Quantum Physics Letters
An International Journal

http://dx.doi.org/10.12785/qpl/040201

Introduction to Quantum Information Theory and Outline of Two Applications to Physics: the Black Hole Information Paradox and the Renormalization Group Information Flow

Fabio Grazioso[1,2,*]

[1] Research associate in science education at Università degli Studi di Napoli Federico II, Italy,
[2] INRS-EMT, 1650 Boulevard Lionel-Boulet, Varennes, Québec J3X 1S2, Canada.

Received: 7 Jun. 2015, Revised: 16 Jul. 2015, Accepted: 17 Jul. 2015
Published online: 1 Aug. 2015

Abstract: This review paper is intended for scholars with different backgrounds, possibly in only one of the subjects covered, and therefore little background knowledge is assumed. The first part is an introduction to classical and quantum information theory (CIT, QIT): basic definitions and tools of CIT are introduced, such as the information content of a random variable, the typical set, and some principles of data compression. Some concepts and results of QIT are then introduced, such as the qubit, the pure and mixed states, the Holevo theorem, the no-cloning theorem, and the quantum complementarity. In the second part, two applications of QIT to open problems in theoretical physics are discussed. The black hole (BH) information paradox is related to the phenomenon of the Hawking radiation (HR). Considering a BH starting in a pure state, after its complete evaporation only the Hawking radiation will remain, which is shown to be in a mixed state. This either describes a non-unitary evolution of an isolated system, contradicting the evolution postulate of quantum mechanics and violating the no-cloning theorem, or it implies that the initial information content can escape the BH, therefore contradicting general relativity. The progress toward the solution of the paradox is discussed. The renormalization group (RG) aims at the extraction of the macroscopic description of a physical system from its microscopic description. This passage from microscopic to macroscopic can be described in terms of several steps from one scale to another, and is therefore formalized as the action of a group. The c-theorem proves the existence, under certain conditions, of a function which is monotonically decreasing along the group transformations. This result suggests an interpretation of this function as entropy, and its use to study the *information flow* along the RG transformations.

Contents

[*] Corresponding author e-mail: fabio.grazioso@emt.inrs.ca

1 Classical information theory

Classical information theory has been introduced by Claude Shannon in 1948 [1, 2]. In this seminal work he has devised a quantitative definition of information content, and then other formal definitions of relevant quantities, in order to allow for a quantitative treatment of those and other related subjects. In the same seminal work he also demonstrated some important theorems which hold for such quantities. In this first section we give a summary of the main concepts of the classical information theory introduced by Shannon.

1.1 Information content

The first important contribution of Shannon has been to address the question: “What is information?”. More precisely, he was looking for a way to measure the amount of information *contained* in a given physical system. This is a rather elusive concept, and it can depend on things difficult to quantify, things such as the context, and the observer background knowledge.

To give an example, we can think at the amount of information contained in human facial expressions. We know at an intuitive level that a big amount of information is contained in a single facial expression (see figure 1), since we sometimes take important decisions based on such informations. But at the same intuitive level we can appreciate how difficult is to quantify this amount. Moreover, the type of information in the example of the facial expressions refers to *emotional states* or *states of consciousness*, and therefore involve some degree of subjectivity in their definition (think e.g. at the famous painting “Mona Lisa” by Leonardo da Vinci, and its enigmatic facial expression, so difficult to define). As usual in science, Shannon has overcome this type of difficulty by first defining clearly the *scope* of his definition. His definition of “content of information” is indeed limited to systems that can be described by a *random variable*.

Figure 1: Examples of facial expressions.

Since we need a precise definition of random variable, following the notation of MacKay [3] we will use the concept of *ensemble*, i.e. the collection of three objects:

$$X \equiv (x, \mathscr{A}_X, \mathscr{P}_X) \qquad (1)$$

where x represents the value of the random variable, $\mathscr{A}_X$ is the set of the possible values it can assume, and $\mathscr{P}_X$ is its *probability distribution* of those values (i.e. the set of the probabilities of each possible value).

1.1.1 Information content of a single outcome

Based on this concept we then introduce the following definition for the *amount of information* gained from the knowledge of a single outcome $x_i \in \mathscr{A}_X$ of the random variable X:

$$h(x_i) \equiv \frac{1}{\log 2} \log \frac{1}{p(x_i)} \qquad (2)$$

where $p(x_i) \in \mathscr{P}_X$ is the probability of the outcome x_i. To give an intuition of this definition we can consider the example of the weather forecast. Let’s simplify, and consider a situation where two only possible weather conditions are possible: *sunny* (☼) and *rainy* (🌧). So, in our example the random variable is “tomorrow’s weather”, the two possible values are $\mathscr{A}_X = \{\text{☼}, \text{🌧}\}$, and there will be a probability distribution $\mathscr{P}_X = \{p(\text{☼}), p(\text{🌧})\}$.

It is worth noting that the definition of Shannon is totally independent from the actual value of the outcome, and only depends on its probability. It is in order to stress this concept that we have used the symbols $\{\text{☼}, \text{🌧}\}$ for the values of the outcome, that are not numerical, and do not appear at all in (2). It is also worth to stress that this definition of “amount of information contained in a single outcome” is a *differential* definition: the difference between the amount of information we possess about the random variable, before and after we know the outcome.

We can illustrate this concept of “differential definition” using the weather variable: in a location where there is a very high probability of sunny weather, with the probability distribution $\mathscr{P}_X = \{p(\text{☼}) = 0.99, p(\text{🌧}) = 0.01\}$, if tomorrow we see sunny weather, we will have learnt very little information. On the other hand, if tomorrow we find rainy weather, we will have gained a lot of useful information, with respect to today.

1.1.2 Information content of a random variable

Using the definition (2) of the information content of a single outcome, we can define the information content of a whole random variable:

$$\begin{aligned} H(X) &\equiv \sum_i p(x_i) h(x_i) \\ &= \frac{1}{\log 2} \sum_i p(x_i) \log \frac{1}{p(x_i)} \end{aligned} \tag{3}$$

This definition can be seen as the *average* of the information gained for each outcome expressed in (2), averaged over all the possible outcomes.

This expression is formally equal (apart from constant factors) to the *entropy* defined in thermodynamics, and Shannon proposed the same name in the context of information theory. This entropy is sometimes called "Shannon entropy", to distingush it from its quantum counterpart, discussed in the following. In the case of a binary variable (i.e. variable with only two possible outcomes) we have:

$$\mathscr{A}_X = \{0,1\} \tag{4a}$$

$$\mathscr{P}_X = \{p, (1-p)\}, \tag{4b}$$

and the entropy of a binary random variable gets the special name of *binary entropy*:

$$H_{(2)} = \frac{1}{\log 2} \left[p \log \frac{1}{p} + (1-p) \log \frac{1}{(1-p)} \right] \tag{5}$$

A plot of the binary entropy as a function of p is shown in figure 2.

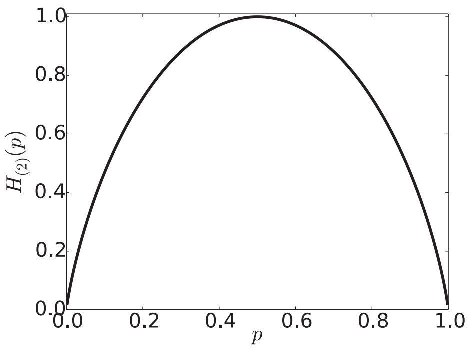


Figure 2: Plot of the entropy of a binary variable (binary entropy) shown in (5).

