Lecture notes for Phys 500 – QM I

*

2014

From: Robert Raussendorf, November 26, 2014.

Abstract. These lecture notes cover material which is not in Sakurai's book. We discuss probabilistic mixtures of quantum states, density operators and their properties, tensor products of Hilbert spaces, completely positive trace-preserving maps (CPTP maps) and their Kraus normal form, entanglement, the no-cloning theorem and the impossibility of superluminal communication in quantum mechanics, and dense coding and teleportation. Furthermore, we discuss the now-famous paper by Einstein, Podolsky and Rosen, entitled "Can quantum mechanics be considered a complete description of physical reality?", local hidden variable models, and the Bell inequalities.

1 Density operators

1.1 Motivation

Consider a stream of spin-1/2 particles where each spin is pointing up (wrt the z-axis) with a probability of 1/2, and pointing down with a probability of 1/2. How do we describe this physical situation?

A suggestion one might come up with is to choose the state with same amplitudes for then spin-up and spin-down components,

$$\frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}}. \tag{1}$$

Indeed, by the Born rule, the probabilities for finding the spin pointing up or down are indeed both equal to 1/2, as required.

However, this is not enough. The suggested state must give the right expectation values and probabilities for *every possible* observable on the given physical system. So let us consider the measurement of $\hat{S}_x$. For a 50/50 mixture of $|\uparrow\rangle$ and $|\downarrow\rangle$—or any classical mixture—the rules of classical probability theory apply, in particular Bayes' rules for conditional probabilities. The probability for finding the "up" result in a measurement of $\hat{S}_x$ equals the probability of having $|\uparrow\rangle_z$ times the conditional probability of finding the "X-up" result in the measurement of $|\uparrow\rangle_z$ plus the probability of having $|\downarrow\rangle_z$ times the conditional probability of finding the "X-up" result in the measurement of $|\downarrow\rangle_z$. Thus, the probability of finding "X-up" for the given mixture should be $1/2 * 1/2 + 1/2 * 1/2 = 1/2$. However, for the state in (1) we find that the same probability should be unity. Contradiction.

We might seek to avoid this contradiction by changing the relative phase in (1) from "+" to some $e^{i\phi}$, but it will not help. While the discrepancy may cease for $\hat{S}_x$, there will always be an observable for which it persists (can you show that?).

1.2 Definition of the density operator

In preparation for resolving the puzzle from the preceding section, it is helpful to introduce the notion of the "trace" $\mathrm{Tr}\, A$ of a operator A.

Definition 1 *Be A a linear operator acting on a Hilbert space $\mathcal{H}$, and $\mathcal{B}$ an orthonormal basis of $\mathcal{H}$. Then*

$$Tr\, A := \sum_{|i\rangle \in \mathcal{B}} \langle i|A|i\rangle. \tag{2}$$

Here a few useful properties of the trace:

1. $\mathrm{Tr}\, A$ is independent of the choice of ONB in Eq. (2).
2. The trace is cyclic, $\mathrm{Tr}\, ABC = \mathrm{Tr}\, CAB$.
3. If linear operators are represented by matrices (as discussed in class), then the operator trace reduces to the ordinary trace for matrices, $\mathrm{Tr}\, M = \sum_i M_{ii}$.

Problem 1. Demonstrate the above three properties of the trace Eq. (2).

We now return to the setting of the Motivation section, or indeed a slightly more general one. Consider the ensemble $\mathcal{E}$ of quantum states,

$$\mathcal{E} = \{p_i,\, |\phi_i\rangle\}.$$

This means that states $|\phi_i\rangle$ are drawn randomly from a set, with probabilities p_i. For this ensemble and for any operator A, the expectation $\langle A\rangle_{\mathcal{E}}$ is

$$\langle A\rangle_{\mathcal{E}} = \sum_i p_i \langle\phi_i|A|\phi_i\rangle.$$

We may now manipulate this expression, as follows

$$\begin{aligned} \langle A\rangle_{\mathcal{E}} &= \textstyle\sum_i p_i \sum_j \langle\phi_i|j\rangle\langle j|A|\phi_i\rangle \\ &= \textstyle\sum_i p_i \sum_j \langle j|A|\phi_i\rangle\langle\phi_i|j\rangle \\ &= \textstyle\sum_j \langle j|A\left(\sum_i p_i|\phi_i\rangle\langle\phi_i|\right)|j\rangle \\ &= \mathrm{Tr}\, A\rho, \end{aligned}$$

where

$$\boxed{\rho := \textstyle\sum_i p_i|\phi_i\rangle\langle\phi_i|}. \tag{3}$$

Definition 2 *For an ensemble $\mathcal{E} = \{p_i,\, |\phi_i\rangle\}$, $\rho := \sum_i p_i|\phi_i\rangle\langle\phi_i|$ is the corresponding density operator.*

If a density operator has an eigenvalue of 1 then it is called *pure*. Otherwise, a density operator is called *mixed*. All pure density operators are of the form $\rho = |\psi\rangle\langle\psi|$ for some state $|\psi\rangle \in \mathcal{H}$ (can you show this?). Thus, they have an ensemble representation with just one state in it. That's the reason for calling such states pure.

Let's see what we have achieved. The just-derived expression

$$\boxed{\langle A\rangle_{\mathcal{E}} = \mathrm{Tr}\, A\rho} \tag{4}$$

for the expectation value of A is the counterpart and generalization of our earlier formula $\langle A\rangle_\psi = \langle\psi|A|\psi\rangle$. Our new expression necessitated the definition of a novel object describing the state of quantum systems, namely the *density operator*. With the new construct in hand, we can now efficiently describe classical mixtures of quantum states. This was previously cumbersome. The density operator ρ is thus a useful generation of the notion of the quantum state $|\psi\rangle$. As Eq. (4) makes clear, all the properties of a given quantum system are encoded in its density operator ρ.

The preceding discussion may prompt a question: If ρ is the fundamental notion of a quantum state rather than $|\psi\rangle$, how could we avoid discussing ρ for so long? We address this question in Section 1.3 below.

We may now specialize to the exact situation of the Motivation section. The 50/50 mixture encountered there was

$$\rho_{50/50} = \frac{1}{2}|\uparrow\rangle\langle\uparrow| + \frac{1}{2}|\downarrow\rangle\langle\downarrow|.$$

Remark. While the density operator ρ for a given ensemble $\mathcal{E}$ is unique, the converse is not true. A given density operator will in general have multiple representations as an ensemble. For example, the above density operator $\rho_{50/50}$ is also described by an ensemble $\mathcal{E}' = \{(1/2, |\leftarrow\rangle), (1/2, |\rightarrow\rangle)\}$ of spins pointing up and down in the x-direction,

$$\rho_{50/50} = \frac{1}{2}|\leftarrow\rangle\langle\leftarrow| + \frac{1}{2}|\rightarrow\rangle\langle\rightarrow|.$$

Properties of density operators. Irrespective of any further specification of the system under consideration, all density operators ρ have the following three important properties:

1. Hermiticity: $\rho^\dagger = \rho$.
2. Normalization: $\mathrm{Tr}\,\rho = 1$.
3. Non-negativity: $\langle\psi|\rho|\psi\rangle \geq 0,\ \forall\,|\psi\rangle \in \mathcal{H}$.

Problem 2. Demonstrate the above three properties, starting from Definition 2 (where all states in the ensemble are normalized to unity).

1.3 Tensor product Hilbert spaces and entanglement

Tensor product Hilbert spaces describe physical systems with a number of independent degrees of freedom, and are thus very important. As an example, consider a chain of n spin-1/2 particles. We already know that the state of each individual particle lives in a Hilbert space $\mathcal{H}_1 = \mathbb{C}^2$. But the combined system is also quantum mechanical, hence must live in a Hilbert space $\mathcal{H}_n$. But how is $\mathcal{H}_n$ constructed?

To approach this question, we choose for each of the n spins 1/2 a basis, say the local eigenbasis of $\hat{S}_z$. Every spin i may point up or down which we write as $|0\rangle_i$ and $|1\rangle_i$ respectively. The special quantum states where each spin is either pointing up or down we may then described by an n-component binary vector $\mathbf{s}$ such that the state of the i-th spin is $|s_i\rangle_i$. The state of all n spins in this configuration is $|\mathbf{s}\rangle$ which we write as

$$|\mathbf{s}\rangle = |s_1\rangle_1 \otimes |s_2\rangle_2 \otimes ... \otimes |s_n\rangle_n. \tag{5}$$

Therein, the symbol "$\otimes$" represents the tensor product between states. It signifies that the states joined by it refer to different degrees of freedom. Between the tensor product and the addition of Hilbert space vectors holds a distributive law,

$$|\alpha\rangle_1 \otimes (|\beta\rangle_2 + |\gamma\rangle_2) = |\alpha\rangle_1 \otimes |\beta\rangle_2 + |\alpha\rangle_1 \otimes |\gamma\rangle_2.$$

Now, the n-particle Hilbert space is the linear span of the basis states Eq. (5),

$$\mathcal{H}_n = \mathrm{span}\big(|\mathbf{s}\rangle,\ \mathbf{s} \in (\mathbb{Z}_2)^n\big).$$

Even if we used a special basis to define it, the n-particle Hilbert space $\mathcal{H}_n$ is a basis-independent construct (show it!), and we write

$$\mathcal{H}_n = \mathcal{H}_1^{(1)} \otimes \mathcal{H}_1^{(2)} \otimes ... \otimes \mathcal{H}_1^{(n)}.$$

Remark. The dimension of $\mathcal{H}_n$ in the example just discussed is $\dim(\mathcal{H}_n) = 2^n$. As a result, the size of the matrices representing operators acting on $\mathcal{H}_n$ grows exponentially fast in the number n of degrees of freedom. This illustrates why quantum mechanics problems are so hard to put on a computer.

Example. To make all this more concrete, lets consider the special case of two tensor factors, i.e. $\mathcal{H}_2^{(1,2)} = \mathcal{H}_1^{(1)} \otimes \mathcal{H}_1^{(2)}$, where each tensor factor is still a two-dimensional Hilbert space. There are $2 \times 2 = 4$ basis states $|\mathbf{s}\rangle$, namely $|(0,0)\rangle = |0\rangle_1 \otimes |0\rangle_2$, $|(0,1)\rangle = |0\rangle_1 \otimes |1\rangle_2$, $|(1,0)\rangle = |1\rangle_1 \otimes |0\rangle_2$, and $|(1,1)\rangle = |1\rangle_1 \otimes |1\rangle_2$. A general state $|\Psi\rangle \in \mathcal{H}_2^{(1,2)} = \mathcal{H}_1^{(1)} \otimes \mathcal{H}_1^{(2)}$ has the form

$$|\Psi\rangle_{1,2} = c_{00}\,|0\rangle_1 \otimes |0\rangle_2 + c_{01}\,|0\rangle_1 \otimes |1\rangle_2 + c_{10}\,|1\rangle_1 \otimes |0\rangle_2 + c_{11}\,|1\rangle_1 \otimes |1\rangle_2.$$

In the above, the local Hilbert space dimension of 2 was chosen only for simplicity. Tensor products can be formed between Hilbert spaces of any dimension, and the dimensions do not need to be the same in different factors. If $\dim(\mathcal{H}_A) = a$ and $\dim(\mathcal{H}_B) = b$ then $\dim(\mathcal{H}_A \otimes \mathcal{H}_B) = ab$.

Entanglement. The *basis states* $|\mathbf{s}\rangle$ in Eq. (5) by their very construction have the property of being tensor product states, $|\mathbf{s}\rangle = |s_1\rangle_1 \otimes |s_2\rangle_2 \otimes ... \otimes |s_n\rangle_n$. General states $|\psi\rangle \in \mathcal{H}_n = (\mathbb{C}^2)^n$ *will not* have this property. That is, in general

$$|\psi\rangle \neq |\alpha\rangle_1 \otimes |\beta\rangle_2 \otimes .. \otimes |\zeta\rangle_n. \tag{6}$$

A quantum state that satisfies the condition Eq. (6) for all local states $|\alpha\rangle$, $|\beta\rangle$, .. , $|\zeta\rangle$ is called *entangled*. Entanglement is a core property of quantum mechanics that sets it apart from classical physics.

