2(\alpha - \beta). \tag{14}$$

Similarly, $\mathbf{P}[P(\alpha) = Q(\beta) = 0] = \frac{1}{2} \sin^2(\alpha - \beta)$. Hence

$$\mathbf{P}[P(\alpha) = Q(\beta)] = \sin^2(\alpha - \beta). \tag{15}$$

So Bell's four-variable inequality reads in this example:

$$\sin^2(\alpha_1 - \beta_1) \leq \sin^2(\alpha_1 - \beta_2) + \sin^2(\alpha_2 - \beta_1) + \sin^2(\alpha_2 - \beta_2), \tag{16}$$

which is clearly violated for the choices $\alpha_1 = 0$, $\alpha_2 = \frac{\pi}{3}$, $\beta_1 = \frac{\pi}{2}$, and $\beta_2 = \frac{\pi}{6}$, see Fig. 3, in which case it reads

$$1 \leq \frac{1}{4} + \frac{1}{4} + \frac{1}{4} . \tag{17}$$

Now we are finished: There does not exist, on any classical probability space, a quadruple P_1, P_2, Q_1, and Q_2 of random variables with the correlations measured in this experiment.

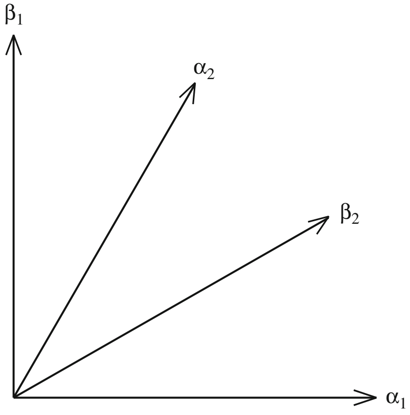


Fig. 3 Directions violating Bell's inequality

Discussion

1. A crucial assumption that goes into Bell's inequality is that it makes sense to compare (cf. (5)) the (possibly random) reactions which a given photon *would* show to different filters, including those it does not actually meet. This assumption is called *realism*; it is made in all classical probabilistic physical theories, but is abandoned in quantum mechanics.
2. A second important assumption, necessary for the validity of Bell's inequality, was mentioned before: The *outcome* on the right (described by $Q(\beta)$ for some β) should not depend on the *angle* α of the polarizer on the left. This assumption is called "locality." In order to justify this assumption, Aspect has made considerable efforts. In his (third) experiment [1], the choice of what to measure on the left (α_1 or α_2) and on the right (β_1 or β_2) was made *during the flight of the photons*, so that any influence which each of these choices might have on *the outcome* on the opposite end would have to travel faster than light. By the causality principle of relativity theory such influences are excluded.
3. The Orsay experiment refutes all imaginable physical theories which are both *local* and *realistic* (cf. 1 and 2 above). Quantum mechanics is local, but not realistic. Its great successes lead us to believe that realism fails for the description of nature. Some prefer to adhere to realism, and so they must give up locality, and hence Einstein causality [4, 9].
4. In our opinion, the phrase "quantum non-locality," which is often heard in the context of Bell's inequalities, signals a misconception. It suggests giving up *both* realism *and* locality. This is too much of a defeat and unnecessary. Quantum mechanics is local. But it describes phenomena which *in a classical theory* could only be explained using some action at a distance.

2.4 The Orsay Experiment as a Card Game

To illustrate the above refutation of local realism more vividly, we shall present the experiment in the form of a card game. Nature can win this game. Can you?

Two players, P and Q, are sitting at a table. They are cooperating to achieve a single goal. There is an arbiter present to deal cards and to count points. On the table there is a board consisting of four squares as drawn in Fig. 4. There are dice and an ordinary deck of playing cards. The deck of cards is shuffled well. (In fact we shall assume that the deck of cards is an infinite sequence of independent cards, chosen fully at random.) First the players are given some time to make agreements on the strategy they are going to follow. Then the game starts, and from this moment on they are no longer allowed to communicate. The following sequence of actions is then repeated many times:

1. The dealer hands a card to P and one to Q. Both look at their own card, but not at the other one's. (The only feature of the card that matters is its color: red or black.)

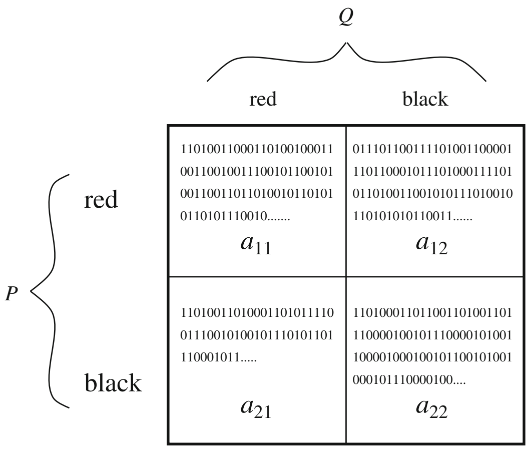


Fig. 4 Board for the Bell game

2. The dice are thrown.
3. P and Q simultaneously say "yes" or "no," according to their own choice. They are free to make their answer depending on any information they possess, such as the color of their own card, the agreements made in advance, the numbers shown by the dice, the weather, the time.
4. The cards are laid out on the table. The pair of colors of the cards determines one of the four squares on the board: These are labeled (red, red), (red, black), (black, red), and (black, black).
5. In the square so determined a 0 or a 1 is written: a 0 when the answers of P and Q have been different, a 1 if they have been the same.

In the course of time, the squares on the board get filled with 0s and 1s. The arbiter keeps track of the percentage of 1s in proportion to the total number of bits in each square; we shall call the time limits of these percentages as the game proceeds: a_{11}, a_{12}, a_{21}, and a_{22}. The aim of the game, for both P and Q, is to get a_{11} larger than the sum of the other three limiting percentages. So P and Q must try to give identical answers as often as they can when both their cards are red, but different answers otherwise.

Proposition 3 (Bell's inequality for the game) *P and Q cannot win the game by classical means, namely*

$$a_{11} \leq a_{12} + a_{21} + a_{22} \ . \tag{18}$$

Proof The best P and Q can do, in order to win the game, is to agree upon some (possibly random) strategy for each turn. For instance, they may agree that P will always say "yes" (i.e., $P_{\text{red}} = P_{\text{black}} =$"yes") and that Q will answer the question "Is my card red?" (i.e., $Q_{\text{red}} =$ "yes" and $Q_{\text{black}} =$"no"). This will lead to a 1 in the (red, red) square or the (black, red) square or to a 0 in one of the other two.

So if the players repeat this strategy indefinitely, on the long run they would get $a_{11} = a_{12} = 1$ and $a_{21} = a_{22} = 0$, disappointingly satisfying Bell's inequality.

The above example is an extremal strategy. There are many (in fact, 16) strategies like this. By the point-wise version (13) of Bell's four-variable inequality, none of these 16 extremal strategies wins the game. Inclusion of the randomness coming from the dice yields a full polytope of random strategies, having the above 16 as its extremal points. But since the inequalities are linear, this averaging procedure does not help. This "proves" our "proposition." Disbelievers are challenged to find a winning strategy. □

Strangely enough, however, nature does provide us with a strategy to win the game, still essentially based on the $\cos^2$ law 2 for photon absorption! Instead of the dice, put a calcium atom on the table. When the cards have been dealt, P and Q put their polarizers in the direction indicated by their cards. If P has a red card, then he chooses the direction $\alpha_1 = 0$ (cf. Fig. 3). If his card is black, then he chooses $\alpha_2 = \frac{\pi}{3}$. If Q has a red card, then he chooses $\beta_1 = \frac{\pi}{2}$. If his card is black, then he chooses $\beta_2 = \frac{\pi}{6}$. No information on the colors of the cards needs to be exchanged. When the calcium atom has produced its photon pair, each player looks whether his own photon passes his own polarizer, and then says "yes" if it does, "no" if it does not. On the long run they will get $a_{11} = 1, a_{12} = a_{21} = a_{22} = \frac{1}{4}$, and thus they win the game.

So the calcium atom, the quantum mechanical die, makes possible what could not be done with the classical die.

3 Toward a Mathematical Model

Coerced by the foregoing considerations, we give up trying to make a classical probabilistic model in order to explain polarization experiments. Instead, we take these experiments as a paradigm for an alternative type of "quantum" probability, to be developed now.

3.1 A Mathematical Description of Polarization

We have discussed (linear) polarization of a light beam. This is completely characterized by a direction in the plane perpendicular to the light beam. So we simply describe states of polarization by different directions in a two-dimensional real plane $\mathbb{R}^2$, or equivalently by unit vectors $\psi \in \mathbb{R}^2$, $\| \psi \| = 1$, pointing in this direction. Actually, since we cannot distinguish between two states which differ by a rotation of π, we shall describe states of polarization by one-dimensional subspaces of $\mathbb{R}^2$. Given two directions of polarization with an angle α between them, spanned by two unit vectors $\psi, \theta \in \mathbb{R}^2$, the probability to find polarization ϑ when a photon is in the state ψ can be expressed as

$$\cos^2\alpha = \langle\psi, \theta\rangle^2, \tag{19}$$

where $\langle\psi, \theta\rangle$ denotes the scalar product between ψ and θ.

In the mathematical model we should distinguish between the physical state of polarization of a photon on the one hand and the filter on the other hand, i.e., the 0–1-valued random variable which asks whether a photon is polarized in a certain direction. This can be done by identifying the random variable with the orthogonal projection P onto the one-dimensional subspace. We can then write

$$\cos^2\alpha \; = \langle\psi, \theta\rangle^2 \; = \langle\psi, P\psi\rangle \; . \tag{20}$$

Since P is 0–1-valued (a photon passes or is absorbed), this probability is equal to the expectation of this random variable:

$$\langle\psi, P\psi\rangle \; = \; \mathbb{E}(P) \; . \tag{21}$$

3.2 The Full Truth About Polarization: The Qubit

In the foregoing description of polarization things were presented somewhat simpler than they are: We considered only linear polarization, thus disregarding circular polarization. The full description of polarization leads to the quantum mechanics of a two-level system or *qubit*:

State of polarization of a photon ≙ one-dimensional subspace of $\mathbf{C}^2$, described by a unit vector ψ spanning this subspace (and determined only up to a phase).

Polarization filter or generalized 0–1-valued random variable ≙ orthogonal projection P onto a complex one-dimensional subspace.

(Also for left- or right-circular polarization there exist physical filters.)

Probability for a photon, described by ψ, to pass through a filter, described by P ≙ $\langle\psi, P\psi\rangle$.

The set of all states is conveniently parametrized by the unit vectors of the form

$$(\cos\alpha, \mathrm{e}^{\mathrm{i}\phi}\sin\alpha) \in \mathbf{C}^2 \, , \quad -\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}, \quad 0 \le \phi \le \pi \; . \tag{22}$$

3.3 Finite-Dimensional Models

The mathematical model that is used by quantum mechanics is the straightforward generalization of the above description. In order to keep things simple, in the following we restrict ourselves to the quantum mechanics in finite dimension. This generalizes the probability theory of systems with only finitely many states. As in classical probability, the generalization to systems with a countably infinite number of states or a continuum of states is analytically more involved.

The model is as follows: *States* correspond to one-dimensional subspaces of $\mathbf{C}^n$, where the dimension n is determined by the model. Again, a state is described conveniently by some unit vector spanning this subspace.

0–1-*Valued random variables* or *events* are described by orthogonal projections onto linear subspaces of $\mathbf{C}^n$. Here also projections onto higher dimensional subspaces make sense.

The *probability* that a measurement of a random variable P on a system in a state ψ gives the value 1 is given by $\langle \psi, P\psi \rangle$.

Note that we do not assume that every orthogonal projection corresponds to a meaningful random variable. Specification of random variables is part of the description of the mathematical model for a given system. In a truly quantum mechanical situation, typically all projections are used. In contrast to this, classical probability is obtained by allowing only very few projections, as follows.

3.4 Finite Classical Models

A finite probability space is usually described by a finite set $\Omega = \{\omega_1, \ldots, \omega_n\}$ and a probability distribution $(p_1, \ldots, p_n)$, $0 \leq p_i \leq 1$, $\sum_i p_i = 1$, such that the probability for ω_i is p_i. A 0–1-valued random variable is a 0–1-valued function on Ω, i.e., a characteristic function χ_A of some subset $A \subseteq \Omega$. In order to describe such a system in our model, we think of $\mathbf{C}^n$ as the space of complex-valued functions on Ω and use the functions δ_i with $\delta_i(\omega_j) = \delta_{i,j}$ as basis. The states of the system, i.e., the points ω_i of Ω, are now represented by the unit vectors δ_i, $1 \leq j \leq n$. The random variable χ_A is identified with the orthogonal projection P_A onto the linear span of the vectors $\{\delta_i : \omega_i \in A\}$. In our basis χ_A becomes a diagonal matrix with a 1 at the ith place of the diagonal if $\omega_i \in A$, and a 0 otherwise. It is obvious that $\omega_i \in A$ if and only if $\chi_A(\omega_i) = 1$ if and only if $\langle \delta_i, P_A\delta_i \rangle = 1$.

Conversely, any set of pairwise commuting projections on $\mathbf{C}^n$ can be diagonalized simultaneously and thus have an interpretation as a set of classical 0–1-valued random variables. Therefore

Classical probability corresponds to sets of pairwise commuting projections.

