

Polytechnic of Turin

Master Degree in Physics of Complex Systems

Identifying classical and quantum causal
relations with transfer entropy

MASTER THESIS

Niccolò Ciolli

Under the supervision of:
Prof. Dr. Davide Girolami
Prof. Dr. Otfried Gühne

Turin, 2024

Acknowledgement

I would like to express my deepest gratitude to my supervisors, Professor Otfried Ghne and Professor Davide Girolami, for their invaluable guidance, support, and encouragement throughout the entire process of researching and writing this thesis.

I would also like to thank all the students in the Theoretical Quantum Optics group in Siegen, especially Ph.D. student Carlos de Gois, for the help and advices during my visiting period at university of Siegen.

Summary

In the realm of physics, causality remains an enigmatic and debated concept. Though numerous measures of causality have been proposed, including Causal Influence [1], Liam’s causality [2], Pearl’s average causal effect [3], Granger’s causality [4], and Transfer entropy [5], a unified consensus on its nature and measurement has yet to be reached.

This thesis focuses on transfer entropy, a model-free metric introduced by Thomas Schreiber, with widespread applications across many disciplines. Notably, transfer entropy has been used in fields like neuroscience ([6], [7]), finance ([8], [9]), and the study of complex systems ([10], [11]).

Our aim is to provide a comprehensive understanding of transfer entropy and explore its application to quantum circuits. We investigate how transfer entropy behaves when data originates from quantum processes, focusing particularly on spin measurements, different initial states, and different channels.

To achieve this, we developed a program ([12]) that can simulate classical and quantum measurements of two interacting observables and calculate the transfer entropy from the obtained time series. For the quantum case, we derive an analytical formula for transfer entropy in our specific setup, enabling us to study its properties and calculate its value without relying on simulation.

Table of Contents

Acknowledgement	i
Abstract	ii
List of Figures	iv
List of Tables	v
1 Transfer entropy: definition and properties.	1
1.1 Information Theory	1
1.2 Definition of <i>transfer entropy</i>	5
1.2.1 Environment and Notation	5
Notation	6
1.2.2 Definition	6
1.2.3 Properties	6
Dependence on the l parameter	7
Transfer entropy's bound	9
1.2.4 Transfer entropy bound saturation	10
1.2.5 Study of interesting processes	15
Transfer entropy critiques	15
Stochastic and deterministic processes	16
2 Transfer entropy on quantum circuits	19
2.1 Introduction	19
2.1.1 Quantum mechanical concepts	19
Density operator	19
Projective measurements	20
Pauli matrices	20
Measuring the spin in the (\hat{n}) direction	21
2.1.2 Composite systems	22
Tensor product	22
Entanglement	22
Schmidt decomposition	22
Channels	23
Kraus-Cirac decomposition of non-local unitary operators	24
No-communication theorem	24
2.1.3 Quantum information	25

	Quantum computation	25
	Quantum entropies	27
2.2	Transfer entropy on quantum processes	28
2.2.1	Simulations	29
	Rotating target qubit measurements	34
2.2.2	Analytical study	35
2.2.3	Properties	40
3	Conclusions and outlooks	45
	References	47

List of Figures

1.1	Venn diagrams for entropy, conditional entropy and mutual information. . .	4
1.2	Example of the notation used to represent two systems	6
1.3	Scheme of sum terms in the bound Eq. 1.17	10
1.4	Transfer entropy varying in function of c and k for the dynamic of Ex. 1.2.1 .	12
1.5	Here Y evolution is random: it assumes a random value at each time-step . .	16
1.6	Here Y evolution is deterministic: it multiplies by -1 the previous value at each time-step	16
2.1	Bloch sphere.	26
2.2	Fifteen curves are represented, on the y-axis we have the value of the transfer entropy while on the x-axis we plot the angle θ of the rotation, which defines the measured observable, from σ_z ($\theta = 0$) to σ_x ($\theta = 1$). The value on the horizontal axis must be multiplied by $\pi/2$	30
2.3	Transfer entropy for state in Eq. 2.7, varying q	32
2.4	Transfer entropy's values when rotating the measurements on the target qubit B from σ_x to σ_y	34
2.5	Data for the entangled initial state $\frac{1}{\sqrt{2}}(00\rangle + 11\rangle)$, rotating measurements on source qubit A . Each point in the surface of the sphere represents a measure of spin in the direction of the versor connecting the center of the sphere with the point. The rotation happens from σ_x to σ_y , then to σ_z , then back to σ_x .	39
2.6	Value of the transfer entropy rotating respectively the measurements on qubit A and B for initial state $ +0\rangle$	40

List of Tables

2.1	Notation used on section 2.2.	29
2.2	Initial states that outputs a certain value of the transfer entropy	31
2.3	Notation for subsection 2.2.3	40

Chapter 1

Transfer entropy: definition and properties.

1.1 Information Theory

Information theory is the mathematical study of the quantification, storage, and communication of information. The key measure in information theory is *entropy*: It quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. It is the building block of many important information theory quantities, which we are going to use extensively in this thesis. For an exhaustive explanation of entropy and entropy related quantities, and how they are linked to quantum information, we refer to the following review [13]. The most commonly used measure of entropy was proposed by Shannon in his renowned article [14]:

Definition 1.1.1. *The entropy H of a variable X is defined as:*

$$H(X) := - \sum_{x \in X} p(x) \log p(x), \quad (1.1)$$

where $p(x)$ is a discrete probability distribution of values x of X .

In the following we report basic properties of Shannon's entropy H :

- It is non-negative.
- $H(X)$ is a continuous function of X .
- $H(X)$ is a symmetric function of its arguments.

For two variables X and Y , the entropy of the joint system is $H(X, Y)$.

We can derive another key quantity, *conditional entropy*: it quantifies the amount of information needed to describe the outcome of a random variable X given that the value of another random variable Y is known.

Definition 1.1.2. *The conditional entropy of X conditioned on Y is*

$$H(X|Y) := - \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(y)} = H(X, Y) - H(Y). \quad (1.2)$$

We list some properties of the above-mentioned quantities:

- Entropy and conditional entropy are both non-negative, as it can be immediately seen, since the logarithm is always negative for arguments less than 1.
- If X and Y are two independent random variables, we have $H(X|Y) = H(X)$, due to the factorization of the probability: $p(x, y) = p(x)p(y)$, the formula for the conditional entropy Eq. (1.2) reduces to the one of the Shannon entropy.
- The maximal entropy of a variable that can take n values is $\log n$ and it is reached when the probability distribution is uniform.

Most of the important quantities in information theory are described with entropy measures. We introduce the Kullback-Leibler Divergence, also called *relative entropy*. It is a quantity used as "distance" between two probability distributions. Let us imagine that a variable has a probability distribution Q , while the real distribution is P . Kullback-Leibler Divergence is a measure that tells us how wrong we are with our assumption. It is often related to the storing and processing of information, which are key concepts in this thesis.

Definition 1.1.3 (Kullback-Leibler Divergence). *Given two probability distributions $p(x)$ and $q(x)$ of values of X :*

$$D_{KL}(p(x)||q(x)) := \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \quad (1.3)$$

To show that the Kullback-Leibler divergence can represent a distance between probability distributions, we prove that it is always non-negative.

Property 1.1.1. *Considering P and Q as probability distributions for the observable X ,*

$$D_{KL}(P||Q) \geq 0 \quad \forall P, Q \quad (1.4)$$

Proof. To prove the above mentioned property, we will use the following inequality

$$-\log x \geq 1 - x.$$

Starting from Eq. (1.3)

$$\sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = \sum_{x \in X} p(x) \left(-\log \frac{q(x)}{p(x)} \right) \geq \sum_{x \in X} p(x) \left(1 - \frac{q(x)}{p(x)} \right).$$

Using the property of probability distributions $\sum_{x \in X} p(x) = 1$,

$$\sum_{x \in X} p(x) \left(1 - \frac{q(x)}{p(x)} \right) = \sum_{x \in X} p(x) - \sum_{x \in X} q(x) = 1 - \sum_{x \in X} q(x)$$

note that non-negativity is obtained when $1 - \sum_{x \in X} q(x) \geq 0$, meaning that it is verified for all Q that represent a probability distribution ($\sum_{x \in X} q(x) = 1$) but also for non-normalized probability distributions with the constrain that $\sum_{x \in X} q(x) \leq 1$. \square

Kullback-Leibler divergence fails as a metric in the usual sense, because it is asymmetric. The asymmetry comes from the fact that we assume one distribution to be the true one, in this case P , since the expectation value is calculated with respect to P . We are calculating the distance of our assumed distribution from the "true" distribution of the variable. For a more detailed explanation, see [15].

To represent and quantify the information shared by two variables, *mutual information* is often used.

Definition 1.1.4 (Mutual Information). *The mutual information between two variables is defined as:*

$$I(X : Y) = \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = H(X) + H(Y) - H(X, Y), \quad (1.5)$$

When we desire to measure the information shared between two systems, after knowing the value of a third additional variable, we use the conditional mutual information:

Definition 1.1.5 (Conditional mutual information). *The conditional mutual information between two variables X and Y , considering an additional variable Z , is defined as:*

$$I(X : Y|Z) := \sum_{x \in X, y \in Y, z \in Z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = H(X|Z) - H(X|Y, Z). \quad (1.6)$$

A convenient and immediate way to visualize entropy quantities is to use Venn diagrams, as we show in Fig. 1.1 Mutual information Eq. (1.5) and conditional mutual information Eq. (1.6) can be written as Kullback-Leibler divergences, thus inheriting non-negativity:

$$I(X : Y) = D_{KL}(p(X, Y)||p(X)p(Y)).$$

Mutual information is equal to the distance between the joint probability distribution of variables X and Y , $p(X, Y)$, and the joint probability distribution assuming that the variables are independent, $p(X)p(Y)$.

$$I(X : Y|Z) = D_{KL}(p(X, Y, Z)||p(X|Z)p(Y|Z)p(Z)), \quad (1.7)$$

with the same meaning as before, after knowing the value of variable Z .

Mutual Information and conditional mutual information are symmetric with respect to X and Y .

Property 1.1.2. *The conditional mutual information is upper bounded by the conditional entropy*

$$I(X : Y|Z) \leq H(X|Z). \quad (1.8)$$

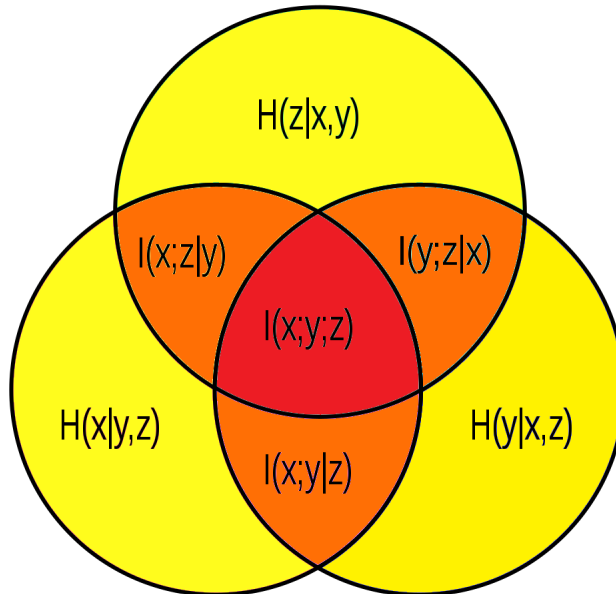


Figure 1.1: Venn diagrams for entropy, conditional entropy and mutual information.

This is clear from the Venn diagram representation shown in Fig. 1.1.

Property 1.1.3. *The conditional entropy of X conditioned on Y vanishes if and only if there exists an analytical function such that $X = f(Y)$:*

$$H(X|Y) = 0 \iff \exists f : f(Y) = X. \quad (1.9)$$

Proof. $f(y) = x$ can be interpreted with probability distributions as: for all x_0 such that $p(x_0) > 0$ there exist a unique y_0 such that $p(y_0|x_0) = 1$.

First we prove the " \Rightarrow " relation. Assuming conditional mutual information to be zero:

$$H(X|Y) = - \sum_{x \in X} \sum_{y \in Y} p(y)p(x|y) \log p(x|y) = 0.$$

Since every summand is non-negative, for the conditional mutual information to vanish each of them must be zero. Imposing this condition, we obtain:

$$p(y) > 0 \implies p(x|y) \log p(x|y) = 0.$$

One has $t \log t = 0 \iff t \in \{0, 1\}$, and the properties of probability distribution assure that $\sum_{x \in X} p(x|y) = 1$, so for each y in the conditional mutual information summation there exists one and only one x s.t. $p(x|y) = 1$. When this happens we can conclude that $x = f(y)$. To extend the property to the whole system we notice that:

$$p(x) > 0 \implies p(y|x) = \delta(x - f(y)) = \begin{cases} 1 & x = f(y) \\ 0 & x \neq f(y) \end{cases}$$

where δ is the Kronecker-delta.

We claim that $X = f(Y)$ with probability 1.