A prototypical entangled state between two spin-1/2 systems A and B is

$$|\Phi\rangle = \frac{|\uparrow\rangle_A \otimes |\downarrow\rangle_B - |\downarrow\rangle_A \otimes |\uparrow\rangle_B}{\sqrt{2}}.$$

We will meet this state in the discussion of "superluminal communication", teleportation, the Bell inequalities (towards the end of this class) and angular momentum theory. In the former three scenarios we will call it a "Bell state" (after John S. Bell of the Bell inequalities, who also wrote the foreword for Sakurai's book), and in the latter a "spin singlet state".

The notion of entanglement can be extended to the more general case of density operators, as follows

Definition 3 *A quantum state (density operator) ρ shared between parties 1,2, .. , n is called separable if it can be written in the form*

$$\rho = \sum_i p_i \, \rho_1(i) \otimes \rho_2(i) \otimes .. \otimes \rho_n(i), \tag{7}$$

where all $\rho_k(i)$ are valid density operators local to the respective parties k, $p_i \geq 0$ for all i, and $\sum_i p_i = 1$. Otherwise ρ is called entangled.

Remark. Separable states can be created by local operation and classical communication (LOCC). Entangled states cannot.

1.4 Partial trace and reduced density operators.

It is useful to introduce the mathematical notion of "partial trace" for composite systems, i.e. tensor product Hilbert spaces. Consider a physical system composed of two parts, S and E,

$$\mathcal{H}_{SE} = \mathcal{H}_S \otimes \mathcal{H}_E.$$

First, in addition to the inner product $\langle \ , \ \rangle$ we already discussed, we can form inner products between states in $\mathcal{H}_E$ and $\mathcal{H}_{SE}$, namely for $|\phi\rangle \in \mathcal{H}_E$ and $|\Psi\rangle \in \mathcal{H}_{SE}$,

$$_E\langle\phi|\Psi\rangle_{SE} \in \mathcal{H}_S.$$

Note that the result of this inner product between two quantum states from different Hilbert spaces is again a quantum state, as opposed to a complex number. This inner product is first defined for tensor product states $|\Psi\rangle_{SE} = |\psi\rangle_S \otimes |\chi\rangle_E$ via

$$_E\langle\phi|\Big(|\psi\rangle_S \otimes |\chi\rangle_E\Big) = \Big(_E\langle\phi|\chi\rangle_E\Big)|\psi\rangle_S.$$

From that, the definition of the new inner product extends to all states $|\Psi\rangle \in \mathcal{H}_{SE}$ by linearity. We can now define the notion of the partial trace on a subsystem.

Definition 4 *Consider a tensor product Hilbert space $\mathcal{H}_{SE} = \mathcal{H}_S \otimes \mathcal{H}_E$, a linear operator A_{SE} acting on that space, and a basis $\mathcal{B}_E$ of $\mathcal{H}_E$. Then,*

$$Tr_E \, A := \sum_{|i\rangle \in \mathcal{B}_E} {}_E\langle i|A_{SE}|i\rangle_E.$$

We may now apply this mathematical definition to the physical object of the density operator.

Definition 5 *Consider a tensor product Hilbert space $\mathcal{H}_{SE} = \mathcal{H}_S \otimes \mathcal{H}_E$, and a quantum state ρ_{SE} living in it. Then,*

$$\rho_S := Tr_E \, \rho_{SE}$$

is the reduced density operator on S. Likewise, $\rho_E := Tr_S \, \rho_{SE}$ is the reduced density operator on E.

Problem 3. Show that if ρ_{SE} is pure then the non-zero eigenvalues of ρ_E and ρ_S and the corresponding degeneracies are the same[1].

To illustrate the usefulness of the notion of the reduced density operator, let us consider a scenario where we have full control over the system S, i.e. we can evolve S unitarily and can measure it, whereas we have no control over and access to E. In fact, that's where these subsystems got their names from, S= "system" and E="environment". In that case, all observables A we could ever measure operate on the tensor factor $\mathcal{H}_S$ only, i.e. their action on $\mathcal{H}_{SE}$ is via $A_{SE} = A_S \otimes I_E$.

For the corresponding expectation values we observe

$$\begin{aligned} \langle A \rangle_\rho &= \mathrm{Tr}\, A_{SE}\rho_{SE} \\ &= \mathrm{Tr}_S \left(\mathrm{Tr}_E (A_S \otimes I_E)\rho_{SE}\right) \\ &= \mathrm{Tr}_S A_S \left(\mathrm{Tr}_E \rho_{SE}\right) \\ &= \mathrm{Tr}_S A_S \rho_S. \end{aligned}$$

That is, the expectation of every measurement can be learned from the reduced density operator ρ_S which is a smaller object that the original density operator ρ_{SE} for the combined system. Note that we could go from the above line 2 to line 3 only because of the trivial action of $A_S \otimes I_E$ on the subsystem E.

We now return to the earlier-posed question of why we could delay discussing density operators for so long while they now seem the more fundamental and more general objects than states $|\psi\rangle$. Recall that we defined the basic laws that govern quantum mechanics, namely the Born rule and the Schrödinger equation, before even mentioning density operators.

For any given quantum state ρ_S living in a Hilbert space $\mathcal{H}_S$, a pure state $|\Psi\rangle_{SE} \in \mathcal{H}_S \otimes \mathcal{H}_E$ is called a *purification* of ρ_S if $\rho_S = Tr_E\, |\Psi\rangle_{SE}\langle\Psi|$.

Theorem 1 *Every density operator ρ has a purification.*

Proof of Theorem 1. By construction. Consider an ensemble $\mathcal{E} = \{p_i, |\psi_i\rangle\}$ representing ρ, an additional Hilbert space $\mathcal{H}_E$ of a dimension that equals the number of states in $\mathcal{E}$, and an orthonormal basis $\mathcal{B}_E$ of $\mathcal{H}_E$. Then, $\sum_{i\Big| |i\rangle \in \mathcal{B}_E} \sqrt{p_i}|\psi_i\rangle \otimes |i\rangle_E$ is a purification of ρ. □

Thus, from a certain perspective, we may say "Nothing new under the sun (here)". All mixed states can be obtained from pure states by tracing out subsystems. This justifies and clarifies why we could talk about the laws of quantum mechanics before introducing density operators. Nevertheless, density operators are useful because they allow us to discuss classical mixtures of quantum states in a simple and elegant way. There are more uses for density operators, as we will soon see.

Problem 4. Given a density operator ρ in $\mathcal{H}_S$, what is the minimal dimension of the auxiliary Hilbert space $\mathcal{H}_E$ such that a purification of ρ can live in $\mathcal{H}_S \otimes \mathcal{H}_E$? (Recall that the ensemble interpretations of any given density operator are not unique.)

[1] You may find the *Schmidt decomposition* useful. Every bipartite pure quantum state $|\Psi\rangle_{AB}$ can be written in the following form:

$$|\Psi\rangle_{AB} = \sum_i \chi_i |\psi_i\rangle_A \otimes |\phi_i\rangle_B,$$

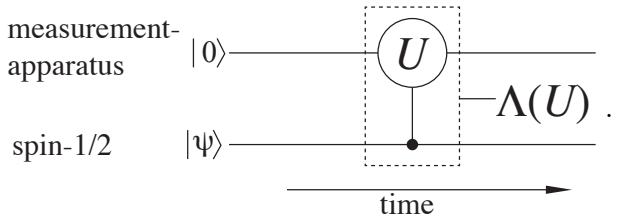
where $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ are sets of orthonormal vectors. *Proof sketch:* This is the polar decomposition of matrices in quantum mechanical guise.

2 Operations on density operators

2.1 Measurement-without-looking revisited

By considering a sequential Stern-Gerlach experiment, we previously established that in quantum mechanics, "measuring and subsequently discarding the outcome" is not the same as "not measuring". We could reason this by appeal to the Born rule, but would now supplement this earlier argument with a more intuitive explanation.

For this purpose, we introduce a different model of a measurement process, a so-called Lüders measurement. This model describes measurement by unitary interaction between a measured system and measurement device, and no state collapse ever takes place. The result of the Lüders measurement is an entangled quantum state in which the the state of the measured system, in the eigenbasis of the measured observable, is correlated with the state of the measurement device. Arguably, such a strict correlation between state of a system and state of the measurement device is all we could ever ask for, since quantum mechanics does not predict outcomes of individual measurement events (only their probability distributions). Concretely, we would depict the Lüders measurement of a spin-1/2 particle as the following quantum circuit:



Therein, the Hilbert space of the measurement system is $\mathbb{C}^2$; i.e., there are two orthogonal states, $|0\rangle$ and $|1\rangle$. The unitary U chosen such that it has the property $U|0\rangle = |1\rangle$, and the non-local operation $\Lambda(U)$ is

$$\Lambda(U) = |\uparrow\rangle_S\langle\uparrow| \otimes I_M + |\downarrow\rangle_S\langle\downarrow| \otimes U_M.$$

It is easily verified that $\Lambda(U)$ is also unitary. It has the property that

$$\Lambda(U)\,(a|\uparrow\rangle + b|\downarrow\rangle)_S \otimes |0\rangle_M = a|\uparrow\rangle_S \otimes |0\rangle_M + b|\downarrow\rangle_S \otimes |1\rangle_M.$$

Thus, the state of the spin in system S becomes strictly correlated with the state of the measurement device. This is the result of the Lüders measurement.

Now, if we commit to never looking at the measurement device, we might as well trace over the system M. The resulting state of the spin system is

$$\rho'_S = |a|^2|\uparrow\rangle_S\langle\uparrow| + |b|^2|\downarrow\rangle_S\langle\downarrow|.$$

It is helpful to look at this in matrix representation (w.r.t the S_z-eigenbasis). The whole procedure of entangling the initial state $\rho_S = |\psi\rangle\langle\psi|$ with a second system M in state $|0\rangle\langle 0|$ by the unitary $\Lambda(U)$, and subsequently tracing over M amounts to

$$\rho_S \cong \begin{pmatrix} |a|^2 & ab^* \\ a^*b & |b|^2 \end{pmatrix} \longrightarrow \rho'_S \cong \begin{pmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{pmatrix}. \tag{8}$$

Note that the off-diagonal elements have vanished. The off-diagonal elements of a density matrix contain the information about the phase relation between the basis states. Since this information is erased, this transformation of a density matrix is called a *dephasing*.

2.2 Weak measurements

2.2.1 Motivation

One problem with the formulation of measurement in quantum mechanics established so far is that measurement is either on or off. Collapse or non-collapse—that is the question. There seems no way of measuring a quantum system only "a bit". A quantum object is changed by being looked at. But then, looking out of what fraction of the corner of an eye still triggers a measurement?

We may speculate about the existence of a critical strength of interaction between measured object and measurement apparatus beyond which a wave function collapse is triggered. If such a threshold was non-zero, it should be possible to establish it as an experimental fact. However, quantum mechanics does not predict such a threshold, and its hypothetical existence would therefore indicate physics beyond quantum theory.

Let's consider the alternative, the threshold being zero. Then, wouldn't even the weakest interaction lead to instantaneous measurement? Would a particle of interstellar dust passing by the backside of the moon, by its gravitational interaction with the 47th electron of a silver atom flying through a Stern-Gerlach apparatus on earth, trigger the collapse of the silver atom's wave function? This seems absurd. No current experiment could detect the gravitational pull of such a light particle so far away. If the weakest disturbance triggers full-blown state collapse, then unimpeded unitary evolution could never occur and its consequences never be observed.

What leaves us with the above two unappealing scenarios as the seemingly only alternatives is the mismatch between the continuity of cause (interaction) and discreteness of effect (measurement). As we observed earlier, unitary evolution cannot be smoothly blended into a projective measurement, remaining a valid quantum operation all the way. There is no physical operation in-between a unitary and a projective measurement. Furthermore, the only quantum operation which is simultaneously a unitary and a projection is the identity, which is a trivial operation from both points of view.

With the novel notion of the density operator under our belt, we need to re-examine the claim just made, namely that "unitary evolution cannot be continuously deformed into measurement". In this way, we will be able to resolve the above impasse. No new postulates of quantum mechanics will be needed, just a simple trick. In fact, we have already come across it.

