2 Quantum Teleportation

Teleportation is the process of moving objects from one place to another, without actually moving the object (as a whole) through the intervening space. This might (or might not) be done by scanning the object down to its (sub)atomic level, transmitting this information to the destination, and rebuilding the object following the "construction plan" (and probably destroying the original, if this has not already happened due to the scanning).

One is tempted to conclude that due to Heisenberg's uncertainty principle, which essentially says that it is impossible to perfectly measure on a quantum level, teleportation is not possible with perfect accuracy. Surprisingly, this is not true, as we show below. However, for it to work, it is necessary that transmitter and receiver have shared an entangled state to start with (which is completely independent of the object to be teleported though).

Consider the four so-called Bell states

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big) \qquad |\Psi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|1\rangle + |1\rangle|0\rangle\big)$$
$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle - |1\rangle|1\rangle\big) \qquad |\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|1\rangle - |1\rangle|0\rangle\big)\,.$$

We assume that Alice and Bob share the first Bell state in that Alice holds/controls the first and Bob the second system of an EPR pair $|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big) \in \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$. Let $|\varphi_E\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathcal{H}_E = \mathbb{C}^2$ be the qubit Alice wants to teleport to Bob. The common state is then given by the 3-qubit state

$$|\psi_{EAB}\rangle = |\varphi_E\rangle \otimes |\Phi^+_{AB}\rangle = (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big)\,.$$

Alice measures the two qubits she controls, i.e., the two systems E and A, in the Bell basis $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$. To compute the result of her measurement, we rewrite the two qubits of Alice in the Bell basis, using the following general identities, which can easily be verified:

$$|0\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle + |\Phi^-\rangle) \qquad |1\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle - |\Psi^-\rangle)$$
$$|0\rangle|1\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle + |\Psi^-\rangle) \qquad |1\rangle|1\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle - |\Phi^-\rangle)$$

The 3-qubit state $|\psi_{EAB}\rangle$ can be written as

$$|\psi_{EAB}\rangle = \frac{1}{2}|\Phi^+\rangle\big(\alpha|0\rangle+\beta|1\rangle\big) + \frac{1}{2}|\Phi^-\rangle\big(\alpha|0\rangle-\beta|1\rangle\big) + \frac{1}{2}|\Psi^+\rangle\big(\beta|0\rangle+\alpha|1\rangle\big) + \frac{1}{2}|\Psi^-\rangle\big(-\beta|0\rangle+\alpha|1\rangle\big)\,.$$

and thus, depending on what Alice observes, the state collapses to one of the following four states.

$$|\Phi^+\rangle\big(\alpha|0\rangle+\beta|1\rangle\big), \quad |\Phi^-\rangle\big(\alpha|0\rangle-\beta|1\rangle\big), \quad |\Psi^+\rangle\big(\beta|0\rangle+\alpha|1\rangle\big), \quad \text{or} \quad |\Psi^-\rangle\big(-\beta|0\rangle+\alpha|1\rangle\big)$$

Note the similarity of Bobs qubit to the original state $|\varphi_E\rangle = \alpha|0\rangle + \beta|1\rangle$. Also note that by the outcome of her measurement, Alice knows in which of the the four states they are, and she can inform Bob by sending two classical bits. Bob can then apply a suitable transformation to his system to obtain the original state: the identity in case Alice observed Φ^+, the transform that maps $|0\rangle$ to $|0\rangle$ and $|1\rangle$ to $-|1\rangle$ in case Alice observed Φ^-, etc.

This procedure can also be appreciated as a perfect encryption of the quantum message $|\varphi_E\rangle$, where the shared EPR pair acts as the shared encryption key, and the two classical bits Alice communicates is the encryption of $|\varphi_E\rangle$: Intercepting these two bits gives no information at all on the message. Indeed, the outcome of Alice's measurement is uniformly distributed, independent of what the state $|\varphi_E\rangle$ is. Note that the key can only be used once, like in the classical perfectly-secure one-time-pad encryption scheme.

3 Nonlocality of Quantum Mechanics

Consider two parties, Alice and Bob, who are far apart and cannot communicate with each other. We pose Alice a "question" $x \in \mathcal{X}$ and Bob a "question" $y \in \mathcal{Y}$, and Alice and Bob are requested to provide replies $a \in \mathcal{A}$ and $b \in \mathcal{B}$, respectively, by performing local computations only, and without communicating. Any possible and impossible behavior of Alice and Bob can be captured by a conditional probability distribution $P_{AB|XY} : \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \to [0,1]$. We call such a conditional probability distribution a **hypercorrelation**. We want to get some understanding on the difference between different physical theories in terms of the hypercorrelations they permit.

We start by introducing the framework of **nonlocal games**. Sometimes, the catchy term **pseudotelepathy** is used in this context. As above, in a nonlocal game Alice and Bob are given "question" x and y, and their joint goal is to provide respective answers a and b such that the quadruple (a, b, x, y) satisfies some well defined (and known) verification predicate $V : \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$: Alice and Bob jointly *win* the game if $V(a, b, x, y) = 1$ and they lose if $V(a, b, x, y) = 0$. What makes the game nontrivial is that, like above, Alice and Bob are not permitted to communicate. Thus, Alice only knows x but not y when she needs to come up with a, and vice versa for Bob. Alice and Bob are allowed to communicate *before* the game starts in order to agree on some strategy, but not anymore once the game has started. The question of interest is of course: what is the maximal winning probability for Alice and Bob (for a given nonlocal game)? Interestingly, it turns out that even though an entangled quantum state does not enable Alice and Bob to communicate (quantum mechanics is **non-signaling**), it still enables them to do better in winning some nonlocal games than with a classical (i.e. non-quantum) strategy. This feature is called the **nonlocality** of quantum mechanics.

3.1 Definitions

We will be interested in the following hypercorrelations.