Whenever $p(x) > 0$:

$$\begin{aligned}
 P(X = f(Y)) &= \mathbb{E}[\delta(x - f(y))] \\
 &= \sum_{y \in Y} \sum_{x \in X} p(x, y) \delta(x - f(y)) \\
 &= \sum_{y \in Y} p(x) \sum_{x \in X} p(x|y) \delta(x - f(y)) \\
 &\stackrel{a)}{=} \sum_{y \in Y} p(x) \sum_{x \in X} p(x|y) \\
 &= \sum_{y \in Y} \sum_{x \in X} p(x, y) = 1
 \end{aligned}$$

where in the a) step, we reduce the summation over all $x = f(y)$, and we saw previously that when $H(X|Y) = 0$ for every $p(y) \neq 0$ there exists a unique x that satisfies this condition. To prove the " \Leftarrow " we assume that there exists $f : X = f(Y)$. This means that each $x \in X$ is a function of an $y \in Y$, obtaining: $p(x|y) = \delta(x - f(y))$. The conditional entropy thus becomes:

$$H(X|Y) = - \sum_{x \in X} \sum_{y \in Y} p(y) \delta(x - f(y)) \log \delta(x - f(y)) = 0.$$

Because $\delta \in \{0, 1\}$ and $t \log t = 0 \iff t \in \{0, 1\}$. □

1.2 Definition of *transfer entropy*

To define the *transfer entropy*, we have to start by Granger's definition of causality [4].

Definition 1.2.1 (Granger null causality principle). *Given two variables X and Y , when past observations of Y do not influence our ability to predict X , then there is no causation effects from Y to X . This can be summarized as a probability distribution:*

$$p(x_{t+1}|x_t, \dots, x_{t-k}) = p(x_{t+1}|x_t, \dots, x_{t-k}, y_t, \dots, y_{t-l}) \quad (1.10)$$

for arbitrary k and l .

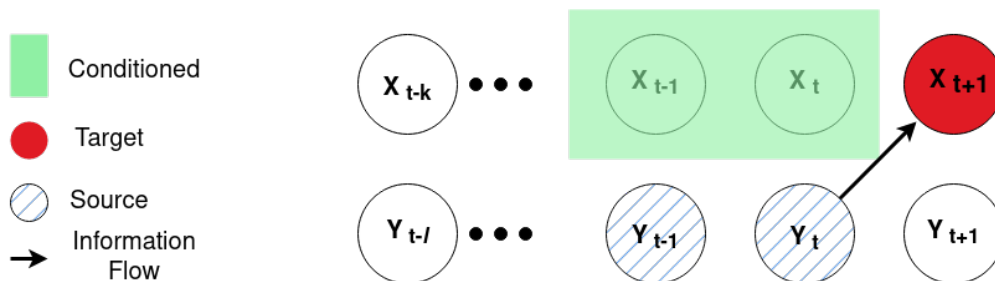
Transfer entropy is a measure, defined by Thomas Schreiber [5], originated from the need to quantify Granger causality. In order to define properly the transfer entropy, we will set up the environment and notation we are going to be working with.

1.2.1 Environment and Notation

We are considering two variables that evolve in time, X and Y , representing our information about object/system/state and X_t, Y_t constitute measurements/realizations of said systems at time t . X_t, Y_t will be referred as variables which represent points in space-time.

A very convenient representation, shown in Fig. 1.2, is a graph where the nodes are space-time points and the arrows the information flow between variables. The highlighted points are the conditioned ones (it will make more sense when we define the transfer entropy), the blue hatched ones are sources of information flow, the red one is the target.

Figure 1.2: Example of the notation used to represent two systems



Notation

We will represent a set of multiple variables in the following way:

- $x_t^k = \{x_{t-k}, \dots, x_{t-1}, x_t\}$ represents the set of variables of system X from time $t - k$ to t so that $p(x_t^k) = p(x_{t-k}, \dots, x_{t-1}, x_t)$ is the joint distribution of the variables
- We define $x_t^0 = x_t$, $x_t^{-1} = \emptyset$
- Note that $\{x_{t+1}, x_t^k\} = x_{t+1}^{k+1}$
- When a parameter of the system is left unchanged, or it does not influence the calculation, it will not be appearing when explicating the argument of functions.

1.2.2 Definition

Now we are ready to present the definition of the transfer entropy:

Definition 1.2.2 (Transfer entropy). *The incorrectness of the assumption made by Granger to define causality in Def. 1.2.1 can be quantified with:*

$$T_{Y \rightarrow X}(t+1, k, l) = \sum_{x_{t+1}^{k+1}, y_t^l} p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1} | x_t^k, y_t^l)}{p(x_{t+1} | x_t^k)}. \quad (1.11)$$

It is said that is a non-linear generalization of Granger causality, because it is zero when Granger non-causality condition (Def. 1.2.1) is true, and it is shown to be equivalent to Granger causality when the variables are Gaussian distributed in the following paper [16]. Transfer entropy is a measure proposed to determine information flow between two systems. Specifically, it measures how useful are the y_t^l variables to predict the value of x_{t+1} given that x_t^k are known. The variables y_t^l assume the role of the considered source of information flow, x_{t+1} the target of such flow and x_t^k are the target's past. In the following we will assume that there is causal influence if and only if the transfer entropy is positive.

1.2.3 Properties

Transfer entropy can be written as a particular conditional mutual information, inheriting the previously discussed properties

$$T_{Y \rightarrow X}(t+1, k, l) = I(X_{t+1} : Y_t^l | X_t^k) \quad (1.12)$$

just by noticing that $p(x_{t+1}, y_t^l | x_t^k) / p(y_t^l | x_t^k) = p(x_{t+1} | y_t^l, x_t^k)$

A brief recap of the previously presented properties of conditional mutual information:

- $T_{Y \rightarrow X} \geq 0$ transfer entropy is non-negative.
- $T_{Y \rightarrow X}(t+1, k, l) \leq H(X_{t+1} | X_t^k)$ transfer entropy is bounded by conditional entropy of the target variable and its past.

Dependence on the l parameter

We can start to look into the properties of transfer entropy by asking how its value changes when an additional source is considered in our calculation. For this purpose we examine the quantity:

$$T_{Y \rightarrow X}(l+1) - T_{Y \rightarrow X}(l).$$

obtaining the behaviour of transfer entropy in function of l .

Property 1.2.1. *The difference between transfer entropy for two consequent l can be written as a Kullback-Leibler divergence*

$$T_{Y \rightarrow X}(l+1) - T_{Y \rightarrow X}(l) = D_{KL}[p(x_{t+1}, x_t^k, y_t^{l+1}) || p(x_{t+1} | x_t^k, y_t^l) p(x_t^k, y_t^l | y_{t-(l+1)}) p(y_{t-(l+1)})] \quad (1.13)$$

This property can be interpreted as follows: adding a new source to the transfer entropy influences it for an amount that is the measure of difference between two probability distributions, one obtained considering the variables x_{t+1}^{k+1} dependent from the new source $y_{t-(l+1)}$ (the probability distribution as a first argument in the Kullback-Leibler divergence) and the second one obtained considering the x_{t+1} variable dependent on (x_t^k, y_t^l) but independent from the last introduced source $y_{t-(l+1)}$ while the set of variables (x_t^k, y_t^l) depends on it.

Proof. We start by rewriting explicitly the transfer entropy formula in Eq. (1.11) for $T_{Y \rightarrow X}(l+1) - T_{Y \rightarrow X}(l)$ and we obtain:

$$\sum_{y_t^{l+1}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1} | x_t^k, y_t^{l+1})}{p(x_{t+1} | x_t^k)} - \sum_{y_t^l} p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1} | x_t^k, y_t^l)}{p(x_{t+1} | x_t^k)}$$

We now use the marginal probability formula $p(x_{t+1}, x_t^k, y_t^l) = \sum_{y_{t-(l+1)}} p(x_{t+1}, x_t^k, y_t^{l+1})$ and separate the summation

$$\sum_{y_t^l} \sum_{y_{t-(l+1)}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1} | x_t^k, y_t^{l+1})}{p(x_{t+1} | x_t^k)} - \sum_{y_t^l} \sum_{y_{t-(l+1)}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1} | x_t^k, y_t^l)}{p(x_{t+1} | x_t^k)},$$

the sum over all variable that are not $y_{t-(l+1)}$ is equal for both terms, so it can be collected, together with the full probability distribution,

$$\sum_{y_t^l} \sum_{y_{t-(l+1)}} p(x_{t+1}, x_t^k, y_t^{l+1}) \left(\log \frac{p(x_{t+1} | x_t^k, y_t^{l+1})}{p(x_{t+1} | x_t^k)} - \log \frac{p(x_{t+1} | x_t^k, y_t^l)}{p(x_{t+1} | x_t^k)} \right)$$

and using logarithm's properties,

$$\sum_{y_t^l} \sum_{y_{t-(l+1)}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1}|x_t^k, y_t^{l+1})p(x_{t+1}|x_t^k)}{p(x_{t+1}|x_t^k)p(x_{t+1}|x_t^k, y_t^l)}.$$

After simplifying $p(x_{t+1}|x_t^k)$, we use Bayes theorem to rewrite the probabilities $p(x_{t+1}|x_t^k, y_t^{l+1})p(x_t^k, y_t^{l+1}) = p(x_{t+1}, x_t^k, y_t^{l+1})$ and we obtain:

$$\sum_{y_t^{l+1}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1}, x_t^k, y_t^{l+1})}{p(x_{t+1}|x_t^k, y_t^l)p(x_t^k, y_t^{l+1})}.$$

We rewrite the probability in the denominator in a more significant way, using Bayes theorem, to get the formula

$$\sum_{y_t^{l+1}} p(x_{t+1}, x_t^k, y_t^{l+1}) \log \frac{p(x_{t+1}, x_t^k, y_t^{l+1})}{p(x_{t+1}|x_t^k, y_t^l)p(x_t^k, y_t^l|y_{t-(l+1)})p(y_{t-(l+1)})}$$

We can recognise the form of a Kullback-Leibler divergence. The only thing that remains to be proven is that the denominator of the log is indeed a probability distribution, i.e. it sums to 1:

$$\begin{aligned} \sum_{y_t^{l+1}} p(x_{t+1}|x_t^k, y_t^l)p(x_t^k, y_t^l|y_{t-(l+1)})p(y_{t-(l+1)}) &= \sum_{y_t^l} p(x_{t+1}|x_t^k, y_t^l) \sum_{y_{t-(l+1)}} p(x_t^k, y_t^{l+1}) = \\ &= \sum_{y_t^l} p(x_{t+1}|x_t^k, y_t^l) \sum_{y_{t-(l+1)}} p(x_t^k, y_t^{l+1}) = \sum p(x_{t+1}, x_t^k, y_t^l) = 1. \end{aligned}$$

Where we just used the previously mentioned Bayes theorem and marginal probability distribution formula. Now we can say that indeed the quantity that we found can be written as a Kullback-Leibler divergence, proving the property. \square

Keeping in mind property 1.2.1, and naming

$$D(l+1) = D_{KL}[p(x_{t+1}, x_t^k, y_t^{l+1})||p(x_{t+1}|x_t^k, y_t^l)p(x_t^k, y_t^l|y_{t-(l+1)})p(y_{t-(l+1)})],$$

transfer entropy can be reformulated as:

$$T_{Y \rightarrow X}(l) = D(l) + T_{Y \rightarrow X}(l-1).$$

We notice that $T_{Y \rightarrow X}(-1) = 0$, where $l = -1$ means that we consider no source (see the notation). Transfer entropy can finally be written as:

$$T_{Y \rightarrow X}(l) = \sum_{i=0}^l D(i). \tag{1.14}$$

The above mentioned Eq. (1.14), displays two behaviours of the transfer entropy:

- Transfer entropy is monotonically increasing as a function of l , since $D_{KL} \geq 0$, from the properties of Kullback-Leibler divergence
- Since $D(l) = T_{Y \rightarrow X}(l) - T_{Y \rightarrow X}(l-1)$, this quantity can be considered as a measure of the strength of information flow due to a specific source (it could be used to model our system more efficiently or to gain more insight about the dynamics)

Transfer entropy's bound

Transfer entropy can be written as a conditional mutual information, and from property 1.1.2 we can write

Property 1.2.2. *Transfer entropy is bounded by the conditional entropy of X_{t+1} and its own past X_t^k .*

$$T_{Y \rightarrow X}(t+1, k, l) \leq H(X_{t+1} | X_t^k) \quad (1.15)$$

To better understand what happens when we consider an additional variable in the past of the target in our calculation, we start by studying how the upper bound varies in function of parameter k . We analyze the difference between two conditional entropies

Property 1.2.3. *The difference between two conditional entropies that variate for a conditioned variable $X_{t-(k+1)}$:*

$$H(X_{t+1} | X_t^{k+1}) - H(X_{t+1} | X_t^k) = -T_{X_{t-(k+1)} \rightarrow X}(k) \quad (1.16)$$

Proof. We start by considering the difference of the two conditional entropies:

$$\begin{aligned} & H(X_{t+1} | X_t^{k+1}) - H(X_{t+1} | X_t^k) = \\ & = - \sum_{x_{t-(k+1)}} \sum_{x_t^{k+1}} p(x_{t+1}, x_t^{k+1}) \log \frac{p(x_{t+1}, x_t^{k+1})}{p(x_t^{k+1})} + \sum_{x_t^k} p(x_{t+1}, x_t^k) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ & \stackrel{a)}{=} - \sum_{x_{t-(k+1)}} \sum_{x_t^{k+1}} p(x_{t+1}, x_t^{k+1}) \log \frac{p(x_{t+1}, x_t^{k+1})}{p(x_t^{k+1})} + \sum_{x_{t-(k+1)}} \sum_{x_t^k} p(x_{t+1}, x_t^{k+1}) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ & \stackrel{b)}{=} - \sum_{x_{t-(k+1)}} \sum_{x_t^{k+1}} p(x_{t+1}, x_t^{k+1}) \left(\log \frac{p(x_{t+1}, x_t^{k+1})}{p(x_t^{k+1})} - \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \right) \\ & \stackrel{c)}{=} - \sum_{x_{t-(k+1)}} \sum_{x_t^{k+1}} p(x_{t+1}, x_t^{k+1}) \log \frac{p(x_{t+1}, x_t^{k+1}) p(x_t^k)}{p(x_t^{k+1}) p(x_{t+1}, x_t^k)} \\ & \stackrel{d)}{=} - \sum_{x_{t-(k+1)}} \sum_{x_t^{k+1}} p(x_{t+1}, x_t^{k+1}) \log \frac{p(x_{t+1} | x_t^k, x_{t-(k+1)})}{p(x_{t+1} | x_t^k)} \\ & \stackrel{e)}{=} -T_{X_{t-(k+1)} \rightarrow X}(k). \end{aligned}$$

- a) marginal probability property was used, $p(x_{t+1}, x_t^k) = \sum_{x_{t-(k+1)}} p(x_{t+1}, x_t^{k+1})$
- b) the term $\sum_{x_{t-(k+1)}} p(x_{t+1}, x_t^{k+1})$ was collected
- c) we applied the property for the difference of logarithms
- d) we used Bayes theorem
- e) we apply the definition of transfer entropy noticing that $x_{t-(k+1)}$ has the role of source in the formulation.

