Quantum Probability
Quantum Information Theory
Quantum Computing

Hans Maassen

Lecture notes of a course to be given in the spring semester of 2004 at the Catholic University of Nijmegen, the Netherlands.

Introduction

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

According to such 'realistic' intuitions all things have their definite place and properties, such as speed, colour, weight, etcetera. Quantum theory, however, refuses to precisely pinpoint them. Now it could be, of course, that quantum theory is just incomplete, that it gives a coarse description of a reality that is actually much finer. If that were the case we should join the heirs of Albert Einstein in their search for a finer mathematical model of physical reality.
However, by the work of John Bell and Alain Aspect it has become clear that the search for such underlying 'hidden variable' models runs into certain difficulties: they must at least allow *action at a distance.* And even if that would not disturb us, (which it does), they have not been very successful in the prediction of new phenomena.
It seems that we must accept the inherent strangeness of quantum theory.

0.1 Quantum Probability

As indicated above, quantum mechanics does not predict the result of physical experiments with certainty, but yields probabilities for their possible outcomes.

Now, the mathematical theory of probability obtained a unified formulation in the 1930's, when Kolmogorov introduced his axioms and defined the universal structure $(\Omega, \Sigma, \mathbb{P})$ of a probability space. However, the mathematical language of probability theory (probability measures and densities, stochastic processes, martingales, Markov chains, ...) for a long time remained completely separated from the mathematical language of quantum mechanics (vectors in a Hilbert space, hermitian operators, unitary transformations, ...).

In the 1970's and 1980's people such as Accardi, Lewis, Davies, Kümmerer, building on ideas of von Neumann's and Segal's concerning algebras of operators, developed a unified framework, a generalized, non-commutative probability theory, in which classical probability theory and quantum mechanics can be discussed together. We shall use their language in this course.

0.2 Quantum Information

In Shannon's (classical) information theory, a single unit of information, the bit, serves to quantify all forms of information, beit in print, in 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 one another. 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'.

0.3 Quantum Computing

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

The idea was developed in the 1980's and 1990's by David Deutsch, Peter Shor, and many others into a flourishing branch of science called 'quantum computing': how to make quantummechanical systems perform calculations more efficiently than ordinary computers can. 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.

0.4 This course

We start with an introduction to quantum probability. No prior knowledge of quantummechanics is assumed; what is needed will be explained in the course.

We begin by demonstrating the 'strangeness' of quantum phenomena by very simple polarization experiments, culminating in Bell's famous inequality, tested in Aspect's equally famous 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, we build up our new probability theory in terms of algebras of operators on a Hilbert space.

Operations on these algebras will then be characterized, and the points where they are at variance with classical operations: what cannot be done with them (copying, coding into classical information, joint measurement of incompatible observables, measurement without perturbing the measured object), and what *can* be done (entangling remote systems, teleportation of this entanglement, sending two bits in a single qubit). Then luring perspectives will be sketched: highly efficient algorithms for sorting, Fourier transformation and factoring very large numbers.

As an example of quantum thinking we shall treat a quantum version of the famous 'three door' or 'Monty Hall' riddle.

We shall introduce the concepts of entropy and information in the classical and the quantum context. We shall describe simple quantum Markov chains and their relation to repeated measurement and quantum counting processes. These will lead to the so-called quantum trajectories: simulation of quantum processes on a (classical) computer.

If time permits we shall go into some of the following topics:

(i) entropic uncertainty relations,
(ii) quantum error correction,
(iii) stochastic Schrödinger equations, and
(iv) ergodicity of quantum trajectories.

1. Why Classical Probability does not Suffice *

1.1 **An experiment with polarisers**

To start with, we consider a simple experiment. In a beam of light of a fixed colour 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 which comes through both filters differs in intensity when the filters are rotated relative to each other. If we fix the first filter and rotate the second, then we see that there is a direction where the resulting intensity is maximal. Starting from this position, and rotating the second filter through an angle α, the light intensity decreases with α until it vanishes for $\alpha = \frac{1}{2}\pi$. Careful measurement shows that the intensity of the light passing the first filter is half the beam intensity (we assume the original beam to be completely unpolarized) and that of the light passing the second filter is proportional to $\cos^2\alpha$. 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 $I_1 = \frac{1}{2}I_0$ and

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

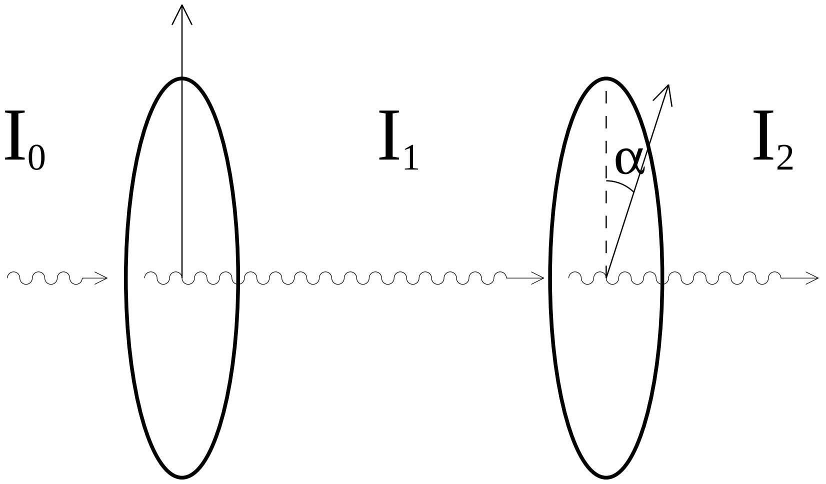

FIG. 1

Now, it has been observed that for extremely low intensities (monochromatic) light comes in small packages, called photons, all of the same energy, (which is independent of the total intensity).

So the intensity must be proportional to the number of photons, and formula (1) has to be given a statistical meaning: a photon passing through the first filter has

[*] This chapter is based on: B. Kümmerer and H. Maassen, *Elements of Quantum Probability*, Quantum Probability Communications, X pp. 73–100.

a probability $\cos^2\alpha$ to pass through the second. So formula (1) only holds on the average, i.e., for large numbers of photons.

If we think along the lines of classical probability, then we may attach to a polarization filter in the direction α a random variable P_α, taking the values 0 and 1, where $P_\alpha(\omega) = 0$ 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 we may write for their correlation:

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

(Here a common notation from probability theory is used, namely, the expression $[P_\alpha = 1 \text{ and } P_\beta = 1]$ stands for the set of those ω for which $P_\alpha(\omega) = 1$ and $P_\beta(\omega) = 1$.)

The following argument shows that this line of reasoning leads into difficulties. Take three polarizing filters $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$, having polarization directions α_1, α_2 and α_3 respectively. We put them on the optical bench in pairs. Then they give rise to random variables P_1, P_2 and P_3 satisfying

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

Proposition (Bell's 3 variable inequality) *For any three 0-1-valued random variables P_1, P_2, and P_3 on a probability space $(\Omega,\mathbb{P})$ the following inequality holds:*

$$\mathbb{P}[P_1 = 1, P_3 = 0] \;\le\; \mathbb{P}[P_1 = 1, P_2 = 0] + \mathbb{P}[P_2 = 1, P_3 = 0].$$

Proof. Write

$$\begin{aligned}\mathbb{P}[P_1 = 1, P_3 = 0] &= \mathbb{P}[P_1 = 1, P_2 = 0, P_3 = 0] + \mathbb{P}[P_1 = 1, P_2 = 1, P_3 = 0] \\ &\le \mathbb{P}[P_1 = 1, P_2 = 0] + \mathbb{P}[P_2 = 1, P_3 = 0].\end{aligned}$$

□

In our example, however, we have

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

Bell's inequality thus reads

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

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

$$\frac{3}{8} \le \frac{1}{8} + \frac{1}{8}.$$

We thus come to the conclusion that classical probability cannot describe this simple experiment!

Remark

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 $\mathbb{P}[P_\alpha = 1] = \frac{1}{2}$, satisfying the requirement that

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

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

$$\mathbb{P}[X \neq Y] = \int_\Omega |X(\omega) - Y(\omega)|\mathbb{P}(d\omega) = \|X - Y\|_1.$$

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

1.2 An improved experiment

A possible criticism to the above argument runs as follows. Are the random variables P_α well-defined? Is it indeed true that for each photon ω and each filter $\mathcal{F}_\alpha$ it is determined whether ω passes through $\mathcal{F}_\alpha$ or not? Could not filter $\mathcal{F}_\alpha$ influence the photon's reaction to filter $\mathcal{F}_\beta$? In fact, it seems quite obvious that it will!

In order to meet this criticism we should do a better experiment. We should let the filters act on each of the photons without influence on each other.
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 behaviour towards 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.

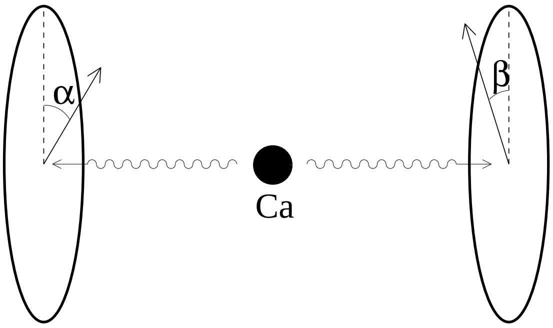

FIG. 2

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 on the left and P_β on the right, then between P_α on the left and P_γ on the right, and finally between P_β on the left and P_γ on the right. The same outcomes are found, violating Bell's three variable inequality, thus strengthening the case against classical probability.

1.3 **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_β 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 an experiment which was actually performed by A. Aspect in Orsay (near Paris) in 1982 [Asp]. 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.

PROPOSITION (Bell's 4 variable inequality) *For any quadruple P_1, P_2, Q_1, and Q_2 of 0-1-valued random variables on $(\Omega,\mathbb{P})$ the following inequality holds:*

$$\mathbb{P}[P_1 = Q_1] \quad \leq \quad \mathbb{P}[P_1 = Q_2] + \mathbb{P}[P_2 = Q_1] + \mathbb{P}[P_2 = Q_2]. \qquad (2)$$

(In fact, by symmetry, neither of these four probablities is larger than the sum of the other three.)

Proof. It is easy to see that for all ω:

$$P_1(\omega) = Q_1(\omega) \Longrightarrow P_1(\omega) = Q_2(\omega) \text{ or } Q_2(\omega) = P_2(\omega) \text{ or } P_2(\omega) = Q_1(\omega)\ .$$

□

Bell's 4-variable inequality can be viewed as the quadrangle inequality with respect to the metric $(X, Y) \mapsto \|X - Y\|_1$.

On the other hand, quantum mechanics predicts (cf. Section 2.4 below), and the experiment of Aspect showed, that one has,

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

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

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

So Bell's 4 variable inequality reads in this example:

$$\sin^2(\alpha_1 - \beta_1) \le \sin^2(\alpha_1 - \beta_2) + \sin^2(\alpha_2 - \beta_1) + \sin^2(\alpha_2 - \beta_2),$$

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

$$1 \le \frac{1}{4} + \frac{1}{4} + \frac{1}{4}.$$

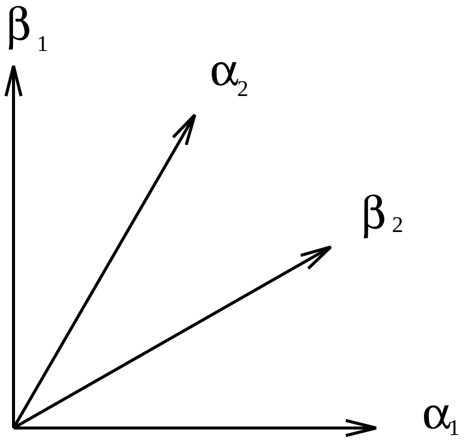

FIG. 3

So 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.

Remarks.

1. When applying the above Proposition to the Orsay experiment, we should keep in mind that a crucial assumption has to be made. It must be assumed that for each $\omega \in \Omega$ the values of $P_j(\omega)$ and $Q_j(\omega)$ are well-defined. This means that in each imagined realization of the world it is determined how each photon will react to any possible filter, including those it does not actually meet. This assumption is typical for classical probabilistic physical theories, but is abandoned in standard quantum mechanics. (Unmeasured quantities like the ones mentioned above are called 'hidden variables' in the literature on the foundations of quantum mechanics.)

2. A second important assumption, also necessary for the applicability of Bell's inequality, is that 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. We shall call

this assumption ‘locality’. In order to justify this assumption, Aspect has made considerable efforts. In his (third) experiment, 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 not possible.

3. Clearly, the above reasoning does not exclude the possibility of an explanation of the experiment in classical probabilistic terms, if one is willing to give up the causality principle. Serious attempts have been made in this direction (e.g. [Boh]).

1.4 The Orsay experiment as a card game

It has now become very difficult for the advocates of classical probability to criticize the experiment. To illustrate this point, we shall again present the experiment, but this time in the form of a card game. Nature can win this game. Can you?

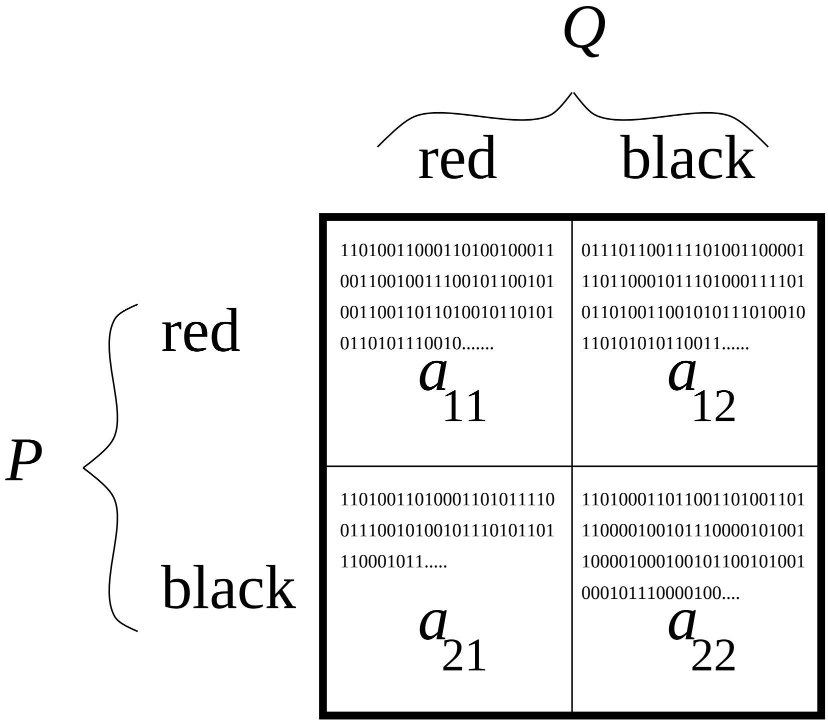

FIG. 4

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 a card 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 colour: red or black.)
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 depend 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, et cetera.
4. The cards are laid out on the table. The pair of colours of the cards determines one of the four squares on the board: these are labelled (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 0's and 1's. The arbiter keeps track of the percentage of 1's in proportion to the total number of digits in each square; we shall call the limits of these percentages as the game stretches out to infinity: 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 anwers as often as they can when both their cards are red, but different answers otherwise.

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

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

'*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 we would repeat this strategy very often, then on the long run we 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, sixteen) strategies like this. By the pointwise version (3) of Bell's 4-variable inequality (recall Section 1.3), none of these sixteen extremal strategies wins the game. Inclusion of the randomness coming from the dice yields a full polytope of random strategies, having the above sixteen 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, solely based on the $\cos^2$ law (1) 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 colours 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 quantummechanical die, makes possible what could not be done with the classical die.

2. Towards a Mathematical Model

2.1 A mathematical description of polarization

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 probability, to be developed now.

We have discussed (linear) polarization of a light beam. This is completely characterized by a direction in the plane perpendicular to the light beam. This suggests that we should describe different directions 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. Moreover, it appears that we cannot physically distinguish between two states which differ by a rotation of π, so we have to describe states of polarizations by one-dimensional subspaces of $\mathbb{R}^2$. (Two unit vectors span the same one-dimensional subspace if they differ only by a sign.) Given two directions of polarization with an angle α between them, spanned by two unit vectors $\psi, \theta \in \mathbb{R}^2$, the transition probability $\cos^2 \alpha$ can be expressed as

$$\cos^2 \alpha = <\psi, \theta>^2$$

where $<\psi, \theta>$ denotes the scalar product between ψ and θ. (Since $\cos^2 \alpha = \cos^2(\pi - \alpha)$, this expression does not depend on the sign of ψ, θ.)

Certainly, in order to come to a 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 filter, (i.e., the random variable), with the orthogonal projection P onto the one-dimensional subspace. We can then write

$$\cos^2 \alpha \; = <\psi, \theta>^2 = <\psi, P\psi> \quad .$$

So we arrive at the following mathematical model:

States of polarization of a photon	$\hat{=}$	one-dimensional subspaces of $\mathbb{R}^2$ described by unit vectors ψ spanning the subspace.
Polarization filters, (i.e., random variables measuring polarization)	$\hat{=}$	orthogonal projections P from $\mathbb{R}^2$ onto the corresponding one-dimensional subspace.
Probability that a photon, described by ψ, passes through the filter described by P	$\hat{=}$	$<\psi, P\psi> = \cos^2 \alpha$.

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

$$< \psi, P\psi > = \mathbb{E}(P) \ .$$

It is important to realize that, although we gave a kind of proof (in Section 1) that polarization experiments *cannot* be described by classical random variables on classical probability spaces, there is *no* logical argument that photons must be described by vectors and filters by projections, as we just did. Indeed, since the beginnings of quantum mechanics there have been many efforts to develop alternative mathematical models. We are going to describe here the traditional point of view of quantum mechanics [Neu]. This will lead to a mathematical model which extends classical probability and up until now has described experiments correctly.

2.2 **The full quantum mechanical 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 2-level system or *qubit*:

State of polarization of a photon $\hat{=}$ one-dimensional subspace of $\mathbb{C}^2$, 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 $\hat{=}$ orthogonal projection P onto a complex one-dimensional subspace.

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

Probability for a photon, described by ψ, to pass through a filter, described by P $\hat{=}$ $< \psi, P\psi >$.

