Quantum entanglement in non-local games, graph parameters and zero-error information theory

Scarpa, G.

Citation for published version (APA):
Chapter 1

Introduction

1.1 Is the world quantum?

Quantum mechanics is so counter-intuitive that it struggled to be accepted by its own creators. Niels Bohr once said [Hei71, page 206]:

“Those who are not shocked when they first come across quantum theory cannot possibly have understood it”.

Albert Einstein was convinced that quantum mechanics is wrong (or at least incomplete). He once asked to Abraham Pais [Pai79, page 907]:

“Do you really believe the moon exists only when you look at it?”.

Why was Bohr shocked and why was Einstein worried about the moon? The concern was about issues with locality and realism. These are two properties of a physical theory that any of us would find obvious and essential. Locality means that an action in a certain place cannot influence the state of a distant object instantaneously, i.e., the effects of an action travel at a finite speed. Realism means that all properties of a physical system have a specific, definite value, even before they are observed (measured). Such things seem so clearly true when looking at “real life”. One cannot call a waiter in a restaurant and immediately get his attention: the sound has to get there and it travels at a finite speed. Also, I am overweight even when I am not standing on the scales!

Einstein, Podolsky and Rosen, in their famous 1935 article [EPR35], observed that quantum physics, in addition to being a theory that explained several classically unexplained phenomena, also predicts the existence of objects with a weird behavior. They exhibited an example of such objects, a bipartite state now known as the EPR pair. This was an example of a large class of systems that are now called entangled. They thought that their example could be used to prove quantum physics wrong or incomplete. Instead, it created a new branch of research.
Let us give an intuitive description of their argument. Since quantum mechanics seems to be consistent with experimental results, it is often assumed to be a complete theory. A physical theory is complete when it perfectly predicts the outcome of a measurement of each definite property of a system. They argue, with the following thought experiment, that such an assumption leads to a contradiction. First, they consider Heisenberg uncertainty principle of quantum mechanics: the values of some pairs of properties, such as the position and the momentum of a particle cannot both be predicted perfectly.\textsuperscript{1} Second, they introduce the above-mentioned EPR pair: a pair of particles that are allowed to interact up to a certain moment and then separated and not allowed to interact anymore. This bipartite quantum system is designed such that if one measures the position of the first particle, then one knows perfectly the position of the second particle without observing it. Similarly, measuring the momentum of the first particle removes any uncertainty on the momentum of the second particle. Therefore, it seems that for the second particle both the position and the momentum have a definite value, even prior to observation. Here comes a “paradox”: if one assumes that quantum theory is complete, then the position and the momentum cannot both have definite values in real life (otherwise there would be no uncertainty principle). On the other hand, quantum mechanics allows to define an EPR pair, which shows that both these quantities have a definite value for the second particle prior to its observation.

There are two explanations of this apparent contradiction. One is that quantum mechanics is wrong, or incomplete. The other is that, somehow, in an EPR pair observing a property of the first particle also determines the value of the property of the second particle, instantaneously. Einstein disliked the second explanation and later called the phenomenon “spooky actions at a distance” [EBB71, page 158]. He believed there is a theory that explains nature while respecting both locality and realism.

The discussion remained open for many years, until Bell in 1964 [Bel64] proposed an experiment that made it possible to test if nature respects locality and realism. He designed a set of measurements of the properties of a bipartite physical system, such that if locality and realism are respected, then the measurements outcomes could not exceed a certain value. This is known as a Bell inequality. The non-locality of the EPR pair in quantum physics allowed to violate a Bell inequality and obtain measurement outcomes that are larger than predicted by classical physics or any other local realistic theory. Experiments of this kind have been done many times, starting from Aspect \textit{et al.} in 1981 [AGR81]. The results are surprising: nature violates a Bell inequality! The tests behaved according to the prediction of quantum physics. But hold your enthusiasm: this does not prove that quantum physics is the correct theory, it just disproves that nature

\textsuperscript{1}For example, if the position of a particle is certain, then quantum theory can only describe the result of a measurement of its momentum probabilistically.
behaves according to classical physics or any local realistic theory. Moreover, there are some loopholes. The problem with these Bell inequality violations is that the amount of violation is very small and this fact allows for (very contrived) classical explanations of the experimental results. These explanations often rely on unlikely combinations of things like noise, malfunction in measurement devices and incorrect post-selection of the results. Researchers have tried to get rid of loopholes in various ways, one of which is to design stronger Bell inequalities. The ideal situation is to get a huge, thus easily noticeable, difference between classical and quantum behavior using very small physical systems in the experiments. Unfortunately, Junge et al. [JPP+10] recently proved that on a bipartite system of local dimension \( n \), the maximum Bell inequality violation is of order \(^2 2^n\). Therefore, to obtain large Bell inequality violations one must work on large systems.

**Our 2 cents** Our contribution to the study of non-locality is 3-fold. In Chapter 2 of this thesis, we achieve the goal of finding very large Bell inequality violations. We design two experiments in the form of non-local games, and we prove that players relying on classical physics are outperformed by quantum players by a factor that is very close to the optimal violation proved by Junge and coauthors. In Chapter 3 and 4 we make use of non-locality and entanglement of physical systems in different settings. We use non-local games to approach mathematical problems on graphs, and we use entanglement to assist and improve communication in some problems of information theory. We proceed now to introduce each chapter individually.

### 1.2 Non-local games

The setting of “games” has been widely used in computer science. Sometimes we face complicated-looking problems, and we seek for a more intuitive equivalent way to describe and work with them. Non-local games are a good example of this: they make the concept of non-locality and Bell inequalities more accessible.

The setting we consider in this thesis is the following, illustrated in Figure 1.1. This is a game between two players, Alice and Bob, and a referee. A referee asks Alice a question \( x \) and asks Bob a question \( y \). He chooses the pair of questions \( x, y \) according to a probability distribution that is known to Alice and Bob. He receives back an answer \( a \) from Alice and an answer \( b \) from Bob. After that, he decides according to the rules if the players win or lose.\(^3\) During the game Alice and Bob are space-like separated: they are put so far away that information, which

\(^2\)The local dimension is the dimension of one part of the bipartite system. The term “order” should be interpreted intuitively as “up to a constant factor”.