Definition 3.1. *Let $P_{AB|XY} : \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \to [0, 1]$ be a hypercorrelation. $P_{AB|XY}$ is called*

- **deterministic** *if there exist functions $f : \mathcal{X} \to \mathcal{A}$ and $g : \mathcal{Y} \to \mathcal{B}$ with the property that $P_{AB|XY}(a, b|x, y) = 1$ if $a = f(x)$ and $b = g(y)$, and $P_{AB|XY}(a, b|x, y) = 0$ otherwise.*
- **classical** *(or* **local***) if $P_{AB|XY}$ is a finite convex linear combination of deterministic hypercorrelations $P^1_{AB|XY}, \ldots, P^L_{AB|XY}$.* [11]
- **quantum** *if there exist quantum systems A and B, a state vector $|\varphi_{\mathsf{AB}}\rangle \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{B}}$, and families $\{\mathbf{E}^x\}_{x\in\mathcal{X}}$ and $\{\mathbf{F}^y\}_{y\in\mathcal{Y}}$ of POVMs $\mathbf{E}^x = \{E^x_a\}_{a\in\mathcal{A}}$ and $\mathbf{F}^y = \{F^y_b\}_{b\in\mathcal{B}}$ such that*

$$P_{AB|XY}(a, b|x, y) = \langle\varphi_{\mathsf{AB}}|(E^x_a \otimes F^y_b)|\varphi_{\mathsf{AB}}\rangle$$

  for all $a \in \mathcal{A}$, $b \in \mathcal{B}$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$.
- **non-signaling** *if*[12]

$$P_{A|XY}(\cdot|\cdot, y) = P_{A|XY}(\cdot|\cdot, y') \;\; \forall\, y, y' \in \mathcal{Y} \quad \text{and} \quad P_{B|XY}(\cdot|x, \cdot) = P_{B|XY}(\cdot|x', \cdot) \;\; \forall\, x, x' \in \mathcal{X}$$

  where $P_{A|XY}$ is naturally defined as $P_{A|XY} = \sum_b P_{AB|XY}(\cdot, b|\cdot, \cdot)$, and similarly $P_{B|XY}$.

[11] This means that there exist $\varepsilon_1, \ldots, \varepsilon_L \geq 0$ with $\sum_\ell \varepsilon_\ell = 1$ and $P_{AB|XY} = \sum_\ell \varepsilon_\ell P^\ell_{AB|XY}$. Hence, a and b are computed by applying *randomized* functions to the respective inputs x and y.

[12] $P_{A|XY}(\cdot|\cdot, y) = P_{A|XY}(\cdot|\cdot, y')$ is to be understood as $P_{A|XY}(a|x, y) = P_{A|XY}(a|x, y')\ \forall\, a \in \mathcal{A}$ and $x \in \mathcal{X}$.

The non-signaling condition guarantees that the hypercorrelation does not permit for instantaneous communication, and as such does not violate relativity. Every of the above classes of hypercorrelations is (strictly) contained in the next one (see Figure 2).

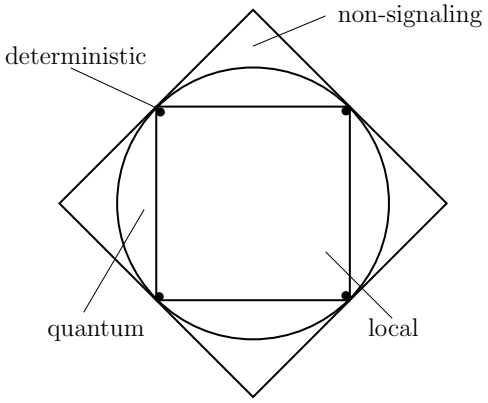

Fig. 2. The different regions of interest.

Formally, a nonlocal game is specified as follows.

Definition 3.2. *A* **nonlocal game** $\mathfrak{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$ *consists of finite sets* $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}$*, a probability distribution* $\pi : \mathcal{X} \times \mathcal{Y} \to [0, 1]$*, and a predicate* $V : \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$*.*

As mentioned above, any (possible or impossible) strategy for Alice and Bob is captured by a hypercorrelation $P_{AB|XY}$. For any nonlocal game $\mathfrak{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$, and for any such strategy/hypercorrelation $P_{AB|XY}$, the probability of Alice and Bob winning the game is given by the value

$$v[P_{AB|XY}](\mathfrak{G}) := \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \pi(x, y) \sum_{\substack{a\in\mathcal{A}\\ b\in\mathcal{B}}} P_{AB|XY}(a, b, x, y)V(a, b, x, y)\,.$$

Definition 3.3. *The* **deterministic***, the* **classical***, the* **quantum***, and the* **non-signaling value** *of a nonlocal game* $\mathfrak{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$ *is respectively defined as*

$$v_{d/c/q/ns}(\mathfrak{G}) := \sup_{P_{AB|XY}} v[P_{AB|XY}](\mathfrak{G})$$

where the sup *is over all deterministic, classical, quantum and non-signaling hypercorrelations* $P_{AB|XY} : \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \to [0, 1]$*, respectively.*

The set of deterministic hypercorrelations (for given finite $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}$) is finite, and the respective sets of classical and non-signaling hypercorrelations are compact. Thus, in those cases the sup is actually a max. Also in the quantum case, the sup is actually a max; but here it is less obvious (and we do not show it here).

By the inclusion property of the four classes of hypercorrelations considered, it follows immediately that $v_d(\mathfrak{G}) \leq v_c(\mathfrak{G}) \leq v_q(\mathfrak{G}) \leq v_{ns}(\mathfrak{G})$ for any nonlocal game $\mathfrak{G}$.

Lemma 3.4. *For any nonlocal game* $\mathfrak{G}$*:* $v_d(\mathfrak{G}) = v_c(\mathfrak{G})$*.*

Proof. It is sufficient to show that $v_d(\mathfrak{G}) \geq v_c(\mathfrak{G})$. Let $P_{AB|XY}$ be a classical hypercorrelation with $v[P_{AB|XY}](\mathfrak{G}) = v_c(\mathfrak{G})$. By definition, $P_{AB|XY}$ is of the form $P_{AB|XY} = \sum_\ell \varepsilon_\ell P^\ell_{AB|XY}$,

where $\varepsilon_1, \ldots, \varepsilon_L \geq 0$ and $\sum_\ell \varepsilon_\ell = 1$, and $P^1_{AB|XY}, \ldots, P^L_{AB|XY}$ are deterministic. By exchanging the order of summation, we obtain that

$$v_c(\mathfrak{G}) = \sum_\ell \varepsilon_\ell \, v[P^\ell_{AB|XY}](\mathfrak{G})$$

from which it follows that there must exist $\ell \in \{1, \ldots, L\}$ with $v_c(\mathfrak{G}) \leq v[P^\ell_{AB|XY}](\mathfrak{G}) \leq v_d(\mathfrak{G})$. This was to be shown. □

In the next section we show that the classical, the quantum, and the non-signaling value are in general strictly separated.

3.2 The CHSH Game

The CHSH game $\mathfrak{G}_{\text{CHSH}} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$ is defined as follows. $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0, 1\}$, i.e., all questions and answers are bits, $\pi(\cdot, \cdot) = \frac{1}{4}$, i.e., x and y are independent and uniformly distributed, and $V(a, b, x, y) = 1$ if and only if

$$a \oplus b = x \wedge y \,,$$

where $\oplus$ is addition modulo 2 and $\wedge$ the logical AND (i.e., multiplication of bits).