2.2.2 Generalized measurements

We recall the Born rule of measurement as discussed earlier in class. Re-expressed in terms of density operators, it reads

$$\text{outcome } i: \ \rho \longrightarrow \rho_i = \frac{P_i \rho P_i^\dagger}{\operatorname{Tr} P_i \rho}, \ \ \text{probability of outcome } i: \ p_i = \operatorname{Tr} P_i \rho. \tag{9}$$

Therein, all P_i are projectors, and $\sum_i P_i = I$.

We now consider a more general (and, for the moment, hypothetical) evolution, namely

$$\text{outcome } i: \ \rho \longrightarrow \rho_i = \frac{A_i \rho A_i^\dagger}{\operatorname{Tr} A_i \rho A_i^\dagger}, \ \ \text{probability of outcome } i: \ p_i = \operatorname{Tr} A_i^\dagger A_i \rho, \tag{10}$$

subject to the constraint that

$$\sum_i A_i^\dagger A_i = I. \tag{11}$$

We also consider the outcome-averaged version of the evolution in Eq. (10), i.e., the evolution resulting from forgetting or ignoring the measurement outcome,

$$\mathcal{A} : \rho \to \sum_i A_i \rho A_i^\dagger. \tag{12}$$

The evolution $\mathcal{A}$ according to Eq. (10) is a so-called generalized measurement, because it contains the projective measurement as a special case (when all A_i are projectors). The operators A_i in Eq. (10) are called *Kraus operators*. The outcome-averaged version Eq. (12) is called a completely positive trace-preserving map (CPTP-map). It has the following properties:

1. It preserves Hermiticity. If $\rho^\dagger = \rho$ then $\mathcal{A}(\rho)^\dagger = \mathcal{A}(\rho)$.
2. It preserves the trace. If $\mathrm{Tr}\rho = 1$ then $\mathrm{Tr}\,\mathcal{A}(\rho) = 1$.
3. It preserves non-negativity. If $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$ then $\langle\psi|\mathcal{A}(\rho)|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$.
4. It preserves non-negativity of composite systems. For any bipartite operator ρ_{AB}, if $\langle\psi|\rho_{AB}|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}_{AB}$ then $\langle\psi|\mathcal{A}_A \otimes I_B(\rho_{AB})|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}_{AB}$.

Maps with Property 3 are called "positive", and maps with Property 4 "completely positive".

Remark. The above Property 4 is strictly stronger than Property 3. There exist maps which are positive but not completely positive. An example is transposition. If applied to a single system, non-negativity is preserved. However, if applied to one part of a composite system, non-negativity is in general *not* preserved. This operation, the "partial transpose" $I_A \otimes T_B$, acts in the following fashion: Consider a bi-partite state $\rho_{AB} = \sum_{ijkl} p_{kl}^{ij} |i\rangle_A\langle j| \otimes |k\rangle_B\langle l|$. Then, $\rho_{AB}^{T_B} = I_A \otimes T_B(\rho_{AB}) = \sum_{ijkl} p_{kl}^{ij} |i\rangle_A\langle j| \otimes |l\rangle_B\langle k|$.

In fact, the failure of the partial transpose to preserve non-negativity can be recast as a criterion for detecting entanglement. Namely, if ρ_{AB} is separable the $\rho_{AB}^{T_B}$ is non-negative (show it!). In dimensions 2×2 and 2×3 the converse is also true. If ρ_{AB} is entangled then $\rho_{AB}^{T_B}$ has at least one negative eigenvalue. In the stated dimensions, this provides a convenient necessary and sufficient condition for the presence of entanglement [1],[2]. Note that it is in general a hard computational problem to detect entanglement by checking whether or not a given density operator can be cast in the form of Eq. (7).

We introduced generalized measurements in Eq. (10). *But do they actually exist as physical operations?* Up to this point, we have no evidence for that, except in the special case of projective measurement. The question is settled by

Theorem 2 *Be $\mathcal{M}$ a generalized measurement, $\mathcal{M}_i : \rho \longrightarrow \rho_i = \frac{A_i \rho A_i^\dagger}{Tr\, A_i \rho A_i^\dagger}$, with probability $p_i = Tr\, A_i^\dagger A_i \rho$ for obtaining the outcome i. If $\sum_i A_i^\dagger A_i = I$ then $\mathcal{M}$ can be realized in quantum mechanics.*

Before proving Theorem 2, we first look at its consequences. Our earlier assertion that "unitary evolution cannot be continuously deformed into measurement" does no longer hold in the new setting. Here is an example: Consider a generalized measurement on a 2-state system with three elements/outcomes

$$\begin{aligned} \text{outcome 1}: \quad A_1 &= cU, \\ \text{outcome 2}: \quad A_2 &= \sqrt{1-c^2}\,|\uparrow\rangle\langle\uparrow|, \\ \text{outcome 2}: \quad A_3 &= \sqrt{1-c^2}\,|\downarrow\rangle\langle\downarrow|, \end{aligned}$$

where $0 \leq c \leq 1$ is a real number, and U a unitary. The above operation satisfies the constraint Eq. (11) for all values of c, and by Theorem 2 is thus a valid generalized measurement. Further, for $c = 0$ it is a projective measurement and for $c = 1$ a unitary. We have thus found a way of continuously interpolating between a projective measurement and a unitary. For the special case of $U = I$, if $c \approx 1$, then the system is "measured only a bit".

Proof or Theorem 2: We reduce the general measurement to a unitary and a projective measurement. For this purpose, we introduce an auxiliary system M in addition to the measured system S. The joint Hilbert space is $\mathcal{H}_{SM} = \mathcal{H}_S \otimes \mathcal{H}_M$, with the dimension of $\mathcal{H}_M$ equal to the number n of Kraus operators A_i in the generalized measurement. We now show that there exists a unitary operation U_{SM} such that the procedure below amounts to the generalized measurement Eq. (10). The procedure consists of the following steps: (i) The initial state $\rho_S \otimes |0\rangle_M\langle 0|$ is put in place, with ρ the state to measure, (ii) the unitary U_{SM} is applied, (iii) the auxiliary system M is measured in the basis $\{|0\rangle, |1\rangle, ..., |n-1\rangle\}$.

The property we require of U_{SM} is that

$$ {}_M\langle k|U_{SM}|0\rangle_M = (A_k)_S, \quad \forall k = 0..n-1. \tag{13}$$

For if this holds then ${}_M\langle k|U_{SM}\left(\rho_S \otimes |0\rangle_M\langle 0|\right)U_{SM}^\dagger|k\rangle_M = \left(A_k \rho A_k^\dagger\right)_S$, and the measurement in step (iii) does indeed produce the required action on ρ.

What remains to show is that the operator U_{SM} with Property (13) can always be chosen unitary. For this, we choose a matrix representation of U_{SM} in a tensor product basis $\mathcal{B} = \{|i\rangle_S \otimes |j\rangle_M\}$, where the local basis for the factor M is the basis in which the measurement of step (iii) is made. U_{SM} can then be represented by a block matrix where the block index refers to the system M and the fine index refers to system S. Specifically, we have

$$U_{SM} \cong \begin{pmatrix} [A_0] & * & .. & * \\ [A_1] & * & & * \\ [A_2] & * & & * \\ \vdots & & & \\ [A_{n-1}] & * & & * \end{pmatrix}. \tag{14}$$

Therein, the entries "$*$" mean "to be filled in". The procedure is insensitive to the values of these entries. Now, can they be chosen such that U_{SM} is unitary? This requires all the columns of $[U_{SM}]$ to be normalized and pairwise orthogonal. For the columns that we have filled in, this is guaranteed by condition Eq. (11), $\sum_i A_i^\dagger A_i = I$. The square matrix can thus be completed to a unitary. □

By considering generalized measurements, we have achieved a continuous interpolation between unitary evolution and projective measurement, which was necessary to establish the notion of a weak measurement. We found that measurements are no longer discrete all-or-nothing affairs. Yes, the particle of interstellar dust behind the moon may measure the silver atom in a Stern-Gerlach apparatus, but only a very little bit. We can safely neglect that effect.

But taking a step back from the proof we just completed, we may again say: Nothing new under the sun! We reduced the desired interpolation between unitary evolution and projective measurement to a composition of the old extremal parts. The trick was to introduce an auxiliary system with which our primary system interacted.

Fun fact. As we have seen, whether a generalized measurement is more general than a projective measurement is a matter of perspective. But here is where generalized measurements really shine:

Consider a quantum system which is prepared in one of two fixed quantum states, namely $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, with $\langle 0|1\rangle = 0$. Is there a measurement that identifies the state, without ever reporting an erroneous outcome?

This cannot be done by a projective measurement, because $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal. However, it can be done by a generalized measurement! The catch is that there will be a third outcome, "don't know". Here is a sketch of how it works: choose $A_1 = c\,|1\rangle\langle 1|$, $A_2 = c|-\rangle\langle -|$, for a suitable $c \in \mathbb{C}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, and A_3 such that the condition Eq. (11) is satisfied, $A_3^\dagger A_3 = I - A_1^\dagger A_1 - A_2^\dagger A_2$. Then, if $|0\rangle$ is prepared, the outcome 1 can never occur. Hence, if the outcome 1 is found, the prepared state must have been $|+\rangle$. Likewise, if the outcome 2 is found then the prepared state must have been $|0\rangle$.

Problem 5. Fill in the details of the above generalized measurement that reliably distinguishes between $|0\rangle$ and $|+\rangle$. In particular, what is the minimal probability of obtaining the "don't know" outcome? (Optimize over the available strategies)

2.3 Decoherence

What we just referred to as a "trick" actually takes place around us all the time. The quantum system existing in the world all by itself, with nothing around it, is a theoretical abstraction that is sometimes well approximated but never achieved. There is always an "environment" around the system: air, electromagnetic radiation, the experimenter, a tram rattling by. This environment is in general also a quantum system, even if all the above examples, with the exception of the EM radiation (laser), have sufficient classical descriptions. The environment interacts with our system of interest, and this interaction can never be perfectly shielded off or controlled.

Let us therefore consider a bipartite system composed of the primary system S and an environment E. We may perform any generalized measurement on S, and in particular may initialize the system S in a pure state unentangled with the environment. This can be done e.g. by a projective measurement. However, we have little control over the environment. As a safe assumption, we may insist that we can neither evolve nor measure the environment altogether. W.l.o.g. we assume E in a pure state. The joint state of the quantum system at time $t = 0$ then is

$$|\Psi(0)\rangle_{SE} = |\psi\rangle_S \otimes |\zeta\rangle_E.$$

We now study the effect of a residual interaction between system and environment, $U_{SE}(t) = e^{-i/\hbar H_{SE} t}$. The time-evolved state $|\Psi(t)\rangle_S E = U_{SE}(t)|\Psi(0)\rangle_{SE}$ will, for $t > 0$, in general be entangled. Since by assumption we cannot learn about the environment, the reduced density operator on S represents all information obtainable about $|\Psi\rangle$,

$$\rho_S(t) = \mathrm{Tr}_E\, |\Psi(t)\rangle_{SE}\langle\Psi(t)|.$$

If, for some t, the joint state $|\Psi(t)\rangle$ is unentangled then $\rho_S(t)$ is pure, and if $|\Psi(t)\rangle$ is entangled then $\rho_S(t)$ is mixed. The loss of purity in the reduced state $\rho_S(t)$ caused by the quantum mechanical interaction with an environment is called *decoherence*. Decoherence is, in generic situations but not always [3], detrimental to quantum properties of the system. Typically, it drives quantum states towards the classical regime, which is characterized by the absence of entanglement within the system.

Let's discuss this for a special scenario, namely where the system S has a so-called pointer basis $\mathcal{P}$ [4], defined by the property that

$$\big[|i\rangle_S\langle i| \otimes I_E, H_{SE}\big] = 0,\ \forall |i\rangle \in \mathcal{P}.$$

The states $|i\rangle \in \mathcal{P}$ of S do not change under evolution by $U_{SE}(t)$. If the initial state is $|\Psi(0)\rangle = (\sum_i \chi_i |i\rangle_S) \otimes |\zeta\rangle_E$, then the time-evolved state $|\Psi(t)\rangle$ becomes $|\Psi(t)\rangle_{SE} = \sum_i \chi_i \, |i\rangle_S \otimes |\zeta_i(t)\rangle_E$, and the corresponding reduced density operator is

$$\rho_S(t) = \sum_{ij} \chi_i \chi_j^* \langle \zeta_j(t) | \zeta_i(t) \rangle \cdot |i\rangle_S \langle j|$$

It is useful to look at two limiting cases, namely (a) the states $|\zeta_i(t)\rangle$ are the same for all i, and (b) the states $|\zeta_i(t)\rangle$ are pairwise orthogonal. Case (a) holds, for example at time $t = 0$. ρ_S is then a rank-one projector, hence the system is in a pure state. Case (b) is typical for all times larger than a critical τ, in systems where the environment explores a large-dimensional Hilbert space. The reason is that for high-dimensional Hilbert spaces, any two random states are, with high probability, almost perfectly orthogonal. The consequence is that ρ_S becomes completely dephased, $\rho_S = \sum_i |\chi_i|^2 |i\rangle_S \langle i|$.