Again as for the information content of a single outcome, we can give some intuition for the definition of the entropy (i.e. information content) of a random variable using the example of the weather forecast. We can notice that in the case of a very biased probability distribution $\mathscr{P}_X = \{p(\text{☼}) = 0.01,\ p(\text{🌧}) = 0.99\}$, although the information content of the very unlikely outcome $h(\text{☼}) = \frac{1}{\log 2} \log \frac{1}{0.01}$ is very high, its weight (i.e. probability) in the average (5) is very small. So we have that *the highest value for the binary entropy is for the uniform probability distribution* $\mathscr{P}_X = \{p(\text{☼}) = 0.5,$ $p(\text{🌧}) = 0.5\}$, so that $p = 1/2$ and all the outcomes are *equiprobable*. It can be shown that this is true not only for the case of a binary variable, but for all the entropies of any random variable. This also explains the constant factor $\frac{1}{\log 2}$ in the definitions of the entropies: it is a normalization factor, so that the maximum entropy is normalized to 1. The factor $\frac{1}{\log 2}$ has also the advantage to make the definitions (3) and (5) independent of the choice of the basis for the logarithms. Alternative and equivalent definitions are:

$$H = -\sum_i p(x_i) \log_2 p(x_i) \tag{6a}$$

$$H_{(2)} = -p \log_2 p - (1-p) \log_2 (1-p). \tag{6b}$$

With this normalization is said that the entropy is measured in *bits*, and the entropy of an unbiased binary variable is 1. Sometimes another normalization is used, where the $\log_2$ is replaced by the natural logarithm $\ln = \log_e$; in this case it is said that the entropy is measured in *nats*.

1.1.3 Comments

We can find an intuitive justification of the definition (2) doing the following observations. First, the probability of two independent variables is the product of the probabilities of each outcome. On the other hand, for the definition (2) of "information from a single outcome" it is reasonable that the information gained from two outcomes from two independent variables is the *sum* of the information gained from each outcome. Thirdly, we have emphasized that the information content only depends on the probability. Given all this, when looking for an expression of the information content, the logarithm of the probability fits all the requirements. The last detail of using the logarithm of the *inverse* of the probability is coming from the requirement that the entropy of a variable has to be maximal (and not minimal) in the case of uniform probability distribution (see figure 2).

1.2 Other important definitions

For the applications we want to introduce in the following sections, we need to define few more quantities. The definitions we need involve two random variables:

$$\{X, \mathscr{A}_X, \mathscr{P}_X\} \tag{7a}$$

$$\{Y, \mathscr{A}_Y, \mathscr{P}_Y\} \tag{7b}$$

1.2.1 Joint entropy

The joint probability $p(x,y)$ is defined as the probability that the variable X has the outcome x *and* the variable Y has the outcome y. Based on this concept, it is easy to define the *joint entropy* of two random variables as:

$$H(X,Y) \equiv \frac{1}{\log 2} \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} \tag{8}$$

It is worth to recall from probability theory that the joint probability is the product of the probabilities in the case of *independent random variables*. So in the case of independent variables the joint entropy is the sum of the entropies.

Complementary to the concept of joint entropy is the definition of *mutual information* of two random variables:

$$I(X:Y) \equiv H(X) + H(Y) - H(X,Y). \tag{9}$$

We can use the intuition that *mutual information is a measure of how much two random variables are not independent*. It is also useful to rephrase this and think that *mutual information is a measure of how much we know about a random variable X if we know about random variable Y*. It is frequently used a graphical representation to visualize the relationship between entropy, joint entropy and mutual information. Instead of the Venn diagrams [4, 5], sometimes misleading, we prefer to use the alternative approach used e.g. by [3], shown in figure 3.

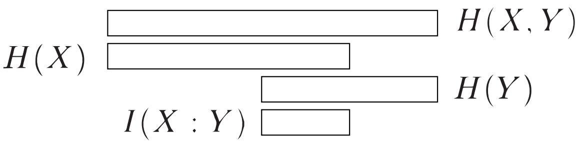


Figure 3: A graphical representation of the relationship between entropy, joint entropy and mutual information.

1.3 Source coding theorem

After having introduced some definitions, we here describe a theorem, called *source coding theorem*.

First, we have to introduce the notion of a source, described as a black box producing sequences of values. The way to model this is to consider those values as the outcomes of random variables. So we consider a sequence of N random variables, and we assume the following hypotheses: that the variables are *independent from each other*, that *the set of possible values is identical* for alle the variables, and finally that *the probability distributions are identical*. This is usually summarized as the N variables being *independent and identically distributed*, or i.i.d..

1.3.1 Typical set

Let's consider a sequence of N i.i.d. *binary* variables. We can write the sequence of variables as $(X_1, X_2, \ldots, X_N) = X^N$, and a single outcome will be a sequence of values as $(x_1, x_2, \ldots, x_N) = x^N$, which in the case of a binary variable can be represented as a sequence of N ones and zeroes. We can call $\mathscr{A}_{X^N}$ the set of all the possible sequences, and we can write it down, (e.g. using the *lexicographic* order) as follows:

$$\begin{array}{c} (0,0,0,0,0,\ldots,0) \\ (1,0,0,0,0,\ldots,0) \\ (0,1,0,0,0,\ldots,0) \\ \vdots \\ (1,1,1,1,1,\ldots,1) \end{array} \tag{10}$$

Given all this, the source coding theorem proves the existence of a subset of $\mathscr{A}_{X^N}$, called *typical set*, with the property that "almost all" the information contained in the random variable is indeed contained in this subset. Moreover, the theorem proves that for a sequence of N i.i.d. variables with entropy $H(X)$, the typical set has $2^{NH(X)}$ elements in it. To be more precise, the theorem can be verbally stated as follows:

Theorem 1(Source coding theorem). *N i.i.d. random variables each with entropy $H(X)$ can be compressed into more than $2^{NH(X)}$ bits with negligible risk of information loss, as $N \to \infty$; conversely if they are compressed into fewer than NH(X) bits it is "virtually certain" that some information will be lost.*

It is of course possible to have a more precise statement, where instead of the "almost all" and "virtually certain" phrases, the proper mathematical expressions, with "the epsilons and the deltas" typical of the mathematical limits are used. For a proof of the theorem see e.g. [3, 6].

1.3.2 Compression

In figure 4 we can see a graphical representation of the typical set, along with the idea that it is possible to label the elements of the typical set. The fundamental idea of compression is that if we use only the $NH(X)$ symbols needed to label the elements of the typical set, instead of using the N symbols of the full sequences, we have a negligible probability to loose information.

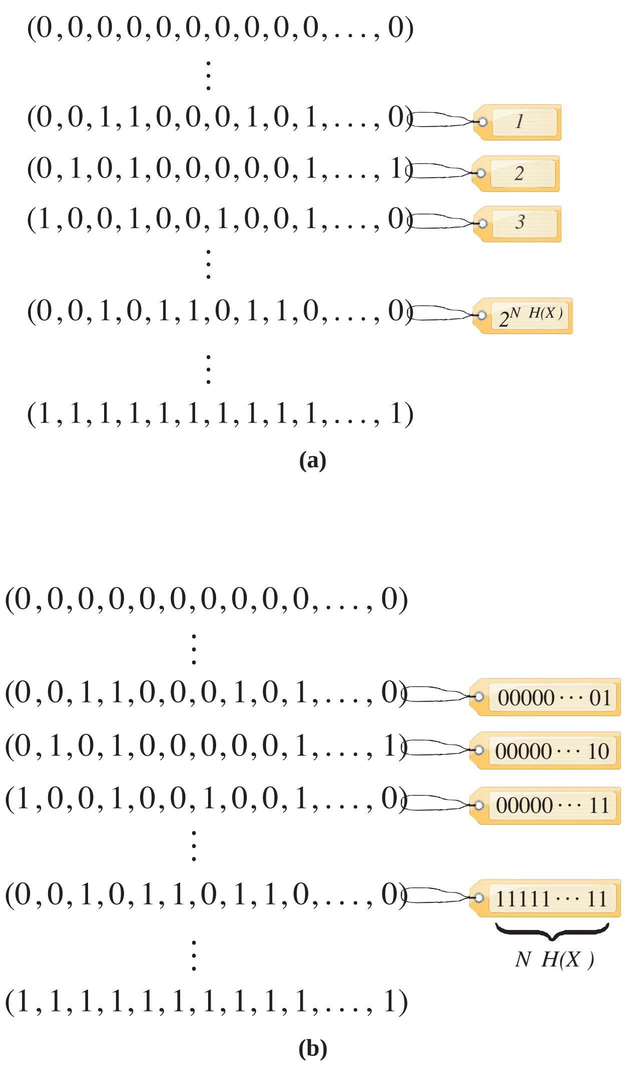


Figure 4: The typical set as a subset of all the possible sequences of N i.i.d. random variables outcomes. (a) The typical set elements can be labeled with a number between 1 and $2^{NH(X)}$. (b) This number can be written with $NH(X)$ binary simbols.