□

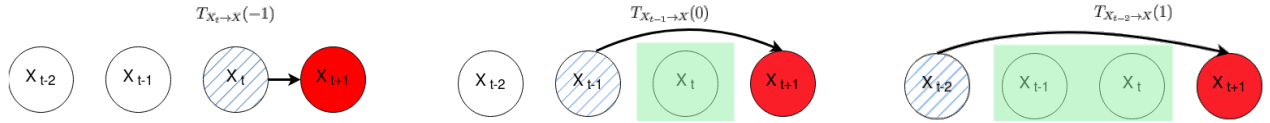


Figure 1.3: Scheme of sum terms in the bound Eq. 1.17

Corollary 1.2.0.1. *The bound of transfer entropy can be rewritten recursively using property 1.2.3 as:*

$$H(X_{t+1}|X_t^k) = H(X_{t+1}) - \sum_{i=0}^k T_{X_{t-i} \rightarrow X}(i-1) \quad (1.17)$$

The term $H(X_{t+1})$ denotes the entropy of the target variable (system X observed at time $t+1$), it represents the uncertainty of X_{t+1} : It can also be seen as the space in X_{t+1} to receive information. Indeed the more information it receives the less uncertain it becomes. The terms in the sum are fluxes of information from a single source variable in its own past, given that all the following variables are known (i.e. conditioned on) as represented in Fig. 1.3. The bound of the transfer entropy the maximal information that a system Y can send to a system X , is the actual information that the target variable can receive, minus all the information it receives from it is past.

1.2.4 Transfer entropy bound saturation

We aim to find the probability distribution that maximizes transfer entropy.

Theorem 1.2.1. *Transfer entropy is equal to its bound, Eq. 1.15, when system X is obtained by a deterministic process from (X_t^k, Y_t^l) , i.e. X_{t+1} is an analytic function of (X_t^k, Y_t^l) . Which means:*

$$p(x_{t+1}, x_t^k, y_t^l) = p(x_t^k, y_t^l).$$

Proof. We equate the transfer entropy to its bound:

$$\begin{aligned} T_{Y \rightarrow X}(k, l) &= H(X_{t+1}|X_t^k) \\ \sum p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1}|x_t^k, y_t^l)}{p(x_{t+1}|x_t^k)} &= - \sum p(x_{t+1}, x_t^k) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ \sum p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1}, x_t^k, y_t^l)p(x_t^k)}{p(x_{t+1}, x_t^k)p(x_t^k, y_t^l)} &= - \sum p(x_{t+1}, x_t^k) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ \sum p(x_{t+1}, x_t^k, y_t^l) \left(\log \frac{p(x_{t+1}, x_t^k, y_t^l)}{p(x_t^k, y_t^l)} - \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \right) &= - \sum p(x_{t+1}, x_t^k) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ \sum p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1}, x_t^k, y_t^l)}{p(x_t^k, y_t^l)} - \sum p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} &= - \sum p(x_{t+1}, x_t^k) \log \frac{p(x_{t+1}, x_t^k)}{p(x_t^k)} \\ \sum p(x_{t+1}, x_t^k, y_t^l) \log \frac{p(x_{t+1}, x_t^k, y_t^l)}{p(x_t^k, y_t^l)} &= 0 \end{aligned}$$

Since $p(x_{t+1}, x_t^k, y_t^l) \geq 0$ and we assume that there is at least one set of value for which $p(x_{t+1}, x_t^k, y_t^l) > 0$, (otherwise we would obtain $0 = 0$, and in this case the theorem is trivially

verified), for the sum to vanish we need to have $p(x_{t+1}, x_t^k, y_t^l) = p(x_t^k, y_t^l)$. Meaning that $x_{t+1} = f(x_t^k, y_t^l)$, because

$$\begin{aligned} p(x_{t+1}, x_t^k, y_t^l) &= p(x_t^k, y_t^l) \\ \frac{p(x_{t+1}, x_t^k, y_t^l)}{p(x_t^k, y_t^l)} &= 1 \\ p(x_{t+1}|x_t^k, y_t^l) &= 1. \end{aligned}$$

The last equation exactly means that $x_{t+1} = f(x_t^k, y_t^l)$. (It was further explained in the proof of property 1.1.3) \square

We now know that deterministic processes saturate the transfer entropy. But as we have seen in Eq. (1.17), the bound of the transfer entropy is still dependent on the k parameter: the known history of our target variable. We may want to know which processes realize the transfer entropy to be equal to its bound and the bound to be maximal. We recall that conditional entropy is maximized by:

$$H(X|Y) \leq H(X) \quad H(X|Y) = H(X) \iff P(X, Y) = P(X)P(Y).$$

Then it is easy to formulate that transfer entropy is maximised when both of the required conditions happen:

$$x_{t+1} = f(x_t^k, y_t^l) \wedge p(x_{t+1}, x_t^k) = p(x_{t+1})p(x_t^k),$$

which is true, but does not give any insight in the processes that leads to a maximal transfer entropy. We present an interesting example found in the book [17] to clarify the problem.

Example 1.2.1. Consider two observables, $X \in \{-1, 1\}$ and $Y \in \{-1, 1\}$. Suppose their initial value is random and then they evolve through the following dynamics:

$$Y_t = -Y_{t-1} \quad X_t = \begin{cases} Y_{t-1} & \text{with probability } \frac{1+c}{2} \\ -Y_{t-1} & \text{with probability } \frac{1-c}{2} \end{cases}$$

Where X_t is equal to Y_{t-1} with probability $p = \frac{1+c}{2}$ or equal to $-Y_{t-1}$ with probability $1 - p$. Parameter $c \in [-1, 1]$. Since the evolution of X is not a deterministic function we can already expect that transfer entropy will not be maximal. We are interested in studying how the transfer entropy varies in function of k (In reference text it is also reported in function of c). Just by looking at the dynamics one could naively infer that:

- $X = f(Y)$, where f is not a deterministic, analytical function.
- X only depend on Y : it is independent on its past.

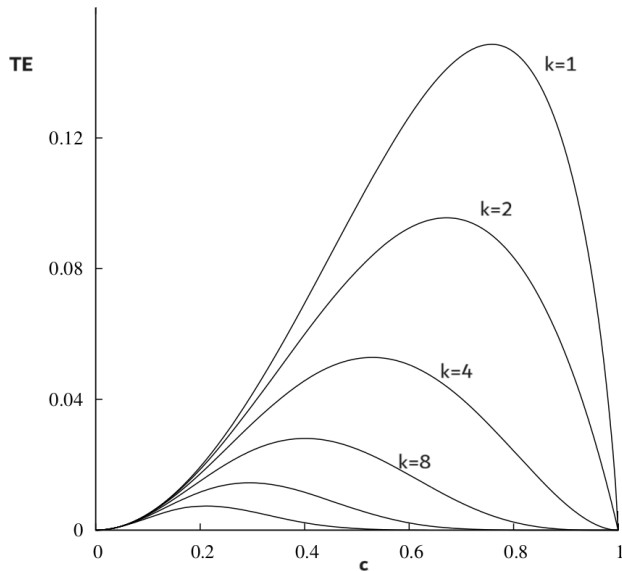


Figure 1.4: Transfer entropy varying in function of c and k for the dynamic of Ex. 1.2.1

However when we compute the transfer entropy varying c and k we obtain the result in Fig. 1.4. We can observe that, when k increases, transfer entropy decreases, and goes to 0 when $k \rightarrow \infty$. It could imply that the bound of the transfer entropy are still dependent on k , i.e. X is dependent on its past: it receives a nonzero flow of information each time the system is conditioned to an additional past variable. This is not immediately clear just by looking at the dynamics. This effect is due to the stored information about Y in X 's past. Indeed, since Y evolves deterministically, once one value is known the full series is given. Having the information about infinite variables in the past of X let us infer the value of Y , making it useless in the prediction of X_{t+1} . There is a flow of information from X 's past to X about the source of causality Y .

We conclude that a description of the dynamic that lead to a maximal transfer entropy rather than a description of the probability distributions of the system is preferable.

Theorem 1.2.2. *Transfer entropy is fully maximized, i.e. it is equal to its maximized bound*

$$T_{Y \rightarrow X}(k, l) = H(X_{t+1})$$

when:

1. $X_{t+1} = f(Y)$, the target variable X_{t+1} is an analytic function of variables of system Y .
2. f is constant with respect to variables of system X .
3. $p(Y) = p(Y_t^l)p(Y/Y_t^l)$: Y_t^l set is independent from its past.
4. The domain of $X_{t+1} = f(Y)$ is disjoint from the domain of any other X_t , $\forall t$.

Proof. From theorem 1.2.1 we know that transfer entropy is equal to its bound when $X_{t+1} = f(X_t^k, Y_t^l)$. During the proof we will assume that $X_{t+1} = f(X_t^k, Y_t^l)$. In order to obtain the maximum bound, it is required for the conditional entropy $H(X_{t+1}|X_t^k)$ to be maximal, which

happens when the two systems are independent: $p(x_{t+1}, x_t^k) = p(x_{t+1})p(x_t^k)$. Firstly we will show that if two variables are independent than they are not function of each other, meaning that $X_{t+1} = f(X_t^k, Y_t^k) = f(Y_t^k)$. Let us proceed for absurd, and suppose that two systems, A and B are independent and there exists a function between two elements $a \in A$, $b \in B$ s.t. $a = f(b)$. Then we have:

$$1 = p(a|b) = \frac{p(a, b)}{p(b)} \stackrel{1)}{=} \frac{p(a)p(b)}{p(b)} = p(a)$$

$$p(a) = 1$$

where in step 1) we used the hypothesis of independence. $p(a) = 1$ means that the only possible result of a realization of system A is a ; which means that any function with domain in B and co-domain in A has value a . Those function can be either constant or independent from B . Proceeding with the proof, we will consider $X_{t+1} = f(Y_t^l)$.

Firstly it is useful to prove a property of conditioned probability distribution for functions. Let A and B be two systems and $a \in A$ and suppose $a_1 = f(B)$ and $a_2 = g(B)$, where f and g are two analytical functions, (All the system B is in the argument, because it does not matter from which elements of B the functions depend on).

$$p(a_1, a_2|B) = 1, \quad p(a_1|B) = 1, \quad p(a_2|B) = 1 \implies p(a_1, a_2|B) = p(a_1|B)p(a_2|B).$$

The probability distributions of distinct variables, obtained from deterministic functions, conditioned on their variables are independent.