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

$$(\cos\alpha, e^{i\phi}\sin\alpha) \in \mathbb{C}^2 \ , \quad \frac{-\pi}{2} \leq \alpha \leq \frac{\pi}{2}, \quad 0 \leq \phi \leq \pi \ .$$

This set can be identified with the points on the unit sphere $S^2 \in \mathbb{R}^3$ when using the polar coordinates $\theta = 2\alpha$ and ϕ. Restricting to states with $\phi = 0$, (which are parametrized by the points of the circle $(\cos 2\alpha, \sin 2\alpha)$ in $\mathbb{R}^2$), we retain the foregoing real description when we identify α with the angle of polarization.

A possible identification of the points of $S^2 \subseteq \mathbb{R}^3$ with physical states, giving the correct values for all probabilities, is as in the picture below:

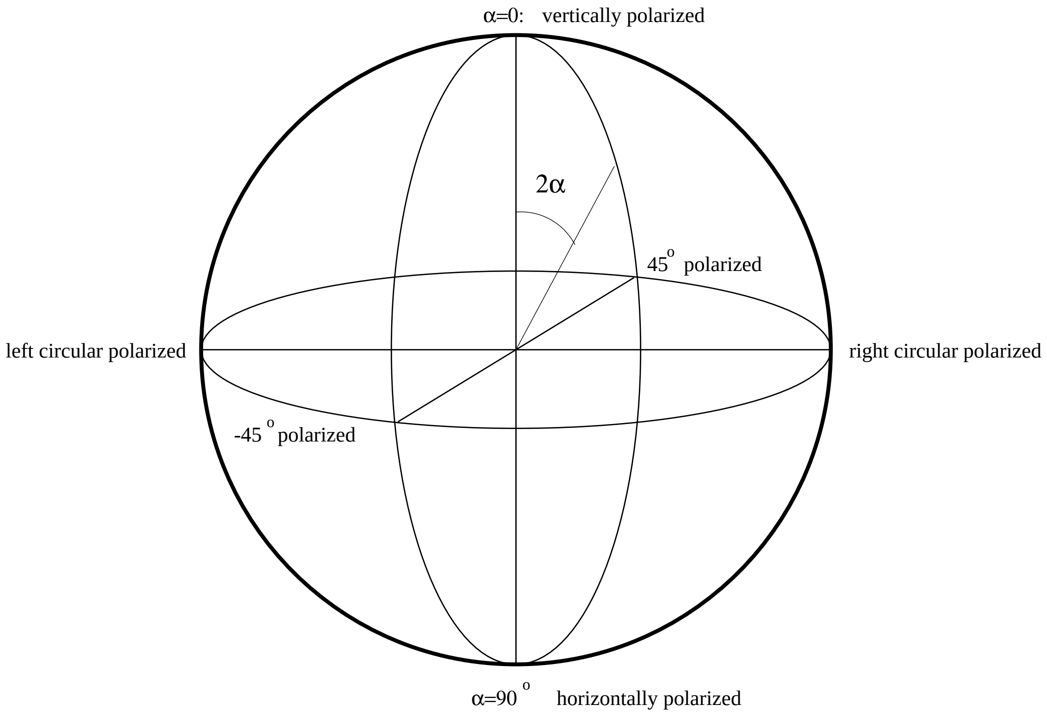

FIG. 5

With these identifications we come to the following mathematical model of polarization of light:

vertically polarized light	$\widehat{=}$	$(1,0) \in \mathbb{C}^2$
horizontally polarized light	$\widehat{=}$	$(0,1) \in \mathbb{C}^2$
light polarized at an angle α to the vertical direction	$\widehat{=}$	$(\cos\alpha, \sin\alpha) \in \mathbb{C}^2$
light polarized at an angle $\alpha = \pm\frac{\pi}{4}$ to the vertical direction	$\widehat{=}$	$\left(\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) \in \mathbb{C}^2$
left-/right-circular polarized light	$\widehat{=}$	$\left(\frac{1}{\sqrt{2}}, \pm\frac{i}{\sqrt{2}}\right) \in \mathbb{C}^2$

and correspondingly

vertical polarizer	$\widehat{=}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
horizontal polarizer	$\widehat{=}$	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$

$$\text{angle-}\alpha\text{-polarizer} \qquad \hat{=} \begin{pmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix}$$

$$\pm\tfrac{\pi}{4}\text{-polarizer} \qquad \hat{=} \begin{pmatrix} \frac{1}{2} & \pm\frac{1}{2} \\ \pm\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$\text{left/right circular polarizer} \qquad \hat{=} \begin{pmatrix} \frac{1}{2} & \mp\frac{i}{2} \\ \pm\frac{i}{2} & \frac{1}{2} \end{pmatrix}$$

2.3 **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, we restrict ourselves to the quantum mechanics of finite dimensions. It generalizes the probability theory of systems with only finitely many states. As in classical probability, the generalization to systems with a countable number of states or a continuum of states is analytically more involved, though conceptually easy.

The model is as follows:

States correspond to one-dimensional subspaces of $\mathbb{C}^n$, 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 are described by orthogonal projections onto a linear subspace of $\mathbb{C}^n$. If the random variable only asks whether the system is in a certain state, then the subspace is one-dimensional. But also projections onto higher dimensional subspaces $\mathcal{K}$ make sense. They answer the question whether the system is in any of the states represented by a unit vector in $\mathcal{K}$. Similar questions for other subsets of states are not allowed!

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

Note that we do not assume that every unit vector $\psi \in \mathbb{C}^n$ describes a state of the system, nor that every orthogonal projection corresponds to a meaningful random variable. Specializing these two sets is part of the description of the mathematical model for a given system. In a truly quantum mechanical situation, typically all possible vectors and projections are used. In contrast to this, a model from classical probability is incorporated into this description as follows.

2.4 **Finite classical models**

A finite probability space is usually described by a finite set $\Omega = \{\omega_1, \ldots, \omega_n\}$ and a probability distribution $(p_1, \ldots, p_n)$, $0 \le p_i \le 1$, $\sum_i p_i = 1$, 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 $\mathbb{C}^n$ 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 i-th 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 $< \delta_i, P_A\delta_i > = 1$.

Conversely, any set of pairwise commuting projections on $\mathbb{C}^n$ 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.

In the above sketch of classical probability an important point is obviously missing: So far we have only considered pure states of the system, a probability distribution $(p_1, \ldots, p_n)$ did not enter the dicussion. How can we describe a situation where a system is in a certain state ψ with probability q and in another state y with probability $1 - q$ $(0 \leq q \leq 1)$?

Obviously, the set of states should be a convex set, containing also the mixed states. In the classical model of probability, the appropriate convex combinations of point measures are taken in order to obtain a new probability measure.

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

$$\sum_i p_i < \psi_i, P\psi_i > \quad .$$

A more convenient description of mixed states is obtained as follows.

For a unit vector $\psi \in \mathbb{C}^n$ denote by Φ_ψ the orthogonal projection onto the one-dimensional subspace generated by ψ. In the physics literature, Φ_ψ is frequently denoted by $|\psi >< \psi|$. By tr denote the trace on the $n \times n$-matrices, summing up the diagonal entries of such a matrix. Then one obtains

$$< \psi, P\psi > = tr(\Phi_\psi \cdot P) \ .$$

Hence

$$\sum_i p_i < \psi_i, P\psi_i > = tr(\sum_i p_i \Phi_{\psi_i} \cdot P) \ = \ tr(\Phi \cdot P) \ ,$$

where $\Phi := \sum_i p_i \Phi_{\psi_i}$.

Being a convex combination of 1-dimensional projections, Φ obviously is a positive (i.e., self-adjoint positive semidefinite) $n \times n$-matrix with $tr(\Phi) = 1$. Conversely, from diagonalizing positive matrices it is clear that any such positive matrix Φ with $tr(\Phi)=1$ can be written as a convex combination of 1-dimensional projections. The set of these matrices forms a closed (even compact) convex set, and its extreme points are precisely the 1-dimensional projections which in turn correspond to pure states, represented also by unit vectors. Therefore it is precisely this class of matrices which represents mixed states. These matrices are frequently called *density matrices.*

Thus, a general mixed state is described by a density matrix Φ and the probability for an observation of P to yield the value 1 is given by $tr(\Phi \cdot P)$.

Remarks

1. Although in this description also pure states are described by 1-dimensional projections, they are not considered as random variables.

2. The decomposition of a density matrix Φ into a convex combination of 1-dimensional projections is by no means unique. The compact convex set of density matrices is far from being a simplex. Indeed, on $\mathbb{C}^2$ it can be affinely 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 $\Phi = \sum_i p_i P_{\delta_i}$ into a convex combination of pure states (within the convex hull of $\{P_{\delta_i} : 1 \leq i \leq n\}$ which is a simplex), then it becomes unique.

4. Physically, a state Φ is completely described by all of its values $tr(\Phi \cdot P)$, 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 $\Phi = \sum_i \frac{1}{n} P_{\delta_i} = \frac{1}{n} \cdot \mathbb{1}$. On the other hand, with the unit vector $\psi = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}) \in \mathbb{C}^n$, we obtain for any subset $A \subseteq \Omega$: $tr(\Phi \cdot P_A) = \frac{1}{n} \cdot |A| = <\psi, P_A \psi>$. Therefore, on the random variables $\{P_A : A \subseteq \Omega\}$, the rank-one-density matrix P_ψ represents the same state as the densitiy matrix $\frac{1}{n} \cdot \mathbb{1}$. Note, however, that P_ψ is not in the convex hull of $\{P_{\delta_i} : 1 \leq i \leq n\}$.

2.5 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 $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4$, where we use the basis

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

with $e_1 = (1,0) \in \mathbb{C}^2$ and $e_2 = (0,1) \in \mathbb{C}^2$. 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

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

Now, the filters $P(\alpha)$ on the left and $Q(\beta)$ on the right, introduced in §1.3, are represented by two-dimensional projection operators on $\mathbb{C}^4$, which are the "2-right amplification" and the "2-left-amplification" of the polarization matrix

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

namely

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

(More about such tensor products will be treated in Section 3.)

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

$P(\alpha)(\mathbb{1}-Q(\beta))$, $(\mathbb{1}-P(\alpha))Q(\beta)$, and $(\mathbb{1}-P(\alpha))(\mathbb{1}-Q(\beta))$. 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”).

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

$$
\begin{aligned}
&<\psi, P(\alpha)Q(\beta)\psi> \\
&= \frac{1}{2}(0,1,-1,0)\times \\
&\times \begin{pmatrix} \cos^2\alpha\cos^2\beta & \cos^2\alpha\cos\beta\sin\beta & \cos\alpha\sin\alpha\cos^2\beta & \cos\alpha\sin\alpha\cos\beta\sin\beta \\ \cos^2\alpha\cos\beta\sin\beta & \cos^2\alpha\sin^2\beta & \cos\alpha\sin\alpha\cos\beta\sin\beta & \cos\alpha\sin\alpha\sin^2\beta \\ \cos\alpha\sin\alpha\cos^2\beta & \cos\alpha\sin\alpha\cos\beta\sin\beta & \sin^2\alpha\cos^2\beta & \sin^2\alpha\cos\beta\sin\beta \\ \cos\alpha\sin\alpha\cos\beta\sin\beta & \cos\alpha\sin\alpha\sin^2\beta & \sin^2\alpha\cos\beta\sin\beta & \sin^2\alpha\sin^2\beta \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \\
&= \frac{1}{2}(\cos^2\alpha\sin^2\beta + \sin^2\alpha\cos^2\beta - 2\cos\alpha\sin\alpha\cos\beta\sin\beta) \\
&= \frac{1}{2}(\cos\alpha\sin\beta - \sin\alpha\cos\beta)^2 \\
&= \frac{1}{2}\sin^2(\alpha-\beta) \quad .
\end{aligned}
$$

3. 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* $\mathbb{P}(S)$ to each of these events. Requirements: the events must form a σ-algebra and the probability measure $\mathbb{P}$ must be σ-additive.
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 Chapter 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:
(i) 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 $\mathcal{H}$.
(ii) The probability function $\mathbb{P} : \mathcal{E} \to [0,1]$ must be σ-additive.

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

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

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

3.1 Hilbert spaces, closed subspaces, and projections

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

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

called the *inner product*, with the following properties:
(i) $\langle \psi, \chi_1 + \chi_2 \rangle = \langle \psi, \chi_1 \rangle + \langle \psi, \chi_2 \rangle$ for all $\psi, \chi_1, \chi_2 \in \mathcal{H}$;
(ii) $\langle \psi, \lambda\chi \rangle = \lambda \langle \psi, \chi \rangle$ for all $\psi, \chi \in \mathcal{H}$ and all $\lambda \in \mathbb{C}$;
(iii) $\overline{\langle \psi, \chi \rangle} = \langle \chi, \psi \rangle$ for all $\psi, \chi \in \mathcal{H}$;
(iv) $\langle \psi, \psi \rangle \geq 0$ for all $\psi \in \mathcal{H}$;
(v) $\langle \psi, \psi \rangle = 0$ implies that $\psi = 0$;
(vi) $\mathcal{H}$ is complete in the norm $\psi \mapsto \|\psi\| := \langle \psi, \psi \rangle^{\frac{1}{2}}$,
i.e. if $\psi_1, \psi_2, \psi_3, \cdots$ is a *Cauchy sequence*:

$$\lim_{n\to\infty} \sup_{m \geq n} \|\psi_n - \psi_m\| = 0 \ ,$$

then there is a vector $\psi \in \mathcal{H}$ such that

$$\lim_{n\to\infty} \|\psi_n - \psi\| = 0 \; .$$

If the conditions (v) and (vi) are not required, we call $\mathcal{H}$ a *pre-Hilbert space.* In a Hilbert space for all vectors ψ, χ the *triangle inequality* is valid:

$$\|\psi + \chi\| \leq \|\psi\| + \|\chi\| \; .$$

In a Hilbert space we have the *Cauchy-Schwarz inequality*:

$$|\langle\psi, \chi\rangle| \leq \|\psi\|\|\chi\| \; .$$

Let $\mathcal{S}$ be a subset of $\mathcal{H}$. By $\mathcal{S}^\perp$ we mean the closed linear subspace of $\mathcal{H}$ given by

$$\mathcal{S}^\perp := \left\{ \, \psi \in \mathcal{H} \, \middle| \, \forall_{\chi\in\mathcal{S}} : \langle\chi, \psi\rangle = 0 \, \right\} \; .$$

By the *linear span* of $\mathcal{S}$, written as $\bigvee \mathcal{S}$, we mean the space of all finite linear combinations of elements of $\mathcal{S}$. Its *closure* $\overline{\bigvee \mathcal{S}}$ is the smallest closed subspace of $\mathcal{H}$ which contains $\mathcal{S}$.

Proposition *3.1. Let $\mathcal{S}$ be a subset of a Hilbert space $\mathcal{H}$. Then every element ψ of $\mathcal{H}$ can be written in a unique way as $\psi_1 + \psi_2$, where*

$$\psi_1 \in \overline{\bigvee \mathcal{S}} \text{ and } \psi_2 \in \mathcal{S}^\perp \; .$$

Moreover,

$$\overline{\bigvee \mathcal{S}} = \mathcal{S}^{\perp\perp} \; .$$

So the map $\psi \mapsto \psi_1$ is an orthogonal projection determined by the set $\mathcal{S}$. Conversely, the Range $P\mathcal{H}$ of any orthogonal projection is a closed linear subspace of $\mathcal{H}$.

Corollary *3.2 Closed linear subspaces of a Hilbert space and orthogonal projections on that space are in one-to-one correspondence.*

Proof of Proposition 3.1. Choose $\psi \in \mathcal{H}$ and let d denote

$$d := \inf_{\vartheta\in\bigvee\mathcal{S}} \|\psi - \vartheta\| \; ,$$

the distance from ψ to the span of $\mathcal{S}$.
Let $\vartheta_1, \vartheta_2, \vartheta_3, \cdots$ be a sequence in $\bigvee \mathcal{S}$ with

$$\lim_{n\to\infty} \|\vartheta_n - \psi\| = d \; .$$

For all $n, m \in \mathbb{N}$ we have by the parallellogram law, which follows from the properties of the inner product,

$$\| \vartheta_n + \vartheta_m - 2\psi \|^2 + \| \vartheta_n - \vartheta_m \|^2 = 2 \left(\| \vartheta_n - \psi \|^2 + \| \vartheta_m - \psi \|^2 \right) .$$

As $n, m \to \infty$, the right hand side tends to $4d^2$. Since $\|\frac{1}{2}(\vartheta_n + \vartheta_m) - \psi\| \geq d$ we must have $\|\vartheta_n - \vartheta_m\| \to 0$. So $\vartheta_1, \vartheta_2, \vartheta_3, \cdots$ is a Cauchy sequence; let ψ_1 be its limit. Then $\psi_1 \in \overline{\bigvee \mathcal{S}}$. Finally we have for all $\chi \in \mathcal{S}$ and all $t \in \mathbb{R}$:

$$\| (\psi_1 + t\chi) - \psi \|^2 = \| \psi_1 - \psi \|^2 + 2t \mathrm{Re} \langle \psi_1 - \psi, \chi \rangle + t^2 \| \chi \|^2 ,$$

and since the left hand side must always be at least d^2, this quadratic function of t must have its minimum at 0. It follows that $\psi_2 := \psi_1 - \psi$ is orthogonal to χ. This proves the first part of the theorem. To prove the second, note that $\mathcal{S}^{\perp\perp}$ is a closed subspace containing $\mathcal{S}$. So $\overline{\bigvee \mathcal{S}} \subset \mathcal{S}^{\perp\perp}$. Conversely suppose that $\psi \in \mathcal{S}^{\perp\perp}$. Then

$$\psi_2 = \psi - \psi_1 \in \mathcal{S}^{\perp\perp} \cap \mathcal{S}^{\perp} = \{0\} ,$$

so $\psi = \psi_1 \in \overline{\bigvee \mathcal{S}}$. □

In this course we will mainly be concerned with finite-dimensional Hilbert spaces. In that case all subspaces are automatically closed, and many of the precautions taken in the proof above are not needed.