2 Quantum Information Theory

If the physical system used as support for the transmission and processing of information is a quantum system, classical information theory is no more valid in all its parts, and a different theory has to be developed: quantum information theory (QIT). As the classical random variable with two possible values (the bit) is the building block of CIT, the quantum random variable with its possible described by vectors of an Hilbert space of dimension two (the qubit) is the building block of QIT (see figure 5). The experimental efforts to implement a qubit in a physical system have already a long history. Among the different approaches we can mention ion traps [7, 8], quantum dots [9, 10], nuclear spins, accessed via nuclear magnetic resonance [11, 12], colour defects in crystals [13, 14] and superconductive structures [15, 16]. In this section we will review the usual axiomatic introduction of quantum mechanics (QM) and the formal tools which are necessary to describe the applications of QIT presented in the following. Among the many references for the axiomatic introduction to quantum mechanics, and the statements of its postulates, we refer mostly to [17].

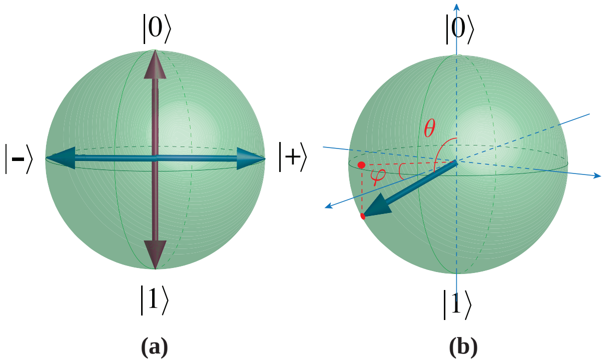


Figure 5: The Block sphere is a two dimensional manifold, and is used to represent the two dimensional Hilbert space of the states of a qubit.

2.1 Mixed states and density operator formalism

The state of a quantum system is represented by an element of an Hilbert space $\mathscr{H}$, of modulus one, which in the Dirac notation can be represented by a "ket" $|\psi\rangle \in \mathscr{H}$. In the case of a qubit (i.e. two-dimensional system) the basis can be represented as $\{|0\rangle, |1\rangle\}$ (called *computational basis*), and the generic state of the qubit will be $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $\alpha, \beta \in \mathbb{C}$, and the link to the angles shown in figure 5b is $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$.

In analogy to the concept of random variable introduced above, we need a formal tool to describe a situation where the state of the quantum system is unknown, and it is only know the set of possible states, with their probability distribution. If a system is in such conditions, it is said to be in a *mixed state*, and the tool to describe mathematically a mixed state is the *density operator*.

2.1.1 Density operator of a pure state

To introduce the density operator, let's first recall some details on linear algebra. The scalar product in the Dirac notation is written as $\langle\phi|\psi\rangle$; if we choose a basis $\{|1\rangle, |2\rangle, \ldots, |n\rangle, \ldots\}$ of the Hilbert space, it is possible to compute the components $\langle i|\psi\rangle = \psi_i$ and $\langle\phi|i\rangle = \phi_i^*$ of the vectors and co-vectors, so to write them as one-column and one-row matrices respectively. In this notation, the scalar product can be seen as a dot product between matrices:

$$\langle\phi|\psi\rangle = (\phi_1, \phi_2, \ldots, \phi_n, \ldots) \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{11a}$$

$$= \sum_i \phi_i^* \psi_i. \tag{11b}$$

But if we invert the order, and write

$$|\psi\rangle\langle\phi| = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} (\phi_1, \phi_2, \ldots, \phi_n, \ldots) \tag{12a}$$

$$= \begin{pmatrix} \psi_1\phi_1^* & \psi_1\phi_2^* & \cdots & \psi_1\phi_N^* \\ \psi_2\phi_1^* & \psi_2\phi_2^* & \cdots & \psi_2\phi_N^* \\ \cdots & \cdots & \cdots & \cdots \\ \psi_N\phi_1^* & \psi_N\phi_2^* & \cdots & \psi_N\phi_N^* \end{pmatrix} \tag{12b}$$

we have a matrix, which can be interpreted as the representation, in the chosen basis, of an *operator defined on the same Hilbert space*.

This was written for two different states $|\psi\rangle$ and $|\phi\rangle$. But using this type of product we can associate to any single vector of the Hilbert space an operator:

$$|\psi\rangle \leftrightarrow$$

$$\leftrightarrow |\psi\rangle\langle\psi| = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} (\psi_1, \psi_2, \ldots, \psi_n, \ldots) \tag{13a}$$

$$= \begin{pmatrix} \psi_1\psi_1^* & \psi_1\psi_2^* & \cdots & \psi_1\psi_N^* \\ \psi_2\psi_1^* & \psi_2\psi_2^* & \cdots & \psi_2\psi_N^* \\ \cdots & \cdots & \cdots & \cdots \\ \psi_N\psi_1^* & \psi_N\psi_2^* & \cdots & \psi_N\psi_N^* \end{pmatrix} \tag{13b}$$

$$\stackrel{\text{def}}{=} \hat{\rho}_\psi \, . \tag{13c}$$

2.1.2 Density operator of a mixed state

When a state of a quantum system can be represented as a vector of an Hilbert space (i.e. a ket in Dirac notation), it is said to be in a *pure state*. But if we want to represent the quantum analog of a random variable, we have to use the concept of mixed state introduced above, where we don't know the state of the system, but only a set of possible states, and their respective probabilities. A mixed state for which all its possible states are equiprobable is said a *maximally mixed state*. It is interesting to point out that whether the system is in a pure or a mixed state depends on both the system *and* the observer, because the knowledge about the system depends also on the observer, and not only on the system itself. The density operators formalism is able to effectively represent this type of states.

Indeed, if the possible states of the system are $\{|\alpha_1\rangle, |\alpha_2\rangle, \ldots, |\alpha_N\rangle\}$, with probabilities $\{p_1, p_2, \ldots, p_N\}$, then the mixed state can be represented as:

$$\sum_{i=1} p_i |\alpha_i\rangle\langle\alpha_i| \, . \tag{14}$$

This can be seen as a linear combination of the density operators associated to the pure states, where the coefficients are the probabilities.