We will now prove that for the process with the properties described in the enunciation of the theorem, $p(x_{t+1}, x_t^k) = p(x_{t+1})p(x_t^k)$, making the conditional entropy maximal. Our objective is to decompose the three probability distributions that form the conditional entropy to explicit how they are influenced by the system Y . Firstly we consider the full distribution $p(x_{t+1}, x_t^k)$ and use the property of marginal distribution to rewrite it:

$$p(x_{t+1}, x_t^k) = \sum_{y_t^{k+l+1}} p(x_{t+1}, x_t^k, y_t^{k+l+1}) = \sum_{y_t^{k+l+1}} p(x_{t+1}, x_t^k | y_t^{k+l+1}) p(y_t^{k+l+1}).$$

We define the set of variables that are argument of the function X_t as \tilde{Y}_t :

$$\tilde{Y}_t = \{y : y \in Y \wedge y \in \text{dom}(X_t = f(Y))\},$$

where $\text{dom}(X_t = f(Y))$ is the domain of the function f that produce X_t . Now, recalling the property mentioned above, we can rewrite the sum as:

$$p(x_{t+1}, x_t^k) = \sum_{\tilde{y}_{t+1}^{k+1}} p(x_{t+1} | \tilde{y}_{t+1}) p(x_t | \tilde{y}_t) \dots p(x_{t-k} | \tilde{y}_{t-k}) p(\tilde{y}_{t+1}^{k+1})$$

Applying the same methodology to the other two probability distributions we obtain

$$p(x_t^k) = \sum_{\tilde{y}_t^k} p(x_{t-k} | \tilde{y}_{t-k}) p(x_{t-k+1} | \tilde{y}_{t-k+1}) \dots p(x_t | \tilde{y}_t) p(\tilde{y}_t^k)$$

$$p(x_{t+1}) = \sum_{\tilde{y}_{t+1}} p(x_{t+1} | \tilde{y}_{t+1}) p(\tilde{y}_{t+1})$$

The last step is to multiply the probabilities and verify which conditions satisfy independence.

$$\begin{aligned} p(x_t^k)p(x_{t+1}) &= \\ &= \left[\sum_{\tilde{y}_t^k} p(x_{t-k}|\tilde{y}_{t-k})p(x_{t-k+1}|\tilde{y}_{t-k+1})\dots p(x_t|\tilde{y}_t)p(\tilde{y}_t^k) \right] \left[\sum_{\tilde{y}_{t+1}} p(x_{t+1}|\tilde{y}_{t+1})p(\tilde{y}_{t+1}) \right] \end{aligned}$$

We will define Y'_t as the set of variables that belong to the set obtained by the intersection of the domain of X_t and the domain of any other X , and X'_t the variables related to these domains

$$\begin{aligned} Y'_t &= \{y : y \in Y \wedge \exists i \text{ s.t. } y \in \tilde{Y}_i \cap \tilde{Y}_i\} \\ X'_t &= \{x : x \in X \wedge \text{dom}(x) \cap Y'_t \neq \emptyset\} \end{aligned}$$

After the product, the only part of the summation that differs is the one that sums over Y'_{t+1} i.e.

$$\sum_{y'_{t+1}} \sum_{y'_{t+1}} p(x'_{t+1}|y'_{t+1})p(x'_{t+1}|y'_{t+1}). \quad (1.18)$$

Indeed from the product of the two probability distributions we get two sums on the overlap of the shared arguments between X_{t+1} and other variables. Conditional probability distribution for variables that are analytical functions conditioned on their domain are a Kronecker delta

$$p(x'_{t+1}|y'_{t+1}) = \delta(X'_{t+1} = f(Y'_{t+1}))$$

knowing that $\delta^2 = \delta$, we obtain

$$\begin{aligned} &\sum_{y'_{t+1}} \sum_{y'_{t+1}} \delta(X'_{t+1} = f(Y'_{t+1})) \\ &= \sum_{y'_{t+1}} 1 = |y'_{t+1}| |Y'_{t+1}| \end{aligned}$$

This term does not appear in the full distribution, so if we want the probability to be independent, we need to eliminate its effects, $|Y'_{t+1}|$ and $|y'_{t+1}|$ must be one, meaning that only one variable belongs to $|Y'_{t+1}|$ and its dimension is 1; there is only one possible result from observing it. A better assumption is considering Y'_t empty: the variables in the domain of X_{t+1} are not part of any other variable domain. In this case there is no overlap of sums and the product can be written as:

$$\sum_{\tilde{y}_{t+1}^{k+1}} p(x_{t-k}|\tilde{y}_{t-k})p(x_{t-k+1}|\tilde{y}_{t-k+1})\dots p(x_{t+1}|\tilde{y}_{t+1})p(\tilde{y}_t^k)p(\tilde{y}_{t+1})$$

The only condition left to have the equality $p(x_{t+1})p(x_t) = p(x_{t+1}, x_t)$ is that $p(\tilde{y}_t^k, \tilde{y}_{t+1}) = p(\tilde{y}_t^k)p(\tilde{y}_{t+1})$: the variables in the domain of X_{t+1} are independent of all variables in other domains. \square

The described dynamic is better understood if we think for example at a stationary process, where at every time step, system X is a function of the same number of past observation of system Y . In this case the conditions in theorem 1.2.2 assume a more understandable form.

1.2.5 Study of interesting processes

Transfer entropy critiques

Thinking about transfer entropy as a measure of causality in every situations can lead to misconceptions. Two are pointed out in the article [18]. We report them below.

Example 1.2.2. Consider the following process:

$$\begin{aligned} Y, X &\in \{0, 1\} \\ X_{t+1} &= X_t \oplus Y_t \\ Y_t &= \text{rand}([0, 1]) \end{aligned}$$

and X_0 and Y_0 are randomly initialized.

$H(X_{t+1})$, the entropy of system X_{t+1} is equal to 1, meaning that it can store a bit of information. The transfer entropy $T_{Y \rightarrow X}$ is also equal to 1, meaning that the process maximizes it, and all the information of the variable X_{t+1} derive from system Y . But we can reverse the logic and calculate the conditional mutual information:

$$I(X_{t+1} : X_t^\infty | Y_t^\infty) = 1.$$

This, in our picture, should represent the *information flow* to X_{t+1} from its past. We have a contradiction, if we think at the transfer entropy as a measure of information flow, because in this case, variable X_{t+1} would receive 2 bit of information while being capable to contain only 1. In short, the 1 bit of reduction in uncertainty $H(X_{t+1})$ should not be localized to either time series. The transfer entropy, however, erroneously localizes this information to its past. It could be inferred that the transfer entropy overestimates information flow.

Example 1.2.3. The next process involves three systems:

$$\begin{aligned} X, Y, Z &\in \{0, 1\}, \\ X_t &= \text{rand}([0, 1]) Y_t = \text{rand}([0, 1]), \\ Z_{t+1} &= X_t \oplus Y_t. \end{aligned}$$

If we calculate the transfer entropy from each system to Z they result $T_{X \rightarrow Z} = T_{Y \rightarrow Z} = 0$, underestimating the influence. Then we could consider to expand the transfer entropy to include a third system by defining the *causal entropy* [19]:

$$C_{X \rightarrow Z | Z, Y} = I(Z_{t+1} : X_t^\infty | Y_t^\infty Z_t^\infty).$$

We obtain:

$$C_{X \rightarrow Z | Z, Y} = C_{Y \rightarrow Z | Z, X} = 1$$

Again, as in the previous case overestimating the flux of information to the target variable. The paper aims at evidencing how transfer entropy and transfer entropy like measures, although being good estimator for the the reduction in uncertainty about one time series conditioned on another, can fails to pinpoint the source of flow and can lead to some misinterpretation of the physics of the process.

Stochastic and deterministic processes

As depicted in theorem 1.2.2, the stochasticity of our system influences transfer entropy. It can affect the value of the transfer entropy and its dependence over the k parameter: the flow of information deriving from its past. To better understand the behaviour of the transfer entropy we code a program that: simulate the evolution of two time series given a certain dynamic and then calculates the transfer entropy over many realization. We report a significant example of the obtained results.

Example 1.2.4. We consider $X, Y \in \{-1, 1\}$ be two systems that evolve according to the following dynamic:

$$X_t = \begin{cases} \text{rand}(\{-1, 1\}) & p \\ Y_{t-1} & 1 - p \end{cases}$$

Meaning that with probability p the value of X is random, and with probability $1 - p$, it is a copy of previous Y . So p is a parameter that represents the randomness of the process. We will study how the behaviour of the transfer entropy changes when the evolution of Y is completely random or it is completely deterministic.

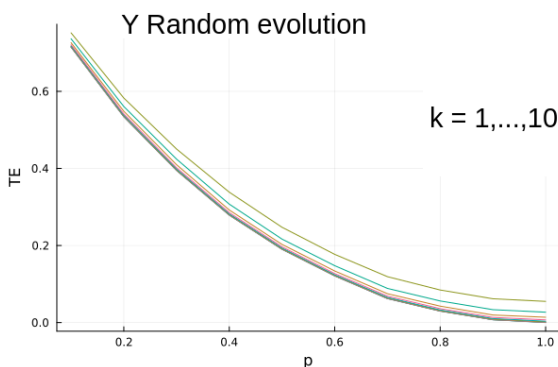


Figure 1.5: Here Y evolution is random: it assumes a random value at each time-step

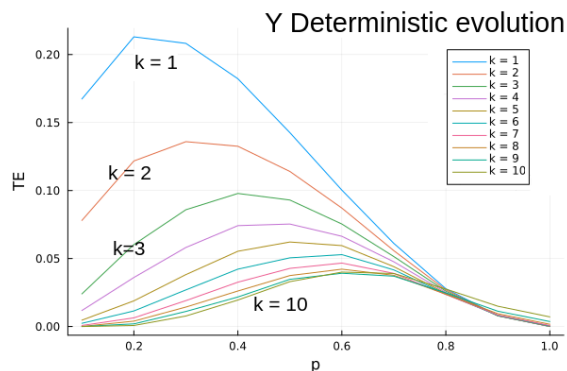


Figure 1.6: Here Y evolution is deterministic: it multiplies by -1 the previous value at each time-step

The graphs are representing the value of the transfer entropy plotted for value of $p \in [0, 1]$ and for increasing k . Considering Fig. 1.5, where Y value is extracted randomly from the set $\{-1, 1\}$, we observe that the transfer entropy is effectively independent of the k parameter, (The slight differences are due to the limited data-set of samples, which create some fluctuation especially for large number of k), and has the expected behaviour in function of p , when $p = 0$ we obtain the greatest value of the transfer entropy that decreases as p increases until it reaches zero when $p = 1$: X evolution is completely randomized. Considering Fig. 1.6, where the evolution of Y is deterministic: we start with a random Y value and then we multiply it by -1 each time-step. As we increase k as expected from our discussions, the value of the transfer entropy decreases, shifting his maximum value towards increasing p . What can be interesting to notice is that increasing k we converge towards a specific curve, with peak around $p \approx 0.6$. The meaning could be that even when the Y dynamic is completely deterministic, if the

process of X is random enough to lose some memory and not be able to recover the Y series even if we have the knowledge of the complete X 's past, but not completely random so that we still have some influence from Y , then there are certain information about the target that can only be predicted with the Y observable.

Chapter 2

Transfer entropy on quantum circuits

2.1 Introduction

In this section we will refresh some concepts of quantum mechanics that are used through this thesis and introduce quantum information theory and quantum computing. In the following we assume basic knowledge of quantum mechanics and linear algebra. Most of the concepts presented can be found in any introductory book on quantum information. For reference, we mainly used the book "Quantum Computation and Quantum Information" by Nielsen and Chuang [20].

2.1.1 Quantum mechanical concepts

Density operator

The state of a system can be represented by a density operator, usually indicated with ρ . We can distinguish two types of density operators, pure and mixed. A pure density operator is an operator that can be expressed as:

$$\rho = |\Psi\rangle\langle\Psi|,$$

where $|\Psi\rangle$ is the state of the system.

A mixed density operator is defined as a linear combination of pure density operators as:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|,$$

where p_i is the probability to have our system in the state $|\psi_i\rangle$.

In the following we list two important properties that characterize the density operator:

- $\text{tr}(\rho) = 1$, the density operator has trace equal to one.
- ρ is a positive operator.

Projective measurements

In quantum mechanics, an observable can be described as an operator, and if we measure it, the state of our system collapses to one eigenvector of the operator. These are called projective measurements. After measuring a state, it is projected onto the eigenvector of the observable operator associated to the eigenvalue corresponding to the measurement result. To describe the process we will use projection operators, which have the following property: $\Pi_i \Pi_i = \Pi_i$.

For each measurement, the probability distribution over the possible outcomes can be computed from the density operator using the following:

Definition 2.1.1 (Born's rule).

$$p(x_i) = \text{tr}(\Pi_i \rho), \quad (2.1)$$

with x_i being the outcome of observable X and Π_i the projector operator.

After the measurement the state changes:

$$\rho \rightarrow \frac{\Pi_i \rho \Pi_i}{\text{tr}(\Pi_i \rho)}.$$

Positive-operator-valued measures (POVMs) are a generalization of projective measurements. POVM is a measure whose values are positive semi-definite operators. It can be described by a set of positive semi-definite Hermitian operators that sum up to the identity:

$$\{A_1, \dots, A_n\}, \quad \sum_{i=1}^n A_i = \mathbb{I}, \quad A_i \geq 0,$$

where the A_i are associated to outcomes i of the measurement. Following definition 2.1.1:

$$p(i) = \text{tr}(A_i \rho),$$

where $p(i)$ is the probability that outcome i is measured.

Pauli matrices

Pauli matrices are a particular set of matrices, usually associated with the spin of a spin-1/2-system. Given an orthonormal basis $\{x, y, z\}$, the spin can be described using Pauli matrices as:

$$S_i = \frac{\hbar}{2} \sigma_i,$$

for $i \in \{x, y, z\}$.

Pauli matrices inherit their properties from commutation and anti-commutation properties of the spin operators:

1. $[\sigma_i, \sigma_j] = \sum_k 2i \epsilon_{ijk} \sigma_k$.
2. $\{\sigma_i, \sigma_j\} = 2\delta_{ij} \mathbb{I}$.
3. $\text{tr}(\sigma_i) = 0, \quad \forall i \in \{x, y, z\}$.