\(^3\)Notice that this setting is different from the usual setting used in game theory. Alice and Bob are not competing, but are collaborating to both win the game.
travels at finite speed, cannot be exchanged between them until they produce the answers.

The value of the game is the maximum winning probability of the players. A Bell inequality as described above, in this setting is an upper bound on the classical value of the game. This means that whatever strategy classical players arrange before the game, they cannot exceed that probability of winning. Sometimes, however, there are quantum strategies that make smart use of entanglement and exceed the upper bound. Therefore, each game has a classical and quantum value, and for some games the quantum value is strictly larger than the classical. In this case we have a Bell inequality violation, and we quantify it by the ratio

\[
\frac{\text{quantum value}}{\text{classical value}}.
\]

Here is an example, due to Peres [Per91]. It is not the simplest and most common example, but it is very intuitive.\(^4\) Consider the magic square from Figure 1.2. It is a 3 × 3 grid that has to be filled in with natural numbers, such that the sum of each row is an even number, and the sum of each column is an odd number. A little thought shows that it is impossible to completely fill in the square.

Now let us base a non-local game on it. The referee lets Alice and Bob arrange a strategy, then separates them. He chooses randomly two numbers \(x, y\) from 1 to 3. Then he asks Alice to give him the numbers \(a_1, a_2, a_3\) of the \(x\)-th row, and Bob to give him the numbers \(b_1, b_2, b_3\) of the \(y\)-th column. The players lose if any of the following happens:

- The intersection of the column and the row is different, i.e., if \(a_y \neq b_x\),
- The sum of Alice’s numbers is not even,
- The sum of Bob’s numbers is not odd.

There are 9 possible question pairs \(x, y\) to this game. The classical players can easily win in 8 out of 9 cases by preparing a square like in Figure 1.2 and

\(^4\)The classic example is the CHSH game, explained in Section 2.1.
by completing the missing entry in two different ways: Alice with a 1 and Bob with a 2. They will win all the question pairs except the questions $x = 3, y = 3$. Therefore, since the referee selects the questions at random, the classical value of the game is at least $8/9$. It is actually equal to $8/9$, because a little thought shows that from a classical strategy that wins in all cases one can derive a completion of the whole square, which is impossible.

Surprisingly, there is a way to win in all 9 cases using entanglement. The players can share a carefully designed entangled state and perform appropriate operations on it to use the non-classical correlations to their advantage. They would always answer correctly even while the referee is sure that they are not communicating (because of their huge distance in space). Therefore in the quantum case the value of this game is 1.

Here is a non-technical intuition on what happens. Alice and Bob share an entangled quantum system before the game starts. During the game, Alice measures her part of the state in a way that depends on her input $x$, obtaining 3 numbers that sum to an even number. Alice’s actions instantly modify the quantum state on Bob’s side at a distance (spooky!). Bob’s measurement depends on $y$ and is designed such that when he reads out the numbers, the intersection with Alice’s row is always the correct number, and together with the other two (random) numbers sums to an odd number.

Basically, non-local correlations are used to “generate on-the-fly” a valid answer to any question pair, without any information being transmitted between Alice and Bob. Such a strategy allows to win the game but not to fill-in a whole magic square, which would be impossible. It is also important to notice that no information about Alice’s input is transferred to Bob during the game, and vice versa. The only thing that Bob knows by measuring his part of the state is that one of his three random numbers is equal to one of Alice’s random numbers, and it is the one in the right position. He knows that it is a wise choice to output such numbers.

**Our contribution** As argued above, the problem with Bell inequality violations, and with the known non-local games, is that the difference between the quantum and classical values is quite small. On the other hand, Junge *et al.* proved that the violation cannot be larger than the local dimension of the shared quantum state.

Our contribution is to exhibit two non-local games that exhibit a violation...
very close to optimal. They are formally discussed in Chapter 2. We informally introduce them here.

The first one is called *Hidden Matching*. The game is inspired by a communication problem in [BYJK08]. It goes as follows. Alice gets a random string of $n$ bits (with $n$ power of 2), and Bob gets a random matching on the $n$ indices.\footnote{A matching on $n$ indices is a set of $n/2$ distinct pairs $(i,j)$. For example, for the numbers between 1 and 8 a possible matching is $\{(1,3),(2,4),(5,8),(6,7)\}$.} Alice outputs a bit string, Bob outputs a pair of indices in his matching and a bit. The winning condition is a bit technical, but a rough intuition is that an analysis on Alice and Bob’s outputs should suggest that Bob knew the sum of Alice’s bits in the positions he declared.

The game is designed to be very difficult to classical players (they can win with probability roughly $\frac{1}{2} + \frac{1}{\sqrt{n}}$) but perfectly playable by Alice and Bob that share a quantum state of local dimension $n$ (they can win with probability 1). We obtain a very large separation by considering the deviation from $\frac{1}{2}$. When quantum players share an entangled state of local dimension $n$, the deviation from $\frac{1}{2}$ is roughly $\sqrt{n}$ times larger than in the classical case.

The second non-local game exhibits a larger separation than the other, but is not perfectly won by quantum players. It is called the *Khot-Vishnoi* game, and it is inspired by an example used by Khot and Vishnoi in [KV05].

Alice and Bob each are given as inputs a set of bit-strings. The two sets are not totally random: Alice’s receives a set of $n$ strings respecting some properties; Bob receives $n$ strings that are obtained from Alice’s by adding a random noise string $z$, which only the referee knows. Alice outputs a string $a$ from her set and Bob outputs a string $b$ from his set. They win the game if Bob’s string is the noisy version of Alice’s string, i.e., $b = a + z$.

This turns out to be an extremely hard problem for classical players. Since for any output of Alice $a$ there is only one good answer for Bob, by playing a random answer their winning probability is $1/n$. Basically, their best strategy is very close in winning probability to giving a random answer, thus the classical value is roughly $1/n$. It can be shown that quantum players, instead, by using $n$ EPR pairs, have a winning probability close to constant (but far from perfect). Therefore, for $n$ large enough, the quantum players sharing $n$ EPR pairs outperform the classical ones by a factor of order roughly $n$.

Thus, this game almost matches the upper bound given by Junge and coauthors. To date, it is the non-local game that shows the largest quantum advantage.