This is an abstract representation of the density operators; if we fix a basis in the Hilbert space, we can write a density operator as a matrix, that will be called *density matix*. A special and not uncommon case is when the set of possible states of a mixed state is an orthonormal basis for the Hilbert space. If we write this orthonormal basis as $\{|1\rangle, |2\rangle, \ldots, |n\rangle, \ldots\}$, and then represent the density matrix associated to a *pure state* in this basis, the matrix elements will be all zero, apart from one single element on the diagonal equal to one, in the position corresponding to the position of the pure state in the basis:

$$\hat{\rho}_n = |n\rangle\langle n| = \begin{pmatrix} 0 \\ \vdots \\ n \\ \vdots \end{pmatrix} (0, \ldots, n, \ldots)$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix} . \tag{15}$$

If we then consider a mixed state such that the possible states are all the elements of the basis:

$$\sum_{i=1} p_i |i\rangle\langle i| \tag{16}$$

its density matrix, represented in this same basis will be diagonal, with the probabilities as diagonal elements:

$$\hat{\rho} = \begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & p_n & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix} . \tag{17}$$

If represented in this basis, non-zero off-diagonal elements indicate that some of the possible states are quantum superpositions of basis states. From the normalization property of the probability distribution it is then easy to see that:

$$\mathrm{Tr}(\hat{\rho}) = \sum_i p_i = 1, \tag{18}$$

where $Tr(\hat{\rho})$ indicates the trace, defined as the sum of the diagonal elements. Since the trace is preserved under change of reference, we can conclude that $Tr(\hat{\rho}) = 1$ is a property of any density matrix. Another property of any density matrix is that the eigenvalues are non-negative. This can be proven rigorously, and can be easily seen in the case of a diagonal density matrix (17), where the eigenvalues have the meaning of probabilities.

2.2 Quantum measurement and quantum complementarity

Continuing with the axiomatic introduction of quantum mechanics, after the concept of mixed states, and the density operators formalism to describe them, we now describe the measurement of the state of a quantum system.

In the following subsections we will give two possible formalizations of the measurement process, namely the *projective measurement*, and the *POVM*. Finally, we will see the concept of *complementarity*.

2.2.1 Projective measurement

A first way to formalize the measurement process is the *projective measurement* or *von Neumann measurement* (see [17, 18]). In this description we associate to the measurement an hermitian operator $\hat{M}$, and its decomposition over the projectors on its eigenspaces:

$$\hat{M} = \sum_m m\hat{P}_m \tag{19}$$

where $\{m\}$, the eigenvalues of $\hat{M}$, are the possible outcomes of the measurement, and the $\{\hat{P}_m\}$ operators are projectors, i.e. satisfy the following properties:

$$\forall m, \hat{P}_m \text{ is hermitian} \tag{20a}$$

$$\forall m, m', \hat{P}_m\hat{P}_{m'} = \delta_{m,m'}\hat{P}_m. \tag{20b}$$

The probability that the outcome of the measurement is m when the system is in the state $|\psi\rangle$ is:

$$p_\psi(m) = \langle\psi|\hat{P}_m|\psi\rangle; \tag{21}$$

and soon after such measurement the state of the system is:

$$\frac{\hat{P}_m|\psi\rangle}{\sqrt{p_\psi(m)}}. \tag{22}$$

From the requirement that the sum of all the probabilities (21) is equal to 1 we have the property of *completeness* for the set of projectors:

$$\sum_m \hat{P}_m = \mathbb{I}. \tag{23}$$

The expectation value of the measurement $\hat{M}$ if the system is in the state $|\psi\rangle$ is:

$$\begin{aligned} E_\psi(\hat{M}) &= \sum_m m\, p_\psi(m) \\ &= \sum_m m\langle\psi|P_m|\psi\rangle \\ &= \langle\psi|\left(\sum_m mP_m\right)|\psi\rangle \\ &= \langle\psi|\hat{M}|\psi\rangle \\ &= \langle\hat{M}\rangle_\psi. \end{aligned} \tag{24}$$

and the standard deviation is:

$$\begin{aligned} \Delta(\hat{M}) &= \sqrt{\langle(\hat{M} - \langle\hat{M}\rangle_\psi)^2\rangle_\psi} \\ &= \sqrt{\langle\hat{M}^2\rangle_\psi - \langle\hat{M}\rangle_\psi^2} \end{aligned} \tag{25}$$

where we have used the compact notation $\langle\psi|\cdot|\psi\rangle = \langle\cdot\rangle_\psi$. Sometimes it is useful to write the projectors as:

$$\hat{P}_m = \hat{M}_m^\dagger\hat{M}_m \tag{26}$$

where $\hat{M}_m$ are called *Krauss operators*. The equations (19)-(23) can be rewritten in terms of the Krauss operators using (26).

2.2.2 POVMs

It is possible to generalize the projective measurement and define the POVM (positive operator-valued measurement [18]), where some of the hypotheses of the projective measurement are relaxed. In particular, we consider the collection of operators that represent the measurement:

$$\{\hat{E}_m\} \tag{27}$$

and relax the hypothesis that those operators are projectors. Similarly to the projective measurement, the probability that the outcome is m if the system is in $|\psi\rangle$ is:

$$p_\psi(m) = \langle\psi|\hat{E}_m|\psi\rangle. \tag{28}$$

Also for the POVM we have the property of completeness: $\sum_m \hat{E}_m = \hat{\mathbb{I}}$, but as a consequence of the (27) not being projectors, is that in general we can not write them in terms of the Krauss operators, as in (26), and therefore for the POVM measurement it is not defined the state of the system *after the measurement*.

A common situation with POVM measurement is when we have a quantum system in a mixed state, where the set of possible states are represented by some vectors of the Hilbert space $\{|\psi_m\rangle\}$, not necessarily orthogonal to each other, and we want a measurement in order *to know in which of the states of the set the system is*. This POVM is represented by the set of operators:

$$\{\hat{E}_m = |\psi_m\rangle\langle\psi_m|\}. \tag{29}$$

These last operators are indeed projectors; however, since the $\{|\psi_m\rangle\}$ are not necessarily orthogonal, this POVM *is not* in general a projective measurement. In this type of POVM, since the set of states does not necessarily form a basis of the Hilbert space, the completeness property has in general to be guaranteed with suitable normalization coefficients.

2.2.3 Quantum complementarity

If we consider the Hilbert space representing the states of the quantum system, each basis can be seen as a different POVM. In particular, an orthogonal basis will correspond to a projective measurement. The preparation and measurement of the quantum state of a physical system can be described in the language of QIT in terms of the encoding and decoding of information by two parties, traditionally called Alice and Bob. The quantum complementarity is then related to the choice of the basis in which each party operates. If we consider the example of a qubit, in figure 6 two different orthogonal bases are shown, the *computational basis* $\{|0\rangle, |1\rangle\}$, and the basis $\{|+\rangle, |-\rangle\}$, where

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{30a}$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{30b}$$

Alice may choose to encode some information in the qubit, using the computational basis $\{|0\rangle, |1\rangle\}$, i.e. she prepares the system in one of the two states of this basis (see figures 5 and 6). The qubit will be then transmitted to Bob, who will perform a measurement to decode the information. If he chooses the diagonal basis $\{|+\rangle, |-\rangle\}$, he will be in the situation where both outcomes of the measurement have 0.5 probability (see figure 6b). To describe this situation in terms of information we can use the concept of mutual information expressed in (9), and say that the mutual information between the (classical) random variable representing Bob's measurement outcome and the (classical) random variable representing the information encoded by Alce, is zero. This means in other terms that the Bob can not access the information of Alice. This situation expresses the concept of quantum complementarity, and based on this concept Charles Bennett and Gilles Brassard in 1984 devised the idea of *quantum cryptography* [19], which over the years has become one of the most developed applications of QIT [20–22, 24].

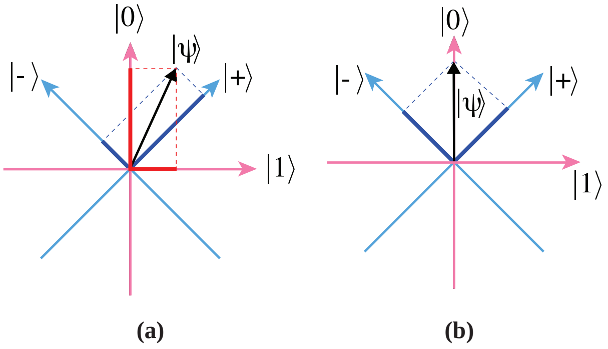


Figure 6: Two orthogonal references in the plane, to represent two different projective measurements: the computational basis $\{|0\rangle, |1\rangle\}$, and the basis $\{|+\rangle, |-\rangle\}$ defined in (30). (a) A generic vector, with its components on the two references. (b) A special case of an eigenvector of the first reference which has equal components on the second reference.