Pauli matrices are explicated in the basis of eigenvectors of σ_z :

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

All Pauli matrices have two eigenvalues -1 and 1 . For a generic σ_i :

$$\begin{aligned} \sigma_i |\uparrow\rangle_i &= +1 |\uparrow\rangle_i, \\ \sigma_i |\downarrow\rangle_i &= -1 |\downarrow\rangle_i. \end{aligned}$$

Usually we use the eigenvectors of σ_z as basis elements, defining the computational basis:

Definition 2.1.2. *The computational basis is defined from the eigenvectors of σ_z :*

$$\begin{aligned} |0\rangle &:= |\uparrow\rangle_z, \\ |1\rangle &:= |\downarrow\rangle_z. \end{aligned}$$

The expressions of the eigenvectors of the Pauli matrices are:

$$\begin{aligned} |\uparrow\rangle_z &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, & |\downarrow\rangle_z &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ |\uparrow\rangle_x &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, & |\downarrow\rangle_x &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \\ |\uparrow\rangle_y &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, & |\downarrow\rangle_y &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \end{aligned}$$

And written in the computational basis:

$$\begin{aligned} |\uparrow\rangle_z &= |0\rangle, & |\downarrow\rangle_z &= |1\rangle, \\ |\uparrow\rangle_x &= |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), & |\downarrow\rangle_x &= |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \\ |\uparrow\rangle_y &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), & |\downarrow\rangle_y &= \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle). \end{aligned}$$

We note that each eigenvector for a specific σ_i can be written as a superposition of the eigenvectors of a different σ_j , in this case σ_z .

Measuring the spin in the (\hat{n}) direction

Spin measurements are represented by the spin operator, which can be described with the Pauli matrices. We will focus on the latter to explain briefly what it means to measure the spin in an arbitrary direction.

Definition 2.1.3. *Given a vector of Pauli matrices $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\hat{n} = (\cos(\theta)\cos(\phi), \cos(\theta)\sin(\phi), \sin(\theta))$ the usual versor in spherical coordinates, a general measurement of spin in the \hat{n} direction is defined as:*

$$M(\theta, \phi) := \vec{\sigma} \cdot \hat{n}.$$

Here we only point out that the eigenvalues of the measures are 1 and -1 for every direction, although the eigenvectors change.

2.1.2 Composite systems

Tensor product

The tensor product is crucial to construct composite systems in quantum mechanics. Suppose we have two Hilbert spaces V, W of dimension m, n respectively. Then their tensor product, which is represented by the operator \otimes , is written as $V \otimes W$ and it is a mn dimensional space. Given $|i\rangle$ and $|j\rangle$ basis for V and W respectively, then the tensor product $|i\rangle \otimes |j\rangle$ is a basis for $V \otimes W$. Usually to represent the tensor product between two vectors $|v\rangle$ and $|w\rangle$ is used the abbreviated notation $|v\rangle \otimes |w\rangle = |vw\rangle$.

Entanglement

A system is defined as *entangled* when the state of the system cannot be factorized. Given two systems A and B , defined on the Hilbert space $H = H_A \otimes H_B$, the most general form for a pure state is:

$$|\phi\rangle = \sum_{ij} c_{ij} |i_A\rangle \otimes |j_B\rangle,$$

where $|i_A\rangle$ and $|j_B\rangle$ are bases for the respective spaces H_A and H_B . A state is said to be separable if there exist complex coefficients a_i and b_j such that:

$$|\phi\rangle = \left(\sum_i a_i |i_A\rangle \right) \otimes \left(\sum_j b_j |j_B\rangle \right) = |\phi_A\rangle \otimes |\phi_B\rangle.$$

When this is not possible we say that the systems are in an entangled state.

The most renown entangled states are the Bell states:

$$\begin{aligned} |\Phi_+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ |\Phi_-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\Psi_+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\Psi_-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned}$$

Schmidt decomposition

In linear algebra, the Schmidt decomposition refers to a way of expressing vectors defined in a space formed as product of two spaces, as the tensor product of vectors belonging to the latter.

Theorem 2.1.1. *Suppose $|\phi\rangle$ is a pure state of a composite system AB . There exists a set of orthonormal vectors $|i_A\rangle \in A$ and $|i_B\rangle \in B$ such that:*

$$|\phi\rangle = \sum_i \lambda_i |i_A\rangle \otimes |i_B\rangle \tag{2.2}$$

with λ_i are real non-negative coefficients and fulfill the normalization $\sum \lambda_i^2 = 1$.

Proof. It is a proof by construction, we consider the state $|\phi\rangle$ written in the basis $|i\rangle \otimes |j\rangle$.

$$|\phi\rangle = \sum_{i,j} c_{ij} |i\rangle |j\rangle,$$

where c_{ij} are the elements of a matrix C . We can rewrite the matrix C using the singular value decomposition (SVD): a linear algebra theorem states that any complex matrix C admits a decomposition $U\Lambda V^\dagger$ such that Λ is a diagonal matrix with non-negative real numbers on the diagonal. Then the elements c_{ij} can be written as $c_{ij} = \sum_{k,l} u_{ik} \lambda_{kl} v_{lj}$ where, since Λ is a diagonal matrix, $\lambda_{kl} = \delta_{kl} \lambda_k$, and the equation becomes:

$$\begin{aligned} |\phi\rangle &= \sum_{i,j} \sum_{k,l} u_{ik} \delta_{kl} \lambda_k v_{lj} |i\rangle |j\rangle = \\ &= \sum_{i,j,k} u_{ik} \lambda_k v_{kj} |i\rangle |j\rangle. \end{aligned}$$

Redefining the terms in the following way:

$$|i_A\rangle = \sum_i u_{ik} |i\rangle \quad |i_B\rangle = \sum_j v_{kj} |j\rangle, \quad \lambda_k = \lambda_i.$$

we re-index the sum:

$$|\phi\rangle = \sum_i \lambda_i |i_A\rangle \otimes |i_B\rangle.$$

If we consider a two qubit space and $|\phi\rangle$ as a state, then we also have the normalization condition $\langle\phi|\phi\rangle = 1$ that reduces the description of the state to one parameter after the vectors $|i_A\rangle |i_B\rangle$ have been determined. \square

Channels

The qubit can evolve due to the action of a unitary operator, which preserves the norm of the state, called *channel*. When a channel acts on a single qubit it is said to be local. We will represent some of the most often used local channels, in the computational basis:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

where X , Y , Z are the Pauli matrices and H is the Hadamard gate. X is often called the *NOT* gate, since it has the effect of exchanging the computational basis vectors.

$$X |0\rangle = |1\rangle, \quad X |1\rangle = |0\rangle.$$

The effect of the Hadamard gate on the computational basis is:

$$H |0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} = |+\rangle, \quad H |1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |-\rangle.$$

Performing a basis change from the Z basis to X basis.

When the system is composite, channels can act on multiple qubits, creating an interaction between them.

In this thesis we will focus on two qubits systems that interact and evolve through a common channel. We list two channels that act on two qubit systems:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad SWAP = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The C-NOT gate, or the control-not gate, is a main example of conditional gate, where the state of a qubit, called control qubit, can vary the operation performed on a second qubit, called target qubit.

The swap gate exchanges the state of the two qubits.

Kraus-Cirac decomposition of non-local unitary operators

We will briefly explain the decomposition of an arbitrary unitary operator U_{AB} acting on two qubits, introduced by Kraus and Cirac in the article [21]. The proof was obtained by construction and it states the following:

Theorem 2.1.2. *For any unitary operator U_{AB} acting on two qubits, there exist local operators U_A, U_B, V_A, V_B , that only act on a qubit such that:*

$$U_{AB} = U_A \otimes U_B U_d V_A \otimes V_B \tag{2.3}$$

where $U_d = e^{-iD}$ with $D = \alpha_x \sigma_x \otimes \sigma_x + \alpha_y \sigma_y \otimes \sigma_y + \alpha_z \sigma_z \otimes \sigma_z$.

Since local unitary operators can be considered as a simple rotation of the state, U_A, U_B, V_A, V_B merely have the effect to change the direction of the spin, and if the spin is measured before and after the evolution, the effect of this operator can be incorporated in the measurements, changing their direction. If we perform a maximization over all measurement angles, then the effect of the local unitary operators vanishes, allowing us to represent any non local two qubit channel with only three parameters.

No-communication theorem

Theorem 2.1.3. *Given a bipartite system AB , there is no local operation on one subsystem that can affect the measurements on the other one.*

Proof. Consider a system that lives on the Hilbert space $H = H_A \otimes H_B$. The state of our system can be described by a density operator of the form:

$$\rho = \sum_i X_i \otimes Y_i,$$

where X_i and Y_i are local operators. The theorem still holds without assuming ρ to be a density matrix, i.e. operators X_i and Y_i could be different from state projection operators.

Suppose that Alice performs a local measurement on qubit A , which can be described as a map defined by Kraus operators K_j such that:

$$\begin{aligned}\Phi(\rho) &= \sum_j (K_j \otimes \mathbb{I}_B) \rho (K_j^\dagger \otimes \mathbb{I}_B), \\ \sum_j K_j K_j^\dagger &= \mathbb{I},\end{aligned}$$

where $\Phi(\rho)$ is the state after Alice's operations. The form of the Kraus operators ($K_j \otimes \mathbb{I}_B$) indicates that Alice only performs local operations: Alice's measurement apparatus does not interact with Bob's system.

The state that Bob observes is the partial trace over Hilbert space H_A of $\Phi(\rho)$. This state can be directly calculated:

$$\begin{aligned}tr_{H_A}(\Phi(\rho)) &= tr_{H_A}(\sum_j (K_j \otimes \mathbb{I}_B) \rho (K_j^\dagger \otimes \mathbb{I}_B)) \\ &= tr_{H_A}(\sum_j (K_j \otimes \mathbb{I}_B) \sum_i X_i \otimes Y_i (K_j^\dagger \otimes \mathbb{I}_B)) \\ &= tr_{H_A}(\sum_j \sum_i (K_j \otimes \mathbb{I}_B) X_i \otimes Y_i (K_j^\dagger \otimes \mathbb{I}_B)) \\ &= tr_{H_A}(\sum_j \sum_i K_j X_i K_j^\dagger \otimes Y_i) \\ &= tr_{H_A}(\sum_i \sum_j K_j K_j^\dagger X_i \otimes Y_i) \\ &= tr_{H_A}(\sum_i X_i \otimes Y_i) \\ &= tr_{H_A}(\rho).\end{aligned}$$

Meaning that Bob can not statistically tell what Alice did. Bob's statistic of the system is unchanged by local operations on Alice system. \square

2.1.3 Quantum information

Quantum computation

The fundamental component of quantum computation is the qubit, an object that lives in a two dimensional Hilbert space. We will use the computational basis as a basis to describe its space. The qubit is described by a state:

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

Imposing the normalization condition $|\alpha|^2 + |\beta|^2 = 1$, it can be expressed as:

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle.$$

The parameters θ and ϕ define a point on the surface of the unit three-dimensional sphere, called the *Bloch sphere*, represented in Fig. 2.1.

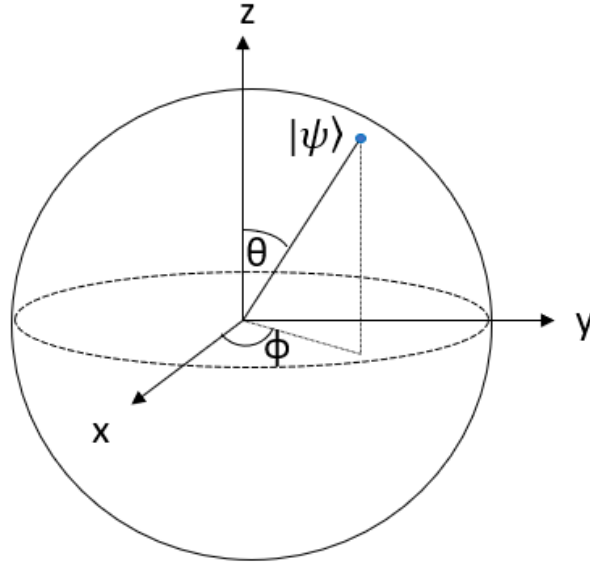


Figure 2.1: Bloch sphere.

Adding a qubit to our system increases the dimension of the Hilbert space by a factor two. A multi qubit system is defined on the Hilbert space resulting from the tensor product of the Hilbert spaces of its qubits. If we want to consider a two qubit system our basis will be composed by four vectors, which are, using the notation $|a\rangle \otimes |b\rangle = |ab\rangle$:

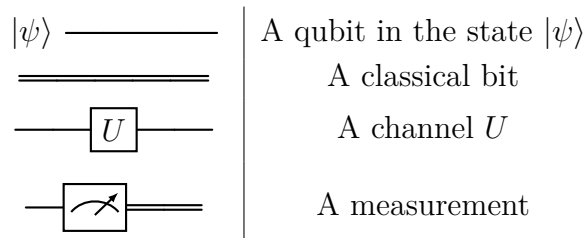
$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}.$$

And the state that describes our system:

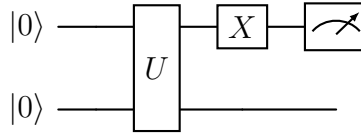
$$|\psi\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle.$$

An important operation that can be done onto our system is measuring, which, as known from quantum mechanics, collapses the state of the qubit into the measured state.

Qubits, channels and measurements are the principal components to construct a quantum circuit. Quantum circuits are used to represent quantum processes. The notation to draw a quantum circuit is:



As an example let us draw a two qubit system where both qubits evolve according to U , then the first qubit' spin is flipped and measured:



Quantum entropies

We start by introducing the Von Neumann entropy. In classical information theory, Shannon entropy measures the uncertainty associated with a probability distribution. Density operators replace probability distributions in quantum mechanics.