### 1.3 Graph parameters

Our first application of non-locality to another area is the study of quantum graph parameters in Chapter 3. In order to explain our contribution, it is essential to define graphs are some important graph parameters.
A graph is a simple yet important mathematical structure. It consists of a set of nodes (also called vertices), and a set of pairs of nodes called edges. Two nodes that form an edge are adjacent or neighbors. In the typical depiction of a graph the nodes are drawn as circles and the edges are lines connecting pairs of them. We can see an example in Figure 1.3. A graph parameter is a quantity that is associated to the graph and usually describes one of its properties. For example, basic graph parameters are its number of nodes and its number of edges. We talk about more interesting ones below.

Why are graphs so important? Basically, because many things can be represented with graphs. For example, we could represent a map, with the nodes being cities and the edges being roads connecting them. But also our family tree, a computer network, the structure of a molecule, the dependencies of software packages, and so on. It is surprising how many problems can be solved by constructing a graph and calculating one of its parameters.

Let us informally explain two of the most studied graph parameters (the formal definitions are in Section 3.2.1).

**Chromatic number** Suppose we are given a big piece of paper with the drawing of a graph. Then we are asked to color each circle using pencils, but never to fill in two adjacent nodes with the same color. Of course, one easy solution is to use one color per node, with a very artistic result. But now suppose each color pencil is very expensive, so we must use as few colors as possible. What is the minimum number of colors that we need to use? This is a graph parameter called the chromatic number. For example, the graph of Figure 1.3 has chromatic number 3, and Figure 1.4 shows an optimal coloring.

It turns out that computing the chromatic number of a graph is hard (it is one of the so-called NP-Hard problems. All known computer programs that solve this problem in general take a huge amount of time in proportion to the graph size. If the graph has $n$ nodes, the best known program will have to perform roughly $2^n n$ elementary operations. This number can be enormous. Even if our computer performs millions of operations per second, this will not prevent us to...
have to wait for ages as soon as we want to color 1000 nodes. One of the most
important open problems in mathematics, the P vs. NP problem, could be solved
if we find an efficient way to color a graph, or prove that there is none. It is hard
even just to decide in general if a given graph has chromatic number equal to 3.

Independence number Suppose we have a wolf, a goat and a cabbage. The
goat wants to eat the cabbage, and the wolf wants to eat the goat. How many of
these can we keep at home? Well, bringing the wolf and the cabbage is optimal
in this case. That was an easy instance of the problem. But now suppose we
have a big aquarium that we want to populate, and we go into a shop to buy fish.
The shop has hundreds of species, but not all species of fish can live with each
other. The shopkeeper gives us a graph where each node is a species of fish, and
two nodes form an edge if the two species are incompatible. How many different
species of fish can we put safely in our aquarium?

This problem is equivalent to finding a subset of the nodes such that no two
nodes are adjacent. Such a subset is called an independent set, and the size of
the largest independent set is a graph parameter called the independence number.
For example, the largest independent set in the graph of Figure 1.4 has 2 nodes.\footnote{Notice that each pair of nodes with the same color form an independent set: this is not a
coincidence, a coloring is a partitioning into independent sets!}
Computing the independence number of a graph is also a hard problem, so the
same discussion as for the chromatic number applies.

Lovász $\vartheta$ number Many famous mathematicians worked on the chromatic
number or the independence number problems (for example Erdős, Lovász, Knuth
and Schrijver).

A new twist on the problem came when the $\vartheta$ number of a graph was defined
in [Lov79]. This is a graph parameter efficiently computable with a semidefinite
program,\footnote{In fact, it started the whole field of semidefinite programming, a subfield of optimization.} and it was introduced for solving a open problem of Shannon in infor-
1.3. **Graph parameters**

information theory: the zero-error capacity of the 5-cycle (we will discuss this kind of problems in Section 1.4). The Lovász $\vartheta$ number lies “sandwiched” between the independence number and the chromatic number of the complement graph,\(^8\) i.e., for all graphs $G$ and their complement $\bar{G}$ it is true that

$$\text{independence number of } G \leq \vartheta \text{ number of } G \leq \text{chromatic number of } \bar{G}.$$ 

The bound works well for some classes of graphs (e.g., the so-called perfect graphs, where the inequalities above are equalities), while the gap is large for other graphs. The quantity is very interesting for computer science, and Knuth dedicated an excellent survey to it [KD93].

**Non-local games based on graph parameters** Researchers have tried to find better bounds for the chromatic and independence numbers. One line of research, formally started by [CMN+07] and [CLMW10] but implicitly present in previous works, uses quantum mechanics to define quantum graph parameters. The main tool we use to define quantum graph parameters is to consider non-local games based on the graph coloring and independent set problems.

We start with the **coloring game** on a graph, illustrated in Figure 1.5a. Suppose Alice and Bob claim that they have a coloring of a graph $G$ that uses $c$ colors, and a referee wants to test this claim. The referee can play the coloring game on $G$ with $c$ colors with Alice and Bob. He lets the players arrange a strategy and then he separates them, such that they cannot communicate with each other anymore. He gives Alice as input a node $x$ of the graph chosen randomly and asks her to color it with one of the $c$ possible colors. Alice gives him the color $a$ as output. The referee gives Bob as input a node $y$ of the graph chosen randomly, and asks him to color with one of the $c$ colors. Bob gives him the color $b$ as output. The referee then checks for the consistency of their outputs. If the players had the same node as input ($x = y$), then they must have given the same color as output ($a = b$). If players had adjacent nodes as input ($x, y$ form an edge), then they must have given different colors as output ($a \neq b$). If the referee finds an inconsistency, the players lose the game.

Can the classical players always win the coloring game on $G$ with $c$ colors? It depends on $c$. If $c$ is equal to or larger than the chromatic number of $G$, then the players can arrange a winning strategy that consists of a correct coloring. On the other hand, if $c$ is strictly smaller than the chromatic number, then the players must lose for at least one input pair. Here is a proof, slightly technical, of the claim. Suppose, towards a contradiction, that players can win the coloring game on $G$ with strictly fewer colors than the chromatic number. Then, we can construct a coloring for $G$ as follows. Note that since for all $x = y$ we have $a = b$, Alice and Bob must follow an equivalent strategy. This implies that Alice must

---

\(^8\)The complement of a graph is a new graph with the same set of nodes, but with an edge between two nodes if and only if there is no edge in the original graph.
output different colors when asked adjacent nodes. Therefore, a proper coloring can be constructed by observing what color Alice outputs for each input.\(^9\) The existence of such a coloring with \(c\) colors contradicts the assumption that the chromatic number is strictly larger than \(c\). Therefore, the chromatic number of the graph \(G\) is the smallest number \(c\) such that Alice and Bob can always win the coloring game on \(G\) with \(c\) colors.