The following can be proven easily by enumerating all possible (deterministic) strategies, i.e., all possible functions $f, g : \{0, 1\} \to \{0, 1\}$.

Proposition 3.5. $v_c(\mathfrak{G}_{\text{CHSH}}) = \frac{3}{4}$.

Surprisingly, by means of a quantum strategy, Alice and Bob can do better. This is sometimes also referred to as a *violation of Bell's inequality*.

Proposition 3.6. $v_q(\mathfrak{G}_{\text{CHSH}}) = \cos(\frac{\pi}{8})^2 = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85$.

We only prove that $v_q(\mathfrak{G}_{\text{CHSH}})$ is at least the claimed value; proving that it cannot be bigger is more involved (and will actually follow from the claims in Section 3.3).

Proof. The shared state is taken to be an EPR pair $|\Phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) \in \mathcal{H}_{\mathsf{A}} \otimes \mathcal{H}_{\mathsf{A}}$, where Alice keeps qubit A and Bob B. If $x = 0$ then Alice measures A in the computational basis $\{|0\rangle, |1\rangle\}$ to obtain a, if $x = 1$ then she measures it in the Hadamard basis $\{H|0\rangle, H|1\rangle\}$. If $y = 0$ then Bob measures A in the Breidbart basis $\{B_0|0\rangle, B_0|1\rangle\}$ to obtain b, if $x = 1$ then he measures it in the Breidbart basis $\{B_1|0\rangle, B_1|1\rangle\}$, where

$$\begin{aligned} B_0|0\rangle &= \cos(\tfrac{\pi}{8})|0\rangle + \sin(\tfrac{\pi}{8})|1\rangle & \qquad B_1|0\rangle &= \cos(\tfrac{\pi}{8})|0\rangle - \sin(\tfrac{\pi}{8})|1\rangle \\ B_0|1\rangle &= -\sin(\tfrac{\pi}{8})|0\rangle + \cos(\tfrac{\pi}{8})|1\rangle & \qquad B_1|1\rangle &= \sin(\tfrac{\pi}{8})|0\rangle + \cos(\tfrac{\pi}{8})|1\rangle \end{aligned}$$

as illustrated in Figure 3. By going through all the cases, it can now be verified that the winning probability is $\cos(\frac{\pi}{8})^2$, as claimed. For example, consider the case $x = 0$, so that Alice measures qubit A is the computational basis. As a result, she observes $a \in \{0, 1\}$ and the joint state collapses to $|a\rangle|a\rangle$. Now, no matter what y is, when Bob measures his qubit, which now is in state $|a\rangle$, in the basis determined by y, he observes $b = a$, and thus $a \oplus b = 0 = x \wedge y$ holds (remember that $x = 0$), with probability

$$p_a = |\langle a|B_y|a\rangle|^2 = \cos(\tfrac{\pi}{8})^2 \,.$$

The case $x = 1$ can be seen equally easily from the picture, although the formal computation is slightly more involved, using that the Hadamard basis can be written as $H|0\rangle = \cos(\frac{\pi}{4})|0\rangle +$

$\sin(\frac{\pi}{4})|1\rangle$ and $H|1\rangle = \sin(\frac{\pi}{4})|0\rangle - \cos(\frac{\pi}{4})|1\rangle$, and using some laws of trigonometry. Here, Alice observes $a \in \{0,1\}$ and the joint state collapses to $H|a\rangle \otimes H|a\rangle$. Now, if $y = 0$ so that Bob measures in basis $\{B_0|0\rangle, B_0|1\rangle\}$, the probability that he observes $b = a$, so that $a \oplus b = 0 = x \wedge y$ is satisfied, equals

$$p_a = |\langle a|HB_0|a\rangle|^2 = |\cos(\tfrac{\pi}{4})\cos(\tfrac{\pi}{8}) + \sin(\tfrac{\pi}{4})\sin(\tfrac{\pi}{8})|^2 = |\cos(\tfrac{\pi}{4} - \tfrac{\pi}{8})|^2 = \cos(\tfrac{\pi}{8})^2 .$$

Finally, if $y = 1$ so that Bob measures in basis $\{B_1|0\rangle, B_1|1\rangle\}$, the probability that he observes $b = a \oplus 1$, so that $a \oplus b = 1 = x \wedge y$ is satisfied, equals

$$p_{a\oplus 1} = |\langle a|HB_0|a \oplus 1\rangle|^2 = |\sin(\tfrac{\pi}{4})\cos(\tfrac{\pi}{8}) + \cos(\tfrac{\pi}{4})\sin(\tfrac{\pi}{8})|^2 = |\sin(\tfrac{\pi}{4} + \tfrac{\pi}{8})|^2 = \cos(\tfrac{\pi}{8})^2 .$$

This proves the claim. □

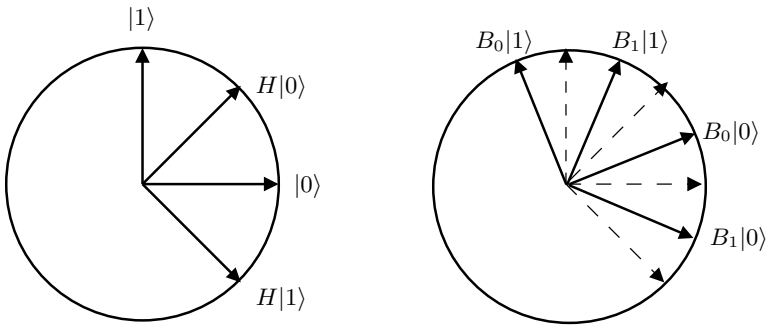

Fig. 3. Quantum strategy for CHSH.

With a suitable non-signaling theory, Alice and Bob can win CHSH with certainty.

Proposition 3.7. $v_{ns}(\mathfrak{G}_{\text{CHSH}}) = 1$.

Proof. Consider the hypercorrelation $P_{AB|XY}$ defined as

$$P_{AB|XY}(u, u \oplus (x \wedge y)|x, y) = \frac{1}{2}$$

for all $u, x, y \in \{0,1\}$, and with the remaining probabilities set to 0. It is easy to verify that this distribution is non-signaling: $P_{A|XY}(\cdot|x,y) = \frac{1}{2} = P_{B|XY}(\cdot|x,y)$ for all $x, y \in \{0,1\}$. And, it gives rise to a winning probability of 1: for any $x, y \in \{0,1\}$, the pairs (a,b) with $a \oplus b \neq x \wedge y$ have 0 probability. □

3.3 Information Causality

The nonlocality of nature is not only a theoretical artifact of quantum mechanics, but it is also supported by experiments, even though, due to noise and losses, these experiments are not yet "loophole-free". A natural question to ask is: why is nature (apparently) nonlocal? Another natural question to ask is: since it seems to be nonlocal, why is it not *more* nonlocal, i.e., why can CHSH be won with probability "only" (roughly) 85% but not with probability 90%, or even with certainty? What is special about these 85%? Is there an explanation for this bound that does not rely on quantum mechanics?