3.5 Mixed States

In the above sketch of quantum probability an important point is still missing: How can we describe a situation where a photon has one polarization with some probability q and in another with probability 1–q? Since states must play the role of probability distributions, this combination should be expressed as a single state of the photon.

In general, if P is any 0–1-valued random variable and $\psi_1, \ldots, \psi_k$ are arbitrary quantum states, each occurring with a probability p_i, $1 \leq i \leq k$, $\sum_i p_i = 1$, then the probability that a measurement of P gives 1 is clearly given by

$$\sum_i p_i \langle \psi_i, P\psi_i \rangle \ . \tag{23}$$

A more convenient description of mixed states is obtained as follows (compare chapter "Hilbert Space Methods for Quantum Mechanics"). For a unit vector $\psi \in \mathbf{C}^n$ denote by ρ_ψ the orthogonal projection onto the one-dimensional subspace generated by ψ. In the physics literature, ρ_ψ is often denoted by $|\psi\rangle\langle\psi|$. Let Tr denote the trace operation (see chapter "Hilbert Space Methods for Quantum Mechanics", Eq. (47)) on the $n \times n$ matrices, summing up the diagonal entries of such a matrix. Then one obtains

$$\langle \psi, P\psi \rangle \ = \ \mathrm{Tr}\,(\rho_\psi\, P) \ . \tag{24}$$

Hence

$$\sum_i p_i \langle \psi_i, P\psi_i \rangle \ = \ \mathrm{Tr}\left(\sum_i p_i \rho_{\psi_i} \cdot P \right) \ = \ \mathrm{Tr}(\rho\, P) \ , \tag{25}$$

where $\rho := \sum_i p_i \rho_{\psi_i}$.

Being a convex combination of one-dimensional projections, ρ is obviously a positive (i.e., self-adjoint positive semi-definite) $n \times n$ matrix with $\mathrm{Tr}\,(\rho) = 1$.

Conversely, from diagonalizing positive matrices it is clear that any such positive matrix ρ with $\mathrm{Tr}\,(\rho) = 1$ can be written as a convex combination of one-dimensional projections. The set of these matrices forms a closed (even compact) convex set, and its extreme points are precisely the one-dimensional projections which in turn correspond to pure states, represented also by unit vectors. Therefore it is this class of so-called *density matrices* which represents mixed states. Thus, a general mixed state is described by a density matrix ρ and the probability for an observation of P to yield the value 1 is given by $\mathrm{Tr}\,(\rho\, P)$.

Remark 2

1. A bounded closed set $\mathcal{U}$ of a real linear space $\mathcal{V}$ is convex if with two points $x_{1,2} \in \mathcal{U}$ it also contains the points $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$, $0 \leq \lambda \leq 1$ of the connecting segment. Those points $x \in \mathcal{U}$ which can be points x_λ of a segment only if $\lambda = 0$ or $\lambda = 1$ are called *extremal*: any point $x \in \mathcal{U}$ can be written as a convex combination of the extremal points of $\mathcal{U}$. Such a representation is unique if and only if $\mathcal{U}$ is a simplex; an n-dimensional simplex is a closed convex set generated by convex combinations of $n + 1$ points $\{x_i\}_{i=1}^{n+1}$ such that the n connecting lines $x_i - x_{n+1}$ are linearly independent. Given a subset $\mathcal{Y} \subseteq \mathcal{V}$, its closed convex hull $\mathcal{U}$ is the smallest convex subset of $\mathcal{V}$ such that $\mathcal{Y} \subset \mathcal{U}$.

2. The decomposition of a density matrix ρ into a convex combination of one-dimensional projections is by no means unique. This point will be further elaborated in Proposition 6. So the compact convex set of density matrices is not a simplex at all. Indeed, on $\mathbf{C}^2$ it can be identified with a full ball in $\mathbb{R}^3$, by taking in $\mathbb{R}^3$ the convex hull of the sphere that was described above.
3. In classical probability the convex set of mixed states is the simplex of all probability distributions. In our picture, if we insist on decomposing a mixed state given by $\rho = \sum_i p_i P_{\delta_i}$ into a convex combination of pure states (within the convex hull of $\{P_{\delta_i} : 1 \le i \le n\}$ which is a simplex), then it becomes unique.
4. Physically, a state ρ is completely described by all of its values $\mathrm{Tr}\,(\rho\, P)$, where P runs through the random variables of the model. Thus, if we consider only subsets of projections, then two different density matrices can represent the same physical state of the system. As a drastic example, consider the classical system $\Omega = \{\omega_1, \ldots, \omega_n\}$ with equidistribution, i.e., $p_i(\omega_i) = \frac{1}{n}$, leading to the density matrix $\rho = \sum_i \frac{1}{n} P_{\delta_i} = \frac{1}{n} \cdot \mathbb{1}$. On the other hand, with the unit vector $\psi = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}) \in \mathbf{C}^n$, we obtain for any subset $A \subseteq \Omega$:

$$\mathrm{Tr}\,(\rho\, P_A) = \frac{1}{n} \cdot |A| = \langle \psi, P_A \psi \rangle \,. \tag{26}$$

   Therefore, on the random variables $\{P_A : A \subseteq \Omega\}$, the rank one density matrix P_ψ represents the same state as the density matrix $\frac{1}{n} \cdot \mathbb{1}$. Note, however, that P_ψ is not in the convex hull of $\{P_{\delta_i} : 1 \le i \le n\}$.

3.6 The Mathematical Model of Aspect's Experiment

As an illustration, we shall now explain the photon correlation in the Orsay experiment, given by the $\cos^2$ law. Note that here we cannot simply refer to the basic $\cos^2$ law of quantum probability, since the filters are acting on two different photons.

The polarization of a pair of photons is described by a unit vector in the tensor product $\mathbf{C}^2 \otimes \mathbf{C}^2 = \mathbf{C}^4$, where we use the basis

$$\begin{aligned}(1,0,0,0) &= e_1 \otimes e_1 =: e_{11},\\ (0,1,0,0) &= e_1 \otimes e_2 =: e_{12},\\ (0,0,1,0) &= e_2 \otimes e_1 =: e_{21},\\ (0,0,0,1) &= e_2 \otimes e_2 =: e_{22},\end{aligned} \tag{27}$$

with $e_1 = (1,0) \in \mathbf{C}^2$ and $e_2 = (0,1) \in \mathbf{C}^2$. For example, in the pure state e_{12} the left-hand photon is vertically polarized and the right-hand photon horizontally. As it turns out, the state of the pair of photons as produced by the calcium atom is described by the state

$$\psi = \frac{1}{\sqrt{2}}(e_{12} - e_{21}). \tag{28}$$

Now, the filters $P(\alpha)$ on the left and $Q(\beta)$ on the right, introduced in Sect. 2.3, are represented by two-dimensional projection operators on $\mathbf{C}^4$, which are the "two-right amplification" and the "two-left-amplification" of the polarization matrix

$$\begin{pmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix}, \tag{29}$$

namely

$$\begin{aligned} P(\alpha) &= \begin{pmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\alpha & 0 & \cos\alpha\sin\alpha & 0 \\ 0 & \cos^2\alpha & 0 & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & 0 & \sin^2\alpha & 0 \\ 0 & \cos\alpha\sin\alpha & 0 & \sin^2\alpha \end{pmatrix}, \end{aligned} \tag{30}$$

$$\begin{aligned} Q(\beta) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} \cos^2\beta & \cos\beta\sin\beta \\ \cos\beta\sin\beta & \sin^2\beta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\beta & \cos\beta\sin\beta & 0 & 0 \\ \cos\beta\sin\beta & \sin^2\beta & 0 & 0 \\ 0 & 0 & \cos^2\beta & \cos\beta\sin\beta \\ 0 & 0 & \cos\beta\sin\beta & \sin^2\beta \end{pmatrix}. \end{aligned} \tag{31}$$

We note that $P(\alpha)$ and $Q(\beta)$ are commuting projections for fixed α and β. It follows that $P(\alpha)Q(\beta)$ is again a projection, as well as the products

$$P(\alpha)(\mathbb{1} - Q(\beta)), \quad (\mathbb{1} - P(\alpha))Q(\beta), \quad (\mathbb{1} - P(\alpha))(\mathbb{1} - Q(\beta)). \tag{32}$$

So we obtain the description of a classical probability space with four states, to be interpreted as

("left photon passes," "right photon passes"),
("left photon passes," "right photon is absorbed"),
("left photon is absorbed," "right photon passes"),
("left photon is absorbed," "right photon is absorbed"). (33)

The probabilities of these four events are found by the actions on $\psi = \frac{1}{\sqrt{2}}(e_{12} - e_{21}) = \frac{1}{2}(0, 1, -1, 0)$ of the four projections. In particular, the probability that both photons pass is given by

$$\langle\psi, P(\alpha)Q(\beta)\psi\rangle = \frac{1}{2}(0, 1, -1, 0)\, M \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} , \tag{34}$$

where, setting $C_a = \cos\alpha$, $C_b = \cos\beta$ and $S_a = \sin\alpha$, $S_b = \sin\beta$, the matrix 4×4 matrix M reads

$$M = \begin{pmatrix} C_a^2 C_b^2 & C_a^2 C_b S_b & C_a S_a C_b^2 & C_a S_a C_b S_b \\ C_a^2 C_b S_b & C_a^2 S_b^2 & C_a S_a C_b S_b & C_a S_a S_b^2 \\ C_a S_a C_b^2 & C_a S_a C_b S_b & S_a^2 C_b^2 & S_a^2 C_b S_b \\ C_a S_a C_b S_b & C_a S_a S_b^2 & S_a^2 C_b S_b & S_a^2 S_b^2 \end{pmatrix} . \tag{35}$$

It thus follows that

$$\begin{aligned} \langle\psi, P(\alpha)Q(\beta)\psi\rangle &= \frac{1}{2}(\cos^2\alpha\sin^2\beta + \sin^2\alpha\cos^2\beta - 2\cos\alpha\sin\alpha\cos\beta\sin\beta) \\ &= \frac{1}{2}(\cos\alpha\sin\beta - \sin\alpha\cos\beta)^2 = \frac{1}{2}\sin^2(\alpha-\beta) . \end{aligned} \tag{36}$$

4 Quantum Probability

In classical probability a model – or *probability space* – is determined by giving a set Ω of outcomes ω, by specifying what subsets $S \subset \Omega$ are to be considered as *events*, and by associating a *probability* $\mathbf{P}(S)$ with each of these events.

Requirements: The events must correspond to subsets from a σ-algebra that is a collection of sets that is closed with respect to all possible (infinite) unions and intersections of its subsets, which are called measurable; further, the probability measure $\mathbf{P}$ must be σ-additive, namely the probability of any union $S = \bigcup_j S_j$ of disjoint measurable subsets, $S_j \cap S_k = \emptyset$, must be the sum of the probabilities of the subsets, $\mathbf{P}(S) = \sum_j \mathbf{P}(S_j)$, and normalized, i.e., $\mathbf{P}(\Omega) = 1$.

In quantum probability we must loosen this scheme somewhat. We must give up the set Ω of sample points: A point $\omega \in \Omega$ in a classical model decides about the occurrence or non-occurrence of all events simultaneously, and this we abandon. Following our polarization example of Sect. 2 we take as *events* certain *closed subspaces* of a *Hilbert space* or, equivalently, a set of *projections*. To all these projections we associate probabilities.

Requirements:

1. The set of $\mathcal{E}$ of all events of a quantum model must be the set of projections in some $*$-*algebra* $\mathcal{A}$ of operators on $\mathcal{H}$.
2. The probability function $\mathbf{P} : \mathcal{E} \to [0, 1]$ must be σ-additive.

According to a theorem of Gleason [6], for $\dim(\mathcal{H}) \geq 3$ this implies that the probabilities are given by a *state* φ on $\mathcal{A}$:

$$\mathbf{P}(E) = \varphi(E) \qquad (E \in \mathcal{A} \text{ a projection})\,. \tag{37}$$

In this section we shall work out the above notions in some detail.

4.1 ∗-Algebras of Operators and States

A *Hilbert space* is a complex linear space $\mathcal{H}$ with a sesquilinear function

$$\mathcal{H} \times \mathcal{H} \to \mathbf{C}: \quad (\psi, \chi) \mapsto \langle \psi, \chi \rangle\,, \tag{38}$$

the *inner* or *scalar* product. (For the defining properties of the inner product and the main facts about Hilbert spaces see chapter "Hilbert Space Methods for Quantum Mechanics".)