Cameron et al. and Avis et al. [CMN\(^+\)07, AHKS06] have found that if the players are allowed to share quantum entanglement, then for some graphs they can win the coloring game using fewer colors than the chromatic number! They defined the quantum chromatic number of a graph as the minimum number of colors such that players that share entanglement can always win the coloring game on \(G\).

Winning the game by using entanglement is not equivalent to produce a valid coloring! Indeed, player produce their answers at random following a correlated probability distribution. How does this work? Here is an intuition (the technical definition is in Section 3.3). The players share an entangled state that, as explained above, exhibits non-classical correlations. Alice performs a measurement on her part of the state that depends on her input \(x\) and outputs the outcome. This measurement is designed to produce a random color from the set of allowed colors as outcome. Alice’s actions influence instantaneously the state on Bob’s side (again, spooky!). Bob performs a measurement on his part of the state that depends on his input \(y\) and outputs the measurement outcome. The measurements are carefully made to give the same (random) outcome for Alice and Bob if \(x = y\) and to give different (random) outcomes if \(x, y\) form an edge. Therefore, Alice and Bob will always win the game, and wisely designed measurements that make use the non-classical correlation allow, for some graphs, to win with a number of colors strictly less than the chromatic number.

A similar discussion can be done for the independence number (see Figure 1.5b). Consider the same non-local setting, where Alice and Bob now are trying to prove that the graph \(G\) contains an independent set of size \(t\). The referee lets them arrange a strategy, then separates them. He gives to Alice the input \(x\) a random number between 1 and \(t\). Alice gives as output a node \(a\), which is meant to correspond to the \(x\)-th element of the independent set. Similarly, the referee gives to Bob as input number \(y\) between 1 and \(t\), and receives a node \(b\) as output. He then checks for consistency: if \(x = y\), then he must have received outputs \(a = b\), and if \(x \neq y\), then he must have received outputs from an independent set, therefore \(a, b\) should not form an edge. A similar analysis as with the coloring game shows that the independence number of a graph is the maximum \(t\) such that the classical players can always win the independent set game on the graph with

\(^9\)A technical note: since the players must always win, we can rule out any use of randomness. Any randomness used to make choices in the strategy must result in winning outputs, therefore the players can just fix the random outcomes and play an equivalent deterministic strategy.
1.3. Graph parameters

(a) Coloring game: Alice and Bob each receive an input from the vertex set $V$ of the graph, and produce one out of $c$ colors as output.

(b) Independent set game: Alice and Bob each receive as input a number from 1 to $t$, and produce an output from the vertex set $V$.

Figure 1.5: Setup of non-local games based on graph parameters.

size-parameter $t$. Cubitt et al. [CLMW10] initiated the study of the quantum version of this game in the context of zero-error information theory (which is the topic of the next section). A more graph-parametric version of the problem has been given in [RM12]. Similarly to the coloring game, also for the independent set game, players that share entanglement have an advantage. For some graphs, they can always win the independent set game with a size-parameter $t$ strictly larger than the independence number. The quantum independence number of a graph is the maximum $t$ such that the quantum players can always win the independent set game on the graph with size-parameter $t$.

These quantum parameters are natural bounds on the classical ones, because the presence of entanglement cannot be a disadvantage for the players. It turns out that the efficiently-computable $\vartheta$ number is “sandwiched” also in between these quantum parameters. Therefore, the quantum parameters are potentially tighter bounds to the classical counterparts than the $\vartheta$ number. However, it is an open problem to determine their computational complexity.

Our contribution Chapter 3 contains our work on quantum graph parameters. We prove properties of the quantum chromatic number, especially in comparison with other graph quantities. We then find a surprising relationship between graphs exhibiting a separation between quantum and classical chromatic number and objects from the foundation of quantum theory known as Kochen-Specker sets. These are sets of vectors with some specific properties that were used to prove the Kochen-Specker theorem, a no-go result that excludes some underlying explanations for quantum mechanical behavior. Kochen and Specker exhibited a finite set of measurements (that are made of the vectors mentioned above) such that it is impossible to choose a pre-determined outcome to each measurement. In short, this construction proves that if there is a classical model based on variables describing the state of quantum system, then those variables cannot be
non-contextual, i.e., independent of the particular measurement. This result is similar to Bell inequalities but without a bipartite setting.

We defined a generalization of these important objects, called projective Kochen-Specker sets. We find that to win the coloring game with fewer colors than the chromatic number, Alice and Bob’s strategies must consist of projective Kochen-Specker sets.

Our other contribution is about the independence number. We find several classes of graphs for which there is a separation between quantum and classical independence numbers. One class is based on our projective Kochen-Specker sets, one is based on the quantum chromatic number and the last one is based on graph states (well-known quantum systems). Moreover, we find that starting from any non-local game, we can construct a graph based on the game description and calculate a bound on the quantum value of the game from the quantum independence number of the graph. Computing this graph parameter appears to be a useful tool for studying non-locality in general.

1.4 Zero-error information theory

Our second application of non-locality is to consider the advantages of entanglement in some information theory problems. The problems are about sending a message from Alice to Bob efficiently in presence of various kinds of noise. Namely, we study the channel problem, the (dual) source problem and their generalization, the source-channel problem. We focus on the zero-error setting, namely the setting where all the information must be transmitted without any possibility of errors. The main reasons are two: first, it has been shown that entanglement is not beneficial in the asymptotic bounded-error setting\textsuperscript{10} [BSST02], and second, the techniques for the zero-error framework come from graph theory.

**Channel problem** The seminal paper by Shannon [Sha56] introduced the concept of zero-error capacity of a noisy channel. This setting is represented in Figure 1.6a.