Complete dephasing of $\rho_S(t)$ in the pointer basis occurs exactly when the states $|\zeta_i(t)\rangle$ become orthogonal. At that point, the environment could *in principle* be measured (although this is infeasible, and we have excluded such a measurement by assumption). The measurement would, besides the state of E, also reveal the state of the system S in the pointer basis. Recall that the complete dephasing of the reduced density operator ρ_S is the effect of a measurement without observing the outcome. This is precisely what the environment does to the system.

We note that to affect a measurement (without looking at the outcome) does not require a conscious decision of any observer. It occurs simply when information of the primary system is copied to the environment, by the interaction between them. In forward reference to the no-cloning theorem we discuss in Section 3.2, note that only information with respect to the pointer basis gets copied to the environment, but no information in any complementary basis. Furthermore, we encounter again a continuous version of measurement (which turns on gradually), and correspondingly a continuous decay of coherences (i.e., off-diagonal matrix elements of ρ_S). When the inner products $\langle \zeta_j(t) | \zeta_i(t) \rangle$ are non-zero but small, the states $|\zeta_i(t)\rangle$, $\zeta_j(t)\rangle$ can be distinguished reasonably well albeit not perfectly.

Problem 6. Consider a star-like network of spins 1/2 where the central spin forms the system and the lateral spins the environment. There are N spins in the environment, and the interaction between system S and environment E is of Ising type, i.e. $H_{SE} = \sum_{k=1..N} J_k/\hbar^2 \, (\hat{S}_z)_S (\hat{S}_z)_{E,k}$. The initial state is $|+\rangle_S \otimes |+\rangle_{E,1} \otimes |+\rangle_{E,2} \otimes .. \otimes |+\rangle_{E,N}$ (all spins point in the positive x-direction).

(a) Consider the special case where there is only a single spin-1/2 in the environment, $N = 1$. Plot the von Neumann entropy $\mathrm{Tr}\, \rho_S(t) \log \rho_S(t)$ of $\rho_S(t)$ vs the time t. What is your interpretation of the plot?

(b) Now consider the of larger number N of spins system in the environment, $N = 2..10$. The coupling strengths J_k are all different, namely $J_{S,k} = J/\sqrt{k}$. Again plot the von Neuman entropy of $\rho_S(t)$ vs t. What changes as N is increased?

(c) Do the same as in (b), but with couplings $J_k = J/k$. What changes in your plot and what remains the same? Can you explain this?

3 Things that quantum mechanics can and cannot do

3.1 Motivation

The empirical Moore's law predicts that CPU processing power doubles every eighteen months. And indeed, the law describes the data since 1970 very well. But it cannot go on over ever. At some point in the not too distant future, according to Moore's law, microprocessor miniaturization will approach the atomic scale. At there very latest, the law must stop then. Quantum effects will become dominant, and will prevent circuitry from operating properly.

Or will they? Can quantum mechanics be used to operate circuits differently and perhaps more efficiently? Can quantum mechanics be the friend of the information scientist, rather than foe? These questions stand at the beginning of the field of quantum information science. People began thinking about them seriously it in the 1980ies. Below we will discuss a few of the early developments. Quantum cryptography and quantum computing will follow later in this course.

3.2 NoGo's: Cloning and superluminal communication

Let us return to the earlier mentioned Bell state, or spin singlet,

$$|\Phi\rangle = \frac{|\uparrow\rangle_A \otimes |\downarrow\rangle_B - |\downarrow\rangle_A \otimes |\uparrow\rangle_B}{\sqrt{2}}. \tag{15}$$

Suppose that this state is prepared between two parties A and B, and party A performs a measurement of its spin in the $\hat{S}_z$-basis (spin up vs down). Then, if the outcome $\uparrow$ is produced, it is clear that the spin at B has been projected into $\downarrow$. Likewise, if the measurement at A produces the outcome $\downarrow$, without any delay (says quantum mechanics) the spin at B will be projected into the $\uparrow$ state. Now imagine party A is set up on earth, and party B on the moon. The previous statements still hold. Quantum mechanics does not constrain the location of the parties in any way, and the above reasoning was independent of those locations. Does that not sound like a superluminal influence, or, in Einstein's words, "spooky action at a distance"?

We will discuss Einstein *et al.*s [5] argument later. At any rate, a sure sign of something superluminal going on in the above Gedankenexperiment would be that information (a Yes or No, an encyclopedia) could be transferred from A to B in no time. So let's think about communication. Does quantum mechanics provide us with superluminal communication?

The task was accomplished if A could choose their measurement outcome, i.e. project their system into the desired state. However, quantum mechanics does not work that way. A measurement is a projection except for the fact that the outcome is random. We cannot choose the outcome we like.

Maybe, our first approach was not quantum mechanical enough. Indeed, what we have so far described and attempted to exploit was the exact anti-correlation between two spins. But how is this different from the following phenomenology: C buys a pair of shoes, picks a random shoe and sends it to A in a parcel, and sends the other one to B. Now, when A opens their box and finds that it is a left (right) shoe, they immediately know that B got the right (left) shoe. Shoes and spins may not the same from every perspective, but from ours they are. So far, we have not exploited quantum mechanics.

But here's an interesting fact about the Bell state Eq. (15). The spins at A and B are anti-correlated in *any* direction (this follows from the spin singlet property, as we will soon discuss). That is, if A and B both measure the observable $\vec{n}\cdot\vec{\hat{S}}$ local to their parties, their respective outcomes will always be opposite, for all $\vec{n}$. This is surely a non-classical effect. So let's see if we can use it for a superluminal communication protocol. Here is a suggestion:

Protocol for the superluminal communication of one bit (patent pending[2])

1. A Bell state has been prepared between the sender A and receiver B prior to their communication.

2. If the bit to transmit has value zero, then A measures their spin in the eigenbasis of $\hat{S}_z$; if the bit has value one, A measures their spin in the eigenbasis of $\hat{S}_x$.

3. B makes a number of copies of the state it now holds.

4. By measuring the copies, B decides whether the state was $|\uparrow, z\rangle$ or $|\downarrow, z\rangle$ (transmitted bit = 0), or whether it was $|\uparrow, x\rangle$ or $|\downarrow, x\rangle$ (transmitted bit = 1).

Judging from the title of this section, there must be something wrong with this protocol.

Indeed there is. The protocol fails in step 3. Quantum states cannot be copied. This fact is established by the following

Theorem 3 ([6]) *Be $|\psi\rangle$ an unknown quantum state, and $|0\rangle$ a fixed "blank" quantum state. Then, the copying operation $C : |\psi\rangle \otimes |0\rangle \longrightarrow |\psi\rangle \otimes |\psi\rangle$ cannot be realized in quantum mechanics simultaneously for all states $|\psi\rangle$.*

Proof of Theorem 3. Assume the copying operation exists. Then, for the input states $|0\rangle$ and $|1\rangle$ it produces the output $|0\rangle\otimes|0\rangle$ and $|1\rangle\otimes|1\rangle$, respectively. Next, we consider the input $(|0\rangle+|1\rangle)/\sqrt{2}$. By linearity of quantum mechanics, the corresponding output follows from the two previous cases,

$$C\left((|0\rangle + |1\rangle)/\sqrt{2}\right) = \frac{1}{\sqrt{2}}C(|0\rangle) + \frac{1}{\sqrt{2}}C(|1\rangle) = \frac{|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle}{\sqrt{2}}.$$

However, the latter is different from $(|0\rangle + |1\rangle)/\sqrt{2} \otimes (|0\rangle + |1\rangle)/\sqrt{2}$, the expected outcome of the copying operation. Contradiction! Hence, the map C is incompatible with quantum mechanics. □

While the no-cloning theorem points to a loophole in the suggested protocol, it does not settle the question of superluminal communication because there might as well be other protocols. This is, however, not the case, as the subsequent argument [7] shows.

Party A wants to send a message $m = 1..d$ to B via a quantum protocol which consists of the following steps.

1. A and B prepare in advance the joint quantum state ρ_{AB}.

2. A applies a general CPTP map $\mathcal{A}_m \otimes I$ to encode the message m.

3. B applies a general measurement Π_r, with $\sum_r \Pi_r^\dagger \Pi_r = I$, to retrieve the message r.

Notes on Step 2: The operation $\mathcal{A}_m$ depends on the message m. Recall that a CPTP map is the outcome-averaged version of a general measurement. B cannot know the outcome of a measurement at A since by assumption the classical signal doesn't have the time to travel.

We are interested in the conditional probabilities $p(r|m)$. In particular, does the probability of obtaining a particular outcome r on the message m that was encoded? Ideal for communication was $p(r|m) = \delta_{mr}$, but can it be achieved? Can any dependence of $p(r|m)$ on m be achieved?

[2]Folklore has it that a patent application was once filed for this protocol.

We assume the following properties about the maps $\mathcal{A}_m$ and Π_r. First, the maps $\mathcal{A}_m$ are all trace-preserving, $\mathrm{Tr}\,\mathcal{A}_m(\rho) = \mathrm{Tr}\,\rho$, c.f. Property 2 of CPTP maps Eq. (12). By linearity of the trace, we can extend this property to composite systems

$$\mathrm{Tr}_A\,\mathcal{A}_m \otimes I(\rho_{AB}) = \mathrm{Tr}_A \rho_{AB}. \tag{16}$$

Physically, this means that the composite system AB does not lose or gain probability when the operation $\mathcal{A}_m$ is applied to one of its parts. For the receiving side B, we assume that the map Π_r is linear,

$$\mathrm{Tr}_A\,\mathcal{A}_m \otimes \Pi_r(\rho_{AB}) = \Pi_r \mathrm{Tr}_A\,\mathcal{A}_m \otimes I(\rho_{AB}). \tag{17}$$

Then, the probability for outcome r given the message m is

$$\begin{aligned} p(r|m) &= \mathrm{Tr}\, I \otimes \Pi_r(\mathcal{A}_m \otimes I(\rho_{AB})) \\ &= \mathrm{Tr}_B\,\Pi_r(\mathrm{Tr}_A(\mathcal{A}_m \otimes I(\rho_{AB})) \\ &= \mathrm{Tr}_B\,\Pi_r(\mathrm{Tr}_A\,\rho_{AB}) \equiv p(r). \end{aligned}$$

Therein, the transition from the first to the second line proceeds by linearity of Π_r, c.f. Eq. (17), and from the second to the third line by preservation of trace Eq. (16).

We thus find that the probability of retrieving out come r at party B is completely independent of the message m. No information is transmitted superluminally. At the level we have probed it here, quantum mechanics is compatible with the theory of relativity.

3.3 Quantum protocols: Dense coding and quantum teleportation

When thinking about quantum mechanical ways of transmitting and processing information, we should perhaps start by choosing an elementary portion, or unit, of quantum information. That is, we are looking for a quantum mechanical counterpart of the classical bit. We usually take this unit to be a quantum bit, or *qubit* for short, which is the general state of a quantum system living in a two-dimensional Hilbert space,

$$|\psi\rangle = \cos\alpha|0\rangle + e^{\mathrm{i}\varphi}\sin\alpha|1\rangle.$$

The qubit is a marriage between standard (classical) information theory and quantum mechanics. From the classical bit it inherits the basis states $|0\rangle$ and $|1\rangle$, and from quantum mechanics the superposition principle.