2.3 von Neumann entropy

In analogy to the definition of *information content* of a classical random variable (Shannon entropy) defined in (3), it is possible to define the von Newmann entropy, in the case of a quantum random variable, in the following way:

$$S(\hat{\rho}) = \frac{1}{\log 2} Tr\left(\hat{\rho}\log\frac{1}{\hat{\rho}}\right). \tag{31}$$

Here $Tr(\cdot)$ represents the trace of the density matrices, and ρ is the density operator representing the random variabile of which S represents the (quantum) information content.

2.3.1 Quantum evolution

To complete the axiomatic framework of quantum mechanics we need one last postulate, about the evolution of a quantum system. It states that the evolution in time of a quantum system is described by an unitary transformation over the Hilbert space describing the states:

$$|\psi(t)\rangle = \hat{U}|\psi(0)\rangle. \tag{32}$$

Here we will not give the details about the actual unitary operator, described by Shrödinger equation.

2.4 Holevo theorem (Holevo bound)

One of the most important results of QIT is the following theorem, called after Alexander Holevo [25]. As for the description of quantum complementarity, this result is best described in terms of the interaction between the two parties Alice and Bob.

Theorem 2(Holevo bound). *Let's suppose that Alice prepares the quantum system in a mixed state described by the density operator $\hat{\rho}_X$, where $X = \{|x_1\rangle, \ldots, |x_n\rangle\}$ are the possible pure states, and $\{p_1, \ldots, p_n\}$ are the corresponding probabilities. Then, Bob performs a measurement, described by a POVM built (as described in section 2.2.2) on the set of pure states $Y = \{|y_1\rangle, \ldots, |y_n\rangle\}$, and we denote y the outcome of this measurement. It is possible to prove that for any such measurements Bob may do there is an upper bound for the* mutual information (9) *between the two random variables X and Y. In particular:*

$$I(X:Y) \leq S(\hat{\rho}) - \sum_x p_x S(\hat{\rho}_x) \tag{33}$$

where $\hat{\rho} = \sum_x p_x \hat{\rho}_x$ is the density operator describing the global mixed state prepared by Alice.

It is worth to stress that from the point of view of Alice (the sender), the information she encodes in the system is a classical information. We can represent it as the integer index labelling the states in the set of quantum states $X = \{|x_1\rangle, \ldots, |x_n\rangle\}$ chosen for the encoding. On the other hand, from the point of view of Bob(the receiver), the system is in a quantum mixed state. The following theorem expresses the relationship between the information contained in those two random variables.

Theorem 3.*Given a classical random variable, encoded in a quantum system using the set of pure states $X = \{|x_1\rangle, \ldots, |x_n\rangle\}$, the relation between the information contained in this classical random variable, and the quantum information contained a mixed quantum state $\hat{\rho}_X$ built with those pure states is:*

$$S(\hat{\rho}) - \sum_x p_x S(\hat{\rho}_x) \leq H(X) \tag{34}$$

the equality being reached in the case $\{|x_1\rangle, \ldots, |x_n\rangle\}$ are all orthogonal vectors.

Because of this second result, we can express the Holevo theorem (33) saying that in a quantum encoding-decoding process the amount of information that Bob can access is in general less than the (classical) information initially encoded by Alice, and that this information can be fully accessed only in the special case where the set of quantum states used for the encoding is orthogonal.

2.5 No-cloning theorem

Another important result of QIT is the no-cloning theorem, introduced by and Wootters, Zurek and Dieks in 1982 [26, 27]. It is a no-go theorem that can be stated very briefly as follows:

Theorem 4(No-cloning). *It is impossible to create an identical copy of an arbitrary* unknown *quantum state.*

The crucial part is the fact that the theorem applies to a situation where the state is unknown.

The theorem can be expressed also in the following alternative statement:

Theorem 5(No-cloning). *Given two states $\{|\psi_1\rangle, |\psi_2\rangle\} \in \mathscr{H}$, which are non-orthogonal, i.e. $0 < |\langle\psi_1|\psi_2\rangle| < 1$, it doesn't exist an unitary transformation defined on two states $\hat{U}: \mathscr{H} \otimes \mathscr{H} \to \mathscr{H} \otimes \mathscr{H}$* such that

$$\hat{U}(|\psi_i\rangle|0\rangle) = |\psi_i\rangle|\psi_i\rangle \tag{35}$$

when i is not known, i.e. when $\psi_i \in \{\psi_1, \psi_2\}$ is unknown.

3 The Black Hole Information Paradox

3.1 Black holes

For the purpose of this review, black holes (BHs) can be briefly described as objects so dense, and with a gravitational field so strong, that on a surface external to them, and called *events horizon*, the escape velocity is higher than the speed of light. This implies that no physical object, not even light itself, can ever leave a BH once it is inside its event horizon.

3.2 Hawking radiation and black hole evaporation

The work of Stephen Hawking in 1974 [28] introduced the notion of the *Hawking radiation* (HR). This phenomenon is in turn due to the phenomenon of *quantum vacuum fluctuations*, that was discussed and theorized at the beginning of the 20th century by the scientists that contributed to develop quantum theory (see e.g. [29, 30]). Quantum vacuum fluctuations are in turn linked to what has been subsequently formalized as the *Heisenberg uncertainty principle* [17, 31], and can be summarized as the continuous and very rapid creation and annihilation of particle-antiparticle couples (see figure 7). Hawking theorized that there is a non-zero probability that a particle-antiparticle couple is generated close enough to the BH's event horizon, so that one of the two particles manages to escape before they re-annihilate while the other is trapped inside the horizon. The net effect is a radiation emitted from the BH while taking some energy from it, and because of the mass-energy equivalence, the phenomenon can be described as the evaporation of the BH. The Hawking radiation has an extremely low intensity, but if the BH is small enough, it can lead to the complete evaporation of the BH in a physically meaningful time, compared to the age of the universe. In its subsequent detailed quanto-mechanic calculations [32, 33], Hawking showed also that the quantum state in which the HR is emitted is a *maximally mixed state* (see section 2.1.2).

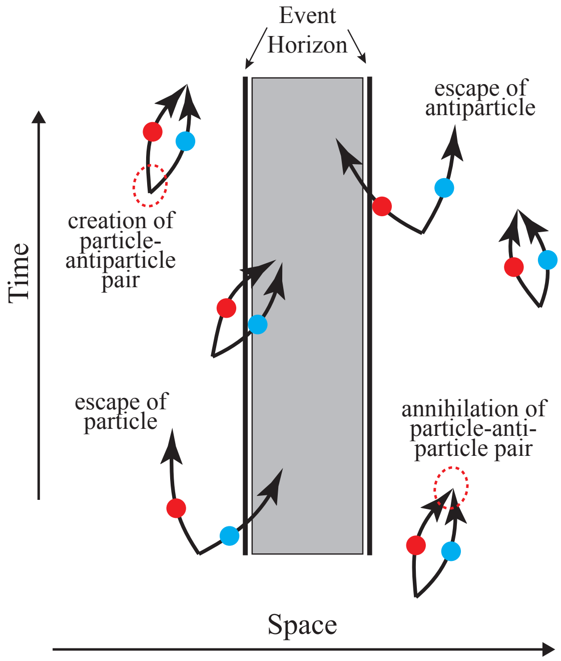


Figure 7: Schematics of the mechanism of quantum vacuum fluctuation and generation of Hawking radiation.