Let $\mathcal{H}$ be a finite-dimensional Hilbert space. By an *operator* on $\mathcal{H}$ we mean a linear map $A : \mathcal{H} \to \mathcal{H}$. Operators can be added and multiplied in the natural way. By the *adjoint* of an operator A we mean the unique operator $A^\dagger$ on $\mathcal{H}$ satisfying

$$\forall \psi, \vartheta \in \mathcal{H} : \quad \langle A^\dagger \psi, \vartheta \rangle = \langle \psi, A\vartheta \rangle\,. \tag{39}$$

The *norm* of an operator A is defined by

$$\| A \| := \sup\left\{ \| A\psi \| \;\middle|\; \psi \in \mathcal{H}, \| \psi \| = 1 \right\}. \tag{40}$$

It has the property

$$\left\| A^\dagger A \right\| = \| A \|^2\,. \tag{41}$$

Exercise 1 *Prove this!*

By a *(unital) ∗-algebra of operators on* $\mathcal{H}$ we mean a subspace $\mathcal{A}$ of the space of all linear maps $A : \mathcal{H} \to \mathcal{H}$ such that $\mathbb{1} \in \mathcal{A}$ and

$$A, B \in \mathcal{A} \quad \Longrightarrow \quad \lambda A,\ A + B,\ A \cdot B,\ A^\dagger \in \mathcal{A}\,. \tag{42}$$

By a *state* on $\mathcal{A}$ we mean a linear functional $\varphi : \mathcal{A} \to \mathbf{C}$ satisfying

1. $\forall A \in \mathcal{A} : \quad \varphi(A^\dagger A) \geq 0$ and
2. $\varphi(\mathbb{1}) = 1$.

We shall call a pair $(\mathcal{A}, \varphi)$ of the above kind a *quantum probability space*.

Examples

1. Let P_1, P_2, ..., P_k be mutually orthogonal projections on $\mathcal{H}$ with sum $\mathbb{1}$. Then their linear span

$$\mathcal{A} := \left\{ \sum_{j=1}^{k} \lambda_j P_j | \lambda_1, \ldots, \lambda_k \in \mathbf{C} \right\} \tag{43}$$

forms a unital $*$-algebra of operators on $\mathcal{H}$. This is basically the classical model of Sect. 2.4.: $\mathcal{A}$ is isomorphic to $\mathbf{C}(\Omega)$, the algebra of all complex functions on the finite set $\Omega = \{1, \ldots, k\}$. If ψ is some vector in $\mathcal{H}$ of unit length, it determines a state φ by

$$\varphi(A) := \langle \psi, A\psi \rangle \,. \tag{44}$$

The probabilities of this classical model are $p_j := \varphi(P_j) = \parallel P_j \psi \parallel^2$. Note that there are many ψ's, and even more density matrices ρ (see Sect. 2.4) determining the same state φ on $\mathcal{A}$.

2. Let $\mathcal{A}$ be the $*$-algebra M_n of all complex $n \times n$ matrices. Let $\varphi(A) := \mathrm{Tr}\,(\rho A)$ with $\rho \geq 0$ and $\mathrm{Tr}\,(\rho) = 1$, as introduced in Sect. 2.4. The state φ is called a *pure* state if $\rho = |\psi\rangle\langle\psi|$ for some unit vector $\psi \in \mathcal{H}$.

   The qubit of Sect. 2.2 corresponds to the case $n = 2$.

   The most general way of representing M_n on a (finite-dimensional) Hilbert space is

$$\mathcal{H} = \mathbf{C}^m \otimes \mathbf{C}^n \quad (m \geq 1); \qquad \mathcal{A} = \left\{ \mathbb{1} \otimes A \,\middle|\, A \in M_n \right\}. \tag{45}$$

3. Let $k, n_1, \ldots, n_k, m_1, \ldots, m_k$ be natural numbers and let the Hilbert space $\mathcal{H}$ be given by

$$\mathcal{H} := \left(\mathbf{C}^{m_1} \otimes \mathbf{C}^{n_1}\right) \oplus \left(\mathbf{C}^{m_2} \otimes \mathbf{C}^{n_2}\right) \oplus \cdots \oplus \left(\mathbf{C}^{m_k} \otimes \mathbf{C}^{n_k}\right) . \tag{46}$$

Let $\mathcal{A}$ be the $*$-algebra given by

$$\mathcal{A} := \left\{ (\mathbb{1} \otimes A_1) \oplus \cdots \oplus (\mathbb{1} \otimes A_k) \,\middle|\, A_j \in M_{n_j} \text{ for } j = 1, \ldots, k \right\}. \tag{47}$$

Let $\psi = \psi_1 \oplus \cdots \oplus \psi_k$ be a unit vector in $\mathcal{H}$ and

$$\varphi(A) := \langle \psi, A\psi \rangle = \sum_{j=1}^{k} \langle \psi_j, A_j \psi_j \rangle \,. \tag{48}$$

If $m_j \geq n_j \forall j$ then every state on $\mathcal{A}$ is of the above form. Otherwise, density matrices may be needed.

In finite dimension Example 1 is the only commutative possibility, Example 2 is the “purely quantum mechanical” situation, and Example 3 is the most general case.

Theorem 1 *Every Abelian, that is commutative, $*$-algebra of operators on a finite-dimensional Hilbert space is isomorphic to $\mathbf{C}(\Omega)$ for some finite Ω.*

This is the finite-dimensional version of Gel’fand’s theorem [3] on commutative (C^*)-algebras.
Proof Since the operators in $\mathcal{A}$ all commute, there exists an orthonormal basis $e_1, \ldots, e_n$ in $\mathcal{H}$ on which they are all represented by diagonal matrices. Then the states $\omega_j : A \mapsto \langle e_j, Ae_j \rangle$ are multiplicative:

$$\begin{aligned} \omega_j(AB) = \langle e_j, ABe_j \rangle = \sum_{i=1}^{n} \langle e_j, Ae_i \rangle \langle e_i, Be_j \rangle = \langle e_j, Ae_j \rangle \langle e_j, Be_j \rangle \\ = \omega_j(A)\omega_j(B) \, . \end{aligned} \tag{49}$$

These states need not all be different; let $\Omega := (\omega_{j_1}, \ldots, \omega_{j_k})$ be a maximal set of different ones. Then the map

$$\iota : \mathcal{A} \to \mathbf{C}(\Omega) : \iota(A)(\omega) := \omega(A) \tag{50}$$

is an isomorphism. The projections of Example 1 are found back as the operators $P_\omega := \iota^{-1}(\delta_\omega)$. □

Exercise 2 *Check that the map ι defined above is indeed an isomorphism of $*$-algebras.*

Definition 1 By the commutant of a set $\mathcal{S}$ of operators on $\mathcal{H}$ we mean the $*$-algebra

$$\mathcal{S}' := \left\{ B : \mathcal{H} \to \mathcal{H} \text{ linear} \,\middle|\, \forall A \in \mathcal{S} : AB = BA \right\} . \tag{51}$$

The algebra generated by $\mathbb{1}$ and $\mathcal{S}$ we denote by alg $(\mathcal{S})$. The center of a $*$-algebra $\mathcal{A}$ is the (commutative) $*$-algebra $\mathcal{Z}$ given by

$$\mathcal{Z} := \mathcal{A} \cap \mathcal{A}' \, . \tag{52}$$

Exercise 3 *Find the center of $\mathcal{A}$ in each of the examples 1, 2, and 3.*

Theorem 2 (Double Commutant Theorem [3]) *Let $\mathcal{S}$ be a set of operators on a finite-dimensional Hilbert space $\mathcal{H}$, such that $X \in \mathcal{S} \Longrightarrow X^\dagger \in \mathcal{S}$. Then*

$$\text{alg}\,(\mathcal{S}) = \mathcal{S}'' \, . \tag{53}$$

Proof Clearly $\mathcal{S} \subset \mathcal{S}''$, and since $\mathcal{S}''$ is a $*$-algebra, we have $\mathrm{alg}\,(\mathcal{S}) \subset \mathcal{S}''$. We shall now prove the converse inclusion. Let $B \in \mathcal{S}''$ and let $\mathcal{A} := \mathrm{alg}\,(\mathcal{S})$. We must show that $B \in \mathcal{A}$.

Step 1: Choose $\psi \in \mathcal{H}$ and let P be the orthogonal projection onto $\mathcal{A}\psi$. Then for all $X \in \mathcal{S}$ and $A \in \mathcal{A}$

$$XPA\psi = XA\psi \in \mathcal{A}\psi \quad \Longrightarrow \quad XPA\psi = PXA\psi\,. \tag{54}$$

So XP and PX coincide on the space $\mathcal{A}\psi$. But if $\vartheta \perp \mathcal{A}\psi$, then $P\vartheta = 0$ and for all $A \in \mathcal{A}$

$$\langle X\vartheta, A\psi\rangle = \langle \vartheta, X^{\dagger}A\psi\rangle = 0\,, \tag{55}$$

so $X\vartheta \perp \mathcal{A}\psi$ as well. Hence $PX\vartheta = 0 = XP\vartheta$, and the operators XP and PX also coincide on the orthogonal complement of $\mathcal{A}\psi$. We conclude that $XP = PX$, i.e., $P \in \mathcal{S}'$. But then we also have $BP = PB$, since $B \in \mathcal{S}''$. So

$$B\psi = BP\psi = PB\psi \in \mathcal{A}\psi\,, \tag{56}$$

and $B\psi$ is of the form $A\psi$ for some $A \in \mathcal{A}$.

Step 2: But this is not sufficient: We must show that $B\psi = A\psi$ for all ψ in a basis for $\mathcal{H}$.

So choose a basis $\psi_1, \ldots, \psi_n$ of $\mathcal{H}$. We define

$$\begin{aligned} \widetilde{\mathcal{H}} &:= \mathcal{H} \oplus \mathcal{H} \oplus \cdots \oplus \mathcal{H} = \mathbf{C}^n \otimes \mathcal{H}\,,\\ \widetilde{\mathcal{A}} &:= \left\{ A \oplus A \oplus \cdots \oplus A \,\middle|\, A \in \mathcal{A} \right\} = \mathcal{A} \otimes \mathbb{1}\,,\\ \widetilde{\psi} &:= \psi_1 \oplus \psi_2 \oplus \cdots \oplus \psi_n\,. \end{aligned} \tag{57}$$

Then $(\widetilde{\mathcal{A}})' = (\mathcal{A} \otimes \mathbb{1})' = \mathcal{A}' \otimes M_n$ and $(\widetilde{\mathcal{A}})'' = (\mathcal{A}' \otimes M_n)' = \mathcal{A}'' \otimes \mathbb{1}$. So $B \otimes \mathbb{1} \in (\widetilde{\mathcal{A}})''$. By step 1 we find an element $\widetilde{A}$ of $\widetilde{\mathcal{A}}$, such that

$$\widetilde{A}\widetilde{\psi} = (B \otimes \mathbb{1})\widetilde{\psi}\,. \tag{58}$$

But $\widetilde{A} \in \widetilde{\mathcal{A}}$ must be of the form $A \otimes \mathbb{1}$ with $A \in \mathcal{A}$, so

$$A\psi_1 \oplus \cdots \oplus A\psi_n = B\psi_1 \oplus \cdots \oplus B\psi_n\,. \tag{59}$$

This implies that $A = B$, hence $B \in \mathcal{A}$. □

Exercise 4 *Find the algebra generated by* $\mathbb{1}$ *and the matrix*

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\,. \tag{60}$$

We give the following proposition without proof. It characterizes the situation of Example 2.

Proposition 4 *If the center of $\mathcal{A}$ contains only multiples of $\mathbb{1}$, then $\mathcal{H}$ and $\mathcal{A}$ must be of the form*

$$\mathcal{H} = \mathbf{C}^m \otimes \mathbf{C}^n, \quad \textit{with} \quad \mathcal{A} = \left\{ \mathbb{1} \otimes A \,\middle|\, A \in M_n \right\}. \tag{61}$$

Proposition 5 *Let $\mathcal{H}$ be a finite-dimensional Hilbert space. Then every $*$-algebra of operators on $\mathcal{H}$ can be written in the form of Example 3.*

Proof The center $\mathcal{A} \cap \mathcal{A}'$ is an Abelian (commutative) $*$-algebra, so Theorem 1 applies, giving a set of projections P_j, $j = 1, \ldots, k$. Then it is not difficult to show that the unital $*$-algebras $P_j \mathcal{A} P_j$ on the Hilbert subspaces $P_j\mathcal{H}$ satisfy the condition of Proposition 4. The statement follows. □

4.2 The Qubit

The simplest non-commutative $*$-algebra is M_2, the algebra of all 2×2 matrices with complex entries. And the simplest state on M_2 is $\frac{1}{2}\mathrm{Tr}$, the quantum analogue of a fair coin.

The events in this probability space are the orthogonal projections in M_2: the complex 2×2 matrices E satisfying

$$E^2 = E = E^\dagger . \tag{62}$$

Let us see what these projections look like. Since E is self-adjoint, it must have two real eigenvalues, and since $E^2 = E$ these must be both 0 and 1. So we have three possibilities:

- Both are 0, i.e., $E = 0$.
- One of them is 0 and the other is 1.
- Both are 1, i.e., $E = \mathbb{1}$.