Now that we have defined a new unit of information, we might want to relate it to the long-known notion of the bit. So, how much classical information fits into a qubit? This is not a straightforward question. We may make the following two observations: (i) Two real numbers $\alpha \in [0, \pi)$ and $\varphi \in [0, 2\pi)$ are required to specify the state of a qubit. Informally speaking, this is an infinite number of bits. However, matters of accuracy and distinguishability come into play. If the angle φ is off by some small amount, we need a (larger) number of copies of the state $|\psi\rangle$ to find out about that. Which brings us to the question of how much information can be extracted from one qubit. (ii) Since the qubit lives in a two-dimensional Hilbert space, every non-trivial projective measurement is composed of rank-1 projectors. There are two outcomes of every such measurement, and any given qubit can only be subjected to one such measurement. Thus, one classical bit of information can be extracted from a projective measurement of one qubit. (For more general measurements, see [8].)

Based on these observations, we may form two hypotheses: (a) A general qubit cannot be faithfully transmitted using a finite number of bits; and (b) A qubit cannot transmit more than a single classical bit. Both hypotheses turn out to be wrong. The key is entanglement.

Dense coding. Let's first turn to the question of how many classical bits can be transmitted by sending one qubit. The answer turns out to be 2 [9]. The proof is constructive–we present a protocol that achieves this rate. To start, we observe that the Bell state Eq. (15) has three cousins who are just as entangled, and with which it forms an orthonormal basis $\mathcal{B}$,

$$\left.\begin{aligned} |\Phi_{00}\rangle &= \tfrac{|0\rangle_A\otimes|0\rangle_B+|1\rangle_A\otimes|1\rangle_B}{\sqrt{2}}, \\ |\Phi_{01}\rangle &= \tfrac{|0\rangle_A\otimes|0\rangle_B-|1\rangle_A\otimes|1\rangle_B}{\sqrt{2}}, \\ |\Phi_{10}\rangle &= \tfrac{|0\rangle_A\otimes|1\rangle_B+|1\rangle_A\otimes|0\rangle_B}{\sqrt{2}}, \\ |\Phi\rangle = |\Phi_{11}\rangle &= \tfrac{|0\rangle_A\otimes|1\rangle_B-|1\rangle_A\otimes|0\rangle_B}{\sqrt{2}}. \end{aligned}\right\} \mathcal{B} \tag{18}$$

We introduce $\sigma_i = \frac{2}{\hbar}\hat{S}_i$, where $i \in \{x, y, z\}$, and note the property

$$\sigma_z|\phi_{00}\rangle = |\Phi_{01}\rangle,\ \sigma_x|\phi_{00}\rangle = |\Phi_{10}\rangle,\ \sigma_y|\phi_{00}\rangle = |\Phi_{11}\rangle,$$

up to a possible global phase (which has no physical significance; why?). Thus, the following protocol [9] transmits two classical bits by sending one qubit:

1. A Bell state $|\Phi_{))}\rangle$ is prepared between a sender A and a receiver B.
2. A encodes a 2-bit message by applying one of the four operations on his end of the state: I, σ_x, σ_y, σ_z.
3. A sends their qubit to B.
4. B measures in the Bell basis $\mathcal{B}$ and thus retrieves the 2-bit message.

Teleportation. We now turn to the question of how many classical bits need to be sent to transmit the state of one qubit. The answer again turns out to be 2, if we are permitted the use of entanglement. The protocol [10] goes like this:

1. A sender A and a receiver B prepare a joint Bell state $|\Phi_{00}\rangle$.
2. Party A now puts in place the state $|\psi\rangle$ to transmit (A holds two qubits at this point).
3. A performs a measurement in the Bell basis $\mathcal{B}$ on their two-qubit system, and thereby obtains the two-bit outcome $\mathbf{o}$.
4. A transmits the two-bit measurement outcome $\mathbf{o}$ to B.
5. Depending on the value of $\mathbf{o}$, B applies one of the four correction operations I, σ_x, σ_y, σ_z to its qubit, and thereby retrieves $|\psi\rangle$.

The proof of correctness of the protocol is left as an exercise. Note that the teleported quantum state $|\psi\rangle$ can be unknown to party A at the beginning of the protocol. If so, it will remain unknown to both parties.

4 Quantum computation

The starting point for the discussion of quantum computation is the question: "*How is the scaling of computational cost affected if we allow for quantum resources?*" Hence, quantum computing is about the efficiency of computation, and not about computability[3]. Furthermore, the emphasis is on 'scaling', i.e. how the computational cost[4] increases as the size of the input to the computation is increased. In particular, quantum computation is not about clock speed.

Computer scientists make a fundamental distinction between algorithms that require resources *polynomial* in the input size vs. algorithms that require resources scaling *exponentially* with the input size. For example, multiplying two prime numbers p and q is algorithmically easy. The computational cost scales quadratically (hence polynomially) in the number of digits of the two numbers. On the other hand, as far as is known, the reverse operation of extracting the prime factors p and q from an integer $N = pq$ is computationally hard (almost exponential in the number of digits of N). The asymmetry in the computational hardness of multiplying prime factors vs. extracting prime factors is the basis of the very widely applied cryptographic protocol RSA.

As it turns out, decomposing an integer $N = pq$ into its prime factors p and q is *not* hard on a quantum computer, and RSA can thus be broken by quantum computers; see Section 4.1.4 below on Shor's algorithm.

The notion of complexity of computation is formalized in computer science in terms of *complexity classes*. There is a sizeable zoo of them. The important ones for the present discussion are P, NP and BQP.

- **P** (classical) is the class of all decision problems which can be solved on a universal classical computer[5] in polynomial time.
- **NP** (classical) is the class of all decision problems for which "Yes" instances have efficiently (=poly-time) verifiable proofs.
- **BQP** (quantum) is the class of all decision problems solvable on a quantum computer in polynomial time, with a success probability[6] of $\geq 2/3$.

Regarding the first two classes, P contains all the computationally easy problems, and $P \subset NP$. Also, from the above $N = pq$ example, one may be led to believe that $P \neq NP$. But that is not a proof (there is only no *known* simple algorithm for factoring). In fact, the question of whether or not $P = NP$ is one of the major open problems in computer science[7].

For the quantum branch of commuter science, it is important to figure out how the quantum complexity class BQP relates to their classical counterparts P and NP. Is $BQP = P$? (If that were the case, there was little room for the usefulness of quantum computers. Grover's data base search, which provides a quadratic speedup, might still be useful.) Is $BQP = NP$? If that were the case, quantum computers would be immensely powerful (unless $P = NP$). The truth is expected to lie somewhere between these two extreme cases, but no alternative is presently provably ruled out. Factoring is not believed to be in P, but its not one of the hardest problems in NP.

[3] This should be expected, since every quantum process, governed by the Schrödinger equation and the Born rule, can be classically simulated, albeit, as far as is known, only inefficiently.

[4] Computational cost may be measured e.g. as runtime $\times$ number of processors used.

[5] The precise definition of 'universal classical computer' herein is 'deterministic Turing machine', which is equivalent to the above discussed classical circuit model.

[6] 2/3 is an arbitrary number here. Any bound $> 1/2$ will serve the same purpose.

[7] Also, it is one of the six remaining millennium problems selected by the Clay Mathematics Institute.

4.1 The circuit model

The classical complexity classes P and NP are defined in reference to a 'standard' model of classical computation, the so-called Turing machine. While a quantum-mechanical version of the Turing machine exists, it is not the most intuitive concept, the quantum complexity class BQP may equivalently be defined in reference to a different model of quantum computation, the circuit model. The circuit model is the most wide-spread model of quantum computation, and may as well be called the 'standard model'. To its description we turn next.

4.1.1 From Maxwell's demon to quantum computation

Maxwell's demon is a thought-experiment invented by the British Physicist James Clerk Maxwell in 1867, to illustrate that the second law of thermodynamics[8] only holds as a statistical law. To drive home his point, Maxwell describes a little 'demon' whose actions hypothetically violate the second law. The demon operates a little door between two halves of a gas container, letting only the slower-than-average molecules pass from left to right, and only the faster-than average molecules pass from right to left. In result the right half of the container should cool down and the left side warm up. This would decrease entropy, and thus violate the second law.

Furthermore, if Maxwell's demon could be constructed as a machine, it would, when coupled with a heat engine, give rise to a perpetuum mobile. So there better be fundamental obstructions to Maxwell's demons in the real world. Among the most famous arguments to rule out Maxwell's demons are those by Szilard and Brillouin. Relevant for this lecture is an argument put forward in 1962 by Rolf Landauer, scientist at the IBM Watson Research Center. Namely, Landauer postulated that the erasure of a single bit of information should carry an energy cost

$$E_{\text{bit-erasure}} \geq \ln 2\, kT. \tag{19}$$

The above inequality is known as Landauer's principle. It resolves Maxwell's paradox in the following way: The demon would naturally remember all the molecules it let pass from left to right, and from right to left. At some point, the demon's memory would be completely filled, and, to proceed, the demon would need to *erase* some of this information. According to Landauer's principle, this erasure costs energy, compensating for any possible energy gain from the two reservoirs.

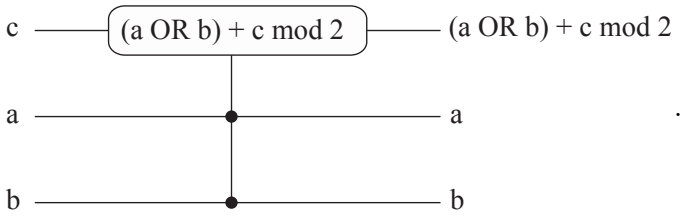
Landauer's principle has itself been criticized and defended, but here is where we stop following the trail of Maxwell's demon. Instead, we ask: "Does Landauer's principle imply a lower bound on the energy consumption of computers?" Surely, we may say, computer memory must be cleared in any sufficiently long computation to make room for later parts of the computation. While this argument seems plausible, it turns out to be a fallacy.

As was realized by Charles Bennett, a co-worker of Landauer's at IBM, computation can proceed without ever erasing a single bit of information. More succinctly, the main characteristic of bit erasure is that it is *irreversible*. Bennett showed that *every computation can be performed in an entirely reversible manner*. Halfway through our path to the circuit model, it is worth examining the construction behind this finding.

The elementary gates of every computation are "OR" and "NEG", where $\mathrm{NEG}(x) = x + 1 \mod 2$. Because *every* computation can be assembled from (many uses) of these gates, they are called "universal". The OR-gate is another example of irreversible operation. Once the gate $o = a \,\mathrm{OR}\, b$ is computed for some bit values a and b, the input cannot be reliably reconstructed

[8] The second law of thermodynamics states that in every naturally occurring thermodynamical process, entropy always increases or at best remains constant, but never decreases.

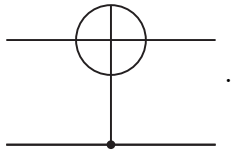
from the output o. A first step towards constructing a reversible computer therefore is to construct a reversible OR-gate. Here it is:



(20)

The gate operation is represented here by its circuit diagram. Any such diagram has an implicit time arrow running from left to right. That is, the variables to the left represent the input to the gate operation, and the variables to the right its output. In the present case, there are three bits of input and three bits of output. Two of the input bits, a and b, are straightforwardly propagated from input to output, while the third bit, c, is updated in a non-trivial manner. We make two observations about the gate in (20). (i) Applying it twice yields the identity operation. Thus, the gate is its own inverse, and, in particular, it has an inverse. (ii) The gate can be used to implement an "OR", namely by setting $c = 0$.

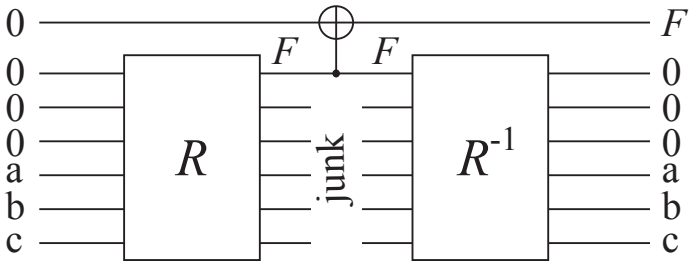
We may apply the same procedure to the NEG-gate, resulting in the so-called controlled NOT-gate (CNOT),



It has the effect of mod-2-adding the bottom bit to the top bit, and letting the bottom bit pass. From the viewpoint or reversibility, there is no need for this gate since already the NOT-gate is reversible. But we will find another use for the CNOT in Bennett's construction.