3.2 $*$-algebras and states

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^* on $\mathcal{H}$ satisfying

$$\forall_{\psi, \vartheta \in \mathcal{H}} : \quad \langle A^* \psi, \vartheta \rangle = \langle \psi, A\vartheta \rangle .$$

The *norm* of an operator A is defined by

$$\|A\| := \sup\{ \, \|A\psi\| \, \big| \, \psi \in \mathcal{H}, \|\psi\| = 1 \, \} .$$

It has the property

$$\|A^* A\| = \| A \|^2 .$$

Exercise. 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

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

By a *state* on $\mathcal{A}$ we mean a linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ satisfying
(i) $\forall_{A\in\mathcal{A}} : \quad \varphi(A^*A) \geq 0\,,$
(ii) $\varphi(\mathbb{1}) = 1\,.$

A pair $(\mathcal{A}, \varphi)$ as described above is called a *quantum probability space.*

Examples

1. Let P_1, P_2, …, P_k be mutually orthogonal projections on $\mathcal{H}$ with $\sum_{j=1}^k P_j = \mathbb{1}$. Then their linear span

$$\mathcal{A} := \Big\{ \sum_{j=1}^{k} \lambda_j P_j \,\Big|\, \lambda_1, \ldots, \lambda_k \in \mathbb{C} \Big\} \;.$$

forms a unital $*$-algebra of operators on $\mathcal{H}$. This is basically the classical model of Section 2.4.: $\mathcal{A}$ is isomorphic to $\mathcal{C}(\Omega)$, 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:

$$\varphi(A) := \langle \psi, A\psi \rangle \;.$$

The probabilities of this classical model are $p_j := \varphi(P_j) = \| \, P_j \psi \, \|^2$. Note that there are many ψ's, and even more density matrices Φ (see Section 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 $\varphi(A) := \mathrm{tr}\,(\Phi A)$ with $\Phi \geq 0$ and $\mathrm{tr}\,(\Phi) = 1$, as introduced in Section 2.4.
The state φ is called a *pure* state if $\Phi = |\psi\rangle\langle\psi|$.
The qubit of Section 2.2 corresponds to the case $n = 2$.
The most general way of representing M_n on a Hilbert space is:

$$\mathcal{H} = \mathbb{C}^m \otimes \mathbb{C}^n \quad (m \geq 1); \qquad \mathcal{A} = \big\{ \, \mathbb{1} \otimes A \,\big|\, A \in M_n \, \big\} \;.$$

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

$$\mathcal{H} := (\mathbb{C}^{m_1} \otimes \mathbb{C}^{n_1}) \oplus (\mathbb{C}^{m_2} \otimes \mathbb{C}^{n_2}) \oplus \cdots \oplus (\mathbb{C}^{m_k} \otimes \mathbb{C}^{n_k}) \;.$$

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

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

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

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

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 quantummechanical' possibility, and Example 3 is the most general case.

THEOREM *3.3 (***Gel'fand***) Every commutative $*$-algebra of operators on a finite-dimensional Hilbert space is isomorphic to $\mathcal{C}(\Omega)$ for some finite Ω.*

Remark. Theorem 3.3 is the finite-dimensional version of Gel'fand's theorem on commutative C*-algebra's.

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:

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

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

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

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

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

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

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

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

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

THEOREM *3.4: (double commutant theorem) Let $\mathcal{S}$ be a set of operators on a finite dimensional Hilbert space $\mathcal{H}$, such that $X \in \mathcal{S} \Longrightarrow X^* \in \mathcal{S}$. Then*

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

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 $A \in \mathcal{A}$:

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

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 $A \in \mathcal{A}$:

$$\langle X\vartheta, A\psi\rangle = \langle \vartheta, X^*A\psi\rangle = 0\ ,$$

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

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

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:

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

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

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

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

$$A\psi_1 \oplus \cdots \oplus A\psi_n = B\psi_1 \oplus \cdots \oplus B\psi_n\ .$$

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

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

Proposition *3.5 If the center of $\mathcal{A}$ contains only multiples of $\mathbb{1}$, then $\mathcal{H}$ and $\mathcal{A}$ must be of the form*

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

Proposition *3.6 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 above.*

Proof. The center $\mathcal{A} \cap \mathcal{A}'$ is an abelian $*$-algebra, so Theorem 3.3 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 3.5. The statement follows. □

3.3 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

$$E^2 = E = E^* .$$

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 both be 0 or 1. So we have three possibilities.

(0) Both are 0; i.e. $E = 0$.
(1) One of them is 0 and the other is 1.
(2) Both are 1; i.e. $E = \mathbb{1}$.

In case (1), E is a one-dimensional projection satisfying

$$\mathrm{tr}\, E = 0 + 1 = 1 \text{ and } \det E = 0 \cdot 1 = 0 .$$

As $E^* = E$ and $\mathrm{tr}\, E = 1$ we may write

$$E = \tfrac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} .$$

Then $\det E = 0$ implies that

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

So the one-dimensional projections in M_2 are parametrised by the unit sphere S_2.

Notation. For $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ let us write

$$\sigma(a) := \begin{pmatrix} a_3 & a_1 - ia_2 \\ a_1 + ia_2 & -a_3 \end{pmatrix} = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3 \;,$$

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

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

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

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

Let us write

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

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

$$\varphi(A) = \mathrm{tr}\,(\rho A) \quad \text{where} \quad \rho = \rho(a) := \tfrac{1}{2}(\mathbb{1} + \sigma(a)), \quad \|a\| \leq 1 \;. \tag{3}$$

We summarise:

Proposition 1.5. *The states on M_2 are parametrised by the unit ball in $\mathbb{R}^3$, as in (3), and the one-dimensional projections in M_2 are parametrised by the unit sphere as in (2). The probability of the event $E(a)$ in the state $\rho(b)$ is given by*

$$\mathrm{tr}\,(\rho(b)E(a)) = \tfrac{1}{2}(1 + \langle a, b\rangle) \;.$$

The events $E(a)$ and $E(b)$ are compatible if and only if $a = \pm b$. Moreover we have for all $a \in S_2$:

$$E(a) + E(-a) = \mathbb{1} \;, \quad E(a)E(-a) = 0 \;.$$

Proof. Calculate. □

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 one 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.)
Particular case: the 'quantum fair coin' is modelled by $(M_2, \frac{1}{2}\mathrm{tr}\,)$.

The quantum coin toss is realised in nature: the spin direction of a particle with total spin $\frac{1}{2}$ behaves in this way.

Photons

There is a second natural way to parametrise the one-dimensional projections in M_2, which is closer to the description of polarisation of photons.
A one-dimensional projection corresponds to a (complex) line in $\mathbb{C}^2$, and such a line can be characterised by its slope, a number $z \in \mathbb{C} \cup \{\infty\}$.

Exercise. Let $f : \mathbb{C} \cup \{\infty\} \to S_2$ be given by

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

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

In particular, the projection $F(\alpha)$ on the line with real slope $\tan\alpha$ with $\alpha \in [0,\pi)$ is given by

$$F(\alpha) = \begin{pmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix} = E(\sin 2\alpha, \sin 2\alpha, \cos 2\alpha)\ . \tag{4}$$

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).

4. Operations on probability spaces

> *'The entire theory of probability is nothing but transforming variables.'*
>
> N.G. van Kampen

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.

4.1 **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. 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

$$\forall_{\omega\in\Omega} : \quad \sum_{\omega'\in\Omega'} t_{\omega\omega'} = 1 \ .$$

Examples.

1. Let τ be a bijection $\Omega \to \Omega'$. We may think of **shuffling** 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

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

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

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

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

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

   Here π is some probability distribution, which we assume to be everywhere nonzero.

It can be shown that every transition matrix can be decomposed as a product of matrices of the types 3, 1 and 2. Such a decomposition is called a *dilation* of the operation in question. See Section 4.3 for an example.

4.2 **Operations on abelian *-algebras**

All the above operations act on the *points* of Ω. In quantum probability, however, there are no such points. So in order to prepare for the introduction of quantum operations, we reformulate the above examples into operations on $*$-algebras and their duals, the spaces of probability distributions.

As before, we denote by $\mathcal{C}(\Omega)$ the $*$-algebra of complex functions on Ω. By $\mathcal{C}(\Omega)^*$ we shall mean the affine space of all probability distributions $(\pi_\omega)_{\omega\in\Omega}$ on Ω, which act on functions by the natural action

$$\pi(f) := \sum_{\omega\in\Omega} \pi(\omega) f(\omega)\ .$$

Then an operation has a *contravariant* action T on the algebra and a *covariant* action T^* on the dual as follows:

$$T : \mathcal{C}(\Omega') \to \mathcal{C}(\Omega) : (Tf')(\omega) := \sum_{\omega'\in\Omega'} t_{\omega\omega'} f'(\omega')\ ;$$
$$T^* : \mathcal{C}(\Omega)^* \to \mathcal{C}(\Omega')^* : (T^*\pi)(\omega') := \sum_{\omega\in\Omega} \pi(\omega) t_{\omega\omega'}.$$

They are related by

$$\forall_{\pi\in\mathcal{C}(\Omega)^*}\forall_{f'\in\mathcal{C}(\Omega')} : \pi(Tf') = \sum_{\omega\in\Omega}\sum_{\omega'\in\Omega'} \pi(\omega) t_{\omega\omega'} f'(\omega') = (T^*\pi)(f')\ .$$

In fact, we shall be a bit sloppy, and sometimes denote by $\mathcal{C}(\Omega)^*$ the *whole* dual space of $\mathcal{C}(\Omega)$, not just the positive normalized functons.

Now we run through the examples again. Let us call $\mathcal{C}(\Omega) : \mathcal{A}$ and $\mathcal{C}(\Omega') : \mathcal{A}'$.

1. Here T is a $*$-isomorphism $\mathcal{A}' \to \mathcal{A}$, and T^* its dual action $\mathcal{A}' \to \mathcal{A}$. Every invertible operation is a $*$-isomorphism. (Check!)
2. Let us denote the operation $\mathcal{A}' \to \mathcal{A}$ associated to a random variable X by j_X:

$$j_X(f') := f' \circ X\ .$$

Then j_X is an injective $*$-homomorphism:

$$j_X(fg) = j_X(f) j_X(g); \quad j_X(f^*) = j_X(f)^*\ .$$

Every injective $*$-homomorphism $j : \mathcal{A}' \to \mathcal{A}$ is of the form j_X for some random variable X. In quantum probability, random variables will be *defined as* $*$-homomorphisms. This is the contravariant version or the *Heisenberg* *picture* of a random variable, whereas the covariant version or the *Schrödinger picture* of a random variable is

$$j_X^* : \pi \mapsto \pi \circ X^{-1} \ .$$

Both describe the operation of *restricting* attention from ω to the values ω' of X.

3. The operation $j_X : \mathcal{A}' \to \mathcal{A}$ above has left inverses: operations $E_X^\pi : \mathcal{A} \to \mathcal{A}'$ of *conditional expectation* with respect to X and some probability distribution $\pi \in \mathcal{C}(\Omega)^*$:

$$E_X^\pi \circ j_X = \mathrm{id}_{\mathcal{A}'} \ .$$

In probability theory this conditional expectation of a function $f \in \mathcal{C}(\Omega)$ is usually denoted as $\mathbb{E}(f|X)$, where the dependence on the probability distribution π is implicit. We shall see below that a conditional expectation can be defined as a *right-invertible operation.* The dual $(E_X^\pi)^*$ is the operation of *stochastic immersion* or *state extension* of a distribution on Ω' to a distribution on the larger space Ω.

Let us now give the algebraic definition of an operation.

Definition. By an *operation* from Ω to Ω' we mean an **affine** map T^* taking probability distributions on Ω to probability distributions on Ω'. Such a map can be extended to a linear map $\mathcal{C}(\Omega)^* \to \mathcal{C}(\Omega')^*$, which we shall denote by the same name T^*. Then T^* is a positive map (i.e. $f \geq 0 \Longrightarrow Tf \geq 0$), which preserves normalisation:

$$\sum_{\omega' \in \Omega'} (T^*\pi)(\omega') = \sum_{\omega \in \Omega} \pi(\omega) \ .$$

By duality an operation T^* brings with it a positive map

$$T : \mathcal{A}' \to \mathcal{A} \quad \text{with} \quad T\mathbb{1}' = \mathbb{1} \ .$$

We usually consider T and T^* as two descriptions of the same operation.

Theorem *4.1. Let (Ω, π) and (Ω', π') be finite classical probability spaces. Suppose that $\pi(\omega) > 0$ for all $\omega \in \Omega$. Let $j : \mathcal{C}(\Omega') \to \mathcal{C}(\Omega)$ and $E : \mathcal{C}(\Omega) \to \mathcal{C}(\Omega')$ be operations such that*

$$E \circ j = \mathrm{id}_{\mathcal{C}(\Omega')} \ \textit{and} \ \pi' \circ E = \pi \ .$$

Then Ω has more points than Ω' and there is a random variable $X : \Omega \to \Omega'$ such that

$$j = j_X \ \textit{and} \ E = E_X \ .$$

We first prove a lemma.

Lemma *4.2.* (**'Abelian Schwartz'**) *For any positive $\mathbb{1}$-preserving map T : $\mathcal{C}(\Omega) \to \mathcal{C}(\Omega')$ and all $f \in \mathcal{C}(\Omega)$ we have:*

$$T(|f|^2) \geq |Tf|^2 \ .$$

Proof. For all $\lambda, \vartheta \in \mathbb{R}$, $f \in \mathcal{C}(\Omega)$:

$$0 \le T\big(|f - \lambda e^{i\vartheta} \cdot \mathbb{1}|^2\big) = T(|f|^2) - 2\lambda \mathrm{Re}\,\big(e^{-i\vartheta} Tf\big) + \lambda^2 \,.$$

So the quadratic function of λ which stands on the right can have at most one zero. Hence for all $\vartheta \in \mathbb{R}$

$$T(|f|^2) \ge \big(\mathrm{Re}\,\big(e^{i\vartheta} Tf\big)\big)^2 \,.$$

The statement follows. □

Proof of the Theorem. For all $g \in \mathcal{C}(\Omega')$:

$$|g|^2 = E \circ j(|g|^2) \ge E(|j(g)|^2) \ge |E \circ j(g)|^2 = |g|^2 \,.$$

So we have equality everywhere; it follows that

$$E\big(j(|g|^2) - |j(g)|^2\big) = 0 \,.$$

Since $\pi' \circ E = \pi$:

$$\pi\big(j(|g|^2) - |j(g)|^2\big) = 0 \,.$$

But since $j(|g|^2) \ge |j(g)|^2$, and π is strictly positive everywhere, we have

$$j(|g|^2) = |j(g)|^2 \,.$$

Now let $h_\omega := j(\delta_{\omega'})$. Then $h_{\omega'} \ge 0$ and

$$h_{\omega'}^2 = j(\delta_{\omega'})^2 = j(\delta_{\omega'}^2) = j(\delta_{\omega'}) = h_{\omega'} \,.$$

So $h_{\omega'}(\omega) = 0$ or 1 for all $\omega \in \Omega$, $\omega' \in \Omega'$. Moreover

$$\sum_{\omega' \in \Omega'} h_{\omega'} = \sum_{\omega' \in \Omega'} j(\delta_{\omega'}) = j(\mathbb{1}) = \mathbb{1} \,.$$

Hence $h_{\omega'} = 1_{S(\omega')}$ for some partition $\left\{\, S(\omega') \,\middle|\, \omega' \in \Omega' \,\right\}$ of Ω. Define

$$X : \Omega \to \Omega' : \omega \mapsto \omega' \quad \text{if} \quad \omega \in S(\omega') \,.$$

Then

$$j(\delta_{\omega'})(\omega) = 1_{S(\omega')}(\omega) = \delta_{\omega'}(X(\omega)) = j_X(\delta_{\omega'})(\omega) \,.$$

It follows that $j = j_X$.

Finally, let $(e_{\omega'\omega})$ denote the transition matrix of $E : \mathcal{C}(\Omega) \to \mathcal{C}(\Omega')$. Then we have for all $\omega', \nu' \in \Omega'$:

$$\sum_{\omega \in S(\nu')} e_{\omega'\omega} = \big(E 1_{S(\nu')}\big)(\omega') = E \circ j(\delta_{\nu'})(\omega') = \delta_{\nu'}(\omega') \,.$$

Hence $e_{\omega'\omega}$ can only be nonzero if $\omega \in S(\omega')$, i.e. if $\omega' = X(\omega)$. And if the latter is the case, then

$$\pi'(\omega') e_{\omega'\omega} = \sum_{\nu' \in \Omega'} \pi'(\nu') e_{\nu'\omega} = (\pi' \circ E)(\omega) = \pi(\omega) \,.$$

Summarising we can conclude that

$$e_{\omega'\omega} = \begin{cases} \frac{\pi(\omega)}{\pi'(\omega')} & \text{if } \omega' = X(\omega); \\ 0 & \text{otherwise.} \end{cases}$$

So $E = E_X$. (See Example 3 of Subsection 4.1.) □

4.3 A dilation

Classical operations can always be decomposed as a product of a stochastic immersion, a 'shuffling' and a restriction. We give a very simple example here. Let $\Omega := \{1,2,3\}$ and let $T : \mathcal{C}_3 \to \mathcal{C}_3$ be given by the matrix

$$\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} .$$

This operation can be decomposed as follows. Let $\Omega' := \Omega \times \{+,-\}$ and X the natural random variable

$$X : \Omega \times \{+,-\} \to \Omega : (\omega, \varepsilon) \mapsto \omega .$$

Let S be given by the cyclic permutation of $\Omega \times \{+,-\}$:

$$\begin{tikzcd}
(1,+) \arrow[r] & (2,+) \arrow[r] & (3,+) \arrow[d] \\
(1,-) \arrow[u] & (2,-) \arrow[l] & (3,-) \arrow[l]
\end{tikzcd}$$

Then we have

$$T = E_X \circ S \circ j_X ,$$

where E_X is taken with respect to the uniform distribution on $\Omega \times \{+,-\}$.

4.4 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.

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

$$\lambda f(\rho) + (1-\lambda) f(\vartheta) = f\big(\lambda\rho + (1-\lambda)\vartheta\big) .$$

This requirement is a consequence of the *stochastic equivalence principle* which states that a system which is in state ρ with probability λ and in state ϑ

with probability $1-\lambda$ can not be distinguished from a system in the state $\lambda\rho + (1-\lambda)\vartheta$.

A map f satisfying (0) 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 of some linear map $T : \mathcal{B} \to \mathcal{A}$. We shall henceforth write T^* instead of f. Of course, T^* must still map $\mathcal{A}^*_{+,1}$ to $\mathcal{B}^*_{+,1}$:

(1) $\operatorname{tr}(T^*\rho) = \operatorname{tr}(\rho)$ for all $\rho \in \mathcal{A}^*$;
If $\rho \geq 0$ then also $T^*\rho \geq 0$.

It would seem at first sight that nothing more can be said a priori about T^*. However, it was realised in the early 1970's by Karl Kraus that the positivity property has to be strengthened in quantum mechanics: if our main system is in a combined state with some other system, then after performing the operation T^* on the main 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 Chapter 1, between the main system and the second system is possible. See Example 4.3 below.
Therefore this stronger form of positivity must be added as a requirement.