In the second case, E is a one-dimensional projection satisfying

$$\mathrm{Tr}\, E = 0 + 1 = 1, \quad \det E = 0 \cdot 1 = 0 . \tag{63}$$

As $E^\dagger = E$ and $\mathrm{Tr}\, E = 1$ we may write

$$E = E(x, y, z) = \tfrac{1}{2} \begin{pmatrix} 1 + z & x - \mathrm{i}y \\ x + \mathrm{i}y & 1 - z \end{pmatrix} . \tag{64}$$

Then $\det E = 0$ implies that

$$\tfrac{1}{4}((1 - z^2) - (x^2 + y^2)) = 0 \quad \Longrightarrow \quad x^2 + y^2 + z^2 = 1 . \tag{65}$$

So the one-dimensional projections in M_2 are parametrized by the unit sphere S_2.
Notation: For $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ let us write

$$\sigma(a) := \begin{pmatrix} a_3 & a_1 - \mathrm{i}a_2 \\ a_1 + \mathrm{i}a_2 & -a_3 \end{pmatrix} = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3\,, \tag{66}$$

where σ_1, σ_2, and σ_3 are the *Pauli matrices*

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{67}$$

We note that for all $a, b \in \mathbb{R}^3$ we have

$$\sigma(a)\sigma(b) = \langle a, b\rangle \cdot \mathbb{1} + \mathrm{i}\sigma(a \times b)\,. \tag{68}$$

We may now write (64) as

$$E(a) := \tfrac{1}{2}(\mathbb{1} + \sigma(a)) \quad (\| a \| = 1)\,. \tag{69}$$

In the same way the possible states on M_2 can be calculated. We find that

$$\varphi(A) = \mathrm{Tr}\,(\rho A) \quad \text{where} \quad \rho = \rho(a) := \tfrac{1}{2}(\mathbb{1} + \sigma(a)), \quad \| a \| \le 1\,. \tag{70}$$

The probability of the event $E(a)$ in the state $\rho(b)$ is given by $\mathrm{Tr}\,(\rho(b)E(a)) = \frac{1}{2}(1 + \langle a, b\rangle)$. The events $E(a)$ and $E(b)$ are compatible if and only if $a = \pm b$. Moreover we have for all $a \in S_2$: $E(a) + E(-a) = \mathbb{1}$, $E(a)E(-a) = 0$.
Interpretation: The state of the qubit is given by a vector b in the three-dimensional unit ball. For every a on the unit sphere we can say with probability 1 that of the two events $E(a)$ and $E(-a)$ exactly one will occur, $E(a)$ having probability $\frac{1}{2}(1 + \langle a, b\rangle)$. So we have a classical coin toss (with probability for heads equal to $\frac{1}{2}(1 + \langle a, b\rangle))$ for every direction in $\mathbb{R}^3$. The coin tosses in different directions are incompatible (see Fig. 5).

The quantum coin toss is realized in nature: apart from photon polarization (see Sect. 3.2), the spin direction of a particle with total spin $\frac{1}{2}$ behaves in this way.

4.3 Photons

There is a second natural way to parametrize the one-dimensional projections in M_2, which is closer to the description of polarization of photons, as treated in Sect. 3.2.

The projection onto the one-dimensional subspace spanned by the unit vector $(\cos\alpha, \mathrm{e}^{\mathrm{i}\varphi}\sin\alpha)$ mentioned in (22) of that section is given by

$$F(\alpha, \varphi) = \begin{pmatrix} \cos^2\alpha & \mathrm{e}^{-\mathrm{i}\varphi}\cos\alpha\sin\alpha \\ \mathrm{e}^{\mathrm{i}\varphi}\cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix}. \tag{71}$$

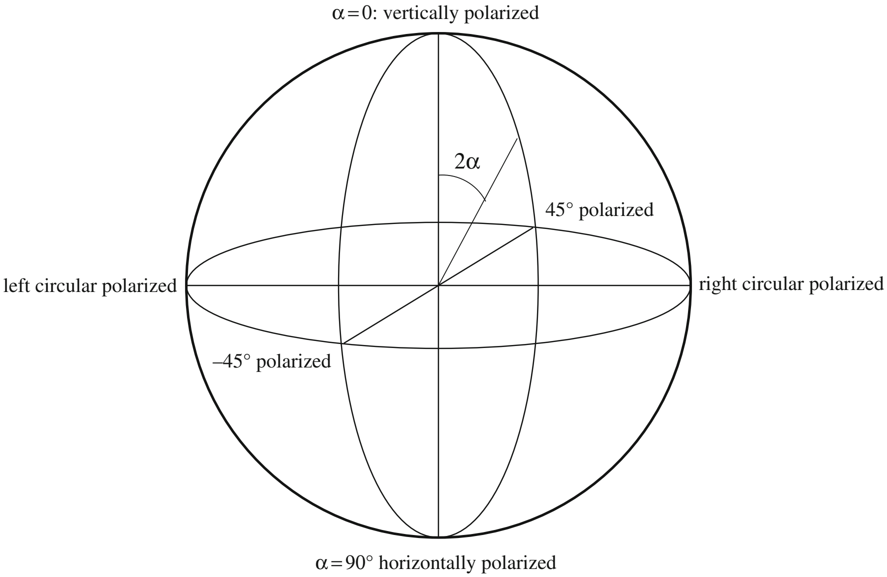


Fig. 5 Bloch sphere of the qubit

Equating this projection to $E(x, y, z)$ in (64) we obtain the relations $x = \sin 2\alpha \cos\varphi$, $y = \sin 2\alpha \sin\varphi$, and $z = \cos 2\alpha$; they define a mapping between the polarization states of a photon and the points of the unit sphere in $\mathbb{R}^3$, called the *Bloch sphere* in this context.

In particular, the projection $F(\alpha, 0)$ onto the line in $\mathbf{C}^2$ with real slope $\tan\alpha$ with $\alpha \in [-\pi/2, \pi/2)$ is given by

$$F(\alpha, 0) = \begin{pmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix} = E(\sin 2\alpha, 0, \cos 2\alpha)\,. \tag{72}$$

Finally, any atomic or molecular system, only two energy levels of which are of importance in the experiment, can be described by some (M_2, φ).

Exercise 5 *Let $f : \mathbf{C} \cup \{\infty\} \to S_2$ be given by*

$$\begin{aligned} f(0) &:= (0, 0, 1)\,; \\ f(\infty) &:= (0, 0, -1)\,; \\ f(re^{i\varphi}) &:= (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta) \\ &\qquad \textit{with} \quad \vartheta = 2\arctan r, \quad r \in (0, \infty), \varphi \in [0, \pi)\,. \end{aligned} \tag{73}$$

Show that $E(f(z))$ is the one-dimensional projection onto the line in $\mathbf{C}^2$ with slope $z \in \mathbf{C}$.

5 Operations on Probability Spaces

Our main objects of study will be *operations* on probability spaces. This means that we shall focus attention on the input–output aspect of probabilistic systems.

5.1 Operations on Classical Probability Spaces

It could be maintained that operations are already the core of *classical* probability. We start with a definition on the level of points.

Definition 2 By an operation from a finite classical probability space Ω to a finite classical probability space Ω' we mean an $\Omega \times \Omega'$ transition matrix, i.e. a matrix $(t_{\omega\omega'})$ of nonnegative numbers satisfying

$$\forall \omega \in \Omega : \quad \sum_{\omega' \in \Omega'} t_{\omega\omega'} = 1 \, . \tag{74}$$

Example 1

1. Let τ be a bijection $\Omega \to \Omega'$. We may think of rearranging a deck of cards, ($\Omega = \Omega' = \{\text{cards}\}$), or the time evolution of a mechanical system ($\Omega = \Omega' =$ phase space), or the shift on sequences of letters, or just some relabeling of the outcomes of a statistical experiment. The associated matrix is

$$t_{\omega\omega'} := \begin{cases} 1 & \text{if} \quad \omega' = \tau(\omega), \\ 0 & \text{otherwise.} \end{cases} \tag{75}$$

2. Let $X : \Omega \to \Omega'$ be surjective. We think of X as an Ω'-valued random variable, where Ω' is usually some subset of $\mathbb{R}$ or $\mathbb{R}^n$ or so. The associated operation is that of "measuring X" or "forgetting everything about ω except the value of X." The associated matrix is again

$$t_{\omega\omega'} := \begin{cases} 1 & \text{if} \quad \omega' = X(\omega), \\ 0 & \text{otherwise.} \end{cases} \tag{76}$$

3. An inverse to the operation of Example 2 is given by

$$t_{\omega'\omega} := \begin{cases} \frac{\pi(\{\omega\})}{\pi(X^{-1}(\{\omega'\}))} & \text{if} \quad \omega' = X(\omega), \\ 0 & \text{otherwise.} \end{cases} \tag{77}$$

   Here π is some probability distribution, which we assume to be everywhere nonzero. This operation describes the immersion of a system Ω' into the larger system Ω.

It can be shown that every transition matrix can be decomposed as a product of matrices of the types 3, 1, and 2. So every operation can be decomposed as an immersion, followed by a rearrangement and a restriction. Such a decomposition is called a *dilation* of the operation in question.

5.2 Quantum Operations

If $\mathcal{A}$ is a unital $*$-algebra describing a quantum system, then we denote by $\mathcal{A}^*$ the dual of $\mathcal{A}$ and by $\mathcal{A}^*_{+,1}$ the positive normalized functionals, i.e., the *states* on $\mathcal{A}$. By $M_n(\mathcal{A})$ we denote the unital $*$-algebra of all $n \times n$ matrices with entries in $\mathcal{A}$. Note that $M_n(\mathcal{A})$ is isomorphic to $M_n \otimes \mathcal{A}$.

Now suppose that we perform a physical operation which takes as input a state on the system $\mathcal{A}$ and yields as its output a state on the system $\mathcal{B}$. Which maps $f : \mathcal{A}^*_{+,1} \to \mathcal{B}^*_{+,1}$ can occur as descriptions of such an operation? We formulate three natural requirements:

1. f must be an affine map. This means that for all $\rho, \theta \in \mathcal{A}^*_{+,1}$ and all $\lambda \in [0, 1]$

$$\lambda f(\rho) + (1 - \lambda) f(\vartheta) = f\big(\lambda\rho + (1 - \lambda)\vartheta\big) \,. \tag{78}$$

   This request comes from the *stochastic equivalence principle* which states that a system which is in state ρ with probability λ and in state ϑ with probability $1 - \lambda$ cannot be distinguished from a system in the state $\lambda\rho + (1 - \lambda)\vartheta$. A map f satisfying this condition can be extended to a unique linear map $\mathcal{A}^* \to \mathcal{B}^*$, since every element of $\mathcal{A}^*$ can be written as a linear combination of (at most four) states on $\mathcal{A}$. So f must be the adjoint (or dual, see chapter "Bipartite Quantum Entanglement", Sect. 3) of some linear map $T : \mathcal{B} \to \mathcal{A}$. We shall henceforth write T^* instead of f.
2. Of course, $f = T^*$ must still map $\mathcal{A}^*_{+,1}$ to $\mathcal{B}^*_{+,1}$ for all $\rho \in \mathcal{A}^*$,

$$\begin{aligned} &\mathrm{Tr}\,(T^*(\rho)) = \mathrm{Tr}\,(\rho)\,, \\ &T^*(\rho) \geq 0 \quad \text{if} \quad \rho \geq 0\,. \end{aligned} \tag{79}$$

3. It would seem at first sight that nothing more can be said a priori about T^*. However, it was realized in the early 1980s by Karl Kraus [7] (see chapter "Hilbert Space Methods for Quantum Mechanics", Sect. 2.4) that the positivity property has to be strengthened in quantum mechanics: If the system under consideration is in a combined state with some other system, then after performing the operation T^* on the former system, the whole combination must still be in some (positive) state. Surprisingly, this is not automatic in the quantum situation, where "entanglement," as treated in Sect. 2 (see chapter "Bipartite Quantum Entanglement" for more details on this point), can occur between the two systems. See Example 2. Therefore this stronger form of positivity must be added as a requirement: For all $n \in \mathbb{N}$

$$\mathrm{id}_n \otimes T^* \quad \text{maps states on} \quad M_n \otimes \mathcal{A} \quad \text{to states on} \quad M_n \otimes \mathcal{B}\,. \tag{80}$$

Requirement (3) is called *complete positivity* of the map T^* (or T for that matter). Summarizing we arrive at the following definition, which we shall formulate in the contravariant, "Heisenberg" picture.

Definition 3 A linear map $T : \mathcal{B} \to \mathcal{A}$ is called an operation (from $\mathcal{A}$ to $\mathcal{B}$!) if the following conditions hold:

1. $T(\mathbb{1}_{\mathcal{B}}) = \mathbb{1}_{\mathcal{A}}$.
2. T is completely positive, i.e., $\mathrm{id}_n \otimes T$ is positive $M_n(\mathcal{B}) \to M_n(\mathcal{A})$ for all $n \in \mathbb{N}$.

Here $M_n(\mathcal{A})$ stands for the algebra of $n \times n$ matrices with entries in $\mathcal{A}$. This algebra is isomorphic to $M_n \otimes \mathcal{A}$.

Example 2 A map which is positive, but not completely positive:
Let $\mathcal{A} := M_2$ and let

$$T^* : \mathcal{A}^* \to \mathcal{A}^* : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & c \\ b & d \end{pmatrix} \tag{81}$$

be the transposition map. Then T^* is linear, positive, and preserves the trace. However, T^* is not completely positive since

$$\mathrm{id}_2 \otimes T^* : \frac{1}{2}\begin{pmatrix} 1&0&0&1 \\ 0&0&0&0 \\ 0&0&0&0 \\ 1&0&0&1 \end{pmatrix} \mapsto \frac{1}{2}\begin{pmatrix} 1&0&0&0 \\ 0&0&1&0 \\ 0&1&0&0 \\ 0&0&0&1 \end{pmatrix}. \tag{82}$$

The matrix on the left is a projection (on the vector $(e_0 \otimes e_0 + e_1 \otimes e_1)/\sqrt{2} \in \mathbf{C}^2 \otimes \mathbf{C}^2$; compare the entangled state of Sect. 2.5), whereas the matrix on the left has eigenvalues $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, and $-\frac{1}{2}$, hence is not a valid density matrix. However, if $\mathcal{A}$ or $\mathcal{B}$ is Abelian, that is commutative, then any positive operator $T : \mathcal{A} \to \mathcal{B}$ is automatically completely positive.