3.3 Black hole paradox

Since it is always possible to prepare the BH, as soon as it forms, in a pure state, and then leave it isolated, the phenomenon of HR leads to a contradiction. Indeed if we consider an isolated BH as an isolated quantum system, according to the postulates of QM seen in section 2.3.1, its evolution should be described by an unitary transformation. But if we consider the process of complete evaporation of the BH, and take into account that the HR is emitted in a mixed state, we would have the evolution of an isolated quantum system from a pure state to a mixed state, in contradiction with that postulate. For what follows it is worth to remember that a maximally mixed state is such that each state of the mixture is equiprobable. So if we describe the final state of the Hawking radiation after the complete evaporation as a quantum random variable, this is in a maximally mixed state, and therefore it has *zero mutual information* with the quantum random variable describing the initial state.

3.3.1 BH paradox in terms of QIT

It is possible to rephrase this contradiction using the concepts of quantum information theory, so to show that contradicting the postulate of unitary evolution of an isolated quantum system is equivalent to contradict the no-cloning theorem introduced in section 2.5.

Let's consider a physical system, containing a certain amount of *information*, dropped into the BH at an early time, and let's ask the question whether this information can in principle be retrieved at a later time or not (see figure 8).

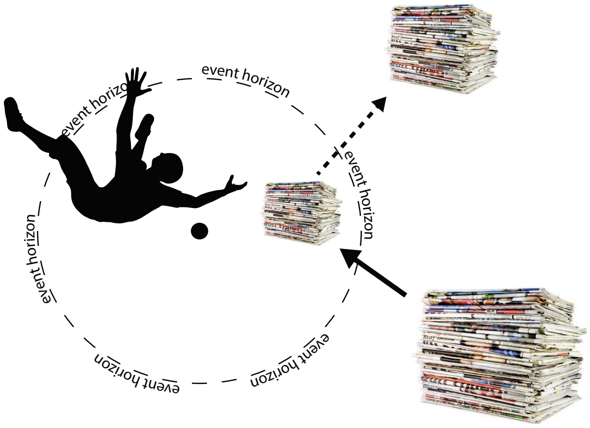


Figure 8: Information falling into the event horizon: can it, even in principle, be retrieved? From the point of view of an in-falling observer, crossing the event horizon has no physical effect, and this suggests that also the information is not destroyed when it falls inside the horizon.

In a deterministic system, following the dynamic equations that describe its evolution, it is in principle possible to reconstruct an earlier state once we fully know the state at a later time (with emphasis on the *full knowledge* of any degree of freedom and their correlations). So, if a BH is well described by quantum mechanics, the answer to the question about the information retrieval should be affirmative, and the Hawking radiation is a good candidate to explain how the information can escape. This in turn would question general relativity, from which the very definition of event horizon descends [34, 35], because by definition nothing can escape the event horizon.
If on the other hand the answer to the question about the information retrieval is negative, then it means that the quantum-mechanical description of the BH and its evolution has to be revised.

Moreover, we can show how, if we accept the notion that somehow the information initially dropped inside the event horizon, eventually escapes via the Hawking radiation, we incur in another problem. Indeed, from the point of view of an in-falling observer, crossing the event horizon has no physical effect. So we can safely assume that the information dropped in the BH still exists intact, inside the event orizon (at least until it reaches the internal singularity of the BH).

Therefore, if the information also escapes, it means that at least a finite time, two copies of the same information exist, inside and outside the event horizon. So this would contradict the no-cloning theorem of section 2.5.

3.3.2 Contributions to the solution from QIT

Although the BH information paradox is still an open problem, QIT has contributed to its comprehension with some important results and insights.

Jacob Bekenstein is one of the leading authors of such line of research [36]. In 1972 he has introduced a *generalized second law* describing the thermodynamics of BHs [37], and in the 1973 he has introduced a definition of BH's *entropy*, as being proportional to its area $\mathscr{A}$ and inversely proportional to the square of Plank's length ℓ_P^2:

$$S_{BH} \propto \frac{\mathscr{A}}{\ell_P^2}. \tag{36}$$

Then, at first Bekenstein [38], and then Bousso [39] have found upper bounds for the BH's entropy. Since the double meaning of the entropy as both a thermodynamic parameter and a measure of the information content of a system (see section 1.1) these results have suggested a information theoretical approach to solve the paradox.

Hayden and Preskyll [40] have used results from quantum error correction, to extend a result already found by Page [**?**]. When the BH is in an advanced stage of its evaporation, more precisely when its entropy is less than half the initial amount, they prove that the information retention time, i.e. the time needed for the information dropped in the event horizon to re-emerge in the Hawking radiation, is relatively short, and in particular:

$$t_{\text{info}} = \mathscr{O}(r_S \log r_S) \tag{37}$$

where r_S is the Schwarzschild radius.

Another contribution to the solution of the BH information paradox, also used by Hayden and Preskyll, is the concept of BH complementarity [42, 43]. This approach considers two possibilities: the information traveling toward the BH from outside, when reaches the event horizon is either transmitted inside or reflected outside. Then, the suggestion is that instead of choosing between those two possibilities, we can accept them both. To solve the conflict with the no-cloning theorem, we assume that, because of the *quantum complementarity* discussed in section 2.2.3 it is impossible for any observer to observe both descriptions, or access both copies of the information. An external observer will see the incoming information being absorbed by the event horizon, and then re-transmitted outside by means of the Hawking radiation, all this process being unitary. The observer falling inside the event horizon from outside will not notice the crossing, and will continue to observe the information that is falling with him. But he will not be able to access the information reflected outside with the Hawking radiation, because that will be encoded in a different basis, such that the mutual information is zero.

Another important result worth to mention is the *holographic principle*, a general result which can be stated as follows: *"Physical processes in a system of $\mathscr{D}$ dimensions are reflected in processes taking place on the $\mathscr{D}-1$ dimensional boundary of that system. There is an equivalence between theories of different sorts written in space-times of different dimensions"* [36, 44].

The fields of QIT, Astrophysics and general relativity have all gained from this interdisciplinary approach; as an example the concept of Generalized Second Law, and the Holographic Principle have also lead to results in QIT. In particular, upper bounds have been proven for the entropy outflow $\frac{\partial S}{\partial t}$, which is a proxy for the communication rate, or information channel capacity [36].

4 The renormalization group information flow

4.1 Description of the RG

The main idea of the renormalization group (RG) is that of a tool to extract the macroscopic description of a physical system (e.g. a field) from its microscopic model. First of all, the change in the descriptions going from the microscopic to the macroscopic model is captured by the change of the interaction constant $g(\mu)$ in the interaction term of the hamiltonian.

This change can be described as the action of an operator $\hat{G}$ applied to the interaction constant:

$$g(\mu_2) = \hat{G}\,[\,g(\mu_1)] \tag{38}$$

where μ_i is a parameter that represents the different scales. Although this transformation is called "renormalization group", it is not formally a group. It is just a "flow of transformations" in the space of all the possible hamiltonians. The main reason why the RG is not a group, is that given a transformation from a small scale description to a large scale description, the inverse transformation is not necessarily defined.