Definition 2.1.4 (Von Neumann entropy). *The Von Neumann entropy of a quantum state ρ is:*

$$S(\rho) := -\text{tr}(\rho \log \rho). \quad (2.4)$$

Since the trace is invariant under unitary transformations, diagonalizing the density matrix, the previous definition 2.1.4, can also be expressed, being λ_i the eigenvalues of the density matrix, as:

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i,$$

where the logarithm is taken to base two.

The basic properties of the Von Neumann entropy are very similar to the classical Shannon entropy's properties. Here we list a few of them:

- The entropy is non-negative and it is 0 if and only if the state is pure.
- In a d -dimensional Hilbert space the entropy is maximized by $\log d$ and its maximal when the density matrix is the one of the completely mixed state: \mathbb{I}/d .
- If a composite system AB is in a pure state, then $S(A) = S(B)$.

As in classical physics, it is useful to define the quantum version of relative entropy

Definition 2.1.5 (Quantum relative entropy).

$$S(\rho||\sigma) = \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma). \quad (2.5)$$

As in the classical case, the quantum relative entropy has the same property of non-negativity, as shown by the Klein's inequality.

Theorem 2.1.4 (Klein's inequality). *The quantum relative entropy is non-negative*

$$S(\rho||\sigma) \geq 0,$$

with equality if and only if $\rho = \sigma$.

Proof. Consider $\rho = \sum p_l |l\rangle \langle l|$, $\sigma = \sum q_j |j\rangle \langle j|$ be orthonormal decomposition for ρ and σ , applying the definition of relative entropy we obtain:

$$S(\rho||\sigma) = \sum_i p_i \log p_i - \sum_i \langle i| \rho \log \sigma |i\rangle.$$

Knowing that:

$$\begin{aligned} \langle i | \rho = \sum_l p_l \langle i | l \rangle \langle l | &= \sum_l p_l \delta_{il} \langle l | = p_i \langle i |, \\ \langle i | \log \sigma | i \rangle &= \sum_j P_{ij} \log q_j \quad \text{where } P_{ij} = \langle i | j \rangle \langle j | i \rangle, \end{aligned}$$

we substitute in the relative entropy equation,

$$S(\rho || \sigma) = \sum_i p_i \log p_i - \sum_{i,j} p_i P_{ij} \log q_j,$$

where we note that $\sum_i P_{ij} = \sum_j P_{ij} = 1$ and $P_{ij} = \delta_{ij}$,

$$S(\rho || \sigma) = \sum_i p_i \left(\log p_i - \sum_j P_{ij} \log q_j \right).$$

Since \log is a strictly concave function, calling $\sum_j P_{ij} \log q_j = s_i$, we have $\sum_j P_{ij} \log q_j \leq \log s_i$ with equality if and only if $\exists j$ s.t. $P_{ij} = 1$. Thus

$$S(\rho || \sigma) \geq \sum_i p_i \log \frac{p_i}{s_i},$$

with equality if and only if $\forall i \exists j$ s.t. $P_{ij} = 1$, meaning that P_{ij} is a permutation matrix; in that case, ρ and σ have the same eigenvectors, just ordered differently.

The previous equation has the same form as the classical relative entropy, for which it is proved that it is always greater than zero, and zero if and only if the two distributions are equal, $p_i = s_i$, meaning that ρ and σ have the same eigenvalues. Remembering that for equality P_{ij} are permutations, then if they have same eigenvalues and eigenvectors the condition implies that the matrices are equal: $\rho = \sigma$. \square

2.2 Transfer entropy on quantum processes

In this section we apply the transfer entropy to data derived from measurements of quantum systems. The quantum circuit that describes our process is:



Where A and B are qubits and U is a channel. B will act as a target and A as a source. The observables that we decide to measure are spins represented by the Pauli matrices. We will use the notation on Tab. 2.1:

We imagine to receive a set of data from measurements, and from them we inquire what are the transfer entropy's properties when the process is quantum. First we will discuss how the transfer entropy varies for different configurations of the circuit 2.6 and later we will retrieve an analytical formula to describe the simulations.

$ 1\rangle_z$	The eigenvector, associated to eigenvalue 1 of the operator σ_z
$\Pi_{1 z}^A = 1\rangle\langle 1 _z \otimes \mathbb{I}_2$	The projector onto $ 1\rangle_z$ that operates in the Hilbert space of qubit A
$\rho_{A1 z} = \frac{\Pi_{1 z}^A \rho \Pi_{1 z}^A}{\text{tr}(\Pi_{1 z}^A \rho)}$	The density matrix after the measurement of observable σ_z of qubit A has result 1

Table 2.1: Notation used on section 2.2.

2.2.1 Simulations

For a better understanding of the results, we will briefly explain how the code [12] that computes the transfer entropy works. Let us imagine that we have a lot of realizations of two time series, A and B . We aim to study the transfer entropy considering system A as a source and B as target, both binary variables (they can only take two values), with $k = l = 1$ and $t = 2$, meaning that we have two possible outcomes from each measurement and that we only consider two instants of time, namely, before and after the application of the channel. The first thing that is going to be calculated, and the only probability that we extract directly from the data, is:

$$p(b_2, a_1, b_1),$$

where b_2, b_1 are the measures of observable B and a_1 the measure of observable A . In order to calculate the probability we use a simple frequentist approach, where for each possible combination of the measurements' outcomes we calculate the probability that it occurs, counting the number of realization, in the following way:

$$p(b_2 = b_2^*, a_1 = a_1^*, b_1 = b_1^*) = \sum_{j=1}^n \delta(b_2^j, b_2^*) \delta(a_1^j, a_1^*) \delta(b_1^j, b_1^*),$$

where δ is the Kronecker function, b_2^j is the j_{th} observation of variable b_2 , n is the total number of different time series and the starred variables are fixed values.

Once the full probability distribution is calculated for each possible combination of outcomes, the rest of required probability distributions can be obtained applying the marginal rule and the Bayes formula in the following way:

$$p(b_2|a_1, b_1) = \frac{p(b_2, a_1, b_1)}{\sum_{b_2} p(b_2, a_1, b_1)} = \frac{p(b_2, a_1, b_1)}{p(a_1, b_1)},$$

$$p(b_2|b_1) = \frac{\sum_{a_1} p(b_2, a_1, b_1)}{\sum_{b_2, a_1} p(b_2, a_1, b_1)} = \frac{p(b_2, b_1)}{p(b_1)}.$$

We simulate the measurement given by the quantum circuit in 2.6. We have the freedom to measure the qubit' spin in a generic direction in the unitary sphere. We do this to emphasize the different behaviours of the transfer entropy over different initial states and how it varies as our measurements rotate.

From the theory we know that after a measurement the state of the qubits collapses into an eigenvector of the measurement operator. Measurements on qubit A and B are described by

operators in different Hilbert spaces, and so they commute (it does not matter the order of the measurements). We will now show results obtained using the C-NOT channel, where we rotate the measurement of the source qubit A from σ_z to σ_x , while we measure the target qubit B on σ_z . We perform the simulation for a comprehensive (but not exhaustive) set of states, which will be presented later. The results for the value of the transfer entropy are reported for all the fifteen initial states with which the simulation is performed in Fig. 2.2.

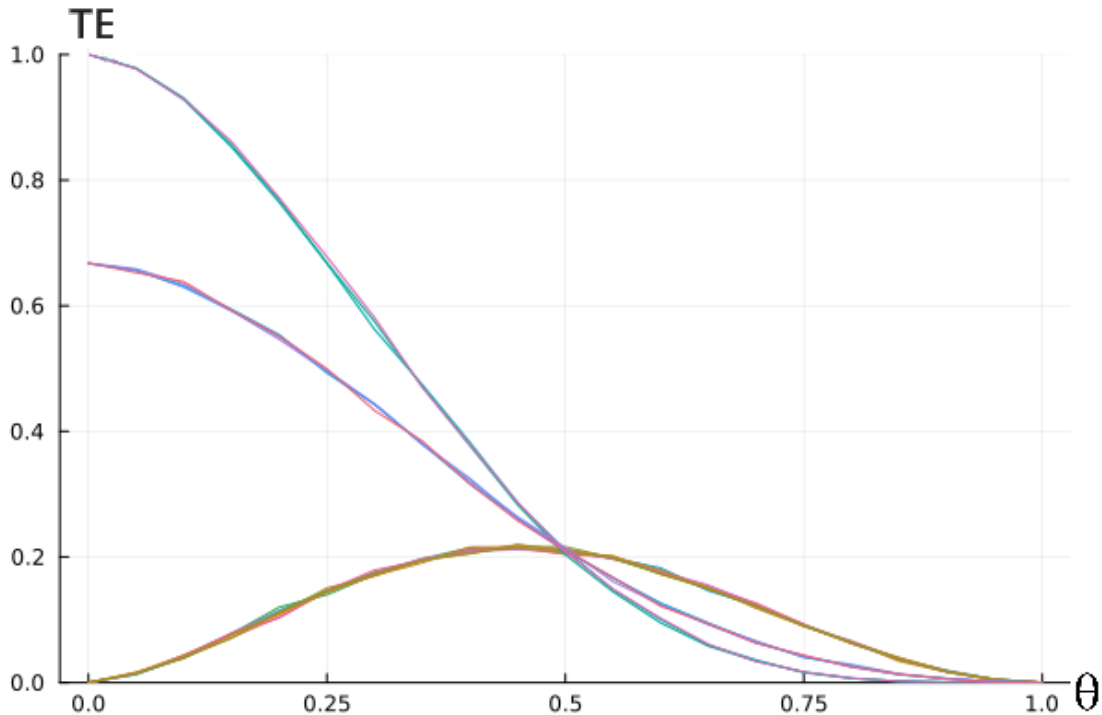


Figure 2.2: Fifteen curves are represented, on the y-axis we have the value of the transfer entropy while on the x-axis we plot the angle θ of the rotation, which defines the measured observable, from σ_z ($\theta = 0$) to σ_x ($\theta = 1$). The value on the horizontal axis must be multiplied by $\pi/2$

We can observe that, the value of the transfer entropy can be divided into three main categories, which depend on the initial state. We classify the curves for their value on the origin, when the measure is performed on σ_z , later we will also try to analytically understand the behaviour of the curve with the changing of the angle, and try to generalize our apparatus. We notice that initial values of the transfer entropy are 1, 0.66 and 0. We report in table 2.2 which states are associated with their respective curve.

We will then give an intuition to why the states reproduce these values of the transfer entropy. Starting from the case where $TE = 1$, we notice that qubit A , the source of our flow of information, is always in the state $|+\rangle$, this state, as explained in the section about quantum concepts, is an eigenstate of σ_x operator:

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

$TE = 1$	$TE = 0.66$	$TE = 0$
$ +0\rangle,$	$\frac{1}{\sqrt{3}}(00\rangle + 01\rangle + 10\rangle)$	$ 00\rangle, 01\rangle, 10\rangle, 11\rangle,$
$ +1\rangle,$	$\frac{1}{\sqrt{3}}(00\rangle + 01\rangle + 11\rangle)$	$\frac{1}{\sqrt{2}} 00\rangle + \frac{1}{\sqrt{2}} 11\rangle,$
$ ++\rangle$	$\frac{1}{\sqrt{3}}(01\rangle + 10\rangle + 11\rangle)$	$\frac{1}{\sqrt{2}} 01\rangle + \frac{1}{\sqrt{2}} 10\rangle,$
	$\frac{1}{\sqrt{3}}(00\rangle + 10\rangle + 11\rangle)$	$ 0+\rangle, 1+\rangle$

Table 2.2: Initial states that outputs a certain value of the transfer entropy

The main point is that the state is totally undetermined with respect to a measurement of σ_z : when measured, the state collapses in $|0\rangle$ or $|1\rangle$ with equal probability. Measuring the state $|+\rangle$ is the same as initializing the value a_1 randomly. As we saw in the previous chapter, the transfer entropy is influenced by the deterministic nature of the source, and it is able to capture fully the flux of information when our source is randomized. We could infer that to obtain a maximal value of the transfer entropy we need an initial state where the source qubit is in an undetermined state respect to its measure, even though this condition is not enough to ensure a maximal value of the transfer entropy as we will see for the next case.

The states where $TE = 0$ are all the ones where source qubit A is in an eigenstate of the measured observable σ_z and the maximally entangled states:

$$\begin{aligned} & \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle, \\ & \frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle. \end{aligned}$$

The reason why for entangled states we obtain zero is that, once the qubit A is measured then also the value of B is certain, so for the same entangled state, each measurement process gives us back the same coupled values of a_1 and b_1 . The transfer entropy, due to the way the probabilities are calculated, can never see the effect of the source on the target, because it always receives the same initial combinations, and attributes the change of the target variable only to its past. The same reasoning is done when the qubit A is prepared in an eigenstate of the measured observable, since the transfer entropy never sees it varying, it will attribute the change, if there are any, to the target's past.