5.3 Examples of Quantum Operations

- Let $U \in M_n$ be unitary. Then the automorphism $T : M_n \to M_n : A \mapsto U^\dagger A U$ is an operation (see Lemma 1).
- The $*$-homomorphism $j : M_k \to M_l \otimes M_k : A \mapsto \mathbb{1} \otimes A$ is an operation (see Lemma 1).
- Let φ be a state on M_k. Then the map $E : M_l \otimes M_k \to M_k : B \otimes A \mapsto \varphi(B)A$ is an operation.

The above examples are to be compared with those in Sect. 4.1. We shall prove their validity in two lemmas.

Lemma 1 *If $\mathcal{A} \subset M_k$ and $T : \mathcal{A} \to \mathcal{B} \subset M_l$ is a $*$-homomorphism, i.e., if for all A, $B \in \mathcal{A}$ we have $T(AB) = T(A)T(B)$ and $T(A^\dagger) = T(A)^\dagger$, then T is completely positive.*

Proof We must show that for all $n \in \mathbb{N}$ the map

$$\mathrm{id}_n \otimes T : \left(A_{ij}\right)_{i,j=1}^{n} \mapsto \left(T(A_{ij})\right)_{i,j=1}^{n} \tag{83}$$

is positive. Indeed, for all $\psi = (\psi_1, \ldots, \psi_n) \in (\mathbf{C}^l)^n$, putting $A = X^\dagger X$ with $X \in M_n(\mathcal{A})$

$$\begin{aligned}\langle \psi, (\mathrm{id}_n \otimes T)(X^\dagger X)\psi\rangle &= \sum_{i,i'=1}^{l} \langle \psi_i, T\big((X^\dagger X)_{ii'}\big)\psi_{i'}\rangle \\ &= \sum_{i,i'=1}^{l} \sum_{j=1}^{n} \langle \psi_i, T\big(X_{ji}^\dagger X_{ji'}\big)\psi_{i'}\rangle \\ &= \sum_{i,i'=1}^{l} \sum_{j=1}^{n} \langle \psi_i, T(X_{ji})^\dagger\, T(X_{ji'})\psi_{i'}\rangle \\ &= \sum_{j=1}^{n} \left\| \sum_{i=1}^{l} T(X_{ji})\psi_i \right\|^2 \geq 0\,.\end{aligned} \tag{84}$$

□

Lemma 2 *Let $\mathcal{A} \subset M_k$, $\mathcal{B} \subset M_l$ and let V be a linear map $\mathbf{C}^l \to \mathbf{C}^k$. Then*

$$T : \mathcal{A} \to \mathcal{B} : A \mapsto V^\dagger A\, V \tag{85}$$

is completely positive.

Proof If $(A_{ij})_{i,j=1}^{n} \in M_n(\mathcal{A})$ is positive, then for all $(\psi_1, \ldots, \psi_n) \in (\mathbf{C}^l)^n = \mathbf{C}^n \otimes \mathbf{C}^l$ we have

$$\begin{aligned}\langle \psi, (\mathrm{id}_n \otimes T)(A)\psi\rangle &= \sum_{i,j=1}^{n} \langle \psi_i, T(A_{ij})\psi_j\rangle = \sum_{i,j=1}^{n} \langle \psi_i, V^\dagger A_{ij}\, V\psi_j\rangle \\ &= \sum_{i,j=1}^{n} \langle V\psi_i, A_{ij} V\psi_j\rangle \geq 0\,.\end{aligned} \tag{86}$$

□

Lemma 2 covers the third case in Example 5.3 since φ can be decomposed into pure states as $\varphi = \sum_i \langle \psi, \cdot\psi\rangle$ and

$$\varphi(B)A = \sum_{i=1}^{l} \lambda_i \langle \psi_i, B\psi_i\rangle A = \sum_{i=1}^{l} \lambda_i V_i^\dagger\, (B \otimes A)\, V_i\,, \tag{87}$$

where $V_i : \mathbf{C}^k \to \mathbf{C}^l \otimes \mathbf{C}^k : \vartheta \mapsto \psi_i \otimes \vartheta$.

5.4 Unraveling Quantum Operations

The following important theorem, together with Proposition 5, characterizes all completely positive maps on finite-dimensional matrix algebras (the version of this result given by Kraus has been discussed in chapter "Quantum Probability and Quantum Information Theory", Sect. 2.4).

Theorem 3 (Stinespring 1955) *Let T be a linear map $M_k \to M_l$. Then T is completely positive if and only if there exist $m \in \mathbb{N}$ and operators $V_1, \ldots, V_m : \mathbf{C}^l \to \mathbf{C}^k$ such that for all $A \in M_k$*

$$T(A) = \sum_{i=1}^{m} V_i^{\dagger} \, A \, V_i \, . \tag{88}$$

We shall give a proof based on a physical argument (cf. [10]). The system is put in an entangled state with a second system, which for convenience we describe by the *opposite algebra* (see below). Then we act on the main system with our operation T, and by complete positivity we get a new state on the pair. Surprisingly, this state fully characterizes the operation T. By decomposing the state into vector states we shall obtain the unraveling we wanted.

Let us first introduce some notation. If $\mathcal{H}$ is a (finite-dimensional) Hilbert space, let $\mathcal{H}'$ denote its *dual*, the space of all linear functionals $\mathcal{H} \to \mathbf{C}$. The elements of $\mathcal{H}'$ are of the form $\overline{\vartheta} : \chi \mapsto \langle \vartheta, \chi \rangle$; in Dirac notation $\overline{\vartheta}$ is denoted as $\langle \vartheta |$. This dual $\mathcal{H}'$ is actually isomorphic to $\mathcal{H}$ itself, but it is convenient to maintain the distinction, as we shall see below. In particular, if $\mathcal{H} = \mathbf{C}^n$, then there is a natural action on $\mathcal{H}'$ of the algebra M_n^t, the *opposite algebra* of M_n, which has the multiplication reflected: $A^t B^t = (BA)^t$. The operator A^t acts on $\overline{\chi}$ as $A^t \overline{\chi} := \overline{\chi} \circ A$.

Now consider the tensor product $\mathcal{H}_{kl} := \mathbf{C}^k \otimes (\mathbf{C}^l)'$ of the Hilbert space $\mathbf{C}^k$ and the dual of $\mathbf{C}^l$. By identifying the vector $\psi \otimes \overline{\vartheta} \in \mathcal{H}_{kl}$ with the operator $|\psi\rangle\langle\vartheta|$: $\chi \mapsto \langle \vartheta, \chi \rangle \cdot \psi$, the Hilbert space $\mathcal{H}_{kl}$ can alternatively be viewed as the space of all operators $\mathbf{C}^l \to \mathbf{C}^k$. On this Hilbert space the algebra $M_k \otimes M_l^t$ acts naturally as follows:

$$A \otimes B^t : \psi \otimes \overline{\vartheta} \mapsto A\psi \otimes B^t \overline{\vartheta} \quad \Big[\approx A|\psi\rangle\langle\vartheta|B \Big] . \tag{89}$$

The space $\mathcal{H}_{ll}$ has a rotation-invariant vector (the so-called fully entangled state on $M_l \otimes M_l^t$), given by

$$\Omega := \frac{1}{\sqrt{l}} \sum_{i=1}^{l} e_i \otimes \overline{e_i} \quad \left[\approx \frac{1}{\sqrt{l}} \sum_{i=1}^{l} |e_i\rangle\langle e_i| = \mathbb{1}_l / \sqrt{l} \right] , \tag{90}$$

for *any* orthonormal basis $e_1, \ldots, e_l$ of $\mathbf{C}^l$. This vector has the property that

$$\begin{aligned}\langle \Omega, (A \otimes B^t)\Omega \rangle &= \frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{l}\langle e_i \otimes \overline{e_i}, (A \otimes B^t)e_j \otimes \overline{e_j}\rangle \\ &= \frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{l}\langle e_i, Ae_j\rangle\langle \overline{e_i}, B^t\overline{e_j}\rangle \\ &= \frac{1}{l}\sum_{i=1}^{l}\sum_{j=1}^{l}\langle e_i, Ae_j\rangle\langle e_j, Be_i\rangle = \frac{1}{l}\mathrm{Tr}\,(AB)\,. \end{aligned} \tag{91}$$

Proof of Stinespring's Theorem The "if" part follows immediately from Lemma 2. For the "only if" part, assume that $T : M_k \to M_l$ is completely positive. Let $\mathcal{H}_{ll} := \mathbf{C}^l \otimes (\mathbf{C}^l)'$ as above and let ω denote the state

$$\omega(X) := \langle \Omega, X\Omega \rangle \tag{92}$$

on $\mathcal{B}(\mathcal{H}_{ll}) \approx M_l \otimes M_l^t$.

Since T is completely positive, the functional ω_T on $\mathcal{B}(\mathcal{H}_{kl}) \approx M_k \otimes M_l^t$, given by

$$\omega_T(A \otimes B^t) := \omega(T(A) \otimes B^t) \tag{93}$$

is also a state. Decompose ω_T into pure states given by vectors $v_1, v_2, \ldots, v_m \in \mathcal{H}_{kl}$:

$$\omega_T(X) = \sum_{i=1}^{m}\langle v_i, Xv_i\rangle\,. \tag{94}$$

Now, as noted above, $v_i \in \mathcal{H}_{kl}$ can be considered as an operator $V_i : \mathbf{C}^l \to \mathbf{C}^k$. We shall show that these operators satisfy requirement (88) of the theorem. Indeed, for all $\psi, \vartheta \in \mathbf{C}^l$

$$\begin{aligned}\sum_{i=1}^{m}\langle \psi, V_i^\dagger\, A\, V_i\vartheta\rangle &= \sum_{i=1}^{m}\langle V_i\psi, AV_i\vartheta\rangle = \sum_{i=1}^{m}\langle v_i, \big(A \otimes (|\overline{\psi}\rangle\langle\overline{\vartheta}|)\big)v_i\rangle_{\mathcal{H}_{kl}} \\ &= \omega_T\big(A \otimes (|\overline{\psi}\rangle\langle\overline{\vartheta}|)\big) = \omega\big(T(A) \otimes (|\overline{\psi}\rangle\langle\overline{\vartheta}|)\big) \\ &= \mathrm{Tr}\,\big(T(A)(|\vartheta\rangle\langle\psi|)\big) = \langle \psi, T(A)\vartheta\rangle\,.\end{aligned} \tag{95}$$

□

The second step is verified by substituting $V_i = \sum_j |\alpha_j^i\rangle\langle\beta_j^i|$ with $\alpha_j^i \in \mathbf{C}^k$, $\beta_j^i \in \mathbf{C}^l$ and realizing that $v_i = \sum_j \alpha_j^i \otimes \overline{\beta_j^i}$.

5.5 Uniqueness of Unravelings

Unraveling (88) is not unique.[1] If the matrices $V_1, \ldots, V_m$ are linearly independent, then they are determined by the completely positive map T up to a transformation of the form

$$V_i' := \sum_{j=1}^{m} u_{ij} V_j \,, \tag{96}$$

where u is a unitary $m \times m$ matrix of complex numbers. In this independent case the number m of terms in the unraveling takes its minimal value, which we shall call the *rank* of the operation T.

In general, any number m of terms, also larger than the rank, can occur in the unraveling of T. But in that case the operators V_i are not linearly independent. In fact, the space $\mathcal{D}$ of *dependencies*, given by

$$\mathcal{D} := \left\{ \lambda \in \mathbf{C}^m \,|\, \sum_{i=1}^{m} \lambda_i^* V_i = 0 \right\} \,, \tag{97}$$

has dimension $m - \mathrm{rank}(T)$ and the matrix u of (96) is a partial isometry with initial space $\mathcal{D}^\perp$ and final space $(\mathcal{D}')^\perp$, where $\mathcal{D}'$ denotes the space of dependencies of the V_i'.

We shall now prove these statements in the context of the decomposition of states. From the proof of Theorem 3 it is clear that they carry over to operations.