(2) For all $n \in \mathbb{N}$ the map $\mathrm{id}_n \otimes T^*$ maps states on $M_n \otimes \mathcal{A}$ to states on $\mathcal{M}_n \otimes \mathcal{B}$.

Requirement (2) 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. 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 $n \in \mathbb{N}$.

Example. 4.3. *A map which is positive, but not completely positive:*
Let $\mathcal{A} := M_2$ and let

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

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

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

The matrix on the left is a projection (on the vector $(e_0 \otimes e_0 + e_1 \otimes e_1)/\sqrt{2} \in \mathbb{C}^2 \otimes \mathbb{C}^2$; compare the entangled state of Section 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, then any positive operator $T : \mathcal{A} \to \mathcal{B}$ is automatically completely positive. (We state this without proof.)

Examples. 4.4. *Some quantum operations*

1. Let $U \in M_n$ be unitary. Then the automorphism $T : M_n \to M_n : A \mapsto U^* AU$ is an operation. (See Lemma 1 below.)
2. The $*$-homomorphism $j : M_k \to M_l \otimes M_k : A \mapsto \mathbb{1} \otimes A$ is an operation. (See Lemma 1 below.)
3. 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 Section 4.1.

Lemma 4.5. *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 $T(A^*) = T(A)^*$, then T is completely positive.*

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

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

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

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

□

LEMMA *4.6. Let $\mathcal{A} \subset M_k$, $\mathcal{B} \subset M_l$ and let V be a linear map $\mathbb{C}^l \to \mathbb{C}^k$. Then*

$$T : \mathcal{A} \to \mathcal{B} : A \mapsto V^* A V$$

is completely positive.

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

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

□

Lemma 2 covers Example 3 of 4.4. since φ can be decomposed into pure states as $\varphi = \sum_i \langle \psi, \cdot \psi \rangle$ and

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

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

The following important theorem, together with Proposition 3.6, characterizes all completely positive maps on finite dimensional $*$-algebras.

THEOREM *4.7. (Stinespring 1955). Let T be a linear map $M_k \to M_l$. The following are equivalent.*
(i) T is completely positive;
(ii) There exist $m \in \mathbb{N}$ and operators $V_1, \ldots V_m : \mathbb{C}^l \to \mathbb{C}^k$ such that for all $A \in M_k$:

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

Moreover, 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

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

where $(u_{ij})_{i,j=1}^m$ *is a unitary* $m \times m$*-matrix of complex numbers.*

We begin with some preparatory work.

Proposition 4.8. *Let* φ *be a state on* $\mathcal{A} := M_k$*. Then any two decompositions of* φ *into* m *pure states*

$$\varphi(A) = \sum_{i=1}^m \langle \psi_i, A\psi_i \rangle = \sum_{i=1}^m \langle \vartheta_i, A\vartheta_i \rangle \tag{5}$$

with linearly independent vectors $\psi_1, \ldots, \psi_m$*, are connected by a transformation of the form*

$$\vartheta_i = \sum_{j=1}^m u_{ij}\psi_j \ .$$

where $(u_{ij})_{i,j=1}^m \in M_m$*.*

Proof. Consider $\psi := (\psi_1, \ldots, \psi_m)$ and $\vartheta := (\vartheta_1, \ldots, \vartheta_m)$ as vectors in $\mathcal{H} := (\mathbb{C}^k)^m = \mathbb{C}^m \otimes \mathbb{C}^k$. Then (5) can be written in the form

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

(So we see that any state can be written as a pure state on some representation of $\mathcal{A}$!) Now, since the vectors $\psi_1, \ldots, \psi_m$ are independent, for any m-tuple $(\chi_1, \ldots, \chi_m) \in \mathcal{H}$ there exists a matrix $A \in \mathcal{A}$ such that for all i we have $A\psi_i = \chi_i$. In other words: $\mathcal{H} = (\mathbb{1} \otimes \mathcal{A})\psi$. Now define $U : \mathcal{H} \to \mathcal{H}$ by:

$$U(\mathbb{1} \otimes A)\psi := (\mathbb{1} \otimes A)\vartheta \ .$$

Then U is isometric, since

$$\begin{aligned} \| U(\mathbb{1} \otimes A)\psi \|^2 &= \| (\mathbb{1} \otimes A)\vartheta \|^2 = \langle (\mathbb{1} \otimes A)\vartheta, (\mathbb{1} \otimes A)\vartheta \rangle = \langle \vartheta, (\mathbb{1} \otimes A^*A)\vartheta \rangle \\ &= \varphi(A^*A) = \| (\mathbb{1} \otimes A)\psi \|^2 \ , \end{aligned}$$

and since U maps $\mathcal{H}$ into $\mathcal{H}$ itself, it must be unitary. What is more, $U \in M_m \otimes M_k$ is actually of the form $u \otimes \mathbb{1}$ for some unitary $u \in M_m$. Indeed, for all $\chi = (\mathbb{1} \otimes X)\psi \in \mathcal{H}$ and all $A \in \mathcal{A}$ we have

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

and therefore $U \in (\mathbb{1} \otimes \mathcal{A})' = M_m \otimes \mathbb{1}$. The statement follows. □

We shall now introduce some useful notation.
Consider the tensor product $\mathcal{H}_{kl} := \mathbb{C}^k \otimes (\mathbb{C}^l)'$ of the Hilbert space $\mathbb{C}^k$ and the dual of $\mathbb{C}^l$. $\mathcal{H}_{kl}$ can be viewed as the space of all operators $\mathbb{C}^l \to \mathbb{C}^k$, but also as a Hilbert space (with natural inner product), on which the algebra $M_k \otimes M_l^T$ can act, in the following way:

$$A \otimes B : \psi \otimes \overline{\vartheta} \mapsto A\psi \otimes B\overline{\vartheta} \quad \left[= A|\psi\rangle \otimes \langle\vartheta|B^T\right] .$$

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

$$\Omega := \sum_{i=1}^{l} e_i \otimes \overline{e_i} \quad \left[\sum_{i=1}^{l} |e_i\rangle \otimes \langle e_i|\right] ,$$

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

$$\begin{aligned}\langle\Omega, (A \otimes B)\Omega\rangle_{\mathcal{H}} &= \sum_{i=1}^{l}\sum_{j=1}^{l} \langle e_i \otimes \overline{e_i}, (A \otimes B)e_j \otimes \overline{e_j}\rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{l}\sum_{j=1}^{l} \langle e_i, Ae_j\rangle\langle\overline{e_i}, B\overline{e_j}\rangle \\ &= \sum_{i=1}^{l}\sum_{j=1}^{l} \langle e_i, Ae_j\rangle\langle e_j, B^T e_i\rangle \\ &= \mathrm{tr}\,(AB^T) = \mathrm{tr}\,(A^T B) ,\end{aligned} \qquad (6)$$

where $\cdot^T$ denotes transposition.

Proof of Stinespring's Theorem. The implication (ii)$\Longrightarrow$(i) follows immediately from Lemma 4.2. For the converse, assume that $T : M_k \to M_l$ is completely positive. Let $\mathcal{H}_{ll} := \mathbb{C}^l \otimes (\mathbb{C}^l)'$ as above, and let ω denote the state

$$\omega(X) := \langle\Omega, X\Omega\rangle$$

on $\mathcal{B}(\mathcal{H}_{ll}) \sim M_l \otimes M_l$. Since T is completely positive, the functional ω_T on $\mathcal{B}(\mathcal{H}_{kl})$ given by

$$\omega_T := (T^* \otimes \mathrm{id}\,)(\omega)$$

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

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

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

$$\begin{aligned}\sum_{i=1}^m \langle \psi, V_i^* A V_i \vartheta\rangle &= \sum_{i=1}^m \langle V_i\psi, A V_i \vartheta\rangle \\ &= \sum_{i=1}^m \langle v_i, \big(A \otimes (\overline{\psi} \otimes \vartheta)\big) v_i\rangle_{\mathcal{H}} \\ &= \omega_T\big(A \otimes (\overline{\psi} \otimes \vartheta)\big) \\ &= \omega\big(T(A) \otimes (\overline{\psi} \otimes \vartheta)\big) \\ &= \langle \Omega, T(A) \otimes (\overline{\psi} \otimes \vartheta)\Omega\rangle \\ &= \mathrm{tr}\,\big(T(A)(\vartheta \otimes \overline{\psi})\big) \\ &= \langle \psi, T(A)\vartheta\rangle \ .\end{aligned}$$

□

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 *4.9. (Cauchy-Schwarz 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 $A \in \mathcal{A}$:*

$$T(A^*A) \geq T(A)^*T(A) \ .$$

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

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

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

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

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

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

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

□

THEOREM *4.10. (Multiplication Theorem) If $T : \mathcal{A} \to \mathcal{B}$ is an operation and $T(A^*A) = T(A)^*T(A)$ for some $A \in \mathcal{A}$, then $T(A^*B) = T(A)^*T(B)$ and $T(B^*A) = T(B)^*T(A)$ for all $B \in \mathcal{A}$.*

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

$$T\big((A^* + \lambda B^*)(A + \lambda B)\big) = T(A)^*T(A) + \lambda T(A^*B + B^*A) + \lambda^2 T(B^*B) ,$$

while by Cauchy-Schwartz

$$\begin{aligned} &T\big((A^* + \lambda B^*)(A + \lambda B)\big) \\ &\qquad\qquad \geq T(A)^*T(A) + \lambda(T(A)^*T(B) + T(B)^*T(A)) + \lambda^2 T(B)^*T(B)) . \end{aligned}$$

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

$$T(A^*B + B^*A) \geq T(A)^*T(B) + T(B)^*T(A) .$$

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 $T(A^*B) = T(A)^*T(B)$ and $T(B^*A) = T(B)^*T(A)$.

□

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

THEOREM *4.11. (Embedding theorem) Let $(\mathcal{A}, \varphi)$ and $(\mathcal{B}, \psi)$ be nondegenerate quantum probability spaces, and let $j : \mathcal{A} \to \mathcal{B}$, $E : \mathcal{B} \to \mathcal{A}$ be operations which preserve the states. If*

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

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

$$P(C_1 B C_2) = C_1 P(B) C_2 \tag{7}$$

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

Following the language used in Section 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 4.1.

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

$$A^*A = E \circ j(A^*A) \geq E(j(A)^*j(A)) \geq E \circ j(A)^* E \circ j(A) = A^*A , \tag{8}$$

so we have equalities here. In particular

$$\psi\big(j(A^*A) - j(A)^*j(A)\big) = \varphi \circ E\big(j(A^*A) - j(A)^*j(A)\big) = 0 ,$$

and as $(\mathcal{B}, \psi)$ is non-degenerate, $j(A^*A) = j(A)^*j(A)$, i.e. j is a *-homomorphism. j is injective since it has the left-inverse E.
But also from (8) we have $E\bigl(j(A)^*j(A)\bigr) = E \circ j(A)^* E \circ j(A)$. The Multiplication Theorem 4.10 then implies that for all $B \in \mathcal{B}$ and $A_1 \in \mathcal{A}$,

$$E(j(A_1)^*B) = E \circ j(A_1)^* E(B) = A_1^* E(B) \ ,$$

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

$$E\bigl(j(A_1)^* B j(A_2)\bigr) = E\bigl(j(A_1)^*B\bigr) E \circ j(A_2) = A_1^* E(B) A_2 \ .$$

Applying j to both sides we find (7).

□

5. 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.

5.1 **'No cloning'**

'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. In a picture:

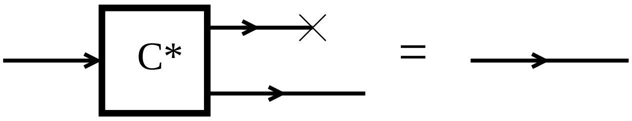

FIG. 6

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

$$(\mathrm{tr} \otimes \mathrm{id}) \circ C^*(\rho) = (\mathrm{id} \otimes \mathrm{tr}) \circ C^*(\rho) = \rho \, . \tag{9}$$

Reformulated in the Heisenberg picture: for all $A \in \mathcal{A}$:

$$C(\mathbb{1} \otimes A) = C(A \otimes \mathbb{1}) = A \, . \tag{10}$$

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 γ be the 'copying' map $\Omega \to \Omega \times \Omega : \omega \mapsto (\omega, \omega)$. This map induces an operation

$$C: \quad \mathcal{C}(\Omega) \times \mathcal{C}(\Omega) \to \mathcal{C}(\Omega): \quad Cf(\omega) := f \circ \gamma(\omega) = f(\omega, \omega) \, .$$

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

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

and the same holds for $C(f \otimes \mathbb{1})$, so (10) is satisfied. In the Schrödinger picture our operation looks as follows:

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

and we see that (9) is satisfied:

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

The following theorem says that this construction is only possible in the abelian case.

Theorem 5.1. *('No cloning') Let $C : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ be an operation. If equation (10) holds for all $A \in \mathcal{A}$, then $\mathcal{A}$ is abelian.*

Proof. (10) implies that for all $A \in \mathcal{A}$:

$$C\big((\mathbb{1} \otimes A)^*(\mathbb{1} \otimes A)\big) = C(\mathbb{1} \otimes A^*A) = A^*A = C(\mathbb{1} \otimes A)^* C(\mathbb{1} \otimes A)$$

Then it follows from the multiplication theorem that for all $A, B \in \mathcal{A}$:

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

□

5.2 **'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:

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

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

THEOREM *5.2. 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 $A \in \mathcal{A}$:

$$\begin{aligned} &A^*A = C \circ D(A^*A) \geq C\bigl(D(A)^*D(A)\bigr) \geq A^*A \\ \text{and} \quad &AA^* = C \circ D(AA^*) \geq C\bigl(D(A)D(A)^*\bigr) \geq AA^* \ , \end{aligned}$$

so that we again have equality everywhere. If $\mathcal{B}$ is abelian, we have $D(A)^*D(A) = D(A)D(A)^*$, so that $A^*A = AA^*$. □

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

5.3 **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 realise 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. $\mathcal{A} \cap \mathcal{A}' = \mathbb{C}\mathbb{1}$), and if we throw away (disregard) this information, then it can not be avoided that some initial states are altered.*

Let us work towards 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

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

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:

$$(\mathrm{id} \otimes \mathrm{tr})M^*(\rho) = \rho \qquad \forall_{\rho \in \mathcal{A}^*} .$$

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

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

where ϑ does not depend on ρ.

In a picture:

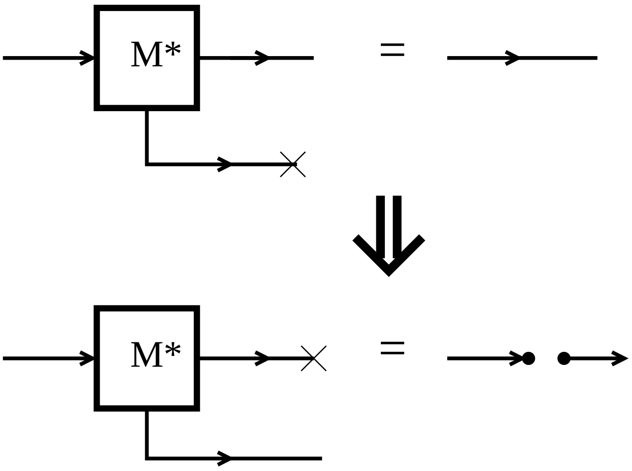

FIG. 7

We again formulate and prove the theorem in the contravariant picture:

THEOREM 5.3. *(Heisenberg's Principle) Let M be an operation $\mathcal{A} \otimes \mathcal{B} \to \mathcal{A}$ such that for all $A \in \mathcal{A}$,*

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

then

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

In particular, if $\mathcal{A}$ is a factor, then $B \mapsto M(\mathbb{1} \otimes B) = \vartheta(B) \cdot \mathbb{1}_{\mathcal{A}}$ for some state ϑ on $\mathcal{B}$.

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

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

But also,

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

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 $\mathbb{C} \cdot \mathbb{1}_{\mathcal{A}}$, i.e. a state on B times $\mathbb{1}_{\mathcal{A}}$. □

5.4 **Random variables and von Neumann measurements**

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

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

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

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

An important case is when $\mathcal{B} = \mathcal{C}(\Omega)$ 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

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

and for $i \neq j$,

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

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

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

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

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

which completely determines j.

Proposition 5.4. *Let $\mathcal{A}$ be a finite-dimensional *-algebra with unit. Then there is a one-to-one corrensponcence between injective *-homomorphisms $j : \mathcal{C}(\Omega) \to \mathcal{A}$ for some $\Omega \subset \mathbb{R}$ and self-adjoint operators $X \in \mathcal{A}$, given by*

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

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

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

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 $p:\mathbb{C}\to\mathbb{C}$ denote the polynomial

$$p(x) := (x-x_1)\cdots(x-x_n)\ .$$

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

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

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

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

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

$$\sum_{i=1}^{n} P_i = 1\!\!1 \quad \text{and} \quad P_iP_k = \delta_{ik}P_k\ .$$

Hence, if we define

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

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

5.5 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 $\mathrm{sp}(X)$, g on $\mathrm{sp}(Y)$:

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

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

A contravariant formulation of these facts is

$$\begin{aligned} M(f \otimes \mathbb{1}) &= f(X) \ ; \\ M(\mathbb{1} \otimes g) &= g(Y) \ . \end{aligned}$$

Theorem 5.5. *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 that on $\mathrm{sp}(Y)$. We apply the multiplication theorem on the measurement operation M, which is supposed to exist. Since

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

we have

$$M\big((x \otimes \mathbb{1})^*(\mathbb{1} \otimes y)\big) = M(x \otimes \mathbb{1})^* M(\mathbb{1} \otimes y) = XY$$

and

$$M\big((\mathbb{1} \otimes y)^*(x \otimes \mathbb{1})\big) = M(\mathbb{1} \otimes y)^* M(x \otimes \mathbb{1}) = YX \ .$$

As $(x \otimes \mathbb{1})^*(\mathbb{1} \otimes y) = x \otimes y = (\mathbb{1} \otimes y)^*(x \otimes \mathbb{1})$, we have $XY = YX$. □

6. Quantum novelties

In the previous chapter 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 later chapters, such as very fast computation and secure cryptography. Here we shall treat 'teleportation' of quantum states and 'superdense coding'.

6.1 Teleportation of quantum states

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

In Section 5.2 ('No classical coding') we have seen that, without any further tools, this is impossible. If Alice were to perform measurements on the qubit, and tell 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 Section 1.3, taking one qubit each home 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 5.1 ('no cloning'). It is for this reason that the procedure is called 'teleportation'.