We want to address this question and give some answer. The goal is to specify some *fundamental principle* that one would expect to hold from a "good" physical theory that should

describe our world, and to show that any physical theory that beats the 85% must necessarily violate this principle.

One candidate for such a fundamental principle is the **non-signaling** principle, which states that distant parties, say Alice and Bob, cannot communicate by performing local operations only on their respective physical systems. In order to communicate, they must *physically send* information. However, as we have seen in the previous section, the non-signaling value of CHSH is 1; thus, beating the 85% does not contradict the non-signaling principle. To explain the apparent bound of 85%, we will generalize the non-signaling principle. The resulting principle is called **information causality**. Informally, it states that if Alice wants to communicate k bits of information then she necessarily must *physically send* k bits of information to Bob. We now make this precise.

We consider two distant parties Alice and Bob, holding respective subsystems of some joint system AB. Note, we consider an arbitrary (non-signaling) physical theory, so we do not know how the system is described nor how it behaves. Now, Alice is given a uniformly distributed n-bit string $U = (U_1, \ldots, U_n)$, i.e., randomly chosen from $\{0,1\}^n$, and Bob's goal is to "access" the bit U_C, where the choice bit $C \in \{1, \ldots, n\}$ is not known to Alice. To this end, Alice performs a local operation (possibly depending on U) on her system A to obtain a k-bit string M, which she then physically sends to Bob, and Bob performs a local operation (possibly depending on C and M) on his system B to obtain a bit W, supposedly equal to U_C (see Figure 4).

Again, for an arbitrary physical theory, we do not know how to describe the behavior of the system AB. Nevertheless, we can consider the hypercorrelation $P_{MW|U(M'C)}$ that is specified by the behavior of AB and by how Alice and Bob make use of it. Note that in the actual game, Bob puts $M' = M$ into the system; however, the physical theory must also specify what happens in case Bob is given $M' \neq M$.

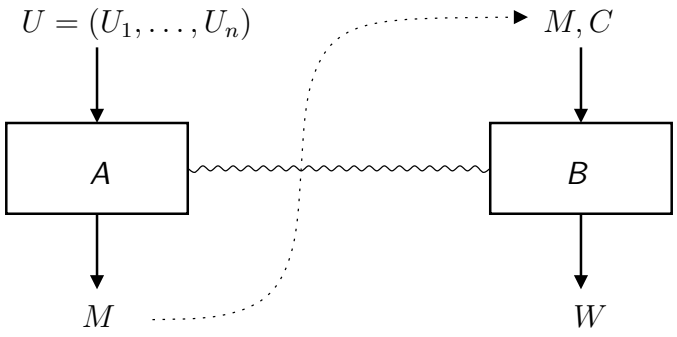

Fig. 4. Accessing U_C with the help of a k-bit message M.

The question of interest is: how many bits of U can Bob access this way? Obviously, by having Alice send $M = (U_1, \ldots, U_k)$, Bob is able to access k bits (even without making use of AB), but not more. Information causality now states that by means of *any* strategy, Bob cannot access more than k bits, i.e., cannot access more bits of U than he physically receives from Alice, so that the above trivial strategy is actually optimal. Before making this formal, we want to discuss potential strategies for which Bob can get *some* but not full information on some bits of U, e.g., for which $W = U_C$ with probability $\frac{3}{4}$ if $C = 1$, $W = U_C$ with probability $\frac{5}{6}$ if $C = 2$, etc. How do we then count the number of bits that Bob can access in such a case? This can be done by means of the **mutual information** measure I (see Appendix B.4). The mutual information $I(W; U_C)$ is a meaningful measure of how much information W gives on

U_C. The number of bits that Bob can access can then be "counted" as

$$\mathfrak{I}(P_{MW|U(M'C)}) = \sum_{c=1}^{n} I(W;U_C|C=c)\,.$$

In some sense, $\mathfrak{I}$ does not count the number of *physical* bits that Bob can access, but the number of *information* bits. In the above trivial strategy, we obviously have that $I(W;U_C|C=c) = 1$ for $c \in \{1,\ldots,k\}$ and $I(W;U_C|C=c) = 0$ for $c \in \{k+1,\ldots,n\}$, such that indeed $\mathfrak{I} = k$.

Definition 3.8. *A hypercorrelation $P_{AB|XY}$ is said to satisfy* **information causality***, if*

$$\mathfrak{I}(P_{MW|U(M'C)}) \leq k$$

for any hypercorrelation $P_{MW|U(M'C)}$ that is implied by $P_{AB|XY}$ and is of the form $U \in \{0,1\}^n$, $C \in \{1,\ldots,n\}$, $M, M' \in \{0,1\}^k$, and $W \in \{0,1\}$.

Being *implied* by $P_{AB|XY}$ should be understood in that if some physical theory permits the hypercorrelation $P_{AB|XY}$ (by means of local operations on a bipartite system), then it also permits $P_{MW|U(M'C)}$. The formal definition is given as follows.

Definition 3.9. *A hypercorrelation $P_{AB|XY} : \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \to [0,1]$* **implies** *a hypercorrelation $P_{A'B'|X'Y'} : \mathcal{A}' \times \mathcal{B}' \times \mathcal{X}' \times \mathcal{Y}' \to [0,1]$ if $P_{AB|XY} \rightsquigarrow P_{A'B'|X'Y'}$, where "$\rightsquigarrow$" is the transitive relation inductively defined by the following set of rules (see also Figure 5):*

- Reflexivity: $P_{AB|XY} \rightsquigarrow P_{AB|XY}$.
- Shared randomness: $P_{AB|XY} \rightsquigarrow P_{RR|\emptyset\emptyset} = P_R$ *for any distribution* $P_R : \mathcal{R} \to [0,1]$.
- Processing in- and outputs: $P_{AB|XY} \rightsquigarrow P_{A'B'|X'Y'} = P_{f(A,x'),g(B,y')|X,Y}(\cdot,\cdot|f_\circ(\cdot), g_\circ(\cdot))$ *for any functions* $f_\circ : \mathcal{X}' \to \mathcal{X}$, $f : \mathcal{A}\times\mathcal{X}' \to \mathcal{A}'$, $g_\circ : \mathcal{Y}' \to \mathcal{Y}$, *and* $g : \mathcal{B}\times\mathcal{Y}' \to \mathcal{B}'$.
- Concurrent composition: *If* $P_{AB|XY} \rightsquigarrow P_{A'B'|X'Y'}$ *and* $P_{AB|XY} \rightsquigarrow P_{A''B''|X''Y''}$ *then also* $P_{AB|XY} \rightsquigarrow P_{A'A''B'B''|X'X''Y'Y''} = P_{A'B'|X'Y'}P_{A''B''|X''Y''}$.
- Sequential composition: *If* $P_{AB|XY} \rightsquigarrow P_{A'B'|X'Y'}$ *and* $P_{AB|XY} \rightsquigarrow P_{A''B''|A'B'}$ *then also* $P_{AB|XY} \rightsquigarrow P_{A'A''B'B''|X'Y'} = P_{A'B'|X'Y'}P_{A''B''|A'B'}$.