In 1954 Murray Gell-Mann and Francis Low published a work on quantum electrodynamics (QED) [45], in which they studied the photon propagator at high energies. They introduced the concept of scaling transformation with a group-like formalism, where the group operator $\hat{G}$ transforms the electromagnetic coupling parameter g:

$$\hat{G}\,[g(\mu_2)] = \left(\frac{\mu_2}{\mu_1}\right)^d \hat{G}\,[g(\mu_1)] \tag{39a}$$

$$g(\mu_2) = \hat{G}^{-1}\left[\left(\frac{\mu_2}{\mu_1}\right)^d \hat{G}\,[g(\mu_1)]\right]. \tag{39b}$$

Equation (39) expresses the requirement that before and after the scaling, the physical laws don't change. So the equation requires that the coupling parameter before and after the scaling changes taking into account the scaling factor $\left(\frac{\mu_2}{\mu_1}\right)^d$. Going from this discrete scaling $\mu_1 \to \mu_2$ to a continuous scaling transformation, it is possible to define a function $\beta(g)$ that expresses the corresponding continuous transformation of the coupling parameter g:

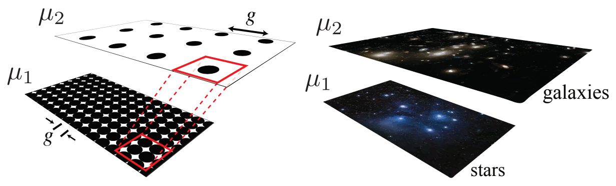
$$\beta\left[g(\mu)\right] = \frac{\partial g(\mu)}{\partial \ln(\mu)}. \tag{40}$$

Between 1974 and 1975 Kenneth Wilson and John Kogut introduced a more general description of this idea [46–48]. In this description, the large scale (macroscopic) behaviour will be linked to the low energy regime of the model, because at long distance only long wavelengths are relevant, while for the microscopic behaviour higher energies will be relevant. With reference to this, in the language of the RG the microscopic, high energy model will be called the *ultraviolet limit*, while the macroscopic, low energy one will be called the *infrared limit*. Another language to express the description at different scales is in terms of *fine graining* and *coarse graining*.

To give an example of the low energy approximation, we can imagine a sinusoidal potential for the microscopic model, and its approximation with a parabolic potential for the macroscopic description. This will be a good description at low energies, i.e. at the bottom of the microscopic potential. However, at high energies this approximation may introduce some divergencies, involving as an example the integration over bigger ranges of energies. Since those divergencies are only due to the approximated description of the potential, this can be corrected introducing a cut-off for the high range of energies. The dynamics of a composite system can be described by the interactions between its components. At a certain scale (graining) μ_1 the physics of that model is described by the hamiltonian of the system, and in particular by its interaction term, i.e. by the interaction constant $g(\mu_1)$. At an bigger scale (coarse graining) μ_2, the components of the lower scale can be "clustered" into a single element of the coarse graining (see section 9a), and the interaction constant is in principle changed. The equations expressing the constrain that: "the physics at different scales has to be the same" are (38) and (39), which express the constrains for the interaction constant $g(\mu_i)$, and another equation that express the constrain between the correlation at different scales, which is the the Callan-Symanzik equation [49–51]:

$$\left[m\frac{\partial}{\partial m} + \beta(g)\frac{\partial}{\partial g} + n\gamma\right] C^{(n)}(x_1,\ldots,x_n;m,g) = 0 \tag{41}$$

where: m is the mass, C is the correlation function between the $(x_1,\ldots,x_n)$ elements of the system, β and γ are two functions that "compensate" the effect of the scale change, in order for the description (i.e. the correlation function) at the different scales to be consistent. In particular β, which we have already seen in (40), captures the change of the coupling constant, while γ captures the change of the field itself.



(a) **(b)**

Figure 9: abstract description of the renormalization group. (a) Two different scales of modelling, with two different interacting constants. (b) An example of such different scales can be found in astrophysics, where the description at the scale of stars (lower image) has an interaction constant different from the description at the scale of galaxies (upper image).

In applying the group transformations, we go from one point of the space (manifold) of all the possible hamiltonians (i.e. in the manifold of the βs and γs) to another. However, there are some points, called *critical points*, or *conformal points*, where the function $g(\mu)$ has its minimum. From another point of view we can think at the manifold of the hamiltonians (each describing a different model for the system, at different scales, with different values of the coupling constant), and then think that the RG transformations describes a flow from one model to the other. The flow always ends at the points that are invariant for this transformation, so those points have to be self-similar. Each of the critical points are characterized by the (minimal) value that the function assumes there, and this value is called the "central charge" of the system.

4.2 The c-function and the link to QIT

The *c-theorem* of Alexander Zamolodchikov [52] individuates, in the case of a two-dimensional renormalizable field, a function which is monotonic along the RG transformations.

This monotonicity suggests an information theoretical meaning for this function, analogue to the information content. [53–55].

Since the seminal result by Zamolodchikov, several authors have worked on c-theorems at dimensions higher than 2 [56–60].

Another approach to the RG is the density matrix renormalization group (DMRG) [61, 62]. Osborne and Nielsen [63] make more explicit the link between DMRG and QIT. A characteristic feature of critical phenomena is the emergence of collective behaviour, and it is conjectured that quantum entanglement is the origin of this cooperative behaviour. DMRG and its explicit quanto-mechanical approach seems the ideal formalism with which to substantiate this conjecture [64, 65].

Finally, a different interdisciplinary approach, not necessarily linked to information theory, is the parallel between the renormalization used in quantum field theory and the renormalization used in thermodynamics and statistical mechanics to describe critical phenomena [46, 66].

Acknowledgements

The author thanks Frédéric Grosshans, Gilles Brassard and Patrick Hayden for the fruitful discussions and inputs on information theory, the latter in particular for having introduced him to the black hole information paradox, and Spyros Sotiriadis for the discussions on the renormalization group.

References

[1] C. E. Shannon, A mathematical theory of communication. Bell System Technical Journal, **27** 379-423 (1948).

[2] C. E. Shannon. A mathematical theory of communication. Bell System Technical Journal, **27** 623-656 (1948).

[3] D. J. C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003).

[4] J. Venn. On the diagrammatic and mechanical representation of propositions and reasonings. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **10**, 1-18 (1880).

[5] J. Venn. On the employment of geometrical diagrams for the sensible representation of logical propositions. Proc. Cambridge Philos. Soc. Math. Phys. Sciences, **1**, 47 (1880).

[6] T. M. Cover and J. A. Thomas. Elements of Information Theory. Wiley Series in Telecommunications and Signal Processing. Wiley (2006).

[7] J. I. Cirac and P. Zoller. Quantum computations with cold trapped ions. Phys. Rev. Lett. **74**, 4091- 4094 (1995).

[8] A. Steane, C. F. Roos, D. Stevens, A. Mundt, D. Leibfried, F. Schmidt-Kaler, and R. Blatt. Speed of ion-trap quantum-information processors. Phys. Rev. A **62**, 042305 (2000).

[9] D. Loss and D. P. DiVincenzo. Quantum computation with quantum dots. Phys. Rev. A **57**, 120-126 (1998).

[10] A. G. Huibers, M. Switkes, C. M. Marcus, K. Campman, and A. C. Gossard. Dephasing in open quantum dots. Physica B: Condensed Matter, **249-251**, 348 - 352 (1998).

[11] D. P. DiVincenzo. Two-bit gates are universal for quantum computation. Phys. Rev. A, **51**, 1015- 1022 (1995).

[12] R. R. Ernst, G. Bodenhausen, and A. Wokaun. Principles of Nuclear Magnetic Resonance inOne and Two Dimensions. Oxford University Press, Oxford (1987).

[13] J. Wrachtrup, S. Y. Kilin, and A. P. Nizovtsev. Quantum computation using the 13c nuclear spins near the single nv defect center in diamond. Optics and Spectroscopy, **91**, 429-437 (2001).

[14] F. Grazioso, B. R. Patton, P. Delaney, M. L. Markham, D. J. Twitchen, and J. M. Smith. Measurement of the full stress tensor in a crystal using photoluminescence from point defects: The example of nitrogen vacancy centers in diamond. Applied Physics Letters, **103**, 101905 (2013).