To understand why the value of the transfer entropy is 0.66 in the states reported in the

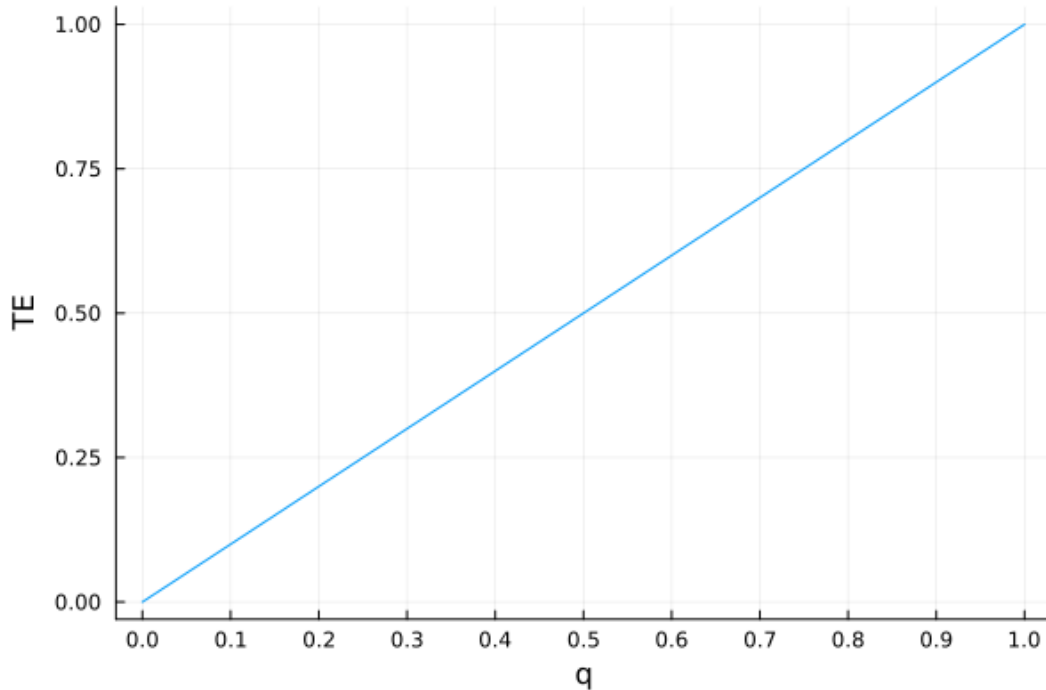


Figure 2.3: Transfer entropy for state in Eq. 2.7, varying q

Tab. 2.2 we can rewrite them in a different form.

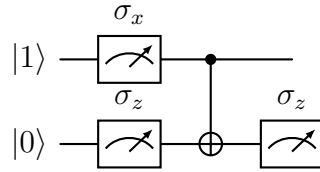
$$\begin{aligned} \frac{1}{\sqrt{3}}(|00\rangle + |01\rangle + |10\rangle) &= \sqrt{\frac{2}{3}}|+0\rangle + \frac{1}{\sqrt{3}}|01\rangle \\ \frac{1}{\sqrt{3}}(|00\rangle + |01\rangle + |11\rangle) &= \sqrt{\frac{2}{3}}|+1\rangle + \frac{1}{\sqrt{3}}|00\rangle \\ \frac{1}{\sqrt{3}}(|01\rangle + |10\rangle + |11\rangle) &= \sqrt{\frac{2}{3}}|+1\rangle + \frac{1}{\sqrt{3}}|10\rangle \\ \frac{1}{\sqrt{3}}(|00\rangle + |10\rangle + |11\rangle) &= \sqrt{\frac{2}{3}}|+0\rangle + \frac{1}{\sqrt{3}}|11\rangle \end{aligned}$$

We see that our states are in a superposition of states where with probability of $2/3$ it collapses on a state that produce a transfer entropy's value of one, and with probability $1/3$ collapses into a state with transfer entropy's value of zero. It suggests a linear dependency between the transfer entropy and the squared coefficients of the state. We observe what happens in Fig. 2.3 in a state with the following structure:

$$\sqrt{q}|+0\rangle + \sqrt{1-q}|01\rangle. \tag{2.7}$$

It is immediately clear that the transfer entropy is linear with the squared coefficients of the initial state. Going back to the Fig. 2.2, we briefly explain why the transfer entropy seems to be 0 for every initial states when σ_x is measured on the source qubit (we have the same identical results for σ_y). We can exploit what happens for an arbitrary initial state.

Example 2.2.1. Suppose to have the following circuit with initial state $|10\rangle$:



The source qubit A , after the measurement will be in an eigenstate of σ_x , in this case with $p = \frac{1}{2}$, which are $|+\rangle$, $|-\rangle$. The first measure on the target qubit B is 1 in this case, (remember how the computational basis is defined). Now if we apply the C-NOT gate we obtain:

$$\begin{aligned}
 CNOT | +0\rangle &= \frac{1}{\sqrt{2}} CNOT |00\rangle + \frac{1}{\sqrt{2}} CNOT |10\rangle = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle \\
 CNOT | -0\rangle &= \frac{1}{\sqrt{2}} CNOT |00\rangle - \frac{1}{\sqrt{2}} CNOT |10\rangle = \frac{1}{\sqrt{2}} |00\rangle - \frac{1}{\sqrt{2}} |11\rangle
 \end{aligned}$$

A measurement on σ_z of the qubit B returns as a result -1 or 1 with equal probability, independently if we measured 1 or -1 for σ_x . Our ability to predict the value of the second measurement of qubit B is completely unaltered by our knowledge of the measurement of σ_x on the source A , yielding:

$$p(b_2|a_1, b_1) = p(b_2|b_1),$$

which is the condition to have transfer entropy's value equal to zero.

Note that using σ_y as measured observable we would have obtained the same result.

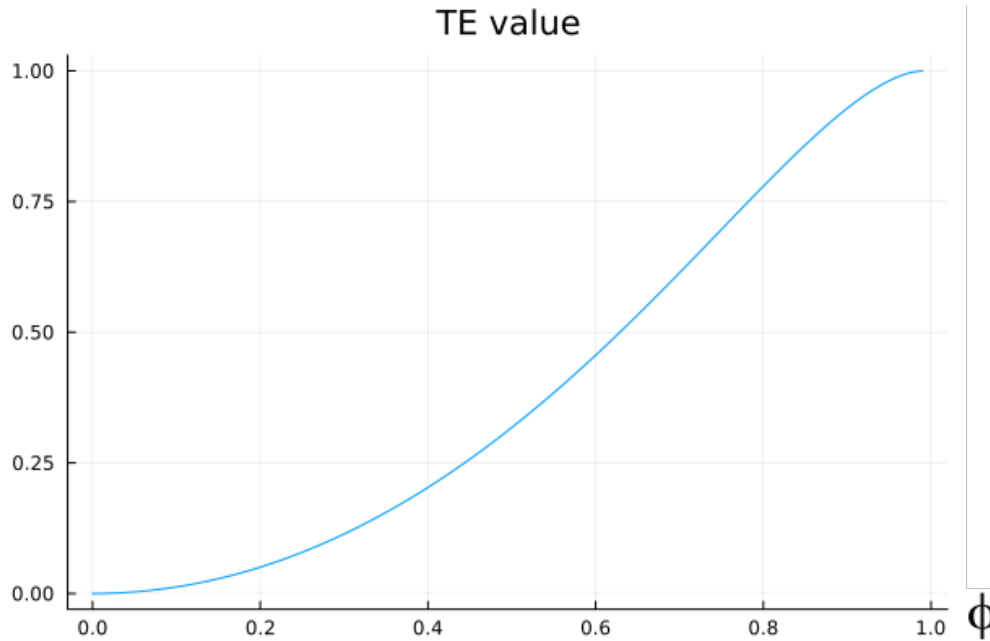


Figure 2.4: Transfer entropy's values when rotating the measurements on the target qubit B from σ_x to σ_y .

Rotating target qubit measurements

If we rotate the measurement over the target qubit B , we can observe some differences; the main one is a different value of the transfer entropy if we measure σ_x or σ_y . In Fig. 2.4 are reported the results from simulations where we measure the target qubit B from σ_x to σ_y with initial state $|+0\rangle$ and gate C-NOT. Recalling that for the same initial state, rotating the measurements on source qubit A we obtained zero value for the transfer entropy for both σ_x and σ_y measurements. To understand this behaviour we recognize the difference between applying the C-NOT gate to σ_x 's eigenvectors and σ_y 's. If the control qubit, in this case A , is in state $|0\rangle$, the C-NOT will not do anything to the states, so we assume that the state of qubit A is $|1\rangle$ and analyze the action of the C-NOT gate when qubit B is in a state corresponding to an eigenvector of σ_x . We will call the C-NOT gate C :

$$\begin{aligned}
 |1+\rangle &= |1\rangle \otimes |+\rangle = \frac{1}{\sqrt{2}} |10\rangle + \frac{1}{\sqrt{2}} |11\rangle, \\
 |1-\rangle &= |1\rangle \otimes |-\rangle = \frac{1}{\sqrt{2}} |10\rangle - \frac{1}{\sqrt{2}} |11\rangle, \\
 C|1+\rangle &= \frac{1}{\sqrt{2}} C|10\rangle + \frac{1}{\sqrt{2}} C|11\rangle = \frac{1}{\sqrt{2}} |11\rangle + \frac{1}{\sqrt{2}} |10\rangle = |1+\rangle, \\
 C|1-\rangle &= \frac{1}{\sqrt{2}} C|10\rangle - \frac{1}{\sqrt{2}} C|11\rangle = \frac{1}{\sqrt{2}} |11\rangle - \frac{1}{\sqrt{2}} |10\rangle = -|1-\rangle.
 \end{aligned}$$

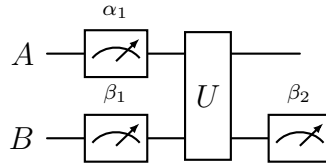
And the state remains unchanged after the action of the C-NOT, if we do not consider an overall phase to the state. For σ_y 's eigenvectors instead

$$\begin{aligned} |1\rangle \otimes |\uparrow\rangle_y &= \frac{1}{\sqrt{2}} |10\rangle + \frac{i}{\sqrt{2}} |11\rangle \\ |1\rangle \otimes |\downarrow\rangle_y &= \frac{1}{\sqrt{2}} |10\rangle - \frac{i}{\sqrt{2}} |11\rangle \\ C |1\rangle \otimes |\uparrow\rangle_y &= \frac{1}{\sqrt{2}} C |10\rangle + \frac{i}{\sqrt{2}} C |11\rangle = \frac{1}{\sqrt{2}} |11\rangle + \frac{i}{\sqrt{2}} |10\rangle = \frac{i}{\sqrt{2}} |1\rangle \otimes |\downarrow\rangle_y \\ C |1\rangle \otimes |\downarrow\rangle_y &= \frac{1}{\sqrt{2}} C |10\rangle - \frac{i}{\sqrt{2}} C |11\rangle = \frac{1}{\sqrt{2}} |11\rangle - \frac{i}{\sqrt{2}} |10\rangle = \frac{-i}{\sqrt{2}} |1\rangle \otimes |\uparrow\rangle_y \end{aligned}$$

We observe that the C-NOT flips the eigenvectors of σ_y in the same way as it does for eigenvectors σ_z , and so we expect a similar value of the transfer entropy when we measure, for source qubit A the spin σ_z and for B , σ_z and σ_y , even though for different initial states we obtain different results.

2.2.2 Analytical study

The purpose of this section is to obtain an analytic formulation of the transfer entropy calculated as described in the previous section. For this purpose we will study the following circuit:



where the measurement of spin are done in an arbitrary angle and $\alpha_1, \beta_1, \beta_2$ represent the tuple of angles necessary to define the direction of the measurements, with α representing measurements on qubit A and β on qubit B , namely:

$$\begin{aligned} \alpha_1 &= (\theta, \phi)_{a_1} & M(\alpha_1) &= \vec{\sigma} \cdot \hat{n}(\alpha_1), \\ \beta_1 &= (\theta, \phi)_{b_1} & M(\beta_1) &= \vec{\sigma} \cdot \hat{n}(\beta_1), \\ \beta_2 &= (\theta, \phi)_{b_2} & M(\beta_2) &= \vec{\sigma} \cdot \hat{n}(\beta_2). \end{aligned}$$

To rewrite the projectors in a general and more convenient way we can exploit property 2.2.1:

Property 2.2.1. *If we have an observable M in a two dimensional Hilbert space, with eigenvalues 1 and -1 , we can write the associated projectors in this form:*

$$|a\rangle \langle a| = \frac{1}{2}(\mathbb{I} + aM)$$

with $a \in \{1, -1\}$

Proof. It is know that:

$$\begin{aligned} M |1\rangle &= |1\rangle, \\ M |-1\rangle &= -|-1\rangle, \\ M &= |1\rangle\langle 1| - |-1\rangle\langle -1|. \end{aligned}$$

In a two dimensional space, the eigenvectors of an observable form a orthonormal basis, so we can write:

$$|1\rangle\langle 1| + |-1\rangle\langle -1| = \mathbb{I}.$$

From the M and \mathbb{I} equation we can find the form of the projector just by summing and subtracting:

$$\begin{aligned} \mathbb{I} + M &= 2|1\rangle\langle 1|, \\ \mathbb{I} - M &= 2|-1\rangle\langle -1|, \end{aligned}$$

which can be easily rewritten in the convenient way shown in the proposition. \square

To keep the notation light we will indicate:

$$\Pi_{a_1|\alpha_1}^A \rightarrow \Pi_{a_1}, \quad \Pi_{b_1|\beta_1}^A \rightarrow \Pi_{b_1}.$$

For the above mentioned property 2.2.1:

$$\Pi_{a_1|\alpha_1}^A \rightarrow \Pi_{a_1} = \frac{1}{2}(\mathbb{I} + a_1 M(\alpha_1)) \otimes \mathbb{I}.$$