Imagine that Alice and Bob communicate through a noisy channel that has an input set $S$ and an output set $V$. Each element $s \in S$ is associated to $V(s)$, a subset of the output set. Such subsets for distinct $s, t \in S$ may overlap. The channel is noisy in the sense that each element $s \in S$, when sent into the channel, can produce as output any element of $V(s)$. Therefore, there are some outputs of the channel on Bob’s side that do not allow Bob to unequivocally distinguish between two inputs on Alice’s side. Two inputs $s, t \in S$ are confusable if there exists an output $v$ such that both of them can produce $v$ as output, i.e., their output sets intersect.

\textsuperscript{10}Where we allow the protocols to make a small probability of mistake while decoding the message.
1.4. Zero-error information theory

Let us provide a simple example of such a channel. Imagine that the input set consists of the numbers from 1 to 6, the output set is \{odd, even, > 3, ≤ 3\} and the obvious rule determines the output subsets. For example, we have that:

- If Alice sends “2”, Bob receives either “even” or “≤ 3”,
- If Alice sends “3”, Bob receives either “odd” or “≤ 3”,
- If Alice sends “5”, Bob receives either “odd” or “> 3”.

Therefore, 2 and 3 are confusable, 3 and 5 are confusable but 2 and 5 are not.

The zero-error channel problem asks the following question:

Alice communicates to Bob through a noisy channel with input set \( S \), output set \( V \) and output subsets \( \{V(s)\}_{s \in S} \). What is the maximum number of bits of information that Alice can send to Bob on average per use of the channel, without any chance of a mistake?

In this introduction, for simplicity, we illustrate the simpler problem of a single use of the channel. To solve this problem, it is useful to introduce the confusability graph of the channel. This is a graph that has node set \( S \) and where \( s, t \in S \) are adjacent if they are confusable. The confusability graph of our simple example is the graph in Figure 1.3. To maximize the number of bits sent with a single use of the channel, Alice and Bob may select an independent set \( I \subseteq S \) of the confusability graph and restrict the inputs to that set. This way, the output subsets of the elements of \( I \) are disjoint, and Bob, upon receiving an output from the channel, can identify without error which input has been sent by Alice. How many bits have been sent? If \( I \) has size \( t \), then sending one out of \( t \) possible messages transfers \( \log_2(t) \) bits of information.\(^{11}\) Therefore, the channel capacity

\(^{11}\)When measuring the amount of information, it is possible to have a non-integer number of bits.
with a single use of the channel is given by the logarithm of the independence number of the confusability graph.

The solution for the original question, the average number of bits sent with many uses of the channel, is known as zero-error Shannon capacity. An intuition for how to solve this problem is the following. Multiple uses of the channel correspond to a larger channel whose confusability graph is given by a graph power (i.e., a bigger graph resulting from a sequence of graph multiplications). For \( n \) that goes from one to infinity, let \( c_n \) be the maximum number of bits sent per each use of the channel, after \( n \) uses. The maximum average capacity of the channel is the maximum\(^{12}\) of the set \( \{c_1, c_2, \ldots \} \). It is not known if the Shannon capacity is computable at all, although the already mentioned \( \vartheta \) number is a useful upper bound.

Sometimes, if Alice and Bob have the additional resource of quantum entanglement, they can exceed the Shannon capacity for a channel. The maximum average number of bits sent in this case is called the entanglement-assisted zero-error Shannon capacity. This quantity was first studied by Cubitt et al. in [CLMW10].

**Source problem** We now consider a different scenario, studied by Witsenhausen [Wit76] in the zero-error setting. This setting is represented in Figure 1.6b.

Alice and Bob are connected to a dual source, that sends an input to Alice from a set \( X \) and some side information to Bob from a set \( U \). Every \( x \in X \) has associated a set \( U(x) \subseteq U \). These are all the elements of \( U \) that could be sent to Bob when \( x \) is sent to Alice. Of course, such sets may intersect. Two inputs \( x, y \) in \( X \) are not uniquely identifiable if their side information subsets intersect.

Consider the simple example we gave above for the channel, in a dual source form. The input set for Alice consists of the numbers from 1 to 6, the side information set for Bob is \( \{\text{odd, even, } > 3, \leq 3\} \) and the obvious rule is used to determine the side information subsets. We have, for example, that:

- If Alice gets “2”, Bob gets either “even” or “\( \leq 3 \)”,
- If Alice gets “3”, Bob gets either “odd” or “\( \leq 3 \)”,
- If Alice gets “5”, Bob gets either “odd” or “\( > 3 \)”.

Therefore, on Bob’s side 2 and 3 are not uniquely identifiable, 3 and 5 are not uniquely identifiable but 2 and 5 are uniquely identifiable.

But the player’s goal is for Alice to send her input to Bob, taking advantage of the side information. So this time the question is:

Alice and Bob have a dual source with Alice’s input set \( X \), Bob’s side information set \( U \) and side information subsets \( \{U(x)\}_{x \in X} \). They have access to a perfect one-way channel from Alice to Bob. What is

\(^{12}\text{More precisely, the supremum.}\)
the minimum number of bits of information that Alice can send on average to Bob per source input, in order for him to recover Alice’s inputs without chance of a mistake?

As with the channel, here we will restrict our discussion to the single source-input and we will use a graph parameter to answer the question.

Define the characteristic graph of the source as the graph with node set $X$ and edges between $x, y \in X$ that are not uniquely identifiable. It turns out that this graph is the same as Figure 1.3. This time, however, the solution to the problem is different. Suppose Alice and Bob agree on a coloring of the graph. When Bob receives a $u \in U$, he can identify the subset $C \subseteq X$ of Alice’s inputs that could be associated with $u$. This subset is a clique of the graph, a set of mutually adjacent nodes (the complementary notion of an independent set). Therefore, in a coloring each of the elements of $C$ must have a different color. Alice can send the color of her input $x$ to Bob, and Bob will have no more doubts. How many bits of information are sufficient for Alice to send? If the chromatic number is $c$, then to communicate one out of $c$ colors, Alice can send $\log_2(c)$ bits.

The discussion for multiple uses of the source is very similar to the one for the channel problem. To the best of our knowledge, the entangled version of this problem has not been discussed before.

**Source-channel problem** Consider the dual source problem, with the additional constraint that Alice and Bob communicate through a noisy channel. Source and channel are defined as before, with related characteristic and confusability graphs. This setting is represented in Figure 1.7. The question is the following:

Alice and Bob have a dual source with Alice’s input set $X$, Bob’s side information set $U$ and side information subsets $\{U(x)\}_{x \in X}$. They have access to a noisy channel from Alice to Bob, with input set $S$, output set $V$ and output subsets $\{V(s)\}_{s \in S}$. What is the minimum ratio number of channel uses/number of source inputs, in order for Bob to recover Alice’s inputs without chance of a mistake?