We illustrate the procedure in a picture.

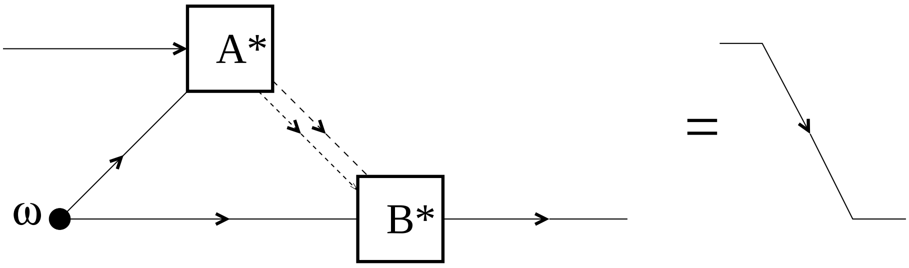

FIG. 8

Here ω is the fully entangled state $X \mapsto \langle \Omega, X\Omega \rangle$ on $M_2 \otimes M_2$ (see the proof of Stinespring's Theorem in Section 4.4).

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*

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

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

The operation performed by Alice has the contravariant description:

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

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

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

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

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

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

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

where $\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, 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 *6.1. The state ω and the operations A and B described above satisfy*

$$(\text{id}_{M_2} \otimes \omega) \circ (A \otimes \text{id}_{M_2}) \circ B = \text{id}_{M_2} \ .$$

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

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

□

In the field of quantum information and computation people are not only interested in finding completely positive maps that perform the tasks set, but are also looking for explicit ways to implement them. In fact teleportation has been carried out succesfully 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:

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

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 *controlled* NOT gate $M_2 \otimes M_2 \longrightarrow M_2 \otimes M_2$ which sandwiches a matrix with

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} .$$

The operation C performs a NOT operation on the first qubit provided that the second is a 1. In diagrams:

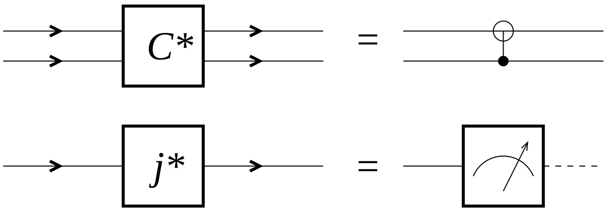

FIG. 9

Check that, using the above building blocks, the procedure of quantum teleportation can be charted as follows:

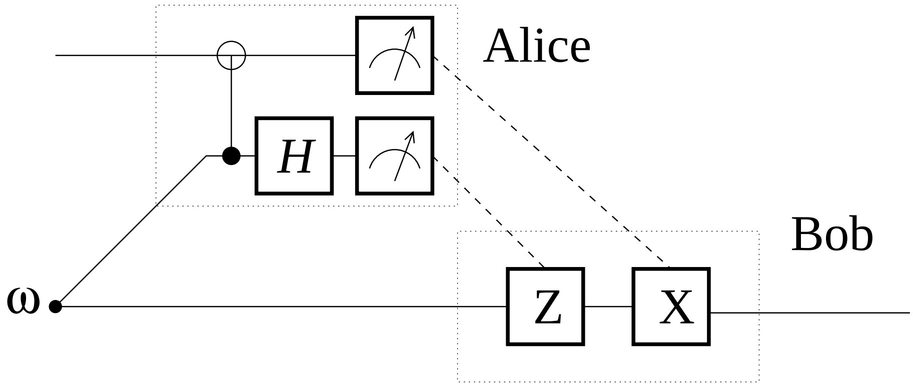

FIG. 10

6.2 Superdense 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-entangled qubit pair.
(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!)

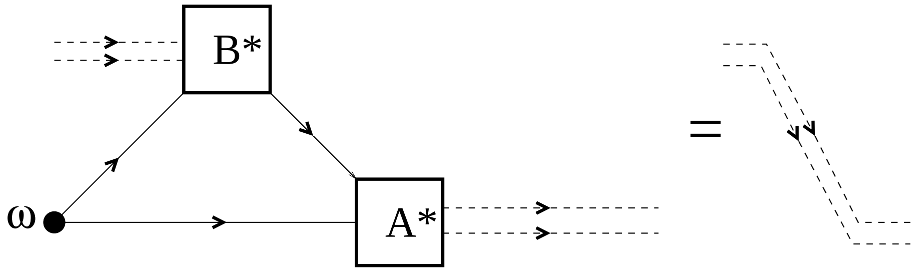

FIG. 11

PROPOSITION 6.2. *Taking ω, A and B as in Proposition 6.1, we have*

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

We leave the proof as an exercise.

Computers, as we know them, are machines that operate on bits. A single (finitary) algorithm or program can be characterised as a function $f : \{0,1\}^p \to \{0,1\}^q$, where p is the length of the input bitstring, and q that of the output. Such a function defines an *operation* (in fact a *-homomorphism) $C : (\mathcal{C}_2)^{\otimes q} \to (\mathcal{C}_2)^{\otimes p}$ given by $C(a) := a \circ f$.
By analogy, we define a *quantum algorithm* as an operation on qubits: a completely positive identity preserving map $Q : (M_2)^{\otimes q} \to (M_2)^{\otimes p}$.

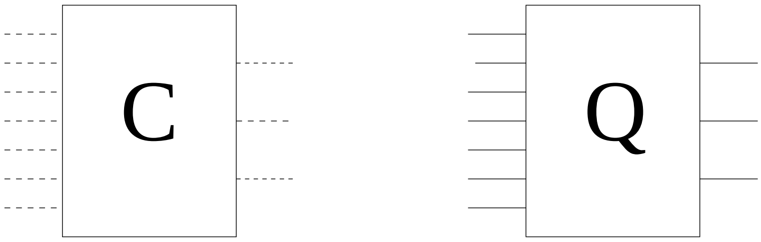

FIG. 12: *A classical and a quantum algorithm*

Since a classical bit can be transformed onto a qubit by embedding, and a qubit can be collapsed to a classical bit by measurement, it follows from this definition that quantum computing is an extension of classical computing: every classical algorith can be performed on a quantum computer.
On the other hand, on the assumption that all operations are physically realizable, it has been shown that some computations can be done in an essentially better way on a quantum computer. The most important motivating example is Shor's algorithm for the factorization of large integers into primes, a feat which potentially threatens present day RSA cryptography.

However, up to now quantum algorithms prove extremely difficult to implement physically. Entangled qubit states, which are the essential new ingredient of quantum computing, are highly vulnerable to decay by interactions with the environment (*'decoherence'*), and for many operations only quite complicated realizations seem to be available. On these grounds, the feasability of quantum computers is still a matter of controversy.

In this chapter we briefly survey some relevant aspects of classical computation, discuss the 'toolbox' of quantum computation, and treat a few of the promising possibilities.

7.1 **Gates**

In classical computational practice algorithms are composed of a large number of standardized operations, which act on just a few bits at a time, and are executed by computer hardware called *gates.* It is known from computer science that only a handful of different gates are needed in order to realize any classical algorithm. For instance NOT-gates, AND-gates, OR-gates, and exlusive OR-gates, (sometimes called XOR-gates), are sufficient, together with the operation of *copying* of bits.

Example. The following classical circuit performs the addition of two two-bit integers: $x_1x_0 + y_1y_0 = z_2z_1z_0$.

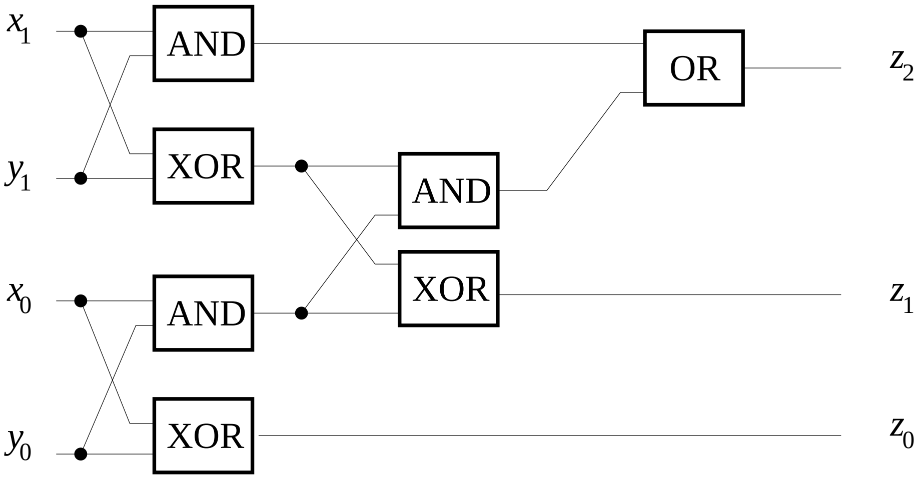

FIG. 13: *An addition circuit*

In the quantum case, 1-qubit operations are divided into *-automorph- isms or *unitary operations* $A \mapsto U^*AU$, where $U \in M_2$ is unitary, and operations which don't preserve pure states, such as *measurement*

$$j : \mathcal{C}_2 \longrightarrow M_2 : (\alpha, \delta) \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$$

and *decoherence*

$$D : M_2 \longrightarrow M_2 : \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha & \varepsilon\beta \\ \overline{\varepsilon}\gamma & \delta \end{pmatrix}, \quad \text{with} \quad |\varepsilon| < 1 .$$

Important unitary operations are

$$X : A \mapsto \sigma_1 A \sigma_1, \quad Y : A \mapsto \sigma_2 A \sigma_2, \quad Z : A \mapsto \sigma_3 A \sigma_3 ;$$

The unitary operators in M_2 are all of the form

$$U = e^{i\alpha} R_a(\|a\|) := \exp(i(\alpha \cdot \mathbb{1} + \sigma(a))),$$
$$\text{with} \quad \sigma(a) := a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3, \quad (a \in \mathbb{R}^3) .$$

The operation $U^* \cdot U$ does not feel the phase α and acts on the Bloch sphere of Section 2.2 as a rotation over an angle $\|a\|$ around the axis $\mathbb{R}a$.

Exercises.

1. Find values for α, ϑ and the unit vector u such that $e^{i\alpha}R_\vartheta(u)$ equals the *phase gate* $S := \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ and the '$\pi/8$ gate' or 'T gate' $T := \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{i} \end{pmatrix}$ respectively.
2. Find values for α, ϑ and u such that $e^{i\alpha}R_\vartheta(u)$ becomes the *Hadamard matrix* $H := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Remark. We shall abuse the letters X, Y, Z, H, S, T by sometimes reading them as the unitary matrices given above, sometimes as the associated operations. This has the slight advantage that diagrams in which these letters occur can be read in terms of vector states in the Schrödinger picture, reading from left to right, or as operations in the Heisenberg picture, reading from right to left. For their action on density matrices in the Schrödinger picture stars have to be added.

The most important two-bit operation used in quantum computing is the controlled NOT gate which we met in Section 6.1. More generally we shall use the the 'controlled U' gate,

$$C_U(M) := \begin{pmatrix} \mathbb{1} & 0 \\ 0 & U^* \end{pmatrix} M \begin{pmatrix} \mathbb{1} & 0 \\ 0 & U \end{pmatrix} , \qquad (M \in M_2 \otimes M_2) .$$

symbolically denoted by the circuit

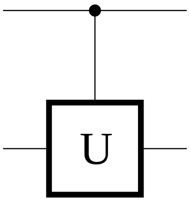

FIG. 14: *Controlled U gate*

LEMMA 7.1. *Every controlled U gate on two qubits can be realised as a combination of two controlled* NOT *gates and some single qubit operations.*

Proof. Write U in terms of Euler angles:

$$U = e^{i\alpha} R_3(\beta) R_2(\gamma) R_3(\delta) .$$

Then introduce the matrices

$$\begin{aligned} A &:= R_3(\beta)R_2\left(\frac{\gamma}{2}\right) \ ; \\ B &:= R_2\left(-\frac{\gamma}{2}\right)R_3\left(-\frac{\delta+\beta}{2}\right) \ ; \\ C &:= R_3\left(\frac{\delta-\beta}{2}\right) \ . \end{aligned}$$

Clearly, $ABC = \mathbb{1}$; but since the operation $X \cdot X$ flips the matrices Y and Z, we have

$$XBX = R_2\left(\frac{\gamma}{2}\right)R_3\left(\frac{\delta+\beta}{2}\right) \ ,$$

so that $e^{i\alpha}AXBXC = U$. This means that

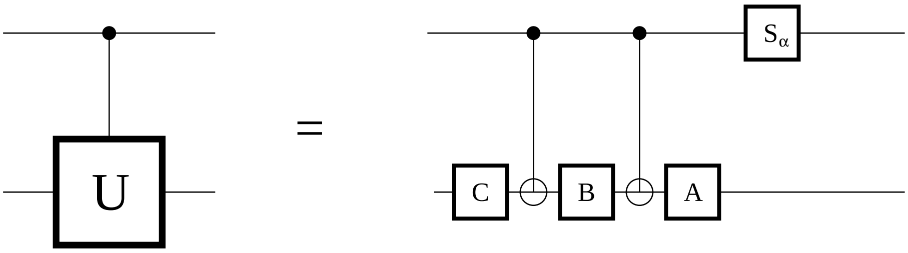

FIG. 15

□

Remark. In quantum computing the distinction between the controlling and the controlled qubit in a 'controlled' operation is unclear. For example we have

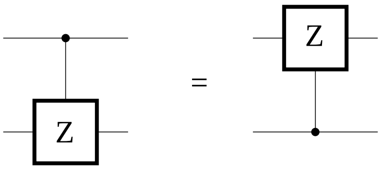

FIG. 16

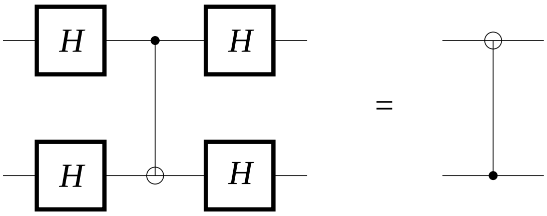

FIG. 17

Exercise. Check this!

7.2 The quantum Fourier transform

Let N be a natural number, and consider the Hilbert space $\mathbb{C}^N$. Let F_N denote the discrete Fourier transform on this space, given by

$$(F_N\psi)(j) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(\frac{2\pi i}{N} jk\right) \psi(k) \; . \tag{11}$$

F_N is a unitary transformation of $\mathbb{C}^N$ and hence defines an operation (a *-automorphism) $F_N^* \cdot F_N$ on the algebra M_N.

Now suppose that N is a power of 2: $N = 2^n$. Then standard binary notation identifies a natural number $j < N$ with a bit sequence $j_{n-1}j_{n-2}\cdots j_1 j_0$ such that

$$j = \sum_{l=0}^{n-1} j_l \, 2^l \; .$$

The space $\mathbb{C}^N$ can be read as the n-fold tensor power of $\mathbb{C}^2$, with the canonical basis

$$e_j = e_{j_{n-1}} \otimes e_{j_{n-2}} \otimes \cdots e_{j_1} \otimes e_{j_0}, \quad (j = 0, \ldots N-1),$$

and F_N becomes an n-qubit operation, which is known as the *Quantum Fourier Transform (QFT).*

PROPOSITION *7.2. The Quantum Fourier Transform maps the product vector $e_j = e_{j_{n-1}} \otimes e_{j_{n-2}} \otimes \cdots e_{j_1} \otimes e_{j_0}$ to the product vector*

$$\begin{aligned} &\frac{1}{\sqrt{2^n}} \big(e_0 + \exp(2\pi i(0.j_0))e_1\big) \otimes \big(e_0 + \exp(2\pi i(0.j_1 j_0))e_1\big) \\ &\qquad\qquad \otimes \cdots \otimes \big(e_0 + \exp(2\pi i(0.j_{n-1}\cdots j_1 j_0))e_1\big) \; , \end{aligned} \tag{12}$$

where $0, j_q \cdots j_1 j_0$ denotes the binary fraction $\sum_{m=0}^{q} j_m 2^{m-q-1}$.

Proof. We calculate

$$\begin{aligned} F_N e_j &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp\left(\frac{2\pi i}{N} k \cdot j\right) e_k \\ &= \frac{1}{\sqrt{2^n}} \sum_{k_0=0}^{1} \cdots \sum_{k_{n-1}=0}^{1} \bigotimes_{l=0}^{n-1} \exp\left(\frac{2\pi i}{2^n} k_l 2^l \cdot j\right) e_{k_l} \\ &= \bigotimes_{l=0}^{n-1} \left(e_0 + \exp\left(2\pi i \frac{j}{2^{n-l}}\right) e_1\right) / \sqrt{2} \\ &= \bigotimes_{l=0}^{n-1} \left(e_0 + \exp\left(2\pi i \frac{j_{n-l-1}\cdots j_1 j_0}{2^{n-l}}\right) e_1\right) / \sqrt{2} \end{aligned}$$

The tensor product should be written from right to left as the running index increases. □

We note that the factors in the tensor product depend on less bits when we move from right to left (with increasing l). This observation enables us to draw a quantum circuit for F_N consisting exclusively of one-bit and two-bit gates.

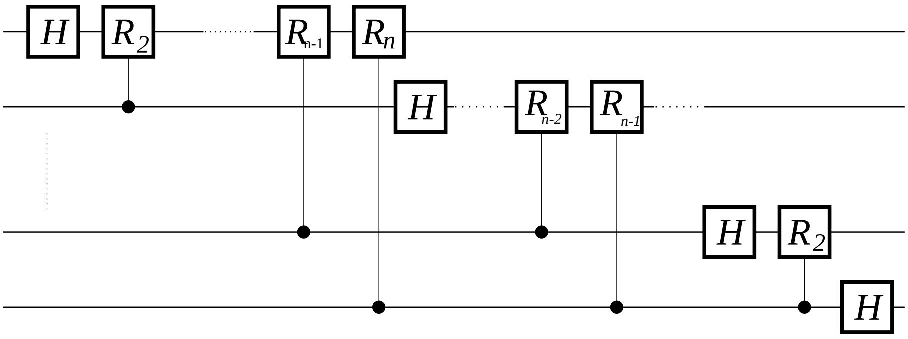

FIG. 18

In order to calculuate the Fourier transform of a vector consisting of 2^n complex numbers, a classical computer which naively evaluates (11) needs $O(2^{2n})$ arithmetic operations. By employing (12) the classical computer can reduce this to $O(n \cdot 2^n)$ operations. This classical algorithm is known as the *Fast Fourier Transform (FFT)* and is widely used in standard computer software.
However, inspection of the diagram in fig. 18 shows that our future quantum computer needs only $\frac{1}{2}n(n+1)$ gates! This is an astounding reduction! Is this really the case? How can we understand this?