It is easy to see that the three classes of hypercorrelations that are of particular interest to us, are closed under implication: if $P_{AB|XY}$ is classical/quantum/non-signaling and if $P_{AB|XY} \rightsquigarrow P_{A'B'|X'Y'}$, then $P_{A'B'|X'Y'}$ is classical/quantum/non-signaling as well.

First we show that quantum mechanics is consistent with information causality.

Proposition 3.10. *Every quantum hypercorrelation satisfies information causality.*

Proof. For a classical and for a quantum hypercorrelation $P_{AB|XY}$, we can actually make use of the fact that we know how to describe the system AB that gives rise to $P_{AB|XY}$. For a classical hypercorrelation, we may assume AB to be common randomness (described by means of random variables A and B with $\mathsf{A} = \mathsf{B}$) and $P_{AB|XY}$ is obtained by local computations, and for a quantum hypercorrelation, AB is a bipartite quantum system (described by a quantum state) and $P_{AB|XY}$ is obtained by local measurements (as specified in Definition 3.1).

Our proof can be appreciated for a *classical* system AB, using the properties of the mutual information as introduced in Appendix B.4. However, the very same proof also works for a *quantum* system AB, based on the fact that the notions of Shannon entropy and mutual information for random variables can be extended to quantum systems (one then usually speaks of the **Von Neumann entropy**). Although not all the properties of the classical Shannon entropy and

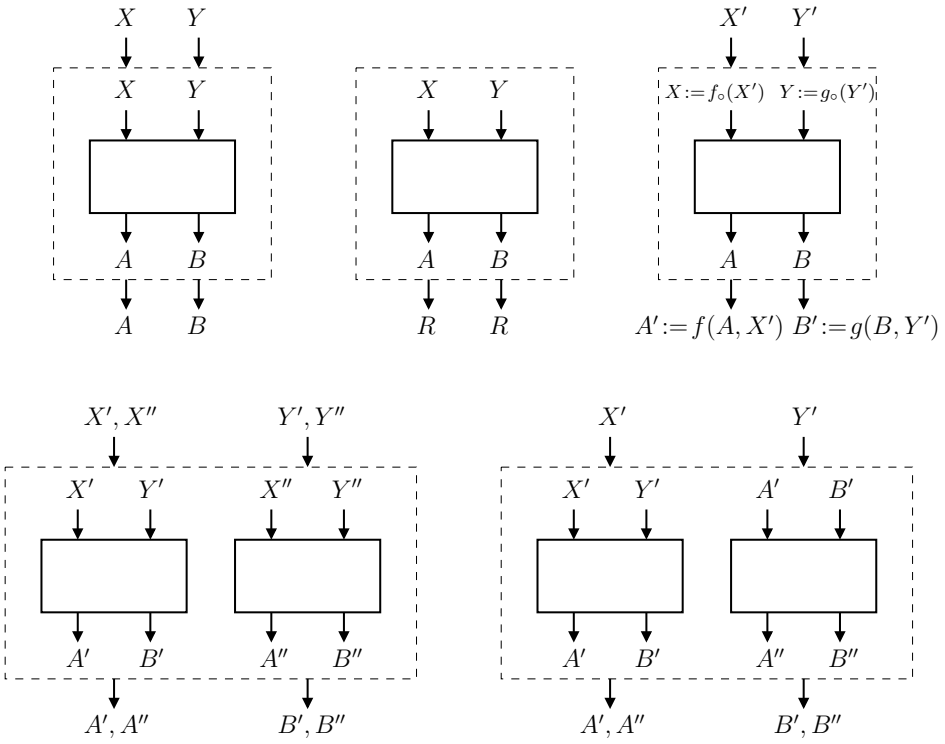

Fig. 5. Basic rules for implication: the inner hypercorrelation(s) implies/-y the outer.

mutual information carry over to the quantum setting (e.g. $I(\mathsf{A};\mathsf{B}) < 0$ becomes possible), all the steps of the proof below do (some with some more involved reasoning though).

By applying the chain rule (see Lemma B.16) twice, by the independence of U from B, and by using some elementary properties of the mutual information, we get that

$$I(U;M\mathsf{B}) = \overbrace{I(U;\mathsf{B})}^{=0} + I(U;M|\mathsf{B}) = I(M;U\mathsf{B}) - \overbrace{I(M;\mathsf{B})}^{\geq 0} \leq I(M;U\mathsf{B}) \leq H(M) \leq k$$

Furthermore, again by using the chain rule twice, using the independence of the different bits of U, and by elementary properties, we obtain

$$\begin{aligned} I(U;M\mathsf{B}) &= I(U_1;M\mathsf{B}) + I(U_2\cdots U_n;M\mathsf{B}|U_1) \\ &= I(U_1;M\mathsf{B}) + I(U_2\cdots U_n;M\mathsf{B}U_1) - I(U_2\cdots U_n;U_1) \\ &= I(U_1;M\mathsf{B}) + I(U_2\cdots U_n;M\mathsf{B}U_1) \\ &\geq I(U_1;M\mathsf{B}) + I(U_2\cdots U_n;M\mathsf{B})\,. \end{aligned}$$

Iterating the above $n-1$ times, we obtain

$$I(U;M\mathsf{B}) \geq \sum_{c=1}^{n} I(U_c;M\mathsf{B}) = \sum_{c=1}^{n} I(U_c;M\mathsf{B}|C=c)$$

where the last inequality follows from the fact that C is chosen independently. Finally, since for every fixed choice for C, W is computed by local operation on M and B, data processing (see Lemma B.10) implies that $I(U_c;W|C=c) \leq I(U_c;M\mathsf{B}|C=c)$ and hence

$$I(U;M\mathsf{B}) \geq \sum_{c=1}^{n} I(U_c;W|C=c) = \sum_{c=1}^{n} I(U_C;W|C=c)\,.$$

This proves the claim. □

Now, we show that any physical theory that beats quantum mechanics in the CHSH game necessarily violates information causality. Consequently, if we consider information causality as a fundamental principle, this explains the bound of roughly 85% in winning CHSH. A hypercorrelation $P_{AB|XY} : \mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y} \to [0,1]$ is **binary** if $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0,1\}$.