Proposition 6 *Let φ be a state on $\mathcal{A} := M_k$ and let two decompositions of φ into pure states be given*

$$\varphi(A) = \sum_{i=1}^{m} \langle \psi_i, A\psi_i \rangle = \sum_{j=1}^{n} \langle \vartheta_j, A\vartheta_j \rangle \,. \tag{98}$$

Let $\mathcal{D} \subset \mathbf{C}^m$ and $\mathcal{D}' \subset \mathbf{C}^n$ denote the dependency spaces of $\psi = (\psi_1, \ldots, \psi_m)$ and $\vartheta = (\vartheta_1, \ldots, \vartheta_n)$, respectively. Then ψ and ϑ are connected by a transformation of the form

$$\vartheta_j = \sum_{i=1}^{m} u_{ji} \psi_i \,, \tag{99}$$

where the $n \times m$ matrix u describes a partial isometry $\mathbf{C}^m \to \mathbf{C}^n$ with initial space $\mathcal{D}^\perp$ and final space $(\mathcal{D}')^\perp$. In particular, if the m-tuple $(\psi_1, \ldots, \psi_m)$ and

[1] This section elaborates on a remark by Mark Fannes and can be skipped in a first reading.

the n-tuple $(\vartheta_1, \ldots, \vartheta_n)$ *are both sequences of independent vectors, then* $n = m$ *and* u *is unitary.*

Proof Consider ψ and ϑ as vectors in $\mathcal{H} := (\mathbf{C}^k)^m = \mathbf{C}^m \otimes \mathbf{C}^k$ and $\mathcal{H}' := (\mathbf{C}^k)^n = \mathbf{C}^n \otimes \mathbf{C}^k$, respectively. Then (98) can be written in the form

$$\varphi(A) = \langle \psi, (\mathbb{1}_m \otimes A)\psi \rangle = \langle \vartheta, (\mathbb{1}_n \otimes A)\vartheta \rangle \,. \tag{100}$$

Let $\Lambda \subset \mathcal{H}$ and $\Lambda' \subset \mathcal{H}'$ be the subspaces consisting of the vectors $(\mathbb{1}_m \otimes A)\psi$ and $(\mathbb{1}_n \otimes A)\vartheta$, respectively, where A runs through the matrix algebra $\mathcal{A} = M_k$. Let $U : \Lambda \to \Lambda'$ be given by

$$U(\mathbb{1}_m \otimes A)\psi := (\mathbb{1}_n \otimes A)\vartheta \,. \tag{101}$$

Then U is well defined, isometric, and onto since

$$\begin{aligned} \| (\mathbb{1}_n \otimes A)\vartheta \|^2 &= \langle (\mathbb{1}_n \otimes A)\vartheta, (\mathbb{1}_n \otimes A)\vartheta \rangle = \langle \vartheta, (\mathbb{1}_n \otimes A^\dagger A)\vartheta \rangle \\ &= \varphi(A^\dagger A) = \| (\mathbb{1}_m \otimes A)\psi \|^2 \,. \end{aligned} \tag{102}$$

We extend U to a map $\mathcal{H} \to \mathcal{H}'$ by putting $U\chi = 0$ for all $\chi \in \mathcal{H}$ which are orthogonal to Λ.

Next, let us show that U is actually of the form $u \otimes \mathbb{1}_k$ for some partial isometry $u : \mathbf{C}^m \to \mathbf{C}^n$. This is equivalent to the statement that for all $A \in M_k$:

$$U(\mathbb{1}_m \otimes A) = (\mathbb{1}_n \otimes A)U \,, \tag{103}$$

which is true since $(\mathbb{1}_m \otimes A)$ leaves $\Lambda^\perp$ invariant, so that both sides vanish on $\Lambda^\perp$. And for $\chi \in \Lambda$, i.e., for $\chi = (\mathbb{1}_m \otimes X)\psi$ with $X \in M_k$, we have

$$\begin{aligned} U(\mathbb{1}_m \otimes A)\chi &= U(\mathbb{1}_m \otimes A)(\mathbb{1}_m \otimes X)\psi = U(\mathbb{1}_m \otimes AX)\psi = (\mathbb{1}_n \otimes AX)\vartheta \\ &= (\mathbb{1}_n \otimes A)(\mathbb{1}_n \otimes X)\vartheta = (\mathbb{1}_n \otimes A)U(\mathbb{1}_m \otimes X)\psi = (\mathbb{1}_n \otimes A)U\chi \,. \end{aligned} \tag{104}$$

It remains to be shown that

$$\Lambda^\perp = \mathcal{D} \otimes \mathbf{C}^k \tag{105}$$

(and analogously $(\Lambda')^\perp = \mathcal{D}' \otimes \mathbf{C}^k$). Clearly, for all $\lambda \in \mathbf{C}^m$ and $\mu \in \mathbf{C}^k$,

$$\langle \lambda \otimes \mu, (\mathbb{1} \otimes A)\psi \rangle = \sum_{i=1}^m \lambda_i^* \langle \mu, A\psi_i \rangle = \left\langle A^\dagger \mu, \left(\sum_{i=1}^m \lambda_i^* \psi_i \right) \right\rangle . \tag{106}$$

It follows that for $\lambda \in \mathcal{D}$ the vector $\lambda \otimes \mu$ is orthogonal to Λ, so we have $\mathcal{D} \otimes \mathbf{C}^k \subset \Lambda^\perp$. To prove the converse inclusion, we first note that the orthogonal projection onto Λ is $U^\dagger U = u^\dagger u \otimes \mathbb{1}_k$, hence $\Lambda = \mathcal{E} \otimes \mathbf{C}^k$ for some subspace $\mathcal{E}$ of $\mathbf{C}^m$. We must show that $\mathcal{E}^\perp \subset \mathcal{D}$. So suppose that $\lambda \perp \mathcal{E}$, so that $\lambda \otimes \mu \perp \Lambda$ for all $\mu \in \mathbf{C}^k$. Putting $A = \mathbb{1}$ in (106) we find that the left-hand side, and hence the right-hand side, is 0 for all μ, so $\sum_{i=1}^m \lambda_i^* \psi_i = 0$ and $\lambda \in \mathcal{D}$. □

5.6 Properties of Quantum Operations

When A and B are operators on a Hilbert space, we mean by $A \geq B$ that the difference $A - B$ is a positive operator. The following is an extremely useful inequality for operations.

Proposition 7 (Cauchy–Schwartz for operations) *Let $\mathcal{A}$ and $\mathcal{B}$ be *-algebras of operators on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ and let $T : \mathcal{A} \to \mathcal{B}$ be an operation. Then we have for all $A \in \mathcal{A}$*

$$T(A^\dagger A) \geq T(A)^\dagger T(A)\,. \tag{107}$$

Proof The operator $X \in M_2 \otimes \mathcal{A}$ given by

$$X := \begin{pmatrix} A^\dagger A & -A^\dagger \\ -A & \mathbb{1} \end{pmatrix} = \begin{pmatrix} A & -\mathbb{1} \\ 0 & 0 \end{pmatrix}^\dagger \begin{pmatrix} A & -\mathbb{1} \\ 0 & 0 \end{pmatrix} \tag{108}$$

is positive. Since T is completely positive and $T(\mathbb{1}) = \mathbb{1}$, it follows that also

$$(\mathrm{id} \otimes T)(X) = \begin{pmatrix} T(A^\dagger A) & -T(A)^\dagger \\ -T(A) & \mathbb{1} \end{pmatrix} \tag{109}$$

is a positive operator. Putting $\xi := \psi \oplus T(A)\psi$ we find that

$$\langle \xi, (\mathrm{id} \otimes T)X\xi \rangle = \langle \psi, \big(T(A^\dagger A) - T(A)^\dagger T(A)\big)\psi \rangle \tag{110}$$

is positive for all $\psi \in \mathcal{H}$. □

Theorem 4 (Multiplication theorem) *If $T : \mathcal{A} \to \mathcal{B}$ is an operation and $T(A^\dagger A) = T(A)^\dagger T(A)$ for some $A \in \mathcal{A}$, then $T(A^\dagger B) = T(A)^\dagger T(B)$ and $T(B^\dagger A) = T(B)^\dagger T(A)$ for all $B \in \mathcal{A}$.*

Proof Take any $B \in \mathcal{A}$ and $\lambda \in \mathbb{R}$. Then

$$T\big((A^\dagger + \lambda B^\dagger)(A + \lambda B)\big) = T(A)^\dagger T(A) + \lambda T(A^\dagger B + B^\dagger A) + \lambda^2 T(B^\dagger B)\,, \tag{111}$$

while by Cauchy–Schwartz

$$\begin{aligned} &T\big((A^\dagger + \lambda B^\dagger)(A + \lambda B)\big) \\ &\quad \geq T(A)^\dagger T(A) + \lambda(T(A)^\dagger T(B) + T(B)^\dagger T(A)) + \lambda^2 T(B)^\dagger T(B))\,. \end{aligned} \tag{112}$$

This inequality holds for all $\lambda \in \mathbb{R}$ which implies

$$T(A^\dagger B + B^\dagger A) \geq T(A)^\dagger T(B) + T(B)^\dagger T(A)\,. \tag{113}$$

Replacing A by iA and B by $-iB$ shows that the opposite inequality also holds, so we have equality. Finally replacing only B by iB shows that $T(A^\dagger B) = T(A)^\dagger T(B)$ and $T(B^\dagger A) = T(B)^\dagger T(A)$. □

In particular, if a Cauchy–Schwartz *equality* holds for an operation T then T is a *-homomorphism.

Theorem 5 (Embedding theorem) *Let $(\mathcal{A}, \varphi)$ and $(\mathcal{B}, \psi)$ be non-degenerate quantum probability spaces and let $j : \mathcal{A} \to \mathcal{B}$, $E : \mathcal{B} \to \mathcal{A}$ be operations which preserve the states. If*

$$E \circ j = \mathrm{id}_{\mathcal{A}}\,, \tag{114}$$

*then j is an injective *-homomorphism and $P := j \circ E$ is a conditional expectation, i.e.,*

$$P(C_1\, B\, C_2) \;=\; C_1\, P(B)\, C_2 \tag{115}$$

for all $C_1, C_2 \in j(\mathcal{A})$ and all $B \in \mathcal{B}$.

Following the language used in Sect. 4.1 we shall call j a *random variable* and P the *conditional expectation with respect to ψ, given j*. Compare the following proof with that of Theorem 4.

Proof For any $A \in \mathcal{A}$ we have by Cauchy–Schwartz

$$A^\dagger A = E{\circ}j\big(A^\dagger A\big) \geq E\big(j(A))^\dagger\, j(A)\big) \geq \big(E{\circ}j(A)\big)^\dagger \big(E{\circ}j(A)\big) = A^\dagger A\,, \tag{116}$$

so we have equalities here. In particular

$$\psi\big(j(A^\dagger A) - j(A)^\dagger\, j(A)\big) = \varphi \circ E\big(j(A^\dagger A) - j(A)^\dagger\, j(A)\big) = 0\,. \tag{117}$$

As $(\mathcal{B}, \psi)$ is non-degenerate, $j(A^\dagger A) = j(A)^\dagger j(A)$, i.e., j is a *-homomorphism. j is injective since it has the left-inverse E.
But also from (116) we have $E\big(j(A)^\dagger\, j(A)\big) = E \circ j(A)^\dagger\, E \circ j(A)$. The Multiplication Theorem 4 then implies that for all $B \in \mathcal{B}$ and $A_1 \in \mathcal{A}$,

$$E(j(A_1)^\dagger B) = E \circ j(A_1)^\dagger\, E(B) = A_1^\dagger\, E(B)\,, \tag{118}$$

and similarly, with $A_2 \in \mathcal{A}$

$$E\big(j(A_1)^\dagger\, Bj(A_2)\big) = E\big(j(A_1)^\dagger\, B\big)E \circ j(A_2) = A_1^\dagger\, E(B)A_2\,. \tag{119}$$

Applying j to both sides we find (115). □

6 Quantum Impossibilities

The result of any physical operation applied on a probabilistic system (quantum or not) is described by a completely positive identity preserving map from the state space of that system to the state space of the resulting system. This imposes strong restrictions on what can be done. Some of these are well-known quantum principles, such as the Heisenberg principle ("no measurement without disturbance"), some are surprising and relatively recent discoveries ("no cloning"), but all of them obtain quite neat formulations in the language of quantum probability.

6.1 No-Cloning

In its original formulation [5, 11] the "No-Cloning Theorem" dealt with the reproduction of nonorthogonal vector states. Here we give an algebraic version, which distinguishes clearly between the classical and the quantum cases.

"Cloning", or – more mundanely – *copying* a stochastic object is an operation which takes as input an object in some state ρ and yields as its output a pair of objects with identical state spaces, such that, if we throw away one of them, we are left with a single object in the state ρ (cf. Fig. 6, which is actually not complete: the same equality should hold with the other output line blocked).

In a formula, for all $\rho \in \mathcal{A}^*_{+,1}$

$$(\mathrm{Tr} \otimes \mathrm{id}) \circ C^*(\rho) = (\mathrm{id} \otimes \mathrm{Tr}) \circ C^*(\rho) = \rho\,. \tag{120}$$

Reformulated in the Heisenberg picture: We call an operation $C : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ a *copying operation* or *copier* if for all $A \in \mathcal{A}$:

$$C(\mathbb{1} \otimes A) = C(A \otimes \mathbb{1}) = A\,. \tag{121}$$

As is well known, copying presents no problem in classical physics or classical probability. Here is an example of a classical copying operation. For simplicity, let us think of the operation of copying n bits. Let Ω denote the space $\{0, 1\}^n$ of all strings of n bits and let $\exp^{(1)}$ be the "copying" map $\Omega \to \Omega \times \Omega : \omega \mapsto (\omega, \omega)$. This map induces an operation

$$C : \quad \mathbf{C}(\Omega) \times \mathbf{C}(\Omega) \to \mathbf{C}(\Omega) : \quad Cf(\omega) := f \circ \exp^{(1)}(\omega) = f(\omega, \omega)\,. \tag{122}$$

Clearly, for all $f \in \mathbf{C}(\Omega)$:

$$C(\mathbb{1} \otimes f)(\omega) = (\mathbb{1} \otimes f)(\omega, \omega) = f(\omega)\,, \tag{123}$$

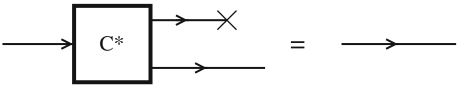


Fig. 6 Definition of a copier

and the same holds for $C(f \otimes \mathbb{1})$, so (121) is satisfied. In the Schrödinger picture our operation looks as follows: for any probability distribution π on Ω,

$$(C^*\pi)(\nu, \omega) = \delta_{\nu\omega}\pi(\omega)\,, \tag{124}$$

and we see that (120) is satisfied:

$$(\mathrm{Tr} \otimes \mathrm{id}\,) \circ C^*(\pi)(\omega) = \sum_{\nu \in \Omega} \delta_{\nu\omega}\pi(\omega) = \pi(\omega)\,. \tag{125}$$

The following theorem says that this construction is only possible in the Abelian (i.e., commutative) case.