Let us compare the argument with the following classical example.
The probability distribution of n coins in a row is given by 2^n positive numbers $\pi_k = \pi_{k_{n-1}k_{n-2}\ldots k_1 k_0}$ with $0 \leq k < 2^n$. Now by tossing the last coin and putting it back in its place, we change its distribution into $T^*\pi$, given by

$$(T^*\pi)_{k_{n-1}k_{n-2}\ldots k_1 k_0} = \tfrac{1}{2}\big(\pi_{k_{n-1}k_{n-2}\ldots k_1 k_0} + \pi_{k_{n-1}k_{n-2}\ldots k_1 (1-k_0)}\big) \; .$$

So our single coin tossing operation T 'performs a calculation' which otherwise would cost us 2^{n-1} additons! Why has no one ever used sequences of coins as an ultrafast classical computer?

The absurdity of this example shows a serious weakness of the Quantum Fourier Transform as performed by the diagram of Fig. 18.: we lack a readout procedure! If we were to just look at the coins (or the qubits in Fig. 18., for that matter), we would see only one bit sequence, without even knowing what its probability was. An accurate measurement of the probabilities of all the possible sequences would involve repeating the experiment many, many times, thus throwing away the whole advantage, and much more. The QFT circuit can only be useful on a single use if it returns some qubit sequence with probability 1. This points at the important difference between the QFT and the ridiculous coin example: the QFT outputs a pure state for every pure input. So there always exists some von Neumann measurement to which it gives a full answer with probability 1. In the coin example this is not the case. The next section shows how to make use of that property.

7.3 Phase estimation

Suppose that we have some quantum circuit which performs a unitary operation with matrix U. Suppose that U has an eigenvalue $u = e^{2\pi i\varphi}$ with eigenvector ψ, which we are able to prepare. Can we build a circuit out of these elements, plus some more quantum gates, that outputs a binary representation of φ?
The answer is 'yes', and the circuit is built as follows. We start with a part that looks like Fig. 19.

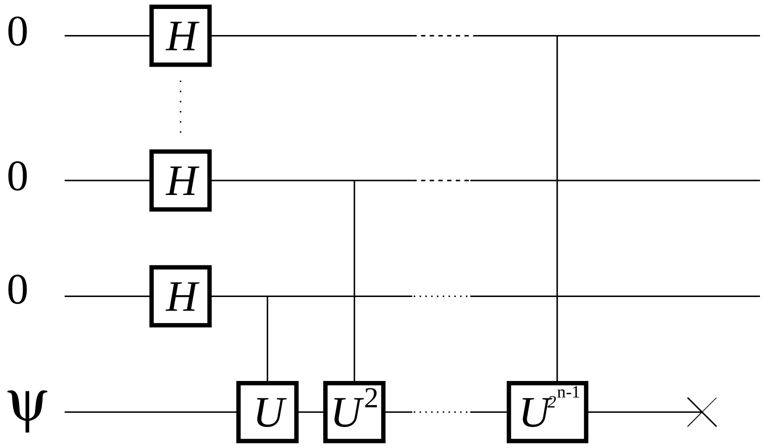

FIG. 19

The result of this first part is the vector

$$\bigotimes_{l=0}^{n-1} \left(e_0 + e^{2\pi i\cdot 2^l \varphi} e_1 \right) / \sqrt{2} = \frac{1}{\sqrt{N}} \sum_{k=0}^{2^n-1} e^{2\pi i k\varphi} e_k \ .$$

Now, if φ is given by the n-bit number $\varphi = 2^{-n}j$ with $j \in \{0, 1, \ldots, 2^n - 1\}$, then this is just the vector $F_{2^n} e_j$. So we only have to connect our circuit to an (inverse) quantum Fourier transformer, in order to obtain $e_j = e_{j_0} \otimes \cdots \otimes e_{j_{n-1}}$. A simple von Neumann measurement on each of the n output channels then reveals the bits j_0, j_1, $\ldots$, j_{n-1} of j, so in fact of the phase φ.

If φ is not a multiple of 2^{-n}, then the algorithm can still be used, yielding n accurate bits of φ with probability $1-\varepsilon$, provided that we use a number t of qubits given by

$$t = n + \left\lceil \log\left(2 + \frac{1}{2\varepsilon}\right)\right\rceil \ .$$

(See [NiC] Section 5.2.1.)

7.4 **Finding the order**

Let N be a (large) integer number between 2^{L-1} and 2^L, and let x be a smaller integer, coprime with N. Then the map

$$u_x : \{0,\dots,N-1\} \to \{0,\dots,N-1\} : y \mapsto xy \ (\mathrm{mod}\, N)$$

is injective and hence invertible. It follows that, if we start with 1, and apply this map repeatedly, we must return to 1 after some number r of iterations. This cycle length is called the *order of x modulo N*.

Let U_x denote the unitary operator $\mathbb{C}^N \to \mathbb{C}^N$ associated to the permutation u_x, and let $\mathcal{H}_x \subset \mathbb{C}^N$ be the r-dimensional subspace spanned by the $U_x^k e_1$. Then the spectrum of the restriction of U_x to $\mathcal{H}_x$ consists of the r-th roots of unity, each having multiplicity 1. This establishes a connection between the order r of x and the eigenvalues of U_x, which permits us to find r using the quantum algorithm of phase estimation treated in §7.3. This program has several aspects to it, which we shall treat in succession.

A. Building controlled U_x-gates out of our knowledge of x and N;

B. finding a suitable input vector ψ for the phase algorithm;

C. distilling r from the measured phase.

A. *Modular exponentiation.* We wish to implement by quantum gates the map

$$e_z \otimes e_y \mapsto e_z \otimes e_{x^z y} \ ,$$

where we count modulo N in the indices. This can be done using unitary extensions of classical gates, put together according to do the binary multiplication $y \mapsto xy$, followed by Euclid's algorithm for the remainder modulo N.

B. *Choosing an input vector.* Ideally we would like to have an eigenvector ψ of U_x to use as an input, but this requires knowledge of the order r itself, which is out of the question. As our second best we can take e_1 as an input, since this vector is proportional to the sum of the eigenvectors of U_x in $\mathcal{H}_x$:

$$e_1 = \frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} \psi_s \ ,$$

where ψ_s, for $s \in \{0, \ldots, r-1\}$ given by

$$\psi_s := \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{\frac{2\pi i}{r} sk} e_{x^k} \ ,$$

is an eigenvector:

$$U_x \psi_s = e^{\frac{2\pi i}{r} s} \psi_s \ .$$

Since the calculations under A are done in a unitary way, preserving linear combinations, the result of the phase estimation is a binary expression for

$$\varphi = \frac{s}{r} \ ,$$

where s is chosen at random from the set $\{0, 1, \ldots, r-1\}$.

C. *Continued fraction expansion.* The final part is a purely classical calculation. Suppose we obtain a $2L$-bit approximation φ' to $\varphi = s/r$, i.e., φ' is a multiple of 2^{-L} satisfying $|\varphi - \varphi'| < 2^{-L}$. Then $\varphi = s/r$ occurs as a term in the continued fraction expansion of φ'. (See [NiC].) This gives us explicit integer values for s' and r' satisfying $s'/r' = s/r$. If we are lucky, which happens a reasonable fraction of the time, then s and r are coprime, so that r' is the actual order of x. This is easily tested by calcululating x^r in the classical way.

7.5 Factorization

The reduction of factoring to order-finding proceeds in two basic steps. The first step is to show that we can compute a factor of N if we can find a non-trivial solution $x \neq \pm 1 (\operatorname{mod} N)$ to the equation $x^2 = 1 (\operatorname{mod} N)$. The second step is to show that a randomly chosen y coprime to N is quite likely to have an order r which is even, and such that $y^{r/2} \neq \pm 1 \ (\operatorname{mod} N)$, thus yielding $x = y^{r/2}$ satisfying the above.
These two steps are embodies in the following two theorems. We shall only prove the easy one, referring the reading to the literature for the other.

Theorem 7.3. *Suppose N is an L bit composite number, and $x \in \{2, \ldots, N-2\}$ is a solution to the equation $x^2 = 1 (\operatorname{mod} N)$. Then at least one of* $\gcd(x-1, N)$ *and* $\gcd(x+1, N)$ *is a non-trivial factor of N, and can be computed using $O(L^3)$ operations.*

Proof. Since $x^2 = 1 (\operatorname{mod} N)$, it must be that N divides $x^2 - 1 = (x+1)(x-1)$, and thus N must have a common factor with $x+1$ or with $x-1$. By the assumption on x this can not be N itself. Using Euclid's algorithm we may compute $\gcd(x-1, N$ and $\gcd(x+1, N)$ and thus obtain a non-trivial factor of N, using $O(L^3)$ operations.

□

THEOREM *7.4. Suppose* $N = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_m^{\alpha_m}$ *is the prime factorization of an odd composite positive integer. Let* x *be an integer chosen at random, subject to the requirements that* $1 \leq x \leq N-1$ *and* x *is coprime to* N. *Let* r *be the order of* x *modulo* N. *Then*

$$\mathbb{P}[r \text{ is even and } x^{r/2} \neq -1 \ (\operatorname{mod} N)] \geq 1 - \frac{1}{2^m} \ .$$

Proof. See [NiC]. □

8. PUBLIC KEY CRYPTOGRAPHY: THE RSA CRYPTOSYSTEM

In 1978 R.L. Rivest, A. Shamir and L. Adleman proposed a public key cryptosystem that has become known as the *RSA system.* Although it lies outside the strict scope of Quantum Computing, being a purely classical algorithm, it is worthwhile treating it here, since its potential vulnerability to Shor's quantum algorithm is such a powerful motivation for the field of Quantum Computing as a whole.

Public key cryptography is the art of devising codes that can be used by the public to encipher messages, but enable only a few authorised persons to decipher them. It is an intriguing fact that such codes are possible at all. The RSA cryptosystem relies on the essential one-way character of multiplication of (large prime) numbers: with the help of a computer it is easy to multiply numbers of — say — 200 digits each, but it is practically impossible to disentangle these factors from their product if you don't know them, even if these factors are prime numbers, and therefore determined in principle by the product.

The motor of the RSA system is an elementary result in number theory: Fermat's 'little' theorem.

THEOREM *8.1* (**Fermat**) *Let* p *be a prime number, and* x *any positive integer, not divisible by* p. *Then*

$$x^{p-1} \equiv 1 \quad (\operatorname{mod} p) \ .$$

Proof. As p is prime, the map

$$\{0, \ldots, p-1\} \to \{0, \ldots, p-1\} : y \mapsto xy \ \ (\operatorname{mod} p)$$

is a permutation. Apart from the trivial cycle $\{0\}$ all cycles of this permutation have the same length, which is the order r of 1. Indeed if $x^k y \equiv y$, then $(x^k - 1)y \equiv 0$, and since p is prime we must have $y \equiv 0$ or $x^k \equiv 1$. Thus r must be a divisor of $p-1$, say $p - 1 = ar$, and

$$x^{p-1} = x^{ar} = (x^r)^a = 1^a = 1 \ .$$

□

Corollary *8.2. Let p and q be prime numbers, and x any positive integer, not divisible by p or by q. Then*

$$x^{(p-1)(q-1)} \equiv 1 \pmod{pq} \ .$$

Proof. By Fermat's Theorem, $x^{(p-1)(q-1)} \equiv 1$ both modulo p and modulo q, hence modulo pq. □

A second result from elementary number theory is also indispensible for an understanding of the RSA system:

Theorem *8.3* (**Euclid's algorithm**) *Given two positive integers a and b it is possible to effectively calculate* $\gcd(a,b)$ *and integers u and v such that*

$$\gcd(a,b) = ua + vb \ .$$

Proof. The algorithm goes as follows. Suppose that $a > b$. Put

$$\begin{aligned} a_0 &:= a \ ; \quad & a_1 &:= b \ ; \\ u_0 &; = 1 \ ; \quad & u_1 &:= 0 \ ; \\ v_0 &; = 0 \ ; \quad & v_1 &:= 1 \ . \end{aligned}$$

Then repeat, starting from $k := 1$ until n, where n is the first value of k for which $a_{k+1} = 0$:

$$\begin{aligned} m_k &:= [a_{k-1}/a_k] \ ; \\ a_{k+1} &:= a_{k-1} - m_k a_k \ ; \\ u_{k+1} &:= u_{k-1} - m_k u_k \ ; \\ v_{k+1} &:= v_{k-1} - m_k v_k \ . \end{aligned}$$

Then at each stage we have

$$\gcd(a_{k-1}, a_k) = \gcd(a,b) \quad \text{and} \quad a_k = u_k a + v_k b \ .$$

In particular for $k = n$. But then $a_{n-1} = m_n a_n$, so that $a_n = \gcd(a,b) = u_n a + v_n b$.

The algorithm is bound to halt since $a_0 > a_1 > a_2 > \ldots$; it is exponentially fast since $a_{k+2} \leq \frac{1}{2} a_k$ for all k. □

8.1 The RSA system

The central authority (let us call her Alice) chooses two large primes p and q, say about 200 digits long. (This is possible! There are reliable and effective stochastic primality tests for large integers.) She keeps these numbers secret, but publishes their product $N := pq$. Then she calculates the secret number

$S := (p-1)(q-1)$ and chooses an integer $1 < e < S$ such that $\gcd(e, S) = 1$. (There are plenty of such numbers, hence she can just choose a number at random and then test coprimality using Euclid's algorithm). This number e, the *encoding cipher* is also published.

If somebody from the public (let us call him Bob) wishes to send a secret message to Alice, he first encodes his message into a number m between 0 and $N-1$, and then calculates the code

$$c := m^e \pmod N ,$$

which he sends to Alice over a public channel. The point now is that, although the calculation of c from m can be done quite easily, by repeated squaring of m, and multiplying only those powers that correspond to a 1 in the binary expansion of e, it is practically impossible to find m back if c, e and N are known. No much better methods are known to do this, than to calculate the e-th power modulo N of all possible messages.

However, for Alice the situation is different. Since she knows S, and since $\gcd(S, e) = 1$, she can use Euclid's extended algorithm to calculate numbers d and b such that

$$de - bS = 1 , \qquad \text{i.e.} \quad de \equiv 1 \pmod S .$$

She then uses the *decoding cipher* d to calculate m from c:

$$c^d = m^{ed} = m^{1+bS} = m \cdot m^{b(p-1)(q-1)} \equiv m \pmod N ,$$

where in the last step Corollary 8.2 is used.

Exercise. What happens if by accident the code m is divisible by p or by q?

9. The quantum Monty Hall problem *

The well-known (classical) *Monty Hall problem* or *three-door problem* is set in the context of a television game show. It can be seen as a two person game, in which a player P tries to win a prize, but a show master (or Quiz master) Q tries to make it difficult for her. This game can be 'quantized', i.e., its key elements can be formulated in a quantum mechanical context, allowing new strategies and new solutions.

* This chapter is based on: Mauro D'Ariano, Richard Gill, Michael Keyl, Reinhard Werner, Burkhard Kümmerer, Hans Maassen: *The quantum Monty Hall problem*, Quantum Information and Computing, 2 (2002) 355-366.

9.1 **The classical Monty Hall problem**

In the last round of a television show, the candidates are given a chance to collect their prize (or lose it) in the following game:

1. Before the show the prize is hidden behind one of three closed doors. The show master knows where the prize is but, of course, the candidate does not.
2. The candidate is asked to choose one of the three doors, which is, however, not opened at this stage.
3. The show master opens another door, and shows that there is no prize behind it. (He can do this, because he knows where the prize is).
4. The candidate can now open one of the remaining doors to either collect her prize or lose.

The question is: should the candidate stick to her original choice or "change her mind" and pick the other remaining door? As a quick test usually shows, most people will stick to their first choice. After all, before the show master opened a door the two doors were equivalent, and they were not touched (nor was the prize moved). So they should still be equivalent. This argument seems so obvious that trained mathematicians and physicists fall for it almost as easily as anybody else. However, the correct solution by which the candidates can, in fact, double their chance of winning, is to always choose the other door. The quickest way to convince people of this is to compare the game with another one, in which the show master offers the choice of either staying with your choice or *opening both other doors*. Anybody would prefer that, especially if the show master courteously offers to open one of the doors for you. But this is precisely what happens in the original game when you always change to the other door.

9.2 **The quantum Monty Hall problem**

We will "quantize" only the key parts of the problem. That is, the prize and the players, as well as their publicly announced choices, will remain classical. The quantum version can even be played in a game show on classical TV.

The main quantum variable will be the position of the prize. It lies in a 3-dimensional complex Hilbert space $\mathcal{H}$, called the *game space*. We assume that an orthonormal basis is fixed for this space so that vectors can be identified by their components, but apart from this the basis has no significance for the game. A second important variable in the game is what we will call the show master's notepad. This might be classical information describing how the game space was prepared, or it might be a quantum system, entangled with the prize. In the latter case, the show master is able to do a quantum measurement on his notepad, providing him with classical information about the prize, without moving the prize, in the sense that the player's information about the prize is not changed by the mere fact that the show master "consults his notepad". A measurement on an auxiliary quantum system, even if entangled with a system of interest, does not alter the reduced state of the system of interest. After the show master has consulted his notepad, we are in the same situation as if the notepad had been a classical system all along. As in the classical game, the situation for the player might change when the show master, by opening a door, reveals to some extent what he saw in his notepad. Opening a door corresponds to a measurement along a one dimensional projection on $\mathcal{H}$.

The game proceeds in the following stages, closely analogous to the classical game:

1. Before the show the game space system is prepared quantum mechanically. Some information about this preparation is given to the show master Q. This can be in the form of another system, called the notepad, which is in a state correlated with the game space.
2. The candidate chooses some one dimensional projection p on $\mathcal{H}$.
3. The show master opens a door, i.e., he chooses a one dimensional projection q, and makes a von Neumann measurement with projections q and $(\mathbb{1}-q)$. In order to do this, he is allowed first to consult his notebook. If it is a quantum system, this means that he carries out a measurement on the notebook. The joint state of prize and notebook then change, but the traced out or reduced state of the prize does not change, as far as the player is concerned. Two rules constrain the show master's choice of q: he must choose "another door" in the sense that $q \perp p$; and he must be *certain not to reveal the prize.* The purpose of his notepad is to enable him to do this. After these steps, the game space is effectively collapsed to the two-dimensional space $(\mathbb{1} - q)\mathcal{H}$.
4. The player P can now choose a one dimensional projection p' on $(\mathbb{1} - q)\mathcal{H}$, and the corresponding measurement on the game space is carried out. If it gives "yes" she collects the prize.