Theorem 3.11. *Any non-signaling binary hypercorrelation $P_{AB|XY}$ that has a CHSH value $v[P_{AB|XY}](\mathfrak{G}_{\text{CHSH}}) > \frac{1}{2} + \frac{1}{2\sqrt{2}}$ violates information causality.*

The proof we give is of algorithmic nature. We treat the given hypercorrelation $P_{AB|XY}$ as a "black-box". It can be queried by parties Alice and Bob, and upon respective queries X and Y it outputs A (to Alice) and B (to Bob) according to the corresponding distribution (independent of any additional information Alice and Bob might hold, when conditioned on X and Y). Given sufficiently many independent such "black-boxes", plus shared randomness, we show how Alice and Bob can beak the information causality bound $\mathfrak{I}(P_{MW|U(M'C)}) \leq k$, where $P_{MW|U(M'C)}$ is determined by Alice and Bob's actions. This way of constructing $P_{MW|U(M'C)}$ is consistent with the notion of implication as defined in Definition 3.9.

Proof. We start the proof by showing that any non-signaling binary hypercorrelation $P_{AB|XY}$ with $v[P_{AB|XY}](\mathfrak{G}_{\text{CHSH}})$ close enough to 1 violates information causality with parameters $n = 2$ and $k = 1$, i.e., Bob can access strictly more than one out of two bits with the help of a 1-bit message. For later convenience, we index the bits of U as $U = (U_0, U_1)$, and thus let C be in $\{0,1\}$.

By assumption, on random input bits X and Y, Alice and Bob can obtain two respective bits A and B that satisfy $A \oplus B = X \cdot Y$ except with probability $1 - \varepsilon$, for some small $\varepsilon \geq 0$. By Lemma 3.12 below, we may actually assume without loss of generality that $P[A \oplus B = X \cdot Y | X = x, Y = y] = 1 - \varepsilon$ for all $x, y \in \{0,1\}$, i.e., that an "error" happens independently of the actual inputs. Thus, the binary random variable $E := A \oplus B \oplus X \cdot Y$, which indicates whether $A \oplus B = X \cdot Y$ or not, satisfies $P_{E|XY}(1|\cdot,\cdot) = \varepsilon$.

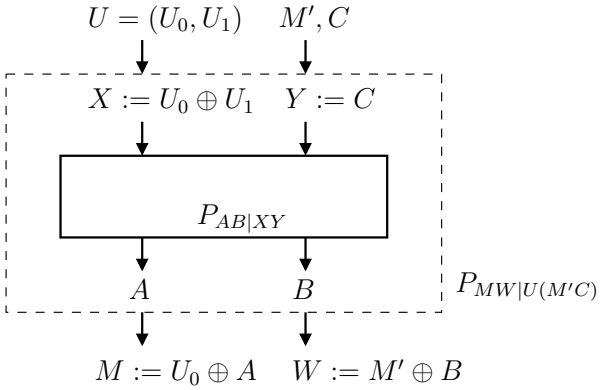

Fig. 6. $P_{MW|U(M'C)}$ implied by $P_{AB|XY}$. Note that M' is only used in the post-processing.

In order to violate information causality, Alice and Bob choose X and Y as follows. Alice chooses $X = U_0 \oplus U_1$, and Bob chooses $Y = C$. As a result, Alice and Bob obtain A and B that satisfy $A \oplus B = X \cdot Y \oplus E$. Alice then sends the 1-bit message $M = U_0 \oplus A$ to Bob, who computes W as $W = M \oplus B$ (see Figure 6). A simple calculation shows that

$$\begin{aligned} W = M \oplus B &= U_0 \oplus A \oplus B = U_0 \oplus X \cdot Y \oplus E \\ &= U_0 \oplus (U_0 \oplus U_1) \cdot C \oplus E = U_0 \cdot (1 \oplus C) \oplus U_1 \cdot C \oplus E = U_C \oplus E\,. \end{aligned}$$

It follows that $P_{W|U_0U_1C}(u_c|u_0,u_1,c) = P_{E|U_0U_1C}(0|u_0,u_1,c) = P_{E|XY}(0|u_0 \oplus u_1, c) = 1-\varepsilon$ for all $u_0, u_1, c \in \{0,1\}$, and thus

$$I(W;U_c|C{=}c) = \mathrm{H}(W|C{=}c) - \mathrm{H}(W|U_c, C{=}c) = 1 - \mathrm{h}(\varepsilon)\,,$$

where h denotes the binary entropy function, and $\mathrm{H}(W|C{=}c) = 1$ holds because U_c is uniformly random and independent of E (conditioned on $C = c$), and therefore W is uniformly random (conditioned on $C = c$).

It now follows that if $\mathrm{h}(\varepsilon) < \frac{1}{2}$ then $I(W;U_0|C{=}0) + I(W;U_1|C{=}1) > 1$, and information causality is violated. A numerical calculation shows that this is the case if $\varepsilon \lesssim 0.11$, i.e., if a CHSH value of (approximately) 0.89 is achieved.

In order to close the gap, we consider $n = 2^\ell$ for an arbitrary $\ell \in \mathbb{N}$, but keep $k = 1$. For convenience, we write $U \in \{0,1\}^n$ as $U = (U_s)_{s\in\{0,1\}^\ell}$ and $C \in \{0,1\}^\ell$. We argue by induction on ℓ that by means of local operations Alice and Bob can obtain bits M and B, respectively, such that $U_C = M \oplus B \oplus E$, where E is such that $P_{E|UC}(1|\cdot,\cdot) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^\ell$. The induction basis, i.e., the case $\ell = 1$, is covered above. For the induction step, we split $U = (U_s)_{s\in\{0,1\}^\ell}$ into $U_0 = (U_{0|s''})_{s''\in\{0,1\}^{\ell-1}}$ and $U_1 = (U_{1|s''})_{s''\in\{0,1\}^{\ell-1}}$, and write C as $C = C'|C''$ where $C' \in \{0,1\}$ and $C'' \in \{0,1\}^{\ell-1}$, and "|" denotes concatenation of strings. By induction hypothesis, Alice and Bob can obtain bits M_0 and B_0, and M_1 and B_1, such that $U_{0|C''} = M_0 \oplus B_0 \oplus E_0$ and $U_{1|C''} = M_1 \oplus B_1 \oplus E_1$, where $P_{E_0|UC}(1|\cdot,\cdot) = P_{E_0|U_0C''}(1|\cdot,\cdot) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^{\ell-1}$, and the corresponding for $P_{E_1|UC}(1|\cdot,\cdot)$.[13] Finally, by applying the base case (using yet another independent instantiation of the hypercorrelation) to (M_0, M_1) and C', Alice and Bob can obtain bits M and B', respectively, such that $M_{C'} = M \oplus B' \oplus E'$, where $P_{E'|M_0M_1UC}(1|\cdots) = P_{E'|M_0M_1C'}(1|\cdots) = \varepsilon$. Furthermore, since independent instantiations of the hypercorrelation are used, $P_{E'E_0E_1|UC} = P_{E'}P_{E_0}P_{E_1}$. Thus, by setting $B = B' \oplus B_{C'}$ and $E = E' \oplus E_{C'}$ (see Figure 7), we have that