With what we have described so far, we can reversibly compute any Boolean function on any input, using up several bits of cleared memory in the state "0". The downside so far is that this computation creates a large number of bits filled with intermediate results, or "junk". We still need to get rid of the junk without using erasure.

As it turns out, the "junk" can be un-computed in a reversible fashion. Denote by R the reversible circuit which takes as input the arguments a, b, c, ... of the computed function F, as well as several auxiliary input bits in the state "0", and outputs $F(a, b, c, ..)$ plus several junk bits. Then, the circuit



(21)

has the desired effect of computing F on the given input, without spoiling auxiliary memory. The auxiliary memory is used in the computation, but afterward returned blank. It may be compared to a catalyst in a chemical reaction. The circuit has 3 parts, namely the reversible circuit R, the

CNOT for copying the result to an extra bit, and the reversible circuit R^{-1} which un-computes the junk bits. R^{-1} consists of the same gates as R but in reverse order (note that the extended OR-gate (20) and NEG are their own inverses).

Circuits of the form (21) provide the starting point for a quantum generalization—the circuit model of quantum computation. Any reversible circuit can be implemented in quantum systems, since the reversible gates used therein are also *unitary*; but a quantum circuit is more general than a reversible circuit in two respects. Namely, (i) it can take superposition states (including entangled states) as input, and (ii) the class of quantum gates is much broader than the class of reversible gates.

We begin by defining the quantum bit—or *qubit*—the elementary unit of quantum information. The qubit can, like the classical bit, be in two distinct (and orthogonal) states $|0\rangle$ and $|1\rangle$. But in addition, all superpositions of the form

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

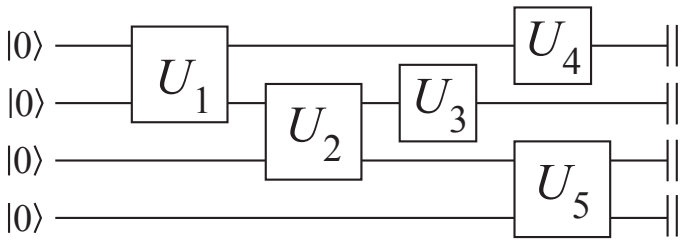
are allowed, subject to the normalization constraint $|\alpha|^2 + |\beta|^2 = 1$. That is, the state of a single qubit can be any state in the Hilbert space $\mathbb{C}^2$. A number $n \geq 2$ of qubits form a "quantum register". The state of a quantum register can be any state in the Hilbet space $\mathbb{C}^{2^n}$. This includes, in particular, entangled states.

The gates that such states are acted upon in a quantum circuit are all unitary, and hence can be realized by evolution according to the Schrödinger equation, given a suitable Hamiltonian. Measurements typically only occur at the end of the computation, to read out the final state of the quantum register. Thus, we have the following correspondence between classical reversible computation and quantum computation in the circuit model:

reversible circuit		**quantum circuit**
bit	-	qubit
reversible gate	-	unitary gate
readout	-	measurement

.

A typical quantum circuit looks something like this (but much bigger):



It begins with the initialization of the quantum register in a fixed and easy-to-prepare state, and ends with the measurement of each qubit in the eigenbasis of σ_z (the so-called computational basis). In-between initialization and readout lies the main part, namely a sequence of unitary gates. As can be seen in the above illustration, some of the gates only act on an individual qubit while other gates have multiple qubits interacting. What are the simplest standardized building blocks of a quantum circuit? I.e., which are the gates that we must require a proper quantum computer to execute? To this question we turn next.

4.1.2 Computational universality

As we discussed above, in classically universal computation there is exist universal sets of gates such that every computation can be built out of gates solely from any such set. Our earlier example was $\{\mathrm{OR}(\cdot,\cdot), \mathrm{NEG}(\cdot)\}$. Another example is $\{\mathrm{NOR}(\cdot,\cdot) = \mathrm{NEG}(\mathrm{OR}(\cdot,\cdot))\}$.

We now construct a quantum counterpart to the notion of a universal set. First, we need to clarify what we want to achieve with a sequence of unitary gates. The benchmark is taken to be the reachability of any transformation in the (special) unitary group $SU(2^n)$, with n the number of qubits. More precisely,

Definition 6 (Quantum universality) *A set G of quantum gates is universal if, for any number n of qubits, every unitary $U \in SU(2^n)$ can be arbitrarily closely approximated by a sequence of gates from G.*

The notion of a 'universal set of gates' sets the standard for what a proper quantum computer needs to be capable of doing. While it is exceedingly rare that a given unitary by itself is computationally useful, a quantum computer capable of executing gates from a universal set is guaranteed to have the full strength of a quantum computer. Without proof, we note that the gate sets

$$\begin{aligned} G_1 &= \{\mathrm{CNOT}_{i,j}, U_k \in SU(2),\, i \neq j, k = 1,..,n\}, \\ G_2 &= \{\mathrm{CNOT}_{i,j}, e^{i\pi/4\sigma_x^{(k)}}, e^{i\pi/8\sigma_z^{(k)}},\, i \neq j, k = 1,..,n\} \end{aligned}$$

are universal. Note that the former set is infinite while the latter is finite.

4.1.3 The Deutsch-Jozsa algorithm

We present here the simplest case of the Deutsch-Jozsa (DJ) algorithm, based on functions of only a single bit[9]. The algorithm has no practical application; rather, it drives home a fundamental point. Specifically, it demonstrates that certain computational problems cannot be solved as fast classically as they can be solved quantum mechanically, thus establishing the inequivalence of the two computational frameworks. This wasn't at all clear in 1985 and 1992 when the DJ algorithm, in different degrees of generality, was found.

The task solved by the DJ algorithm is the following: Consider a function $f : \mathbb{Z}_2 \longrightarrow \mathbb{Z}_2$, which is either constant ($f(0) = f(1)$), or balanced ($f(1) = f(0) + 1 \mod 2$). Querrying an oracle which evaluates the function f, *how many oracle calls are required to verify whether f is constant or balanced?* For clarification, an oracle is a device we can probe with questions. The oracle answers those questions, but we have no idea what's going on inside it.

Classically, two oracle calls are required. The function must be evaluated on both possible inputs to decide the question.

Quantum-mechanically, a single oracle call suffices. This assumes we have a quantum oracle accepting calls in superposition. Moreover, we require the oracle to be a deterministic quantum operation, i.e., a unitary. The quantum oracle is a slight generalization of the two-qubit CNOT gate which we have already met. It has the following action

$$|a\rangle \otimes |b\rangle \longrightarrow |a\rangle \otimes |b + f(a) \mod 2\rangle. \tag{22}$$

As can be easily verified, this gate action is indeed unitary, for every function f. The DJ algorithm proceeds by querying the oracle with the input state $\frac{|0\rangle+|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle-|1\rangle}{\sqrt{2}}$, and then measuring the first qubit in the eigenbasis of σ_x.

[9] It is a deterministic version of Deutsch's original algorithm of 1985.

To see what's going on, let's first check what the oracle (22) returns when queried with the input $|0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}$:

$$|0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \overset{(22)}{\longrightarrow} |0\rangle \otimes \frac{|0 + f(0) \mod 2\rangle - |1 + f(0) \mod 2\rangle}{\sqrt{2}} = (-1)^{f(0)}|0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

Note that neither the state of the first nor the second qubit have changed, only the overall phase. This is called a phase kickback.

Likewise, when the quantum oracle is queried with the input state $|1\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}$, we find

$$0\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \overset{(22)}{\longrightarrow} (-1)^{f(1)}|1\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

Again, the action of the oracle amounts to the kickback of a global phase. By querying the oracle in a superposition of the two above states, the kickback phases become relative and thus measurable,

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \overset{(22)}{\longrightarrow} (-1)^{f(0)} \frac{|0\rangle + (-1)^{f(1)-f(0)}|1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

Thus, the information of whether f is constant or balanced can indeed be inferred from the measurement of the first qubit in the basis $\big\{(|0\rangle + |1\rangle)/\sqrt{2}, |0\rangle - |1\rangle)/\sqrt{2}\big\}$, after a single call of the quantum oracle.

4.1.4 Shor's factoring algorithm

Shor's factoring algorithm breaks the cryptography system RSA. RSA is in widespread use today, e.g. for encrypting email and sending credit card information.

RSA. Suppose Allice (A) wants to purchase a book at some internet store "Books.com" (B) via internet, and for this purpose needs to send her credit card information. We assume that there is an eavesdropper tapping into the communication between A and B. How can an eavesdropper be prevented from learning Alice's secrets?–Here is the RSA protocol:

1. B picks two very large prime numbers, p and q, and computes $N = pq$. N is sent to A, but anyone can listen. N is public information, while p and q are not.

2. A uses N to encode her credit card number, or whatever piece of secret information, $N : c \to \overline{c}$, and sends $\overline{c}$ to B. Again this information is sent over a public channel, so everyone can listen.

3. B decodes $\overline{c}$, using p and q.

Let's review what's public and what is private in this protocol: The prime factors p and q are private, i.e., only known by B. Their product $N = pq$ is public. The point of the protocol is that the public information N is sufficient to *encode* information, but the private information p, q is required to *decode* it. So, using the public N, everyone can send encrypted messages around, but only B can read them.

But how is the public information N any less than the private information p, q? Once N is known, its prime factors are uniquely specified and can be computed explicitly. However, as far as

is known, decomposing numbers into their prime factors, even if there are only two of them, is a very hard computational problem. This is in stark contrast to the inverse problem of multiplying (prime) numbers, which is computationally efficient. The problem of decomposing an integer into prime factors is *not even provably hard*. It is just that, after much trying, no efficient classical algorithm has been found. The RSA crypto system is built on the *computational assumption* that factoring is hard.

This assumption is known to fail when we can apply a quantum computer to the factoring problem. The runtime of Shor's quantum algorithm is

$$T_{\mathrm{QM}} = O\left(\mathrm{ld}(N)^2\right), \tag{23}$$

where $\mathrm{ld}(N)$ is the number of digits of N. This complexity is similar to multiplying numbers on a classical machine. To the contrary, the best known classical algorithm for factoring has a runtime

$$T_{\mathrm{class}} = O\left(\exp\left(\mathrm{ld}(N)^{1/3}\right)\right). \tag{24}$$

This is much worse. The scaling is not quite exponential but worse than any polynomial function in $\mathrm{ld}(N)$.

Shor's algorithm. Recall that the task is to factor an integer $N = pq$ into its prime factors p and q. Around the try quantum algorithm, there is a some classical pre- and post processing. The whole procedure consists of the following steps:

1. Choose an integer m and compute $\gcd(m, N)$. If the gcd is 1, then go to step 2. otherwise you are done! Note that the gcd can be efficiently computed by Euklid's algorithm.

2. Use a quantum computer to determine the period P of the function $a \mapsto m^a \mod N$.

3. Check whether P is even or odd. If P is odd, then go back to step 1 (this happens with a probability of 1/4). If P is even then continue.

   By definition of the period P we have $m^P \mod N = 1$. Therefore,

   $$\left(m^{P/2} - 1\right)\left(m^{P/2} + 1\right) \mod N = 0.$$

   Now, if $\left(m^{P/2} - 1\right) \mod N = 0$ then go back to step 1 (this happens with a probability of 1/2). Otherwise continue to step 4.

4. Compute $d = \gcd\left(m^{P/2}, N\right)$ and output it. d is a prime factor of N.

The quantum part: finding the period P. Shor's quantum algorithm uses two quantum registers, i.e., two sets of quantum bits,

$$\begin{aligned} |\text{register 1}\rangle &\in \mathcal{H}_1 = (\mathbb{C}^2)^{2L}, \\ |\text{register 2}\rangle &\in \mathcal{H}_2 = (\mathbb{C}^2)^{L}, \end{aligned}$$

where $L = \lceil \mathrm{ld}(N) \rceil$. Thus, register 1 can hold a number of size $\sim N^2$, and register 2 a number of size N. We assume that register 1 is initialized in the state

$$|\text{register 1}\rangle = \frac{1}{2^L} \sum_{a=0}^{2^{2L}-1} |a\rangle,$$

and register 2 is prepared in the state $|\mathbf{0}\rangle$, i.e., all belonging qubits are individually prepared in the state $|0\rangle$. The quantum algorithm then consists of four steps:

1. *Modular exponentiation.* The operation of modular exponentiation is defined on the eigenstates of the σ_z-basis (=the computational basis) by

$$|a\rangle \otimes |c\rangle \mapsto |a\rangle \otimes |c + m^a \mod N\rangle. \tag{25}$$

By linearity of QM, the operation is then defined on all input quantum states. It is easily verified that it is unitary. Since modular exponentiation is classically efficient on a single input, by the reversible computing techniques described in the previous section, the quantum modular exponentiation operation Eq. (25) can be efficiently realized.