Theorem 6 ("No-cloning") *Let $\mathcal{A}$ be a $*$-algebra of operators on a (finite-dimensional) Hilbert space. Then $\mathcal{A}$ admits a copying operation if and only if $\mathcal{A}$ is Abelian.*

Proof If $\mathcal{A}$ is Abelian, by Gel'fands Theorem 1, $\mathcal{A}$ is isomorphic to $\mathbf{C}(\Omega)$ for some finite set Ω, and the above construction of a copier applies. Conversely, suppose that $C : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ is a copying operation. Then (121) implies that for all $A \in \mathcal{A}$

$$C\big((\mathbb{1} \otimes A)^\dagger(\mathbb{1} \otimes A)\big) = C(\mathbb{1} \otimes A^\dagger A) = A^\dagger A = C(\mathbb{1} \otimes A)^\dagger C(\mathbb{1} \otimes A). \tag{126}$$

Then it follows from the Multiplication Theorem 4 that for all $A, B \in \mathcal{A}$

$$\begin{aligned} AB &= C(A \otimes \mathbb{1})C(\mathbb{1} \otimes B) = C\big((A \otimes \mathbb{1})(\mathbb{1} \otimes B)\big) \\ &= C\big((\mathbb{1} \otimes B)(A \otimes \mathbb{1})\big) = C(\mathbb{1} \otimes B)C(A \otimes \mathbb{1}) = BA\,. \end{aligned} \tag{127}$$

□

6.2 No Classical Coding

Closely related to the above is the rule that "quantum information cannot be classically coded": It is not possible to operate on a quantum system, extracting some information from it, and then from this information reconstruct the quantum system in its original state:

$$\rho \in \mathcal{A}^* \overset{C^*}{\longmapsto} \pi \in \mathcal{B}^* \overset{D^*}{\longmapsto} \rho \in \mathcal{A}^*\,. \tag{128}$$

We formulate this theorem in the contravariant ("Heisenberg") picture:

Theorem 7 *Let $\mathcal{A}$ and $\mathcal{B}$ be *-algebras and let $C : \mathcal{B} \to \mathcal{A}$ and $D : \mathcal{A} \to \mathcal{B}$ be operations, ("coding" and "decoding"), such that $C \circ D = \mathrm{id}_{\mathcal{A}}$. Then if $\mathcal{B}$ is Abelian, so is $\mathcal{A}$.*

Proof We have for all $A \in \mathcal{A}$

$$A^\dagger A = C \circ D(A^\dagger A) \geq C\big(D(A)^\dagger D(A)\big) \geq A^\dagger A \tag{129}$$

and

$$AA^\dagger = C \circ D(AA^\dagger) \geq C\big(D(A)D(A)^\dagger\big) \geq AA^\dagger\,, \tag{130}$$

whence equality holds everywhere. If $\mathcal{B}$ is Abelian, then $D(A)^\dagger D(A) = D(A)D(A)^\dagger$ and $A^\dagger A = AA^\dagger$. □

Exercise 6 *Prove that, if $A^\dagger A = AA^\dagger$ for all $A \in \mathcal{A}$, then $\mathcal{A}$ is Abelian.*

6.3 The Heisenberg Principle

The *Heisenberg principle* states – roughly speaking – that no information on a quantum system can be obtained without changing its state. In this form, the statement is not so interesting: If we realize that the *state* of the system expresses the expectations of its observables, given the information we have on it, it is no wonder that this state changes once we gain information!

A more precise formulation is the following:

> *If we extract information from a system whose algebra $\mathcal{A}$ is a factor (i.e., $\mathcal{A} \cap \mathcal{A}' = \mathbf{C}\mathbb{1}$), and if we throw away (disregard) this information, then still it cannot be avoided that some initial states are altered.*

Let us work toward a mathematical formulation: A *measurement* is an operation performed on a physical system which results in the extraction of information from that system, while possibly changing its state. So a measurement is an operation

$$M^* : \mathcal{A}^* \to \mathcal{A}^* \otimes \mathcal{B}^*\,, \tag{131}$$

where $\mathcal{A}$ describes the physical system and $\mathcal{B}$ the output part of a measurement apparatus which we couple to it. $\mathcal{A}^*$ consists of states and $\mathcal{B}^*$ of probability distributions on the outcomes. So $\mathcal{B}$ will be commutative, but we do not need this property here. Now suppose that no initial state is altered by the measurement:

$$(\mathrm{id} \otimes \mathrm{Tr})M^*(\rho) = \rho \qquad \forall \rho \in \mathcal{A}^*\,. \tag{132}$$

Suppose also that $\mathcal{A}$ is a factor. We claim that no information can be obtained on ρ:

$$(\mathrm{Tr} \otimes \mathrm{id})M^*(\rho) = \vartheta\,, \tag{133}$$

where ϑ does not depend on ρ. Figure 7 symbolically expresses this fact.
We again formulate and prove the theorem in the contravariant picture:

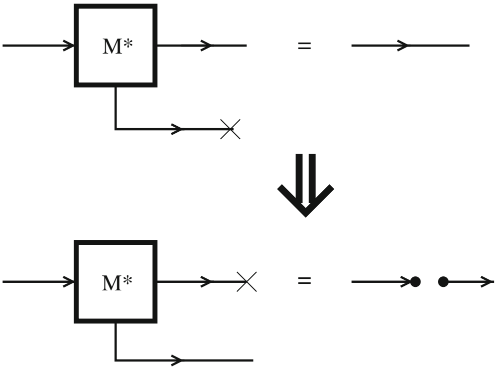


Fig. 7 The Heisenberg principle

Theorem 8 (Heisenberg's principle) *Let M be an operation $\mathcal{A} \otimes \mathcal{B} \to \mathcal{A}$ such that for all $A \in \mathcal{A}$,*

$$M(A \otimes \mathbb{1}) = A\,, \tag{134}$$

then

$$M(\mathbb{1} \otimes B) \in \mathcal{A} \cap \mathcal{A}'\,. \tag{135}$$

In particular, if $\mathcal{A}$ is a factor, then for some fixed state ϑ on $\mathcal{B}$

$$M(\mathbb{1} \otimes B) = \vartheta(B) \cdot \mathbb{1}_{\mathcal{A}}. \tag{136}$$

We note that (136) implies (133), since for all ρ on $\mathcal{A}$ and all $B \in \mathcal{B}$

$$\big((\mathrm{Tr} \otimes \mathrm{id}\,)M^*\rho\big)(B) = \rho\big(M(\mathbb{1} \otimes B)\big) = \rho\big(\vartheta(B)\mathbb{1}_{\mathcal{A}}\big) = \vartheta(B)\,. \tag{137}$$

Proof As in the proof of the "no cloning" theorem we have by the multiplication theorem for all $A \in \mathcal{A}$, $B \in \mathcal{B}$

$$M(\mathbb{1} \otimes B) \cdot A = M(\mathbb{1} \otimes B)M(A \otimes \mathbb{1}) = M(A \otimes B)\,. \tag{138}$$

But also,

$$A \cdot M(\mathbb{1} \otimes B) = M(A \otimes \mathbb{1})M(\mathbb{1} \otimes B) = M(A \otimes B)\,. \tag{139}$$

So $M(\mathbb{1} \otimes B)$ lies in the center of $\mathcal{A}$. If $\mathcal{A}$ is a factor, then $B \mapsto M(\mathbb{1} \otimes B)$ is an operation from B to $\mathbf{C} \cdot \mathbb{1}_{\mathcal{A}}$, i.e., a state on B times $\mathbb{1}_{\mathcal{A}}$. □

6.4 Random Variables and von Neumann Measurements

Following the suggestion made in Sect. 4.2 (in particular case 2), we define a *random variable* to be a *-homomorphism from one algebra $\mathcal{B}$ to a (larger) algebra $\mathcal{A}$:

$$\mathcal{A} \xleftarrow{j} \mathcal{B} . \tag{140}$$

In the covariant ("Schrödinger") picture this describes the operation j^* of *restriction to* the subsystem $\mathcal{B}$:

$$\mathcal{A}^* \xrightarrow{j^*} \mathcal{B}^* . \tag{141}$$

An important case is when $\mathcal{B} = \mathbf{C}(\Omega)$ for some finite set Ω; then j is to be viewed as an Ω*-valued random variable*. Let $\Omega = \{x_1, \ldots, x_n\}$. Then $j(1_{\{x_i\}})$ is a projection, P_i say, in $\mathcal{A}$, with the properties that

$$\sum_{i=1}^{n} P_i = \sum_{i=1}^{n} j(1_{\{x_i\}}) = j(\mathbb{1}_{\mathcal{B}}) = \mathbb{1}_{\mathcal{A}} \tag{142}$$

and for $i \neq j$,

$$P_i P_k = j(1_{\{x_i\}}) j(1_{\{x_k\}}) = j(1_{\{x_i\}} \cdot 1_{\{x_k\}}) = 0 . \tag{143}$$

We interpret P_i as the event "the random variable described by j takes the value x_i." Note that j can be written as

$$j(f) = j\left(\sum_{i=1}^{n} f(x_i) 1_{\{x_i\}}\right) = \sum_{i=1}^{n} f(x_i) P_i . \tag{144}$$

In particular, if $\Omega \subset \mathbb{R}$, then j defines a Hermitian operator

$$j(\mathrm{id}) = \sum_{i=1}^{n} x_i P_i =: X , \tag{145}$$

which completely determines j.

Proposition 8 *Let $\mathcal{A}$ be a finite-dimensional *-algebra with unit. Then there is a one-to-one correspondence between injective *-homomorphisms $j : \mathbf{C}(\Omega) \to \mathcal{A}$ for some finite $\Omega \subset \mathbb{R}$ and self-adjoint operators $X \in \mathcal{A}$, given by*

$$j(\mathrm{id}) = X . \tag{146}$$

Proof If j is a *-homomorphism $\mathbf{C}(\{x_1, \ldots, x_n\}) \to \mathcal{A}$ with $x_1, \ldots, x_n$ real, then

$$X := j(\mathrm{id}) = \sum_{i=1}^{n} x_i j(1_{\{x_i\}}) =: \sum_{i=1}^{n} x_i P_i \tag{147}$$

is a Hermitian element of $\mathcal{A}$. Conversely, if $X \in \mathcal{A}$ is Hermitian, then let $x_1, \ldots, x_n$ be its eigenvalues. Let $p : \mathbf{C} \to \mathbf{C}$ denote the polynomial

$$p(x) := (x - x_1) \cdots (x - x_n) \tag{148}$$

and let, for $i = 1, \ldots, n$, the (Lagrange interpolation) polynomial p_i be given by

$$p_i(x) := \frac{p(x)}{(x - x_i)p(x_i)} . \tag{149}$$

Then $p_i(x_k) = \delta_{ik} p_k$, so we have on the spectrum $\mathrm{sp}(X) = \{x_1, \ldots, x_n\}$ of X

$$\sum_{i=1}^{n} p_i = 1 \quad \text{and} \quad p_i \cdot p_k = \delta_{ik} p_k . \tag{150}$$

It follows that the projections $P_i := p_i(X)$, with $i = 1, \ldots, n$, lie in the algebra $\mathcal{A}$ and satisfy

$$\sum_{i=1}^{n} P_i = \mathbb{1} \quad \text{and} \quad P_i P_k = \delta_{ik} P_k . \tag{151}$$

Hence, if we define

$$j(f) := \sum_{i=1}^{n} f(x_i) P_i , \tag{152}$$

then j is a *-homomorphism with the property that $j(\mathrm{id}) = X$. Clearly, different X's correspond to different j's. □

6.5 The Joint Measurement Apparatus

Let X and Y be self-adjoint elements of the *-algebra $\mathcal{A}$. We consider X and Y as random variables taking values in the spectra $\mathrm{sp}(X)$ and $\mathrm{sp}(Y)$.