In the simplified single-source inputs scenario, the setting is as follows. Alice receives an input $x$, and sends a sequence of inputs $s_1, \ldots, s_k$ to Bob through the channel. Bob uses his side information $u$ together with the channel outputs $v_1, \ldots, v_k$ to recover $x$ without errors.

The answer to the question above is the source-channel rate, a function of the characteristic graph of the source and the confusability graph of the channel. This time, however, the solution does not depend on graph parameters but on graph homomorphisms. A homomorphism between two graphs is an edge-preserving map from the node set of the first graph to the node set of the second graph:
adjacent nodes of the first graph are mapped to adjacent nodes of the second graph. When solving the source-channel problem, Alice and Bob find a homomorphism from the characteristic graph of the source to the complement of the confusability graph of (multiple uses of) the channel. An intuition for this is that they want to map non-uniquely-identifiable source inputs (edges in the source graph) to non-confusable channel inputs (non-edges in the channel graph).

Notice that, as intuition suggests, the source problem and channel problem can be seen as special cases of the source-channel problem. The source problem is the source-channel problem with a non-noisy binary channel (therefore one channel use per bit); the channel problem is the source-channel problem with a completely uncorrelated source (therefore Bob has no side information).

To the best of our knowledge, the entangled version of this problem has not been discussed before.

Our contribution In Chapter 4, we define the entangled version of the source-channel problem. We allow Alice and Bob to share entanglement and we give a rigorous mathematical description of the new setting. We prove many properties of our problem, including a relation with the $\vartheta$ number and with quantum graph parameters of Chapter 3. We also show that the entangled channel problem of [CLMW10] is a special case of our setup.

Then, we construct source-channel pairs where quantum players have a large advantage over classical players by extending a known class of graphs.

Remarkably, we use the celebrated quantum teleportation scheme [BB84] in the proofs of some of our results. To the best of our knowledge, this is the first time quantum teleportation was used as a tool in the zero-error channel capacity setting.
1.5 Basics of quantum theory

We now move to the technical part of this chapter, a brief introduction to quantum mechanics. We assume familiarity with the basic concepts of linear algebra, which are summarized in [NC00, Section 2.1]. Since this thesis is focused on non-locality and its applications, we will not discuss quantum circuits or quantum computing in detail. We will instead concentrate on the concepts of quantum state, evolution and measurement, especially in a bipartite setting. These notions are crucial to understand non-locality presented in Chapter 2 and its applications presented in Chapters 3 and 4.

State vectors A quantum system of finite dimension $d$ lives in a $d$-dimensional complex inner product space, denoted by $\mathbb{C}^d$. Its quantum state is described by a unit vector in $\mathbb{C}^d$, called state vector (or pure state).

Let us illustrate these concepts with an example. A simple yet important quantum system is the quantum bit, frequently abbreviated as qubit. The quantum bit is the basic unit of quantum information, in the same way as the bit is the unit of classical information.

A classical bit can take value 0 or 1 and can be implemented by any classical physical system that can be in two distinct states, e.g., the presence/absence of voltage in a wire, the reflection/absorption of light by the surface of a laser disc, the orientation of the magnetic field in hard drives. Analogously, a quantum bit can take values 0 and 1 and can be represented by any quantum physical system that can be in two distinct classical states. For example, consider the polarization of a photon: we can associate the value 0 to vertical polarization and 1 to horizontal polarization. These are two classical states, but a quantum system can be in a superposition of both.

In quantum mechanics, we can describe a quantum bit with a unit vector in $\mathbb{C}^2$. Let us define the two classical states as $\left|0\right\rangle = [1,0]^T$ and $\left|1\right\rangle = [0,1]^T$, following Dirac’s notation.13 These two orthonormal vectors are often called the standard basis or computational basis. The quantum bit can be in any state of the form

$$\left|\psi\right\rangle = \alpha_0 \left|0\right\rangle + \alpha_1 \left|1\right\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix},$$

where $\alpha_0, \alpha_1$ are complex numbers satisfying $|\alpha_0|^2 + |\alpha_1|^2 = 1$. Thus, the state $\left|\psi\right\rangle$ of the quantum bit is a linear combination of the possible classical values, where the coefficients $\alpha_0, \alpha_1$ are called amplitudes and intuitively are “weights” indicating how much the qubit is towards 0 or 1.

---

13Dirac notation is standard in quantum information theory. A column vector $\psi$ is written in the “ket” notation $\left|\psi\right\rangle$, while its complex conjugate transpose is written in the “bra” notation $\langle\psi|$. At first it may seem confusing, but it turns out to be very convenient in formulas. For example, given any two vectors $\psi, \phi$ of the same dimension, their “bra-ket” is their inner product $\langle\psi|\phi\rangle$ (a scalar) and their “ket-bra” is their outer product $\left|\psi\right\rangle\langle\phi|$ (a matrix).
Now consider two quantum systems, the first living in $\mathbb{C}^d$ and the second living in $\mathbb{C}^{d'}$. We can describe these two systems together as a larger quantum system living in the tensor space $\mathbb{C}^d \otimes \mathbb{C}^{d'}$. For example, consider two qubits. Their state can be described as a superposition of the basis states $\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$, where $\otimes$ denotes the tensor product. They are often abbreviated as $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ or even more concisely as $\{00, 01, 10, 11\}$.

Thus, two qubits can be in any state of the form

$$\alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle,$$

with $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. In general, an $n$-qubit system is a unit vector that is a superposition of the $2^n$ basis states:

$$\sum_{i \in \{0,1\}^n} \alpha_i |i\rangle.$$

**Evolution** The evolution of $d$-dimensional quantum systems is described by unitary matrices in $\mathbb{C}^{d \times d}$. These are the transformations that preserve the norm of the vectors. Intuitively, these are the allowed operators because they map a state vector to another state vector, i.e., preserve the norm 1.