As in the classical case, the question is: how should the player choose the projection p' in order to maximize her chance of winning?

From the classical case it seems likely that choosing $p' = p$ is a bad idea. So let us say that the *classical strategy* in this game consists of always switching to the orthogonal complement of the previous choice, i.e., to take $p' = \mathbb{1}-q-p$. Note that this is always a projection because, by rule 3, p and q are orthogonal one dimensional projections. We will analyze this strategy in Section ?, which turns out to be possible without any specification of how the show master can guarantee not to stumble on the prize in step 3.

There are two main ways the show master can satisfy the rules. The first is that he chooses randomly the components of a vector in $\mathcal{H}$, and prepares the game space in the corresponding pure state. He can then just take a note of his choice on a classical pad, so that in stage 3 he can compute a vector orthogonal to both the direction of the preparation and the direction chosen by the player. Q's strategies in this case are discussed in Section 9.5. The second and more interesting way is to use a quantum notepad, i.e., another system with observable algebra $\mathcal{N}$, and to prepare a "maximally entangled state" on $M_3 \otimes \mathcal{N}$. Then until stage 3 the position of the prize is completely undetermined in the strong sense only possible in quantum mechanics, but the show master can find a safe door to open on M_3 by making a suitable measurement on $\mathcal{N}$. Q's strategy in this case is discussed in Section 9.6.

9.3 **A mathematical formulation**

We now formulate the rules of the game in terms of operations on $*$-algebras.

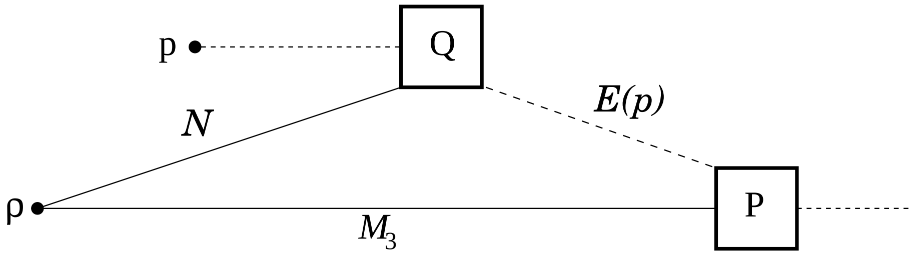

FIG. 6

1. The quiz master Q prepares the prize and his notebook in a state ρ on $\mathcal{N} \otimes M_3$, which is known to everyone, also to the player P. In this state all 'doors' should be equally probable:

$$\forall_{A \in M_3} : \quad \rho(\mathbb{1} \otimes A) = \tfrac{1}{3}\mathrm{tr}\,(A)\ . \tag{13}$$

2. P chooses p from the set $\mathcal{P}_1(M_3)$ of one-dimensional projections in M_3. We think of p as the 'north pole', and we define the 'equator' as the set

$$\mathcal{E}(p) := \left\{\, q \in \mathcal{P}_1(M_3) \,\middle|\, q \perp p \,\right\}\ .$$

3. Q performs an operation $Q : C\bigl(\mathcal{E}(p)\bigr) \to \mathcal{N}$. He must make sure not to reveal the prize:

$$\rho(Q \otimes \mathrm{id}\,)(q) = 0\ , \tag{14}$$

where q is an abbreviation for the map $\mathcal{E}(p) \to M_3 : q \mapsto q$, an element of $C(\mathcal{E}(p) \to M_3) \cong C(\mathcal{E}(p)) \otimes M_3$. Note that such an operation can only exist due to some correlation between the prize and the notebook.

4. P performs a von Neumann measurement P of a projection p', depending on q; i.e. she actually measures a projection $p' \in C(\mathcal{E}(p) \to M_3)$, taking values in $\mathcal{P}_1(M_3)$, to be considered as a projection in $C(\mathcal{E}(p)) \otimes M_3$.

The diagram makes clear that Q does not interact with the prize. We forbid this for the following reason: If measuring the prize would be allowed, then reshuffling it could not be excluded, since in quantum mechanics measurement and manipulation cannot be clearly distinguished.

The input which P is facing now, is the state $\rho\circ(Q\otimes\mathrm{id}\,)$ on $C(\mathcal{E}(p))\otimes M_3$, i.e. a stochastic projection q correlated with a stochastic state on M_3. For the sequel it will be useful to give an alternative, more picturesque description of this state. The map $\mathcal{E}(p)\to\mathbb{C}: f\mapsto\rho(Q\otimes\mathrm{id}\,)(f\otimes\mathbb{1})=\rho(Q(f)\otimes\mathbb{1})$ is a state on the C^*-algebra $C(\mathcal{E}(p))$, and hence determines a probability measure w satisfying

$$\rho(Q(f)\otimes\mathbb{1})=\int_{\mathcal{E}(p)}f(r)w(dr)\;.$$

If B is a Borel subset B of $\mathcal{E}(p)$, $w(B)$ is the probability that Q will choose an element $q\in B$. We shall write $w(B)$ also as $\rho(Q(B)\otimes\mathbb{1})$. Now let E denote an event in M_3. Then we have for all Borel subsets B of $\mathcal{E}(p)$ that $\rho(Q(B)\otimes E)\leq w(B)$, hence by the Radon-Nikodym theorem there exists a density $q\mapsto\rho_q(E)$ with the property that

$$\rho(Q(B)\otimes E)=\int_B\rho_q(E)w(dq)\;. \tag{15}$$

We see that $\rho_q(E)$ is the probability of E, given that the projection q is selected by Q. So ρ_q is the *conditional state* on the prize, given that Q points at the door q. Since the door is empty anyway, it does not matter if Q opens it or not. We shall assume he does not.

9.4 **The classical strategy**

PROPOSITION *9.1. Whatever operation Q the quiz master performs, the player can always attain a probability 2/3 to win the prize.*

Proof. Let P perform a measurement of the projection $p':=\mathbb{1}-p-q\in\mathcal{E}(p)\otimes M_3$. Then the probability for P to win is given by

$$\begin{aligned}\rho(Q\otimes\mathrm{id}\,)(p')&=\rho(Q\otimes\mathrm{id}\,)(\mathbb{1}-p-q)\overset{(14)}{=}\rho(Q\otimes\mathrm{id}\,)(\mathbb{1}\otimes(\mathbb{1}-p))\\&=\rho(Q(\mathbb{1})\otimes(\mathbb{1}-p))=\rho\big(\mathbb{1}\otimes(\mathbb{1}-p)\big)\\&\overset{(13)}{=}\tfrac{1}{3}\mathrm{tr}\,(\mathbb{1}-p)=\tfrac{2}{3}\;.\end{aligned}$$

□

9.5 **Classical notepads**

In this section we consider the case that the show master records the prepared direction of the prize on a classical notepad. We will denote the one dimensional projection of this preparation by r. Then when he has to open a door q, he needs to choose $q \perp r$ and $q \perp p$. This is always possible in a three dimensional space. But unless $p = r$, he has no choice: q is uniquely determined. This is the same as in the classical case, only that the condition "$p = r$", i.e., that the player chooses *exactly* the prize vector typically has probability zero. Hence Q's strategic options are not in the choice of q, but rather in the way he randomizes the prize positions r, i.e., in the choice of a probability measure μ on the set of pure states. This amounts to the preparation

$$\rho_\mu(f \otimes A) := \int_{\mathcal{P}_1(M_3)} f(r)\mathrm{tr}\,(rA)\mu(dr)\ .$$

It would seem that the best the player can do is to use the classical strategy, and win $2/3$ of the time. However, this turns out to be completely wrong!

Preparing along the axes

Suppose the show master decides that since the player can win as in the classical case, he might as well play classical too, and save the cost for an expensive random generator. Thus he fixes a basis with projections p_1, p_2, p_3, and chooses each one of the basis vectors with probability $1/3$:

$$\rho(f \otimes A) := \tfrac{1}{3}(f(p_1)A_{11} + f(p_2)A_{22} + f(p_3)A_{33})\ .$$

Then $\rho(\mathbb{1} \otimes A) = \frac{1}{3}\mathrm{tr}\,(A)$, and there seems to be no giveaway. In fact, the two can now play the classical version, with P choosing likewise a projection along a basis vector.

But suppose she does not, and chooses instead the projection

$$p = \tfrac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

along the vector $(1, 1, 1)/\sqrt{3}$. Then if the prize happens to be prepared in the direction $(1, 0, 0)$, the show master has no choice but to choose for q the unique projection orthogonal to these two, which is along $\chi = (0, 1, -1)$. So when Q announces his choice, P only has to look which component of the vector is zero, to *find the prize with certainty!*

This might seem to be an artifact of the rather minimalistic choice of probability distribution. But suppose that Q has settled for any *arbitrary finite collection of vectors* Ψ_α and their probabilities. Then P can choose a vector Φ which lies in none of the two dimensional subspaces spanned by two of the Ψ_α. This is possible, even with a random choice of Φ, because the union of these two dimensional subspaces has measure zero. Then, when Q announces the projection q, P will be able to reconstruct the prize vector with certainty: at most one of the Ψ_α can be orthogonal to q. Because if there were two, they would span a two dimensional subspace, and together with Φ they would span a three dimensional subspace orthogonal to q, which is a contradiction.

Of course, any choice of vectors announced with floating point precision is a choice from a finite set. Hence the last argument would seem to allow P to win with certainty in every realistic situation. However, this only works if she is permitted to ask for q at any desired precision. So by the same token (fixed length of floating point mantissa) this advantage is again destroyed.

This shows, however, where the miracle strategies come from: by announcing q, the show master has not just given the player $\log_2 3$ bits of information, but an infinite amount, coded in the digits of the components of q (or the vector χ).

Preparing real vectors

The discreteness of the probability distribution is not the key point in the previous example. In fact there is another way to economize on random generators, which proves to be just as disastrous for Q. The vectors in $\mathcal{H}$ are specified by three complex numbers. So what about choosing them real for simplicity? An overall phase does not matter anyhow, so this restriction does not seem to be very dramatic.

Here the winning strategy for P is to take $\Phi = (1, i, 0)/\sqrt{2}$, or another vector whose real and imaginary parts are linearly independent. Then the vector $\chi \perp \Phi$ announced by Q will have a similar property, and also must be orthogonal to the real prize vector. But then we can simply compute the prize vector as the outer product of real and imaginary part of χ.

For the vector Φ specified above we find that if the prize is at $\Psi = (\Psi_1, \Psi_2, \Psi_3)$, with $\Psi_k \in \mathbb{R}$, the unique vector χ orthogonal to Φ and Ψ is connected to Ψ via the transformations

$$\begin{aligned} \chi &\propto (\Psi_3,\ -i\Psi_3,\ -\Psi_1 + i\Psi_2) \\ \Psi &\propto (-\mathrm{Re}\,\chi_3,\ \mathrm{Im}\,\chi_3,\ \chi_1)\ , \end{aligned}$$

where "$\propto$" means "equal up to a factor", and it is understood that an overall phase for χ is chosen to make χ_1 real.

Uniform distribution

The previous two examples have one thing in common: the probability distribution of vectors employed by the show master is concentrated on a rather small set of pure states on $\mathcal{H}$. Clearly, if the distribution is more spread out, it is no longer possible for P to get the prize every time. Hence it is a good idea for Q to choose a distribution which is as uniform as possible. There is a natural definition of "uniform" distribution in this context, namely the unique probability distribution on the unit vectors, which is invariant under arbitrary unitary transformations. Is this a good strategy for Q?

PROPOSITION *9.2. If Q prepares the prize in a pure state chosen uniformly at random, keeping a classical note of this pure state, then P cannot raise his probability of winning above 2/3.*

In other words, the pair of strategies: "uniform distribution for Q and classical strategy for P" is an equilibrium point of the game. We do not know yet, whether this equilibrium is unique, in other words: If Q does not play precisely by the uniform distribution: can P always improve on the classical strategy? We suppose that the answer to this question is yes, to find a proof of this conjecture has turned out, however, to be a hard problem.

Remark. Although this is clearly an optimal situation for the show master, it must be noted that he needs in principle an infinitely large notebook, since he has to record in it the pure state in which the prize was prepared with infinite accuracy. Any finite notebook falls into the trap of Section 9.3.

Proof. Let μ_0 be the rotation invariant probability measure on $\mathcal{P}_1(M_3)$. The ρ_{μ_0} is rotation iinvariant in the sense that for all $f \in C(\mathcal{P}_1(M_3))$, $A, U \in M_3$, U unitary:

$$\rho(f_U, U^*AU) = \rho(f, U) \ ,$$

where $f_U(r) := f(U^*rU)$. But then, for all $p \in \mathcal{P}_1(M_3)$, all U commuting with p and $g \in C(\mathcal{E}(p))$:

$$\rho_{\mu_0}(Q \otimes \mathrm{id}\,)(g_U \otimes U^*AU) = \rho_{\mu_0}(Q \otimes \mathrm{id}\,)(g \otimes A) \ .$$

Formulating this in terms of the conditional state ρ_q we obtain that

$$\int_{\mathcal{E}(p)} g(U^*qU)\mathrm{tr}\,(\rho_q U^*AU)w(dq) = \int_{\mathcal{E}(p)} g(q)\mathrm{tr}\,(\rho_q A)w(dq) \ .$$

It follows that for all $q \in \mathcal{E}(p)$:

$$U\rho_q U^* = \rho_{UqU^*} \ . \tag{16}$$

In particular, if $Uq = qU$ we have $U\rho_q = \rho_q U$. So ρ_q commutes with every U commuting with p and q:

$$\rho_q \in \{p, q\}'', \quad \text{hence} \quad \rho_q = \alpha_q p + \beta_q q + \gamma_q(\mathbb{1} - p - q) \ .$$

Now, $\beta_q = \mathrm{tr}\,(q\rho_q) = 0$ by (14) and (15); α_q does not depend on q because of (16); $\alpha_q + \gamma_q = \mathrm{tr}\,(\rho_q) = 1$ and

$$\alpha_q = \alpha = \int \mathrm{tr}\,(p\rho_q)w(dq) = \rho_{\mu_0}(Q \otimes \mathrm{id}\,)(\mathbb{1} \otimes p) = \rho_{\mu_0}(\mathbb{1} \otimes p) = \tfrac{1}{3}\mathrm{tr}\,(p) = \tfrac{1}{3} \ .$$

We find that

$$\rho_q = \frac{1}{3}p + \frac{2}{3}(\mathbb{1} - p - q) \ .$$

Hence the classical strategy for P is clearly optimal. □

9.6 **A quantum notebook**

Finally consider the situation where the quiz master posesses a quantum 'notebook', which is an exact copy of the prize.

PROPOSITION *9.3. If Q prepares the prize and the notebook in the maximally entangled state ω of Section 4.4, then again P cannot raise his probability of winning above 2/3.*

Proof. Let $p = p_1$, p_2 and p_3 be orthogonal projections in M_3. Q is the following action: Measure p_2; if yes then point at p_3, else at p_2. So

$$Qg := g(p_2)(p_1 + p_2) + g(p_3)p_2 \ .$$

Let p' be any strategy for P. Then

$$\begin{aligned} \omega(Q \otimes \mathrm{id}\,)(p') &= \omega(p'(p_2) \otimes (p_1 + p_3) + p'(p_3) \otimes p_2) \\ &= \tfrac{1}{3}\mathrm{tr}\,\big(p'(p_2)^T(p_1 + p_3)\big) + \tfrac{1}{3}\mathrm{tr}\,\big(p'(p_3)^T p_2\big) \\ &\leq \tfrac{1}{3}\mathrm{tr}\,p'(p_2) + \tfrac{1}{3}\mathrm{tr}\,p'(p_3) = \tfrac{2}{3} \ . \end{aligned}$$

□

We note that the show master can only avoid needing an infinite amount of information by using a quantum notebook. So here is another situation where quantum information is superior to classical information.

We admit that this is partly due to the very harsh rule (14) that Q is not allowed to risk revealing the prize, which forces him to betray its location. An interesting version of the game would be to put no such restriction on Q's behaviour, but to give the prize away whenever it appears behind the door opened by Q. Classically, Q has nothing to gain in taking this risk, but in the quantum situation he may profit from this possibility!

10. Quantum Markov chains

The notion of a Markov chain has been generalised by Kümmerer outside the context of probability, to something that is meaningful in an arbitrary category. The aim of this generalisation was to motivate the definition of a 'quantum Markov chain' given by Accardi, Frigerio and Lewis in the early 1980's. As a by-product the older 'unitary dilations' of Sz.-Nagy and Foias were incorporated into the same scheme.

The main idea is, that a Markov chain is a dilation (a 'blowing up') of a semigroup of arbitrary operations to a group of invertible operations, in the same way as a dissipative evolution in physics can always be extended to a world that evolves reversibly.