On the given input, the modular exponentiation produces the state (of both quantum registers) is

$$|\Psi_{\text{step 1}}\rangle = \frac{1}{2^L} \sum_{a=0}^{2^{2L}-1} |a\rangle \otimes |m^a \mod N\rangle.$$

2. *Measurement of register 2.* The second register is measured in the computational basis. Denote the outcome by $\mathbf{s} = (s_1, s_2, .., s_L)$.[10] The resulting state of the quantum registers is

$$|\Psi_{\text{step 2}}\rangle \sim \sum_{a|\, m^a \mod N = \mathbf{s}} |a\rangle \otimes |\mathbf{s}\rangle.$$

Because of the periodicity of the function $a \mapsto m^a$, we may rewrite this as

$$|\Psi_{\text{step 2}}\rangle \sim \sum_{n \in \mathbb{N} | 0 \le a_0 + nP < 2^{2L}} |a_0 + nP\rangle \otimes |\mathbf{s}\rangle,$$

where a_0 is defined via $m^{a_0} = \mathbf{s}$. The key point to note here is that that $\langle i| \otimes \langle \mathbf{s}|\Psi_{\text{step 2}}\rangle$ is a periodic function in i with period P, composed of a bunch of δ-peaks.

It shall be noted that measuring the quantum register in the computational basis (=eigenbasis of the individual σ_z's) reveals no information about the period. This arises because of the random offset $a_0(\mathbf{s})$ in the peak locations. In a first such measurement, we obtain a random peak location $a_0(\mathbf{s}) + nP$, for some n. After that measurement the state of the quantum register is destroyed, and the computation needs to start over. In a second run, the measurement of register 2 will lead an outcome $\mathbf{s}'$, and, correspondingly, the subsequent measurement of a peak position in register 1 will yield $a_0(\mathbf{s}') + n'P$, for some $n' \in \mathbb{N}$. Since the offsets are in general random an distinct in runs 1 and 2, nothing can be learned about P from those measurements.

3. Instead, before measuring in the computational basis, we perform a quantum Fourier transform on register 1. The quantum Fourier transform (QFT) is defined is defined via

$$|j\rangle \mapsto \frac{1}{M} \sum_{k=0}^{M-1} e^{2\pi i/M\, jk} |k\rangle,$$

where, in our case, $M = 2^{2L}$. The QFT is a close analogue of the classical Fourier transform, and its circuit implementation, which is based on the fast Fourier transform, is efficient.

[10] The outcome $\mathbf{s}$ does not matter for the remainder of the algorithm, and indeed, the measurement does not even need to be performed. However, to understand what's going on, its easier to assume that the measurement is performed.

The effect of the QFT on $|\Psi_{\text{step 2}}\rangle$ with the amplitudes $\langle i| \otimes \langle \mathbf{s}|\Psi_{\text{step 2}}\rangle \sim \sum_n \delta(a_0(\mathbf{s}) + nP, i)$ is to produce another periodic function composed of δ-peaks, but in the Fourier transform the peak locations are integer multiples of

$$P' = \frac{2^{2L}}{P}. \tag{26}$$

Note in particular that the dependence of the peak locations on $a_0(\mathbf{s})$ has disappeared! The dependence on $a_0(\mathbf{s})$ has moved into an exponential factor $e^{2\pi i a_0(\mathbf{s})/2^{2L}\, k}$ in the wave function $\Psi(k)$ after QFT, which does not affect the measurement probabilities in the computational basis.

Now, repeated runs of the algorithm yield a collection of integer multiples of P', one multiple in each run of the algorithm. Now, having a few such multiples (with unknown integer multipliers), P' can with high probability of success be extracted by computing the gcd of these P'-multiples. The original period P can then be obtained from P' via Eq. (26).

Remark: The analysis of the last step in Shor's factoring algorithm above is greatly simplified. As is apparent from Eq. (26), if P is an integer (which it is) then P' will in general not be. In other words, the δ-peaks in the state of register 1 after the QFT will be necessarily broadened. In fact, the size $2L$ of this register was chosen to limit the extent of this broadening, but the broadening cannot be eliminated altogether. Thus, at the level of integer spacings, the measured multiples of P' have errors. Now, the computation of gcd-s does not tolerate any such errors, and it can therefore not be straightforwardly applied. A variant that can cope with the unavoidable imperfections in the measured multiples of P' is the method of continued fractions; See e.g. [21].

4.2 The adiabatic model

Returning to our earlier discussion of complexity classes, an algorithmic problem is called *hard* w.r.t. a given complexity class if it is at least as hard as any problem in this class. That is, there must be a polynomial-time reduction from any problem in the class to the given problem. An algorithmic problem is called *complete* w.r.t. a given class if it is simultaneously hard and in the class. It is natural to ask whether NP has complete problems.

The answer to this question is 'Yes', and in fact, there is a fairly natural such problem, 3SAT. Consider m binary variables satisfying n clauses (=constraints), where each clause depends on no more than three of the variables. That is, we have a system of equations $f_i(x_{j(i)}, x_{k(i)}, x_{l(i)}) = 0$, with $i = 1, .., n$, and all f_i are Boolean functions. Does this system of equations have a solution?

Here is a physical method to find out. Build a Hamiltonian on a Hilbert space $(\mathbb{C}^2)^m$,

$$H_f = g \sum_{i=1}^{n} F_i,$$

where $F_i = \sum_{\mathbf{x}\in\mathbb{Z}_2^m} f_i(x_{j(i)}, x_{k(i)}, x_{l(i)})|\mathbf{x}\rangle\langle\mathbf{x}|$, and g is some arbitrary positive coupling constant. This Hamiltonian looks remotely physically feasible, since it only requires 3-body interactions[11]. All there seems to do now is to wait for the physical system to settle into its ground state. Then, one may measure the ground state in the computational basis $\{|\mathbf{x}\rangle, \mathbf{x} \in \mathbb{Z}_2^m\}$ and check whether it has zero energy/ satisfies all the constraints.

[11] It is possible to further reduce the required interactions to two-body, which makes the resulting Hamiltonian really feasible. But, from our present birds-eye perspective, these are practical details.

But will the system settle into its ground state? If so, how long might that take? The idea of adiabatic quantum computation [11] is to "help" the quantum system find the ground state of the Hamiltonian H_f. This proceeds not by an optimized cooling schedule but rather by way of the adiabatic theorem.

Theorem 4 (Adiabatic evolution) *A physical system remains in its instantaneous eigenstate if the Hamiltonian is changed very slowly, and at all times there is a gap between the energy of the instantaneous eigenstate and the remainder of the instantaneous energy spectrum.*

Based on adiabatic evolution, the idea of adiabatic quantum computation is to begin in the ground state of a Hamiltonian H_i which is known and easy to prepare, and then gradually change the Hamiltonian from H_i to H_f. This may e.g. be accomplished by the time-dependent Hamiltonian

$$H(t) = \left(1 - \frac{t}{T}\right) H_i + \frac{t}{T} H_f,$$

such that $H(0) = H_i$ and $H(T) = H_f$. If T is large enough such that the evolution is adiabatic during the entire evolution, then the system starts in its ground state at $t = 0$, remains in its instantaneous ground state throughout, and, at $t = T$, arrives in the desired ground state of H_f.

So, can adiabatic computation solve NP-complete problems? Well, we have not discussed the fine print yet. We still need to resolve the question of how slow is slow enough for adiabaticity. An approximate condition for adiabaticity is [15]

$$\hbar \frac{|\langle E_n | \frac{d}{dt} | E_m \rangle|}{|E_n - E_m|} \ll 1. \tag{27}$$

We can see from this expression that the condition for adiabaticity becomes hard to satisfy when energy eigenvalues come very close to another at some point in the evolution. These so-called avoided crossings typically do happen, and the resulting energy gaps have a tendency of becoming smaller and smaller with increasing system size. This, in turn, requires T to become large with increasing system size. Thus, the need to satisfy the adiabaticity condition introduces a scaling of runtime T with problem size.

The question now is whether this scaling of runtime is more favourable then it would be in other models of computation. To decide this in the general case is a (likely too) hard physics problem. Solutions are known only in specific cases and for certain ranges of parameters. And bear in mind that the discussion so far pertains to the idealized scenario of zero temperature. At finite temperature, the possibility of thermal excitation further complicates the picture, and in general makes it harder to arrange for the system to remain in its ground state.

Further reading on adiabatic quantum computation.

- The adiabatic model (at zero temperature) and the circuit model are computationally equivalent (=equally powerful) [12]. Either model can simulate the other with polynomial overhead.
- The validity of the adiabatic theorem under its stated conditions has recently been questioned [13]. This created a flurry of activity in sorting out what these conditions should be, and in result, such conditions are now known at any imaginable level of rigour [14]-[16].
- An important question is whether narrow avoided crossings can be avoided. No says [17]. Yes (with qualifications) says [18]. Some propose to boost small gaps by coding [19],[20].

5 The Bell inequalities

5.1 Can quantum mechanics be considered complete?

Consider two particles in an entangled state. Quantum mechanics tells us that, if the first particle is measured, then not only the first, but also the second particle ends up in a state that depends on the measurement outcome obtained. Specifically, for a spin singlet of two spins 1/2, if the spin of the first particle is measured in any direction, the spin of the second particle will be oriented anti-parallel with the spin of the first. The situation is the same independent of whether the two particles are in close proximity, or one is on earth and the other is on the moon. Albert Einstein was uncomfortable with that thought. He called it "spooky action at a distance".

The paper by Einstein, Podolsky and Rosen (EPR) entitled "Can quantum mechanics be considered a complete description of physical reality?" [5] is Einstein's disbelief cast into a formal argument. The paper does not call into doubt the correctness of quantum mechanics, but rather its completeness. Can quantum mechanics predict everything that is in principle predictable?

EPR start out by explaining what they consider "physical reality" and "completeness" of a theory describing physical phenomena. Physical reality, they state, consists of "elements of reality" which they characterize as follows:

> We shall be satisfied with the following criterion, which we regard as reasonable. *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.*

Their requirement for completeness of a physical theory is the following:

> Whatever the meaning assigned to the term *complete*, the following requirement for a complete theory seems to be a necessary one: *every element of the physical reality must have a counterpart in the physical theory.* We shall call this the condition of completeness.

In the following, we present a slightly simplified version of the EPR argument, due to David Bohm, that works with spin degrees of freedom rather than the original position and momentum. Consider a spin-1/2 system with its spin pointing in the positive z-direction, $|\psi\rangle = |\uparrow_z\rangle$. That is, $|\psi\rangle$ is an eigenstate of $\hat{S}_z$,

$$\hat{S}_z|\psi\rangle = \frac{\hbar}{2}|\psi\rangle.$$

Thus, according to the above criterion, $S_z = \hbar/2$ is an element of reality. Furthermore, since $|\psi\rangle$ is not an eigenstate of $\hat{S}_x$, the outcome of a measurement of $\hat{S}_x$ can not be predicted with certainty, according to the formalism of quantum mechanics. Thus, S_x is not an element of reality, if quantum mechanics is complete. We are thus left with two alternatives:

(1) Quantum mechanics is not complete.

(2) If the operators corresponding to two physical quantities do not commute, then the two quantities cannot both correspond to elements of reality.