By a *joint measurement* M^* of these random variables we mean an operation that takes a state ρ on $\mathcal{A}$ as input and yields a probability distribution π on $\mathrm{sp}(X) \times \mathrm{sp}(Y)$ as output, in such a way that for all functions f on $\mathrm{sp}(X)$ and g on $\mathrm{sp}(Y)$

$$\rho(f(X)) = \sum_{x\in\mathrm{sp}(X)} \sum_{y\in\mathrm{sp}(Y)} \pi(x,y)f(x)\,, \tag{153}$$

$$\rho(g(Y)) = \sum_{x\in\mathrm{sp}(X)} \sum_{y\in\mathrm{sp}(Y)} \pi(x,y)g(y)\,. \tag{154}$$

A contravariant formulation of these requirements is

$$M(f\otimes\mathbb{1}) = f(X)\,, \tag{155}$$

$$M(\mathbb{1}\otimes g) = g(Y)\,. \tag{156}$$

Theorem 9 *If two random variables X and Y allow a joint measurement operation, then they commute.*

Proof Let us denote by x the identity function on $\mathrm{sp}(X)$ and by y on $\mathrm{sp}(Y)$. We apply the multiplication theorem on the measurement operation M, which is supposed to exist. Since

$$M\big((x\otimes\mathbb{1})^\dagger\,(x\otimes\mathbb{1})\big) = M(x^2\otimes\mathbb{1}) = X^2 = M(x\otimes\mathbb{1})^\dagger\,M(x\otimes\mathbb{1})\,, \tag{157}$$

we have

$$M\big((x\otimes\mathbb{1})^\dagger\,(\mathbb{1}\otimes y)\big) = M(x\otimes\mathbb{1})^\dagger\,M(\mathbb{1}\otimes y) = XY \tag{158}$$

and

$$M\big((\mathbb{1}\otimes y)^\dagger\,(x\otimes\mathbb{1})\big) = M(\mathbb{1}\otimes y)^\dagger\,M(x\otimes\mathbb{1}) = YX\,. \tag{159}$$

$$\text{As } (x\otimes\mathbb{1})^\dagger\,(\mathbb{1}\otimes y) = x\otimes y = (\mathbb{1}\otimes y)^\dagger\,(x\otimes\mathbb{1}),\ \text{ we have } XY = YX. \tag{160}$$

□

7 Quantum Novelties

In the previous section we saw certain strange limitations that quantum operations are subject to. Let us now look at the other side of the coin: some surprising possibilities. We leave treatment of the really sensational features to other contributions in this volume, such as very fast computation and secure cryptography. Here we shall treat “teleportation” of quantum states and “dense coding.”

7.1 Teleportation of Quantum States

Suppose that Alice wishes to send to Bob the quantum state ρ of a qubit over a (classical) telephone line.

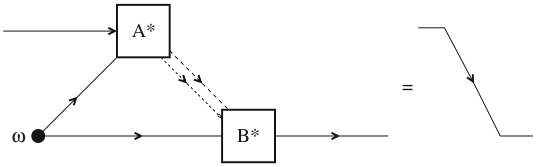


Fig. 8 Teleportation based on shared entanglement

In Sect. 6.2 we have seen that, without any further tools, this is impossible. If Alice performed measurements on the qubit and told the results to Bob over the telephone, these would not enable Bob to reconstruct the state ρ. However, suppose that Alice and Bob have been together in the past, and that at that time they have created an entangled pair of qubits, as introduced in Sect. 2.3, each taking one qubit with them. It was discovered in 1993 by Bennett, Wootters, Peres, and others that by making use of this shared entanglement, Alice is indeed able to transfer her qubit to Bob. Of course, she cannot avoid destroying the original state ρ in the process; otherwise Alice and Bob would have copied the state ρ, which is impossible by Theorem 6.1 ("no cloning"). It is for this reason that the procedure is called "teleportation."

We illustrate the procedure in Fig. 8.

Here ω is the fully entangled state $X \mapsto \langle \Omega, X\Omega \rangle$ on $M_2 \otimes M_2$ (see the proof of Theorem 3, Stinespring's theorem).

The procedure runs as follows. Alice possesses two qubits, one from the entangled pair and one which she wishes to send to Bob. She performs a von Neumann measurement on these two qubits along the four *Bell projections*

$$Q_{00} := \frac{1}{2}\begin{pmatrix} 1&0&0&1\\ 0&0&0&0\\ 0&0&0&0\\ 1&0&0&1 \end{pmatrix}, \qquad Q_{01} := \frac{1}{2}\begin{pmatrix} 1&0&0&-1\\ 0&0&0&0\\ 0&0&0&0\\ -1&0&0&1 \end{pmatrix}, \tag{161}$$

$$Q_{10} := \frac{1}{2}\begin{pmatrix} 0&0&0&0\\ 0&1&1&0\\ 0&1&1&0\\ 0&0&0&0 \end{pmatrix}, \qquad Q_{11} := \frac{1}{2}\begin{pmatrix} 0&0&0&0\\ 0&1&-1&0\\ 0&-1&1&0\\ 0&0&0&0 \end{pmatrix}. \tag{162}$$

The operation performed by Alice has the contravariant description:

$$A : \mathbf{C}_2 \otimes \mathbf{C}_2 \to M_2 \otimes M_2 : \quad A(e_i \otimes e_j) := Q_{ij}\,. \tag{163}$$

The two bits Alice obtains in this way – (i, j) say – she sends to Bob over the telephone. He then takes his own qubit from the entangled pair, and if $j = 1$ performs the "phase flip" operation

$$Z : \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{00} & -\rho_{01} \\ -\rho_{10} & \rho_{11} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{164}$$

and if $j = 0$ he does nothing. Then, if $i = 1$ he performs the "quantum NOT" operation

$$X : \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{11} & \rho_{10} \\ \rho_{01} & \rho_{00} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{165}$$

and if $i = 0$ he does nothing. In the Heisenberg picture, the result of Bob's actions is the operation

$$B : M_2 \to \mathbf{C}_2 \otimes \mathbf{C}_2 \otimes M_2 : \quad M \mapsto M \oplus \sigma_3 M \sigma_3 \oplus \sigma_1 M \sigma_1 \oplus \sigma_2 M \sigma_2 \,, \tag{166}$$

where $\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 := \begin{pmatrix} 0 & -\mathrm{i} \\ i & 0 \end{pmatrix}$, and $\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are Pauli's spin matrices. Bob ends up with a qubit in exactly the same state as Alice wanted to send.

We formulate this result in the Heisenberg picture.

Proposition 9 *The state ω and the operations A and B described above satisfy*

$$(\mathrm{id}_{M_2} \otimes \omega) \circ (A \otimes \mathrm{id}_{M_2}) \circ B = \mathrm{id}_{M_2} \,. \tag{167}$$

Proof We just calculate for $M \in M_2$:

$$\begin{aligned}
M &\overset{B}{\longmapsto} M \oplus \sigma_3 M \sigma_3 \oplus \sigma_1 M \sigma_1 \oplus \sigma_2 M \sigma_2 \\
&\overset{A \otimes \mathrm{id}}{\longmapsto} (Q_{00} \otimes M) + (Q_{01} \otimes \sigma_3 M \sigma_3) + (Q_{10} \otimes \sigma_1 M \sigma_1) + (Q_{11} \otimes \sigma_2 M \sigma_2) \\
&= \frac{1}{2} \begin{pmatrix} M + \sigma_3 M \sigma_3 & 0 & 0 & M - \sigma_3 M \sigma_3 \\ 0 & \sigma_1 M \sigma_1 + \sigma_2 M \sigma_2 & \sigma_1 M \sigma_1 - \sigma_2 M \sigma_2 & 0 \\ 0 & \sigma_1 M \sigma_1 - \sigma_2 M \sigma_2 & \sigma_1 M \sigma_1 + \sigma_2 M \sigma_2 & 0 \\ M - \sigma_3 M \sigma_3 & 0 & 0 & M + \sigma_3 M \sigma_3 \end{pmatrix} \\
&= \left(\begin{array}{ccccccccc} m_{00} & 0 & 0 & 0 & | & 0 & 0 & 0 & m_{01} \\ 0 & m_{11} & 0 & 0 & | & 0 & 0 & m_{10} & 0 \\ 0 & 0 & m_{11} & 0 & | & 0 & m_{01} & 0 & 0 \\ 0 & 0 & 0 & m_{00} & | & m_{01} & 0 & 0 & 0 \\ - & - & - & - & | & - & - & - & - \\ 0 & 0 & 0 & m_{10} & | & m_{11} & 0 & 0 & 0 \\ 0 & 0 & m_{01} & 0 & | & 0 & m_{00} & 0 & 0 \\ 0 & m_{01} & 0 & 0 & | & 0 & 0 & m_{00} & 0 \\ m_{10} & 0 & 0 & 0 & | & 0 & 0 & 0 & m_{11} \end{array} \right) \overset{\mathrm{id} \otimes \omega}{\longmapsto} \begin{pmatrix} m_{00} & m_{01} \\ m_{10} & m_{11} \end{pmatrix} = M \,.
\end{aligned} \tag{168}$$

□

Teleportation has been carried out successfully in the lab by Zeilinger et al. in Vienna in 1997 using polarized photons, and by other experimenters using different techniques later.

For the sake of such experiments explicit operations have been developed that form the "building blocks" of the diversity of quantum operations needed. For example the operation performed by Alice to prepare the teleportation of a qubit can be decomposed into an interaction and a measurement. Let j be the ordinary measurement operation of a qubit:

$$j : \mathbf{C}_2 \to M_2 : \qquad (f_0, f_1) \mapsto \begin{pmatrix} f_0 & 0 \\ 0 & f_1 \end{pmatrix} . \tag{169}$$

Let H denote the *Hadamard gate*, which acts on states or observables by multiplication on the left and on the right by the *Hadamard matrix* $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and let C denote the CNOT gate (compare chapter "Bipartite Quantum Entanglement", Sect. 2) with matrix $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. The operation C performs a NOT operation on the first qubit provided that the second is a 1, which is as shown in Fig. 9.

Check that, using the above building blocks, the procedure of quantum teleportation can be charted as in Fig. 10.

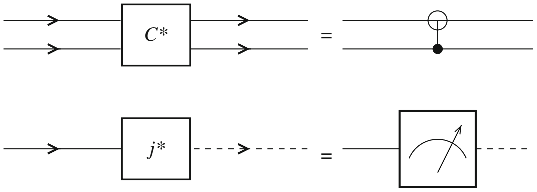


Fig. 9 Conventional signs used for the C and j operations

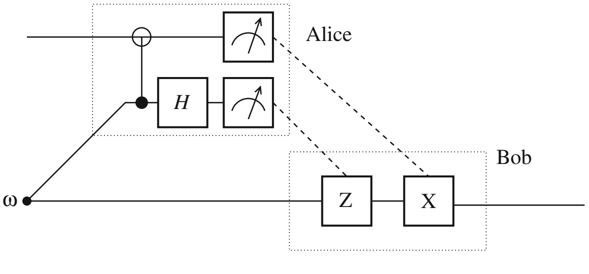


Fig. 10 More detailed scheme of teleportation

7.2 Dense Coding

We have seen that Alice can "teleport" a qubit using two classical bits, given a pre-entangled qubit pair. A kind of converse is also possible: Bob can communicate two classical bits to Alice by sending her a single qubit, again given a shared pre-

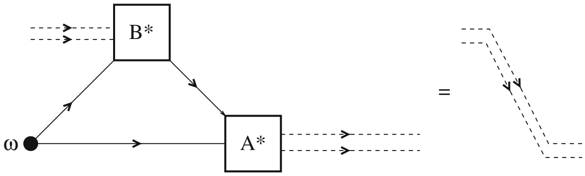


Fig. 11 Superdense coding: two bits in a single photon

entangled qubit pair (Fig. 11). (We have interchanged the roles of Alice and Bob here because it turns out that in that case they can continue using exactly the same equipment as they used for teleportation!)

Proposition 10 *Taking ω, A, and B as in Proposition 9, we have*

$$(\mathrm{id}_{\mathbf{C}_2\otimes\mathbf{C}_2} \otimes \omega) \circ (B \otimes \mathrm{id}_{M_2}) \circ A = \mathrm{id}_{\mathbf{C}_2\otimes\mathbf{C}_2} . \tag{170}$$

We leave the proof as an exercise.

References

1. Aspect, A., Dalibard, J., Roger, G.: Phys. Rev. Lett. **49**, 1804 (1982) 70, 72
2. Bell, J.S.: Physics **1**, 195 (1964) 68
3. Bratteli, O., Robinson, D.W.: Operator Algebras and Quantum Statistical Mechanics I. Springer, New York (1981) 83
4. Bohm, D.: Phys. Rev. **85**, 189 (1952) 72
5. Diecks, D.: Phys. Lett. A **92**, 271 (1982) 98
6. Dvurecenskij, A.: Gleason's theorem and its applications. In: Mathematics and Its Applications, vol. 60, p. 348. Kluwer, Dordrecht (1992) 81
7. Kraus, K.: *States, Effects and Operations*. Lect. Notes Phys. **190**. Springer, Berlin (1983) 89
8. Kümmerer, B., Maassen, H.: Elements of Quantum Probability. In: Hudson, R.L., Lindsay, J.M. (eds.) Quantum Probability Communications X, pp. 73–100. World Scientific, Singapore (1998) 67
9. Nelson, E.: Dynamical Theories of Brownian Motion. Princeton University Press, Princeton (1967) 72
10. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000) 92
11. Wootters, W.K., Zurek, W.H.: Nature **299**, 802 (1982) 98