10.1 **Classical Markov chains on finite state spaces**

Let Ω be a finite set, and let T be a transition probability matrix on Ω with invariant probability distribution φ:

$$T = (t_{ij})_{i,j\in\Omega}, \qquad t_{ij} \in [0,1] \; ;$$

$$\forall_{i\in\Omega} : \sum_{j\in\Omega} t_{ij} = 1, \qquad \forall_{j\in\Omega} : \sum_{i\in\Omega} \varphi_i t_{ij} = \varphi_j \; .$$

(See Section 4.1 and 4.2 for this notation.)
It is a well-known result in probability theory that such a structure (Ω, φ, T) determines a *stationary Markov chain*, i.e. a sequence of random variables

$$X_0, X_1, X_2, \ldots$$

on a probability space $(\widehat{\Omega}, \Sigma, \mathbb{P})$ such that for all $i, j \in \Omega$ and $n \in \mathbb{N}$,

$$\begin{aligned} &\mathbb{P}[X_0 = i] = \varphi_i \quad \text{and} \\ &\mathbb{P}[X_{n+1} = j | X_0 = i_0, \ldots, X_n = i_n = i] = \mathbb{P}[X_{n+1} = j | X_n = i] = t_{ij} \; . \end{aligned}$$

Now, as φ is stationary, the sequence of random variables can be extended to a two-sided infinite sequence

$$\ldots, X_{-2}, X_{-1}, X_0, x_1, X_2, \ldots$$

and we may assume that Σ is generated by these random variables. Because of this *minimality* assumption, and since the process $(X_n)_{n\in\mathbf{Z}}$ is stationary, there exists a map $\tau : \widehat{\Omega} \to \widehat{\Omega}$ connecting the random variables:

$$X_{n+1} = X_n \circ \tau \; .$$

A concrete choice for $\widehat{\Omega}$ is $\Omega^{\mathbf{Z}}$, the space of all infinite paths through Ω, and in this case τ is simply the left shift:

$$\tau(\omega)_n = \omega_{n+1} \; .$$

In the above familiar situation we now distinguish two 'layers', which become more clear if we formulate matters in terms of algebras, as we have done all the time in these notes.
Let $\mathcal{A} := C(\Omega)$. Then the triple $(\mathcal{A}, \varphi, T)$ forms the 'upper layer'.
In the 'lower layer', let $\widehat{\mathcal{A}}$ denote the algebra $L^\infty(\widehat{\Omega}, \Sigma, \mathbb{P})$, generated by functions of the form $(X_n)_{n\in\mathbf{Z}}$ because of our minimality assumption. Let us consider the left shift as an operator on the functions on $\widehat{\Omega}$:

$$\widehat{T} : \widehat{\mathcal{A}} \to \widehat{\mathcal{A}} : \quad \big(\widehat{T}g\big)(\omega) := g\big(\tau(\omega)\big) \; .$$

We have now arranged our notation in such a way that $(\mathcal{A}, \varphi, T)$ and $(\widehat{\mathcal{A}}, \widehat{\varphi} := \mathbb{P}, \widehat{T})$ are very similar structures; we call them both (stochastic) *dynamical systems.* An important difference is, however, that $\widehat{T}$ is invertible, and $(\widehat{T}^n)_{n\in\mathbb{Z}}$ is a *group* of automorphisms of $(\widehat{\mathcal{A}}, \widehat{\varphi})$, whereas $(T^n)_{n\in\mathbb{N}}$ is an arbitrary *semigroup* of operations on $(\mathcal{A}, \varphi)$.
Moreover, $(\mathcal{A}, \varphi, T)$ can be *embedded* into $(\widehat{\mathcal{A}}, \widehat{\varphi}, \widehat{T})$ by the map

$$j : \mathcal{A} \to \widehat{\mathcal{A}} : \quad f \mapsto f \circ X_0 \; .$$

(In fact there is such an embedding $j_n : f \mapsto f \circ X_n$ for all $n \in \mathbb{Z}$, which can be written as $\widehat{T}^n \circ j$.)
Associated to j there is a conditional expectation (see Section 4.2)

$$E : \widehat{\mathcal{A}} \to \mathcal{A} : \quad g \mapsto \mathbb{E}(g|X_0) \; ,$$

which is the left-inverse of $j : E \circ j = \mathrm{id}_{\mathcal{A}}$.
Now, the dynamical systems $(\mathcal{A}, \varphi, T)$ and $(\widehat{\mathcal{A}}, \widehat{\varphi}, \widehat{T})$ are related by

$$E\big(f(X_n)\big|X_0\big) : i_0 \mapsto \sum_{i_0\in\Omega} \sum_{i_1\in\Omega} \cdots \sum_{i_n\in\Omega} t_{i_0 i_1} \cdots t_{i_{n-1} i_n} f(i_n) = (T^n f)\,(i_0) \; ,$$

for all $n \in \mathbb{N}$. In other words, since $f(X_n) = f \circ X_0 \circ \tau^n = \widehat{T}^n(j(f))$:

$$T(f) = E \circ \widehat{T}^n \circ j(f) \; .$$

In brief, we say that $(\widehat{\mathcal{A}}, \widehat{\varphi}, \widehat{T})$ with embedding j is a dilation of $(\mathcal{A}, \varphi, T)$ if the following diagram commutes:

$$\begin{tikzcd}
(\mathcal{A}, \varphi) \arrow[r, "T^n"] \arrow[d, "j"'] & (\mathcal{A}, \varphi) \\
(\widehat{\mathcal{A}}, \widehat{\varphi}) \arrow[r, "\widehat{T}^n"] & (\widehat{\mathcal{A}}, \widehat{\varphi}) \arrow[u, "E"']
\end{tikzcd} \quad , \qquad (n \geq 0) \; .$$

where $\widehat{T}$ is an automorphism.

10.2 **Excursion into categories**

A *category* is a class of *objects* between which *morphisms* or simply *arrows* are defined in such a way that the existence of arrows $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$ implies the existence of an arrow $g \circ f : \mathcal{X} \to \mathcal{Z}$, and that for all objects $\mathcal{Y}$ there is a special arrow, called $\mathrm{id}_{\mathcal{Y}}$, with the property that for any $f : \mathcal{X} \to \mathcal{Y}$ we have $f \circ \mathrm{id}_{\mathcal{X}} = f$ and for any $g : \mathcal{Y} \to \mathcal{Z}$ we have that $\mathrm{id}_{\mathcal{Y}} \circ g = g$. The operation $\circ$ of composition of arrows must be associative: $f \circ (g \circ h) = (f \circ g) \circ h$.

A morphism $f : \mathcal{X} \to \mathcal{Y}$ is called a *section* if there exists a morphism $g : \mathcal{Y} \to \mathcal{X}$ (called a *left inverse* of f) with the property that $g \circ f = \mathrm{id}_{\mathcal{X}}$. It is called a *retraction* if there exists $h : \mathcal{Y} \to \mathcal{X}$ (called a *right inverse* of f), such that $f \circ h = \mathrm{id}_{\mathcal{Y}}$. If such g and h both exist, they must be equal, and are called the *inverse* of f. In that case f itself is an *isomorphism* of $\mathcal{X}$ and $\mathcal{Y}$. If $\mathcal{X} = \mathcal{Y}$, then f is called an *automorphism* of $\mathcal{X}$.

With this terminology we may now define our Markov chains.

Definition. Let T be a morphism $\mathcal{X} \to \mathcal{X}$. By a *dilation* of T we mean an automorphism $\widehat{T}$ of some object $\widehat{\mathcal{X}}$, together with a section $J : \mathcal{X} \to \widehat{\mathcal{X}}$ with left-inverse $P : \widehat{\mathcal{X}} \to \mathcal{X}$ such that the following diagram commutes for all $n \in \mathbb{N}$.

$$\begin{tikzcd}
\mathcal{X} \arrow[r, "T^n"] \arrow[d, "J"'] & \mathcal{X} \\
\widehat{\mathcal{X}} \arrow[r, "\widehat{T}^n"] & \widehat{\mathcal{X}} \arrow[u, "P"']
\end{tikzcd}$$

The dilation is called *minimal* if for all objects $\mathcal{Y}$ and all arrows $f, g : \widehat{\mathcal{X}} \to \mathcal{Y}$:

$$\left(\forall_{n\in\mathbb{N}} : f \circ \widehat{T}^n \circ J = g \circ \widehat{T}^n \circ J\right) \quad \Longrightarrow \quad f = g \; .$$

It is called a *Markov* dilation if there exists an object $\mathcal{V}$ (the 'past' or 'Vergangenheit') and a retraction $P_{(-\infty,0]} : \widehat{\mathcal{X}} \to \mathcal{V}$ with right-inverse $J_{(-\infty,0]}$ such that

$$J_{(-\infty,0]} \circ P_{(-\infty,0]} \circ \widehat{T}^n \circ J = \begin{cases} J \circ T^n & \text{if } n \geq 0, \\ \widehat{T}^n \circ J & \text{if } n \leq 0. \end{cases}$$

10.3 **Hilbert spaces with contractions**

We now apply the above definitions to the category whose objects are Hilbert spaces and whose morphisms are contractions.

It is not difficult to check that in this category the sections are isometries, retractions are orthogonal projections, and isomorphisms are unitary operators. The following theorem was proved in the 1950's by Sz.-Nagy and Foias.

Theorem *8.1. Let $\mathcal{H}$ be a Hilbert space. Every contraction $C : \mathcal{H} \to \mathcal{H}$ has a unique minimal ('unitary') dilation $(\widehat{\mathcal{H}}, U; J)$:*

$$\begin{tikzcd}
\mathcal{H} \arrow[r, "C^n"] \arrow[d, "J"'] & \mathcal{H} \\
\widehat{\mathcal{H}} \arrow[r, "U^n"] & \widehat{\mathcal{H}} \, . \arrow[u, "P=J^*"']
\end{tikzcd}$$

This dilation is automatically Markovian.

Before giving a general proof, let us first consider the simplest case: the contraction $\mathbb{C} \to \mathbb{C} : z \mapsto cz$ where $-1 < c < 1$. We can realise the commutative diagram for $n = 1$ by coupling $\mathcal{H} = \mathbb{C}$ to a second copy of $\mathcal{H}$, i.e. $\widehat{\mathcal{H}} := \mathbb{C} \oplus \mathbb{C}$. So we choose the embedding $J : z \mapsto z \oplus 0$, and then rotate the main component away with

$$U_1 := \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} ,$$

where α must be chosen such that $\cos\alpha = c$.
However, if we try to use the same trick for $n = 2$ we obtain instead of $c^2 = \cos^2\alpha$:

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha \; .$$

The second unwanted term $-\sin^2\alpha$ is an 'echo', it has bounced back into the embedded space $J\mathcal{H}$ in the second step. This echo has to be repressed. This can be done by making our auxiliary space larger, and shifting away the used part. Repeating this method indefinitely, and extending to both sides, our auxiliary space becomes the Hilbert space $l^2(\mathbb{Z})$, on which the right shift operator S acts. Thus our full dilation is given by:

$$\widehat{\mathcal{H}} := \mathbb{C} \oplus l^2(\mathbb{Z}), \quad U := U_1 S, \quad J : z \mapsto z \oplus 0, \quad P := J^* \; .$$

Proof of the theorem.
Existence: The above example generalises to arbitrary contractions $C : \mathcal{H} \to \mathcal{H}$ as follows. Let $\widehat{\mathcal{H}} := \mathcal{H} \oplus \bigoplus_{n=-\infty}^{\infty} \mathcal{H}_n$, where each $\mathcal{H}_n$ is a copy of $\mathcal{H}$. Let $U_1 : \widehat{\mathcal{H}} \to \widehat{\mathcal{H}}$ leave all the spaces $\mathcal{H}_n$ with $n \neq 0$ invariant, and act on $\mathcal{H} \oplus \mathcal{H}_0$ as follows:

$$U_1 := \begin{pmatrix} C & -\sqrt{\mathbb{1} - C^*C} \\ -\sqrt{\mathbb{1} - CC^*} & C^* \end{pmatrix} .$$

Let S be the right shift on $\bigoplus_{n=-\infty}^{\infty} \mathcal{H}_n$, and let $U := U_1 S$. Then the diagram commutes for all $n \geq 0$. The above construction need not be minimal, but minimality is reached by restricting to the space generated by the images of the operators $U^n J$.
Uniqueness: The statement to be proved is that any two minimal dilations must be unitarily equivalent. So let $(\widehat{\mathcal{H}}, U; J)$ and $(\widetilde{\mathcal{H}}, \widetilde{U}; \widetilde{J})$ be minimal dilations of $(\mathcal{H}, C)$. We define a linear map $V_0 : \widehat{\mathcal{H}} \to \widetilde{\mathcal{H}}$ by

$$V\left(U^n \circ J(\psi)\right) := \widetilde{U}^n \circ \widetilde{J}(\psi) \; ; \qquad (n \in \mathbb{Z}, \psi \in \mathcal{H}); .$$

We claim that this V_0 is well-defined and isometric. Indeed, for all $m \in \mathbb{N}$, $\lambda_1 \ldots \lambda_m \in \mathbb{C}$ and $\psi_1, \ldots \psi_m \in \mathcal{H}$ we have

$$\begin{aligned} \left\| V_0 \left(\sum_{j=1}^{m} \lambda_j U^j \circ J(\psi_j) \right) \right\|^2 &= \sum_{j=1}^{m} \sum_{k=1}^{m} \overline{\lambda_j} \lambda_k \langle \widetilde{U}^j \circ \widetilde{J}(\psi_j), \widetilde{U}^k \circ \widetilde{J}(\psi_k) \rangle \\ &= \sum_{j=1}^{m} \sum_{k=1}^{m} \overline{\lambda_j} \lambda_k \langle \psi_j, \widetilde{J}^* \widetilde{U}^{k-j} \widetilde{J}(\psi_k) \rangle \; , \end{aligned} \tag{17}$$

and since

$$\widetilde{J}^*\widetilde{U}^{k-j}\widetilde{J} = \begin{cases} C^{k-j} & \text{if } k \geq j, \\ (C^*)^{j-k} & \text{if } k \leq j, \end{cases}$$

the left hand side of (17) is the same with or without V_0. Since both dilations are minimal, it follows that V_0 extends to a unitary map $V : \widehat{\mathcal{H}} \to \widetilde{\mathcal{H}}$ satisfying $VJ = \widetilde{J}V$ and $VU = \widetilde{U}V$: a unitary equivalence of the two dilations.
The Markov property: Let $\mathcal{H}_{(-\infty,0]}$ be the Hilbert subspace of $\widehat{\mathcal{H}}$ spanned by the vectors $U^{-m} \circ J\psi$ with $m \geq 0$ and $\psi \in \mathcal{H}$. Let $P_{(-\infty,0]}$ be the orthogonal projection onto $\mathcal{H}_{(-\infty,0]}$. We must show that for all $n \geq 0$ and $\psi \in \mathcal{H}$:

$$P_{(-\infty,0]}U^n J\psi = (J \circ P)U^n J\psi \, .$$

Equivalently, for all $m, n \geq 0$ and $\psi, \vartheta \in \mathcal{H}$:

$$\langle U^{-m}J\vartheta, (\mathbb{1} - J \circ P)U^n J\psi\rangle = 0 \; ;$$

i.e.,

$$\langle \vartheta, J^* \left(U^{m+n} - U^m JJ^* U^n\right) J\psi\rangle = 0 \, .$$

But since $J^*U^kJ = C^k$ for $k \geq 0$, this is just the semigroup property

$$C^{m+n} = C^m C^n \, .$$

□

10.4 Probability spaces with transition operators

Let us now consider the category whose objects are algebras $\mathcal{A}$ of bounded functions on probability spaces with their natural states φ and whose morphisms are positive linear maps T preserving expectations and the constant function 1. In this category sections are random variables, retractions are conditional expectations, and isomorphisms are given by invertible measure-preserving maps.
This is actually the category in which Markov chains were first defined, and we may formulate without further ado:

THEOREM *8.2. Every morphism $T : (\mathcal{A}, \varphi) \to (\mathcal{A}, \varphi)$ has a unique minimal Markov dilation $(\widehat{\mathcal{A}}, \widehat{\varphi}, \widehat{T}; j)$.*

This is a well-known result. Probablists usually hardly distinguish between the matrix of transition probabilities and the full Markov chain. Here, however, it is useful to indicate how in the uniqueness proof both the Markov property and the commutativity of the algebras are needed.

Proof of uniqueness. Let $(\widehat{\mathcal{A}},\widehat{\varphi},\widehat{T};j)$ and $(\widetilde{\mathcal{A}},\widetilde{\varphi},\widetilde{T};\widetilde{j})$ be minimal Markov dilations of $(\mathcal{A},\varphi,T)$. We define an algebra homomorphism V_0 by requiring that for all $n\in\mathbb{Z}$ and $f\in\mathcal{A}$:

$$V_0\left(\widehat{T}^n\circ j(f)\right)=\widetilde{T}^n\circ\widetilde{j}(f)\;.$$

We shall show that V_0 extends isometrically to all linear combinations of products of the algebra elements of the form $\widehat{T}^n\circ j(f)$. By minimality of the two dilations it will then follow that V_0 extends to an isomorphism V of $\widehat{\mathcal{A}}$ and $\widetilde{\mathcal{A}}$ which carries j over to $\widetilde{j}$ and $\widehat{T}$ to $\widetilde{T}$.

The isometric property of V_0 can be proved in the same way as in the case of Theorem 8.1: by showing that

$$\widehat{\varphi}\left(\widehat{T}^{n_1}\circ j(f_1)\cdots\widehat{T}^{n_k}\circ j(f_k)\right)\tag{18}$$

is the same for both dilations. This calculation runs as follows.

Since $\widehat{\mathcal{A}}$ is commutative, we can put the numbers $n_1,\ldots,n_k$ in (18) in increasing order. For $n\in\mathbb{N}$ let $\mathcal{A}_{(-\infty,n]}$ be generated by the functions $\widehat{T}^m\circ j(f)$ with $m\le n$ and $f\in\mathcal{A}$, and let $E_{(-\infty,n]}$ denote the associated conditional expectation. The Markov property then says that for $m\ge n$ and $f\in\mathcal{A}$ we have

$$\begin{aligned}E_{(-\infty,n]}\widehat{T}^m\circ j(f)&=E_{\{n\}}\widehat{T}^m\circ j(f)\\&=\widehat{T}^n\circ j\circ E\circ\widehat{T}^{m-n}\circ j(f)\\&=\widehat{T}^n\circ j\bigl(T^{m-n}(f)\bigr)\;.\end{aligned}$$

We can thus reduce the expectation of the product (18) to

$$\begin{aligned}&\widehat{\varphi}\circ E_{(-\infty,n_{k-1}]}\left(\widehat{T}^{n_1}\circ j(f_1)\cdots\widehat{T}^{n_{k-1}}\circ j(f_{k-1})\widehat{T}^{n_k}\circ j(f_k)\right)\\&=\widehat{\varphi}\left(\widehat{T}^{n_1}\circ j(f_1)\cdots\widehat{T}^{n_{k-1}}\circ j(f_{k-1})E_{(-\infty,n_{k-1}]}\left(\widehat{T}^{n_k}\circ j(f_k)\right)\right)\\&=\widehat{\varphi}\left(\widehat{T}^{n_1}\circ j(f_1)\cdots\widehat{T}^{n_{k-1}}\circ j(f_{k-1})\widehat{T}^{n_{k-1}}\circ j(T^{n_k-n_{k-1}}(f_k))\right)\\&=\widehat{\varphi}\left(\widehat{T}^{n_1}\circ j(f_1)\cdots\widehat{T}^{n_{k-1}}\circ j\bigl(f_{k-1}T^{n_k-n_{k-1}}(f_k)\bigr)\right)\;.\end{aligned}$$

Continuing inductively we find that the expectation (18) equals

$$\varphi\left(f_1\cdot T^{n_2-n_1}\left(f_2\cdot T^{n_3-n_2}\left(f_3\cdots T^{n_k-n_{k-1}}(f_k)\right)\right)\right)\;,$$

a quantity which indeed does not depend on the dilation. □