$$M \oplus B \oplus E = M \oplus B' \oplus B_{C'} \oplus E' \oplus E_{C'} = M_{C'} \oplus B_{C'} \oplus E_{C'} = U_{C'|C''} = U_C$$

where

$$\begin{aligned} P_{E|UC}(1|\cdot,\cdot) &= P_{E'E_{C'}|UC}(0,1|\cdot,\cdot) + P_{E'E_{C'}|UC}(1,0|\cdot,\cdot) = P_{E'}(0)P_{E_{C'}}(1) + P_{E'}(1)P_{E_{C'}}(0) \\ &= (1-\varepsilon)\big(\tfrac{1}{2} - \tfrac{1}{2}(1-2\varepsilon)^{\ell-1}\big) + \varepsilon\big(\tfrac{1}{2} + \tfrac{1}{2}(1-2\varepsilon)^{\ell-1}\big) = \tfrac{1}{2} - \tfrac{1}{2}(1-2\varepsilon)^{\ell}\,. \end{aligned}$$

This completes the induction proof.

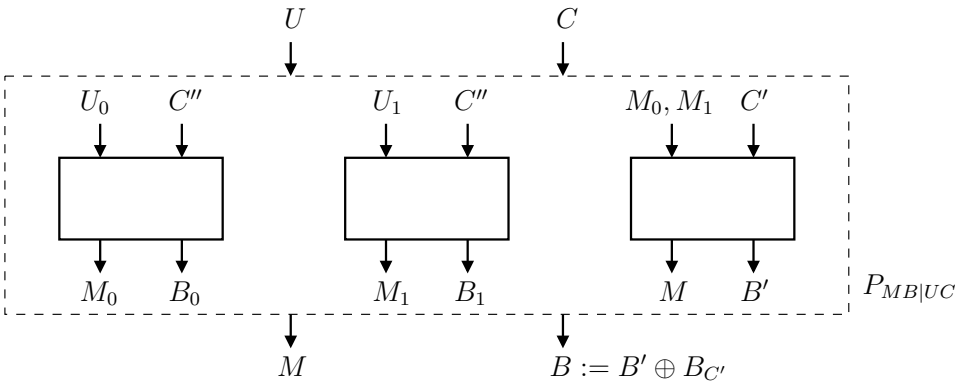

Fig. 7. $P_{MB|UC}$ implied by $P_{AB|XY}$, such that $W = M \oplus B$ violates information causality.

[13] The first equality should be understood as $P_{E_0|UC}(1|u,c) = P_{E_0|U_0C''}(1|u_0,c)$ for all $u = (u_0,u_1)$ and $c = c'|c''$, and it holds because U_1 and C' are (jointly) independent of U_0, C'' and E_0.

It remains to argue that this proves the claim of the theorem. By communicating the bit M to Bob, Bob can compute $W = M \oplus B$ which satisfies that $P_{W|UC}(u_c|u,c) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^\ell$ for all u and c. It follows that

$$I(W;U_C|C{=}c) = \mathrm{H}(W|C{=}c) - \mathrm{H}(W|U_c,C{=}c) = 1 - \mathrm{h}\big(\tfrac{1}{2} - \tfrac{1}{2}(1-2\varepsilon)^\ell\big) \geq \frac{(1-2\varepsilon)^{2\ell}}{2\ln(2)}\,,$$

where we use that $1 - \mathrm{h}(\frac{1+q}{2}) \geq \frac{q^2}{2\ln(2)}$ (see Lemma 3.13 below), and hence

$$\mathfrak{I} = \sum_{c\in\{0,1\}^\ell} I(W;U_C|C{=}c) \geq \frac{2^\ell(1-2\varepsilon)^{2\ell}}{2\ln(2)}\,.$$

Therefore, if $2(1-2\varepsilon)^2 > 1$, which is equivalent to $\varepsilon < \frac{1}{2} + \frac{1}{2\sqrt{2}}$, then $\mathfrak{I} > 1$ for large enough ℓ, and thus information causality is violated. □

Lemma 3.12. *Any non-signaling binary hypercorrelation $P_{AB|XY}$ implies a non-signaling binary hypercorrelation $P_{A'B'|X'Y'}$ with*

$$P[A' \oplus B' = X' \cdot Y' | X'{=}x', Y'{=}y'] = v[P_{AB|XY}](\mathfrak{G}_{\text{CHSH}})$$

for all $x', y' \in \{0,1\}$.

Note that for the original hypercorrelation, $P[A \oplus B = X \cdot Y] = v[P_{AB|XY}](\mathfrak{G}_{\text{CHSH}})$ only holds *on average* over the randomly chosen X and Y. For the implied hypercorrelation, it holds *for any* choice.