With those alternatives in mind, we now consider two spins 1/2 in a singlet state,

$$|\Psi\rangle = \frac{|\uparrow\rangle_A|\downarrow\rangle_B - |\downarrow\rangle_A|\uparrow\rangle_B}{\sqrt{2}}. \tag{28}$$

Using this setup, EPR now prove that quantum mechanics is incomplete. The proof proceeds as follows: Assume that the above alternative (1) is wrong. Consider a measurement of system A. If A is measured in the x-basis, the spin at B is projected into an eigenstate of $\hat{S}_x$ which can be predicted with certainty. $\hat{S}_x^{(B)}$ therefore is an element of reality. Likewise, if A is measured in the z-basis, the spin at B is projected into an eigenstate of $\hat{S}_z$, and $\hat{S}_z^{(B)}$ is an element of reality. Furthermore, since the measurement of A does not in any way affect the system at B, those observables must have corresponded to elements of reality already before the measurement. That is, both $\hat{S}_z$ and $\hat{S}_x$ correspond to elements of reality. Thus, alternative (2) is wrong.

To summarize, (1) is wrong implies (2) is wrong, and we are left with no alternative. Contradiction. Alternative (1), namely that QM is incomplete, must therefore be correct. □

Towards the end of their paper, EPR offer a route to circumvent their conclusion (for Bohm's version as discussed above, please replace $Q \longrightarrow S_z$ and $P \longrightarrow S_x$):

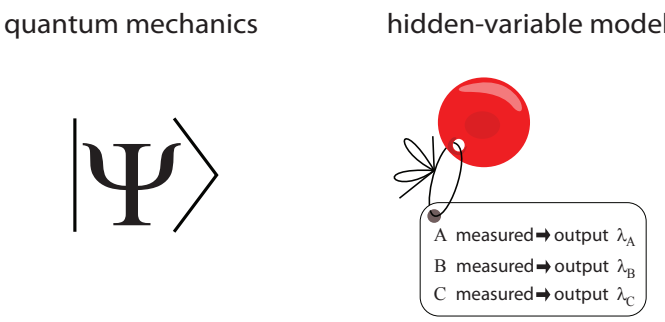
> One could object to this conclusion on the grounds that our criterion of reality is not sufficiently restrictive. Indeed, one would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted*. On this point of view, since either one or the other, but not both simultaneously, of the quantities P and Q can be predicted, they are not simultaneously real. This makes the reality of P and Q depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality could be expected to permit this.

Most of the scientific community today seems at ease with the latter "unreasonable" definitions of physical reality that make the reality of properties of system B dependent on what happens on system A. The EPR paper went almost unnoticed for three decades but then became increasingly influential. This is, perhaps, in part due to its last paragraph,

> While we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible.

5.2 Hidden variable models

The candidate descriptions of the physical world hinted at in the last paragraph of the EPR paper are now called "hidden variable models". Pictorially, we may represent a hidden variable model (HVM) like this



In an HVM, a physical system (e.g. a particle) is described not by a quantum state, but by the particle "with a catalogue attached to it". In this catalogue all observables that can possibly be measured on the particle are listed, and it is specified which outcome should be outputted when a particular measurement is made. The catalogue attached to the particle furthermore carries a serial number, λ. Catalogues with different numbers λ can assign different values to the observables listed, and the serial numbers are distributed according to some probability distribution $p(\lambda)$. In this way, hidden variable models attempt to mimic the probabilistic character of quantum measurement.

Note that HVMs are essentially classical theories. No amplitudes exist, but only probability distributions. A further crucial difference is that in quantum mechanics, the measurement outcome is brought about by the act of measurement. In a hidden variable model, it has existed all the time, and is merely revealed by the measurement.

5.3 Bell inequalities

Hidden variable models, at least as they were introduced above, do not seem particularly plausible theories of reality. However, if one is not fond of them, one surely needs a more substantial rejoinder than individual taste. It turns out that no general such argument can be given. HVMs are viable. Bohmian mechanics is an example.

However, this changes with the seemingly most innocent additional assumptions. One of the obstructions to the viability of hidden variable models is locality. Locality means here that if the given HVM describes a composite system with parts A and B, say, then the "catalogue" only contains observables which are either local to A or local to B. That is, only the outcomes of local measurements are predicted by a local HVM. However, these outcomes can still be correlated. Local HVMs are ruled out by the celebrated Bell inequalities. Somewhat ironically, their discoverer John Bell was *very* fond of hidden variable models.

Let's first consider a scenario where a hidden variable model does work. We go back to the singlet state of the EPR argument (in Bohm's version), and consider simultaneous measurement on the two spins A and B in the z- or x-basis. There are catalogues with four distinct serial numbers needed to describe the situation by an HVM (all outcomes to be multiplied by $\hbar/2$),

λ	$S_z^{(A)}$	$S_z^{(B)}$	$S_x^{(A)}$	$S_x^{(B)}$
1	+1	-1	+1	-1
2	+1	-1	-1	+1
3	-1	+1	+1	-1
4	-1	+1	-1	+1

This table describes accurately the perfect anti-correlation of spin measurements on system A and B, if the measurements are along the same axis (either both x or both z). Furthermore, if the

probability distribution is $p(\lambda) \equiv 1/4$, then also the probability distributions for the outcomes of all four local measurements are correctly reproduced.

However, there are other cases where no local HVM can reproduce the predictions of quantum mechanics. To such a scenario we turn now. Instead of Bell's original version, we present here an inequality due to Clauser, Horne, Shimony and Holt (CHSH). We consider a bipartite scenario with two measurement settings per party, namely the settings a and a' for party A, and a and b' for party B. The possible outcomes are ± 1 for each measurement.

In the quantum scenario, we associate observables $A(a)$, $A(a')$, and $B(b)$, $B(b')$ with those settings, and we assume that we are working with some quantum state $|\Psi\rangle$ that we are free too choose. The correlation between the measurement of $A(a)$ and $B(b)$ is

$$C(a,b) = \langle \Psi | A(a) \otimes B(b) | \Psi \rangle.$$

In a local hidden variable model, the same situation is described as follows. For each of the four measurement settings, there is a corresponding outcome, e.g. $A(a', \lambda)$, which depends on the setting, and also on the serial number λ of the catalogue. The value of the above correlation then is

$$C(a,b) = \int d\lambda \, p(\lambda) A(a,\lambda) B(b,\lambda).$$

In the following, we are interested in a particular correlation that involves all four possible combinations of bi-partite measurement settings, namely

$$C = C(a,b) + C(a',b) + C(a,b') - C(a',b').$$

For the HVM, we can constrain the value the correlation C takes. Namely

$$\begin{aligned} C &= \int d\lambda \, p(\lambda) \big(A(a,\lambda)B(b,\lambda) + A(a,\lambda)B(b',\lambda) + A(a',\lambda)B(b,\lambda) - A(a',\lambda)B(b',\lambda) \big) \\ &= \int d\lambda \, p(\lambda) \big(A(a,\lambda) \left[B(b,\lambda) + B(b',\lambda) \right] + A(a',\lambda) \left[B(b,\lambda) - B(b',\lambda) \right] \big). \end{aligned}$$

Now, the terms in the brackets, $[B(b,\lambda) - \pm B(b',\lambda)]$ take values ± 2 or 0. Furthermore, if one is ± 2 the other is zero. Since the A's in front of those expressions take values ± 1, the entire expression in the round bracket is ≤ 2, for all values of λ. This expression is now averaged according to some probability distribution $p(\lambda)$. Since $p(\lambda)$ is non-negative, the HVM prediction is

$$C \leq 2. \tag{29}$$

This is the CHSH inequality.

Let us now turn to the quantum-mechanical description. Note that we have so far not specified the quantum state and the measured observables, apart from the fact that they should have eigenvalues ± 1 only. For the state, we choose the spin singlet Eq. (28). As for the observables, we choose

$$A(a) = \sigma_z^{(A)}, \; A(a') = \sigma_x^{(A)}, \text{ and } B(b) = \frac{\sigma_x^{(B)} + \sigma_z^{(B)}}{\sqrt{2}}, \; B(b') = \frac{\sigma_x^{(B)} - \sigma_z^{(B)}}{\sqrt{2}}.$$

For this setting we find that quantum mechanics predicts

$$C = 2\sqrt{2}. \tag{30}$$

This prediction is incompatible with the CHSH inequality (29), and thus with all local hidden variable models.

Experimental tests of the Bell/ CHSH inequalities have been performed with increasing sophistication, and they have decided in favour of quantum mechanics. See [24] for the first conclusive such experiment. Local hidden variable models are thus ruled out as descriptions of physical reality.

A Historical note on teleportation and dense coding

The above protocols of dense coding and teleportation are the exact opposite of technically hard. Yet they are published in Physical Review Letters – how come? Their value is not in mastering a tedious and long-standing problem. Rather they moved into new territory, saw some unexpected phenomenology, and opened up a new field[12].

David Mermin, whom you may know from a condensed matter physics class, was a referee for the teleportation and dense coding papers. He made his reviews publicly available [22] in 2003. His report on the dense coding paper is particularly telling, and I have reprinted it here from [22].

```
Bennett and Wiesner, "Communication via one-and two-particle.  .  ." LT4749

Your question was:  Does this qualify as "strikingly different" enough to publish?
I have never read anything like it, and I have read a lot on EPR, though far from
everything ever written.  So as far as I know it is different.
But strikingly?  The argument is very simple, so shouldn't the point be obvious?
After reading the paper I put it aside and spent the next week working hard on
something totally unrelated.  Every now and then I would introspect to see if some
way of looking at the argument had germinated that reduced it to a triviality.  None
had.  Last night I woke up at 3am, fascinated and obsessed with it.  Couldn't get back
to sleep.  Thats my definition of "striking".
So I say its strikingly different and I say publish it.
```

References

[1] Asher Peres, *Separability Criterion for Density Matrices*, Phys. Rev. Lett. **77**, 14131415 (1996).

[2] Michal Horodecki, Pawel Horodecki, Ryszard Horodecki, *Separability of Mixed States: Necessary and Sufficient Conditions*, Physics Letters A **223**, 1-8 (1996).

[3] F. Verstraete, M. Wolf and J.I. Cirac, *Quantum computation and quantum-state engineering driven by dissipation*, Nature Physics 5, 633 - 636 (2009).

[4] Wojciech H. Zurek, *Decoherence, einselection, and the quantum origins of the classical*, Reviews of Modern Physics **75**, 715 (2003).

[5] A. Einstein, B. Podolsky, and N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. 47, 777 (1935).

[6] Wootters, William; Zurek, Wojciech, *A Single Quantum Cannot be Cloned*, Nature **299**, 802803 (1982).

[7] D. Bruss, G.M. D'Ariano, C. Macchiavello, and M.F. Sacchi, *Approximate quantum cloning and the impossibility of superluminal information transfer*, Phys. Rev. A **62** 062302 (2000).

[8] Alexander S. Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, Problems of Information Transmission **9**, 177 (1973).

[9] C. Bennett and S.J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. Phys. Rev. Lett. **69**, 2881 (1992).

[12] In fact, the teleportation paper is now a PRL milestone, and a "free to read" article.

[10] C.H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres and W.K. Wootters, *Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels*, Phys. Rev. Lett. **70**, 1895 (1993).

[11] E. Farhi *et al.*, arXiv:quant-ph/0001106v1.

[12] D. Aharonov *et al.*, SIAM Journal of Computing **37**, 166-194 (2007).

[13] P. Marzlin and B. Sanders, Phys. Rev. Lett. **93**, 160408 (2004).

[14] M.S. Sarandy, L.-A. Wu, D.A. Lidar, arXiv:quant-ph/0405059v3.

[15] M.H.S. Amin, Phys. Rev. Lett. **102**, 220401 (2009).

[16] Sabine Jansen, Mary-Beth Ruskai, Ruedi Seiler, J. Math. Phys. **48**, 102111 (2007).

[17] Boris Altshuler, Hari Krovi, and Jeremie Roland, Proc. Natl. Acad. Sci. USA, **107**, 1244612450 (2010).

[18] N. Dickson and M.H.S. Amin, Phys. Rev. Lett. **106** 050502 (2011).

[19] Stephen P. Jordan, Edward Farhi, Peter W. Shor, Phys. Rev. A **74**, 052322 (2006).

[20] Kristen L. Pudenz, Tameem Albash, and Daniel A. Lidar, Nature Communications **5**, 3243 (2014).

[21] Samuel Lomonaco, arXiv:quant-ph/0010034v1.

[22] N.D. Mermin, *Copenhagen Computation: How I Stop Worrying and Love Bohr*, arXiv:0305088 (quant-ph); IBM Journal of Research and Development, Vol. **48**, No. 1, 53-62 (2004).

[23] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).

[24] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett **49**, 1804 (1982).