[15] J. E. Mooij, T. P. Orlando, L. Levitov, L. Tian, C. H. van der Waal, and S. Lloyd. Josephson persistent-current qubit. Science, **285**, 1036-1039 (1999).

[16] T. Yamamoto, Y. A. Pashkin, O. Astafiev, Y. Nakamura, and J. S. Tsai. Demonstration of conditional gate operation using superconducting charge qubits. Nature, **425**, 941-944, 10 (2003).

[17] C. Cohen-Tannoudji, B. Diu, and F. Laloe. Quantum Mechanics (vol.1),. WILEY-VCH, wiley-vch edition, (2005).

[18] M. A. Nielsen and Isaac L. Chuang. Quantum information and computation. Quantum In- formation and Computation, (2000).

[19] C. H. Bennett and G. Brassard. Quantum Cryptography: Public key distribution and coin tossing. Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, 175, Bangalore, (1984).

[20] C. H. Bennett. Quantum cryptography using any two nonorthogonal states. Physical Review Letters, **68**, 3121 (1992).

[21] F. Grosshans, G. Van Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier. Quantum key distribution using gaussian-modulated coherent states. Nature, **421**, 238-241, 01 (2003).

[22] V. Scarani, A. Acín, G. Ribordy, and N. Gisin. Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations. Phys. Rev. Lett., **92**, 057901 (2004).

[23] F. Grazioso and F. Grosshans. Quantum-key-distribution protocols without sifting that are resistant to photon-number- splitting attacks. Phys. Rev. A, **88**, 052302 (2013).

[24] V. C. Usenko and F. Grosshans. Unidimensional continuous-variable quantum key distribution. arXiv:1504.07093.

[25] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. Problemy Peredachi Informatsii, **9**, 3-11 (1973).

[26] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. Nature, **299**, 802-803 (1982).

[27] D. Dieks. Communication by EPR devices. Physics Letters A, **92**, 271-272 (1982).

[28] S. W. Hawking. Black hole explosions? Nature, **248**, 30-31 (1974).

[29] P. Debye. Interferenz von rüntgenstrahlen und wärmebewegung. Annalen der Physik, **348**, 49- 92 (1913).

[30] W. Nernst. Über einen versuch von quantentheoretischen betrachtungen zur annahme stetiger energie änderungen zurückzukehren. Verhandlungen der Deutschen Physikalischen Gesellschaft, **4**, (S 83) (1916).

[31] W. K. Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. Zeitschrift für Physik, **43**, 172- 198 (1927).

[32] S. W. Hawking. Particle creation by black holes. Communications in mathematical physics, **43**, 199-220 (1975).

[33] S. W. Hawking. Breakdown of predictability in gravitational collapse. Physical Review D, **14**, 2460 (1976).

[34] C. W. Misner, K. S. Thorne, and John A. Wheeler. Gravitation. W. H. Freeman and Company, (1973).

[35] N. D. Birrell and P. C. W. Davies. Quantum fields in curved space. Cambridge University Press, (1984).

[36] J. D. Bekenstein. Black holes and information theory. Contemporary Physics, **45**, 31-43 (2004).

[37] J. D. Bekenstein. Black holes and the second law. Lettere Al Nuovo Cimento (1971-1985), **4**, 737-740 (1972).

[38] J. D. Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. Physical Review D, **23**, 287 (1981).

[39] R. Bousso. The holographic principle. Rev. Mod. Phys., **74**, 825-874 (2002).

[40] P. Hayden and J. Preskill. Black holes as mirrors: quantum information in random subsystems. Journal of High Energy Physics, **2007**, 120 (2007).

[41] D. N. Page. Average entropy of a subsystem. Phys. Rev. Lett., **71**, 1291-1294 (1993).

[42] L. Susskind, L. Thorlacius, and J. Uglum. The stretched horizon and black hole complementarity. Phys. Rev. D, **48**, 3743-3761 (1993).

[43] G. ’t Hooft. On the quantum structure of a black hole. Nuclear Physics B, **256**, 727 - 745 (1985).

[44] G. ’t Hooft. The holographic principle. arXiv preprint hep-th/0003004 (2000).

[45] M. Gell-Mann and F. E. Low. Quantum electrodynamics at small distances. Phys. Rev., **95**, 1300-1312 (1954).

[46] K. G. Wilson and J. Kogut. The renormalization group and the ε expansion. Physics Reports, **12**, 75-199 (1974).

[47] K. G. Wilson. Renormalization group methods. Advances in Mathematics, **16**, 170-186 (1975).

[48] K. G. Wilson. The renormalization group: Critical phenomena and the Kondo problem. Rev. Mod. Phys., **4**, 773-840 (1975).

[49] C. G. Callan. Broken scale invariance in scalar field theory. Phys. Rev. D, **2**, 1541-1547 (1970).

[50] K. Symanzik. Small distance behaviour in field theory and power counting. Communications in Mathematical Physics, **18**, 227-246 (1970).

[51] K. Symanzik. Small-distance-behaviour analysis and wilson expansions. Communications in Math- ematical Physics, **23**, 49-86 (1971).

[52] A. B. Zamolodchikov. Irreversibility of the flux of the renormalization group in a 2d field theory. JETP lett, **43**, 730-732 (1986).

[53] S. M. Apenko. Information theory and renormalization group flows. Physica A: Statistical Me- chanics and its Applications, **391**, 62-77 (2012).

[54] C. Bény and T. J. Osborne. Information geometric approach to the renormalisation group. arXiv:1206.7004 (2012).

[55] C. Bény and T. J. Osborne. Information loss along the renormalization flow. Verhandlungen der Deutschen Physikalischen Gesellschaft, **48**, 1 (2013).

[56] J. L. Cardy. Is there a c-theorem in four dimensions? Physics Letters B, **215**, 749 - 752 (1988).

[57] I. Jack and H. Osborn. Analogs of the c-theorem for four-dimensional renormalizable field theories. Nuclear Physics B, **343**, 647 - 688 (1990).

[58] A. Cappelli, D. Friedan, and J. Latorre. c-theorem and spectral representation. Nuclear Physics B, **352**, 616 - 670 (1991).

[59] T. Appelquist, A. G. Cohen, and M. Schmaltz. A new constraint on strongly coupled field theories. Phys. Rev. D, **60**, 045003 (1999).

[60] H. Casini, M. Huerta, R. C. Myers, A. Yale. Mutual information and the F-theorem. arXiv preprint hep-th/1506.06195 (2015).

[61] S. R. White. Density matrix formulation for quantum renormalization groups. Phys. Rev. Lett., **69**, 2863-2866 (1992).

[62] R. M. Noack and S. R. White. The density matrix renormalization group. In Density-Matrix Renormalization, 27-66. Springer (1999).

[63] T. J. Osborne and M. A. Nielsen. Entanglement, quantum phase transitions, and density matrix renormalization. Quantum Information Processing, **1**, 45-53 (2002).

[64] D. Aharonov. Quantum to classical phase transition in noisy quantum computers. Phys. Rev. A, **62**, 062311 (2000).

[65] S. Sachdev and B. Keimer. Quantum criticality. Physics Today, **64**, 29 (2011).

[66] G. Parisi. Statistical field theory, volume 28. Perseus Books New York, (1998).

Fabio Grazioso has graduated at Università degli Studi di Napoli Federico II (Italy) and has received the DPhil degree at University of Oxford (United Kingdom). He has held the position of postdoctoral fellow at École Normale Supérieure de Cachan (France), and at the Université de Montréal (Canada). He is currently postdoctoral fellow at INRS-EMT (Canada), and research associate in science education at Università di Napoli Federico II (Italy). His research interests are all centered around quantum information theory. In particular he has worked in projects involving experimental and theoretical quantum optics, low temperature spectroscopy of crystal color defects, experimental and theoretical quantum key distribution and quantum complexity theory.