Proof. $P_{A'B'|X'Y'}$ can be obtained from $P_{AB|XY}$ with the help of shared randomness and local computation as follows. The shared randomness consists of random and independent bits ΔX and ΔY. Setting $X = X' \oplus \Delta X$, $Y = Y' \oplus \Delta Y$, $A' = A \oplus X \cdot \Delta Y$ and $B' = B \oplus \Delta X \cdot Y \oplus \Delta X \cdot \Delta Y$ (see Figure 8), we see that

$$X' \cdot Y' = (X \oplus \Delta X) \cdot (Y \oplus \Delta Y) = X \cdot Y \oplus X \cdot \Delta Y \oplus \Delta X \cdot Y + \Delta X \cdot \Delta Y = A' \oplus B'$$

if and only if $A \oplus B = X \cdot Y$. Therefore, for random and independent X and Y,

$$P[A' \oplus B' = X' \cdot Y' | X'{=}x', Y'{=}y'] = P[A \oplus B = X \cdot Y | X'{=}x', Y'{=}y'] = P[A \oplus B = X \cdot Y]$$

where the latter equality holds because A, B, X, Y is independent of X' and Y'. The claim follows because $v[P_{AB|XY}](\mathfrak{G}_{\text{CHSH}}) = P[A \oplus B = X \cdot Y]$. □

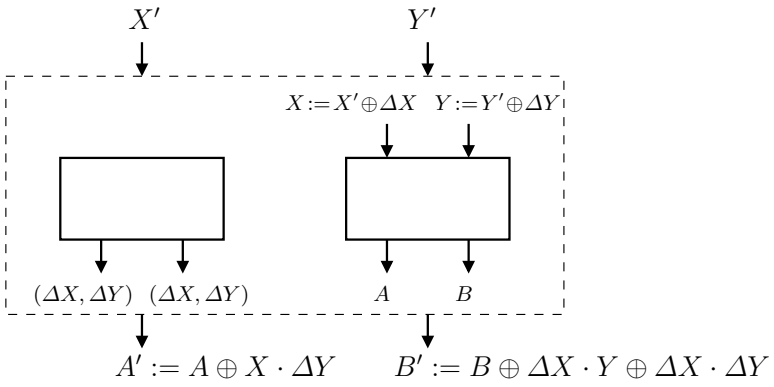

Fig. 8. Randomizing the inputs.

Lemma 3.13. *For all* $0 \leq q \leq \frac{1}{2}$ *it holds that* $1 - \mathrm{h}(\frac{1+q}{2}) \geq \frac{q^2}{2\ln(2)}$.

Proof (sketch). Both terms vanish for $q = 0$. Furthermore, the derivative of the left hand side equals $\frac{1}{2\ln(2)} \ln\bigl(\frac{1+q}{1-q}\bigr)$. By the mean value theorem, it is thus sufficient to show that this is not smaller than the derivative of the right hand side, $\frac{q}{\ln(2)}$. By the monotonicity of the exponential function, this is equivalent to showing that

$$\exp(2q) \leq \frac{1+q}{1-q}\,.$$

The latter can easily be shown by using the power series representation $\exp(2q) = \sum_n \frac{(2q)^n}{n!}$. □

4 Basic Concepts – Continued

4.1 Partial Trace and General Quantum Operations

Consider a joint quantum system AB, whose (mixed or pure) state is described by density matrix $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. In what state is, say, system B when considered as a single system? The answer is the **reduced** density matrix $\rho_B = \mathrm{tr}_A(\rho_{AB})$, where the **partial trace** tr_A is defined below. We also say that we **trace out** system A.

Definition 4.1. *The* **partial trace** $\mathrm{tr}_A : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_B)$ *is the unique linear function with*

$$\mathrm{tr}_A\bigl(|\varphi_A\rangle|\varphi_B\rangle\langle\psi_A|\langle\psi_B|\bigr) = \mathrm{tr}_A\bigl(|\varphi_A\rangle\langle\psi_A| \otimes |\varphi_B\rangle\langle\psi_B|\bigr) = \langle\psi_A|\varphi_A\rangle|\varphi_B\rangle\langle\psi_B|$$

for all $|\varphi_A\rangle, |\psi_A\rangle \in \mathcal{H}_A$ *and* $|\varphi_B\rangle, |\psi_B\rangle \in \mathcal{H}_B$. *Similarly,* $\mathrm{tr}_B \colon \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_A)$ *is defined.*

In other words, for an arbitrary density matrix $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ (which can be written as $\rho_{AB} = \sum_{i,j} |i\rangle\langle j| \otimes \sigma_{ij}$ where $\{|i\rangle\}_{i\in I}$ is the canonical basis), the partial trace $\mathrm{tr}_A(\rho_{AB})$ adds up the block-diagonal elements of ρ_{AB}; it is not too hard to see that $\mathrm{tr}_A(\rho_{AB}) \in \mathcal{D}(\mathcal{H}_B)$. When ρ_{AB} is clear from the context, then we may simply write ρ_B instead of $\mathrm{tr}_A(\rho_{AB})$.

The justification why the reduced density matrix $\rho_B = \mathrm{tr}_A(\rho_{AB})$ should be "the right" description of the system B follows from the observation that measuring part B of the joint state ρ_{AB} produces the same outcome distribution as when measuring ρ_B:

Proposition 4.2. *Measuring part* B *of a joint state* $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *produces the same outcome distribution as when measuring* $\rho_B = \mathrm{tr}_A(\rho_{AB}) \in \mathcal{D}(\mathcal{H}_B)$ *(wrt. to the same measurement).*

Proof. Due to linearity, it suffices to prove the claim for a pure state $\rho_{AB} = |\varphi\rangle\langle\varphi|$, where we may write $|\varphi\rangle = \sum_j \alpha_j |j\rangle|\psi_j\rangle$ for some orthonormal basis $\{|j\rangle\}_j$ of $\mathcal{H}_A$, where $\sum_j |\alpha_j|^2 = 1$ and $\||\psi_j\rangle\| = 1$. We can thus write

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) = \mathrm{tr}_A(|\varphi\rangle\langle\varphi|) = \sum_{j,k} \alpha_j \bar{\alpha}_k \langle j|k\rangle |\psi_k\rangle\langle\psi_j| = \sum_j |\alpha_j|^2 |\psi_j\rangle\langle\psi_j|\,.$$

Since we are merely interested in the outcome, but not the post-measurement state, we use the POVM formalism. Thus, when measuring part B of the original state, i is observed with probability

$$\begin{aligned}\mathrm{tr}\bigl((\mathbb{I} \otimes E_i)\rho_{AB}\bigr) &= \mathrm{tr}\bigl((\mathbb{I} \otimes E_i)|\varphi\rangle\langle\varphi|\bigr) = \sum_{j,k} \alpha_j \bar{\alpha}_k \,\mathrm{tr}\bigl((\mathbb{I} \otimes E_i)(|j\rangle\langle k| \otimes |\psi_k\rangle\langle\psi_j|)\bigr) \\ &= \sum_{j,k} \alpha_j \bar{\alpha}_k \,\mathrm{tr}\bigl(|j\rangle\langle k| \otimes E_i|\psi_k\rangle\langle\psi_j|\bigr) = \sum_{j,k} \alpha_j \bar{\alpha}_k \,\mathrm{tr}(|j\rangle\langle k|)\,\mathrm{tr}(E_i|\psi_k\rangle\langle\psi_j|)\end{aligned}$$