Let us go back to our simple example to illustrate this new concept. While the only allowed operations on a classical bit are to flip the bit or to leave it untouched, there are infinitely many transformations that one can do on a quantum bit. We will see later a very interesting transformation. For now, let us just show the unitary transformation equivalent to a bit-flip, that is

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We have that

$$X|\psi\rangle = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix}.$$

Thus, applying the $X$ unitary on a quantum bit swaps its amplitudes for $|0\rangle$ and $|1\rangle$. Intuitively this means that if the qubit was “weighted more towards zero” it will now be “weighted more towards one”.

Other important unitary matrices that we use later are the **Hadamard matrix** and the **Pauli matrices**. The $(2 \times 2)$ Hadamard matrix is defined as

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

We can easily construct Hadamard matrices that act on systems whose dimension is a power of two, for example $n$ qubits. One can obtain such a Hadamard matrix...
1.5. Basics of quantum theory

by tensor products of $2 \times 2$ Hadamard matrices. The matrix $H^{\otimes n}$ acts as follows on basis states:

$$H^{\otimes n}|i\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in \{0,1\}^n} (-1)^{i \cdot j} |j\rangle,$$

where $i \cdot j$ is the bitwise inner product of $i$ and $j$. Hadamard matrices play an important role in Chapters 2 and 4. The Pauli matrices are the $2 \times 2$ matrices $X$ (which we have already seen), the identity matrix $I$, together with

$$Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \text{ and } Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$ 

They are important in Section 3.5 mainly because of the so-called Pauli group. This is a group of 16 matrices ($I, X, Y, Z$ with coefficients $\pm 1$ and $\pm i$) under matrix multiplication.

**Measurement**  Unfortunately, we cannot directly access the state vector of a unknown quantum system. It would be nice if we could: for example, we could have extremely efficient communication by encoding a huge amount of data in the amplitudes of a single quantum bit. One of the most widely known counterintuitive features of quantum mechanics is that “observation changes the state of the system”. Let us see how it works, starting with our qubit example.

Measurements describe an experiment on a quantum state. Suppose we have a qubit in the state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and we measure it according to the standard basis, i.e., we check with the aid of a measurement device if this qubit is in the classical state $|0\rangle$ or $|1\rangle$. Quantum mechanics predicts that we observe outcome “0” with probability $|\alpha_0|^2$ and “1” with probability $|\alpha_1|^2$. That is why we need the state vector to be a unit vector: the squares of the amplitudes are probabilities, and they must sum to one! Quantum mechanics says that the state, after the measurement outcome “$i$”, collapses to the classical state $|i\rangle$, therefore losing all the information contained in the amplitudes.

For most quantum algorithms, this is the end of the story. All we need is to manipulate the amplitudes of the qubits in a smart way and measure according to the standard basis in order to obtain the desired results. In this thesis, however, it is often crucial to use a more general form of measurement, known as Positive Operator-Valued Measure (POVM). A $t$-outcome POVM is a collection

$$M = \{E_i \in \mathbb{C}^{d \times d} : i \in [t]\}$$

\(^{14}\text{In general, a Hadamard matrix is a square matrix } A \in \{-1,1\}^{\ell \times \ell} \text{ that satisfies } AA^T = \ell I. \text{ It is conjectured that there is a Hadamard matrix acting on spaces of every dimension that is multiple of 4. We do not know constructions for all such cases. (See Section 4.2.2.) Also, note that in the original definition, Hadamard matrices are } \pm 1 \text{-valued matrices. They have a normalization coefficient when used as quantum operations, to make them unitary.}\)
Chapter 1. Introduction

of positive semidefinite matrices \( E_i \equiv M_i^\dagger M_i \) that satisfy \( \sum_{i=1}^{t} E_i = I \), where \( I \) is the identity matrix. If we perform a \( t \)-outcome measurement \( \mathcal{M} \) on a \( d \)-dimensional system with state vector \( |\phi\rangle \), then we observe a random variable \( \lambda \) over the set \([t]\) whose probability distribution is given by

\[
\Pr[\lambda = i] = \langle \phi | M_i^\dagger M_i | \phi \rangle.
\]

In the event that \( \lambda = i \), we say that we get measurement outcome \( i \).

Before giving an example, let us define a simpler and important kind of measurement: the projective measurement. This is a special case of POVM where all elements \( M_i \) are projectors.\(^\text{16}\) Therefore, for all \( i \) we have \( E_i = M_i^\dagger M_i = M_i^2 = M_i \). Moreover, in a projective measurement all the elements are pairwise orthogonal, i.e., for all \( i \neq j, M_i M_j = 0 \). We will use this kind of measurement and its properties extensively in Chapter 3. After a projective measurement gave outcome \( i \), the state collapses to a new state

\[
\frac{M_i |\phi\rangle}{\| M_i |\phi\rangle \|},
\]

where \( \| \cdot \| \) denotes the \( \ell_2 \) norm of a vector. Notice that in this case, the state vector need not be a classical state, but it can still be in a quantum superposition.

Now we go back to our example, and illustrate how the measurement according to the standard basis can be seen as a projective measurement. Consider a qubit in state \( |\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \) with real amplitudes. Let \( \mathcal{M} = \{ |0\rangle \langle 0|, |1\rangle \langle 1| \} \). Then, the probability of observing “0” can be written as

\[
\langle \psi | (|0\rangle \langle 0|)^\dagger |0\rangle \langle 0| |\psi \rangle = |\alpha_0|^2,
\]

as explained before. It also turns out that the state after the measurement is\(^\text{17}\)

\[
\frac{|0\rangle \langle 0| |\psi\rangle}{\| |0\rangle \langle 0| |\psi\rangle \|} = \left[ \begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array} \right] \left[ \begin{array}{c} \alpha_0 \\ \alpha_1 \end{array} \right] = \left[ \begin{array}{c} 1 \\ 0 \end{array} \right] = |0\rangle.
\]

Bipartite systems and entanglement In this thesis we often consider bipartite settings, where we usually have two players, called Alice and Bob, that share a quantum system. If Alice has a \( d \)-dimensional system and Bob has a \( d' \)-dimensional system, they share a system in \( \mathbb{C}^d \otimes \mathbb{C}^{d'} \). Each party has direct

\(^{15}\)A matrix is positive semidefinite if and only if all its eigenvalues are non-negative.

\(^{16}\)Projectors are Hermitian matrices that satisfy \( M^2 = M \).

\(^{17}\)Up to a global phase. If two state vectors differ only by a scalar of absolute value 1, then they are physically indistinguishable.
1.5. Basics of quantum theory

access to his/her part of the state only. Therefore, if the parties are separated and unable to communicate, they can only perform local operations, i.e., the unitary matrices are in tensor product form $A \otimes B$, where $A$ acts on $\mathbb{C}^d$ and $B$ acts on $\mathbb{C}^{d'}$. Moreover, if Alice performs a measurement $\{E_i\}_{i \in [k]}$ and Bob $\{F_j\}_{j \in [\ell]}$ on a shared system in state $|\phi\rangle$, the probability of outcome “$i$” for Alice and “$j$” for Bob is $\langle \phi | E_i \otimes F_j | \phi \rangle$. After a projective measurement, the state collapses as explained above.

The state itself need not be in a product form. Indeed, there exist states in $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ that cannot be written as $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$. These states are called entangled. One famous example, which we mentioned in Section 1.1, is the EPR pair, named after the authors of [EPR35]:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

States of the form $|\psi\rangle^\otimes k$, i.e., $k$ EPR pairs, are very important for the results of Chapters 2 and 3.18

**Density matrices** There are two equivalent ways of representing and working with quantum states. The one we have just seen is based on state vectors, while the second one is based on density matrices. Both formulations have their own merits. The first is arguably simpler and deals very well with describing quantum algorithms, while the second is almost essential in multipartite settings, in order to easily describe sub-systems and mixtures of quantum states.19 As many authors in the field, we will switch between the two formalism when convenient. That is why we introduce both of them.

The set of possible states of a $d$-dimensional quantum system is formed by the $d \times d$ complex positive semidefinite matrices whose trace equals 1. These

18 At this point, let us clarify a common abuse of notation regarding EPR pairs. Strictly speaking, it is not correct that two EPR pairs are equal to the state $|\psi|^\otimes 2$. One must take into account who possesses which qubit and rearrange the qubits after performing the tensor product. Let us denote by $|i\rangle_A$ a basis state of Alice’s system, and by $|i\rangle_B$ a basis state of Bob’s system. Then we have

$$|\psi\rangle \otimes |\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B) \otimes \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)$$

$$= \frac{1}{2}(|0\rangle_A \otimes |0\rangle_B \otimes |0\rangle_A \otimes |0\rangle_B) + |0\rangle_A \otimes |0\rangle_B \otimes |1\rangle_A \otimes |1\rangle_B +$$

$$+ (|1\rangle_A \otimes |1\rangle_B \otimes |0\rangle_A \otimes |0\rangle_B) + (|1\rangle_A \otimes |1\rangle_B \otimes |1\rangle_A \otimes |1\rangle_B)$$

$$= \frac{1}{2}(|00\rangle_A|00\rangle_B + |01\rangle_A|01\rangle_B + |10\rangle_A|10\rangle_B + |11\rangle_A|11\rangle_B),$$

where the last equality follows from the rearrangement of the qubits.

19 Mixtures are simply probability distributions on quantum states. Any similarity with “mixed strategies” from game theory should not come as a surprise. Indeed, the term was introduced by von Neumann, who worked on both fields.
are called density matrices. A pure state \( |\psi\rangle \) has density matrix \( |\psi\rangle\langle\psi| \). A mixed state that is in pure state \( |\psi_i\rangle \) with probability \( p_i \) has density matrix \( \sum_i p_i |\psi_i\rangle\langle\psi_i| \). Distinct mixtures of pure states may have the same density matrix.

An operation on \( \rho \in \mathbb{C}^{d \times d} \) is a mapping \( \rho \mapsto U \rho U^\dagger \) induced by the unitary \( U \). A \( t \)-outcome measurement \( M \) on a system in state \( \rho \) is a collection of positive semidefinite matrices \( E_i = M_i^\dagger M_i \), as described above. By performing a measurement, one observes a random variable \( \lambda \) over the set \([t]\) whose probability distribution is given by

\[
\Pr[\lambda = i] = \Tr(E_i \rho) = \Tr(M_i \rho M_i^\dagger).
\]

After the measurement with outcome \( i \), the state collapses to

\[
\frac{M_i \rho M_i^\dagger}{\Tr(M_i \rho M_i^\dagger)}.
\]

The possible states of a pair of quantum systems \((A, B)\) of dimensions \( d \) and \( d' \) respectively are the trace-1 positive semidefinite matrices in \( \mathbb{C}^{d \times d} \otimes \mathbb{C}^{d' \times d'} \). A useful way to describe a part of the system is with a linear operator called the partial trace, which we introduce now. For matrices \( A \in \mathbb{C}^{d \times d} \) and \( B \in \mathbb{C}^{d' \times d'} \) define \( \Tr_A(A \otimes B) = \Tr(A)B \) and \( \Tr_B(A \otimes B) = A \Tr(B) \), and extend these definitions linearly to all matrices of \( \mathbb{C}^{d \times d} \otimes \mathbb{C}^{d' \times d'} \).

The pair of systems \((A, B)\) is said to be in an entangled state if it is in a state \( \rho \) which is not a convex combination of states of the form \( \rho_A \otimes \rho_B \). A pure state \( \rho \) in \( \mathbb{C}^{d \times d} \otimes \mathbb{C}^{d' \times d'} \) is said to be maximally entangled if \( \Tr_A(\rho) = \Tr_B(\rho) = \frac{1}{d} I \).

Now we consider the setting where Alice and Bob hold (possibly entangled) quantum systems \( A \) and \( B \), respectively, and they each perform a measurement.

Suppose that the pair \((A, B)\) is in the state \( \rho \) and that Alice performs a \( t \)-outcome measurement \( M = \{E_i\}_{i \in [t]} \) on \( A \). Then, the probability that Alice gets measurement outcome \( i \) equals \( p_i = \Tr((E_i \otimes I) \rho) \). Moreover, in the event that Alice gets measurement outcome \( i \), Bob’s system \( B \) is left in the state \( \rho_i = \Tr_A((E_i \otimes I) \rho) / p_i \). If Bob now performs an \( r \)-outcome measurement \( M' = \{F_j\}_{j \in [r]} \) on \( B \), then the probability that he gets outcome \( j \in [r] \) equals \( \Tr(F_j \rho_i) \).