As explained, the code to calculate the transfer entropy starts with calculating the full probability distribution $p(b_2, a_1, b_1)$. We rewrite the probability distribution in the transfer entropy's formula using Born's rule 2.1.1. The probability of the first measure is straightforward. Suppose we start in a state ρ , we apply the rule obtaining the density matrix:

$$\begin{aligned} \rho_{a_1} &= \frac{\Pi_{a_1} \rho \Pi_{a_1}}{\text{tr}(\Pi_{a_1} \rho)}, \\ \text{with } p(a_1) &= \text{tr}(\Pi_{a_1} \rho). \end{aligned}$$

We can calculate the state and the probability of the next measurement, keeping in mind that the projectors for the two different qubit commute, since they act on different Hilbert spaces. We obtain the density matrix after the second measurement:

$$\rho_{b_1} = \frac{\Pi_{b_1} \rho_{a_1} \Pi_{b_1}}{\text{tr}(\Pi_{b_1} \rho_{a_1})},$$

with probability given by:

$$\begin{aligned} \text{tr}(\Pi_{b_1} \rho_{a_1}) &= \text{tr}\left(\Pi_{b_1} \frac{\Pi_{a_1} \rho \Pi_{a_1}}{\text{tr}(\Pi_{a_1} \rho)}\right) = \frac{\text{tr}(\Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1})}{\text{tr}(\Pi_{a_1} \rho)} = \frac{p(b_1, a_1)}{p(a_1)} \\ &= p(b_1|a_1). \end{aligned}$$

Combining all the formulas together:

$$\begin{aligned}\rho_{b_1} &= \frac{\Pi_{b_1} \rho_{a_1} \Pi_{b_1}}{p(b_1|a_1)} = \frac{\Pi_{b_1} \frac{\Pi_{a_1} \rho \Pi_{a_1}}{p(a_1)} \Pi_{b_1}}{p(b_1|a_1)} = \frac{\Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1}}{p(b_1|a_1)p(a_1)} \\ &= \frac{\Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1}}{p(b_1, a_1)}.\end{aligned}$$

Afterwards the channel operator is applied to the density matrix:

$$\rho_U = U \rho_{b_1} U^\dagger.$$

The last measure will give us a result with the following probability:

$$\begin{aligned}p(b_2|b_1, a_1) &= \text{tr}(\Pi_{b_2} \rho_U) = \\ &= \frac{\text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger)}{p(b_1, a_1)}\end{aligned}$$

We notice that:

$$p(b_2, b_1, a_1) = p(b_2|b_1, a_1)p(b_1, a_1) = \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger).$$

This is the starting point for the computation of the transfer entropy. We also observe that:

$$\begin{aligned}\sum_{b_2} p(b_2, a_1, b_1) &= \sum_{b_2=\pm 1} \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_1 U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) + \text{tr}(\Pi_{-1} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_1 U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger + \Pi_{-1} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}((\Pi_1 + \Pi_{-1}) U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\mathbb{I} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \text{tr}(\Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1}) \\ &= p(b_1, a_1)\end{aligned}$$

Where we used cyclic property of the trace and known properties of projectors.

To calculate the transfer entropy we require three probability distributions:

$$p(b_2, a_1, b_1), \quad p(b_2|a_1, b_1) = \frac{p(b_2, a_1, b_1)}{p(a_1, b_1)}, \quad p(b_2|b_1) = \frac{\sum_{a_1=\pm 1} p(b_2, a_1, b_1)}{p(b_1)}.$$

We already know how to obtain $p(b_2, a_1, b_1)$ and $p(b_2|a_1, b_1)$, $p(b_1)$ is easily calculated from the initial state and the Born rule, while $p(b_2, b_1) = \sum_{a_1=\pm 1} p(b_2, a_1, b_1)$ has to be calculated.

In order to obtain this probability we expand the sum:

$$\begin{aligned}\sum_{a_1=\pm 1} p(b_2, a_1, b_1) &= \\ &= \sum_{a_1=\pm 1} \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_1 \rho \Pi_1 \Pi_{b_1} U^\dagger) + \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{-1} \rho \Pi_{-1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_1 \rho \Pi_1 \Pi_{b_1} U^\dagger + \Pi_{b_2} U \Pi_{b_1} \Pi_{-1} \rho \Pi_{-1} \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_{b_2} U \Pi_{b_1} [\Pi_1 \rho \Pi_1 + \Pi_{-1} \rho \Pi_{-1}] \Pi_{b_1} U^\dagger).\end{aligned}$$

We analyze the central term and apply property 2.2.1:

$$\begin{aligned}\Pi_1 &= \frac{1}{2}(\mathbb{I} + M(\alpha_1)) \otimes \mathbb{I} \\ \Pi_{-1} &= \frac{1}{2}(\mathbb{I} - M(\alpha_1)) \otimes \mathbb{I}\end{aligned}$$

studying the two terms independently:

$$\begin{aligned}\Pi_1 \rho \Pi_1 &= \\ &= \frac{1}{2}(\mathbb{I} + M(\alpha_1)) \otimes \mathbb{I} \rho \frac{1}{2}(\mathbb{I} + M(\alpha_1)) \otimes \mathbb{I} = \\ &= \frac{1}{4}(\mathbb{I} + M(\alpha_1) \otimes \mathbb{I}) \rho (\mathbb{I} + M(\alpha_1) \otimes \mathbb{I}) = \\ &= \frac{1}{4}(\rho + M(\alpha_1) \otimes \mathbb{I} \rho + \rho M(\alpha_1) \otimes \mathbb{I} + M(\alpha_1) \otimes \mathbb{I} \rho M(\alpha_1) \otimes \mathbb{I}), \\ \Pi_{-1} \rho \Pi_{-1} &= \\ &= \frac{1}{2}(\mathbb{I} - M(\alpha_1)) \otimes \mathbb{I} \rho \frac{1}{2}(\mathbb{I} - M(\alpha_1)) \otimes \mathbb{I} = \\ &= \frac{1}{4}(\mathbb{I} - M(\alpha_1) \otimes \mathbb{I}) \rho (\mathbb{I} - M(\alpha_1) \otimes \mathbb{I}) = \\ &= \frac{1}{4}(\rho - M(\alpha_1) \otimes \mathbb{I} \rho - \rho M(\alpha_1) \otimes \mathbb{I} + M(\alpha_1) \otimes \mathbb{I} \rho M(\alpha_1) \otimes \mathbb{I}).\end{aligned}$$

The sum explicated will read:

$$\Pi_1 \rho \Pi_1 + \Pi_{-1} \rho \Pi_{-1} = \frac{1}{2} \rho + \frac{1}{2} (M(\alpha_1) \otimes \mathbb{I}) \rho (M(\alpha_1) \otimes \mathbb{I}).$$

Introducing this result in the equation of the probability, we obtain:

$$\begin{aligned}\sum_{a_1=\pm 1} p(b_2, a_1, b_1) &= \text{tr}(\Pi_{b_2} U \Pi_{b_1} [\Pi_1 \rho \Pi_1 + \Pi_{-1} \rho \Pi_{-1}] \Pi_{b_1} U^\dagger) = \\ &= \text{tr}(\Pi_{b_2} U \Pi_{b_1} [\frac{1}{2} \rho + \frac{1}{2} (M(\alpha_1) \otimes \mathbb{I}) \rho (M(\alpha_1) \otimes \mathbb{I})] \Pi_{b_1} U^\dagger) = \\ &= \frac{1}{2} \text{tr}(\Pi_{b_2} U \Pi_{b_1} \rho \Pi_{b_1} U^\dagger) + \frac{1}{2} \text{tr}(\Pi_{b_2} U \Pi_{b_1} (M(\alpha_1) \otimes \mathbb{I}) \rho (M(\alpha_1) \otimes \mathbb{I}) \Pi_{b_1} U^\dagger),\end{aligned}$$

omitting $\otimes \mathbb{I}$:

$$p(b_2, b_1) = \frac{1}{2} \text{tr}(\Pi_{b_2} U \Pi_{b_1} \rho \Pi_{b_1} U^\dagger) + \frac{1}{2} \text{tr}(\Pi_{b_2} U \Pi_{b_1} M(\alpha_1) \rho M(\alpha_1) \Pi_{b_1} U^\dagger)$$

We present all the probability distributions' formulas together with the transfer entropy equation, for the sake of clarity:

$$TE = \sum_{b_2, a_1, b_1}^{\pm 1} p(b_2, a_1, b_1) \log \frac{p(b_2|a_1, b_1)}{p(b_2|b_1)} \quad (2.8)$$

$$p(b_2, a_1, b_1) = \text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger) \quad (2.9)$$

$$p(b_2|a_1, b_1) = \frac{\text{tr}(\Pi_{b_2} U \Pi_{b_1} \Pi_{a_1} \rho \Pi_{a_1} \Pi_{b_1} U^\dagger)}{\text{tr}(\Pi_{a_1} \Pi_{b_1} \rho)} \quad (2.10)$$

$$p(b_2|b_1) = \frac{\frac{1}{2} \text{tr}(\Pi_{b_2} U \Pi_{b_1} \rho \Pi_{b_1} U^\dagger) + \text{tr}(\Pi_{b_2} U \Pi_{b_1} M(\alpha_1) \rho M(\alpha_1) \Pi_{b_1} U^\dagger)}{\text{tr}(\Pi_{b_1} \rho)} \quad (2.11)$$

The previous equation 2.8, can be generalized up to 10 parameters:

- Two angles for each of the three measurements, leading up to six parameters.
- One angle to describe our initial state, that adds one parameter, for the Schmidt decomposition.
- Three parameters that describe any two qubit interacting unitary due to the Kraus-Cirac decomposition.

We report the results obtained rotating the measurement on source qubit A and rotating both measurements for B at the same angle, with a fixed initial state, when applying the C-NOT gate. We confront the simulation's results with the ones obtained by the analytic formula. We project the value of the transfer entropy on the Bloch sphere where every point on the surface indicates a measure of the spin in that direction and the color indicates the value of the transfer entropy.

We report in Fig. 2.5 the data from the simulations as a scatter plot over the sphere, while the data from the analytic formula are represented as continuous lines. This is done to differentiate them but mainly because using the analytic formula allows us to compute the transfer entropy way faster than simulating. We obtain the same results between simulations and analytic

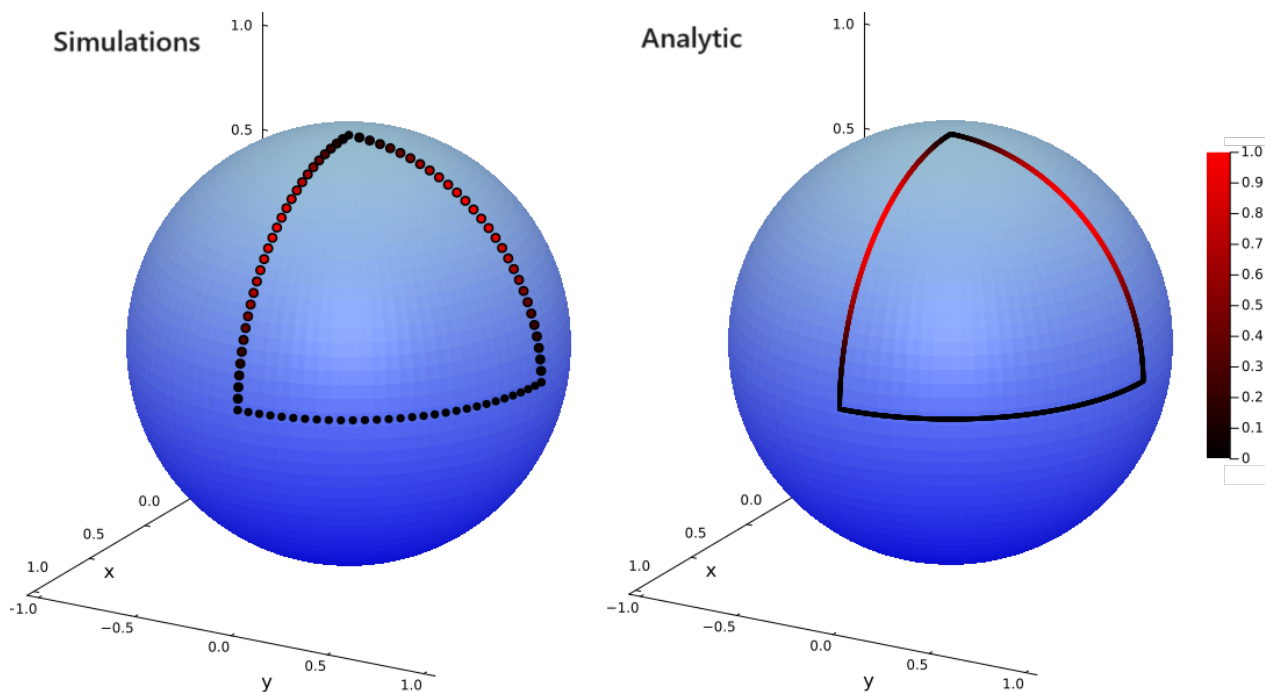


Figure 2.5: Data for the entangled initial state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, rotating measurements on source qubit A . Each point in the surface of the sphere represents a measure of spin in the direction of the versor connecting the center of the sphere with the point. The rotation happens from σ_x to σ_y , then to σ_z , then back to σ_x .

formula, for different states and rotating over all angles, proving, at least computationally the validity of the analytic formula. We can now exploit the speed of the formula 2.8 to compute values of the transfer entropy for all points in the sphere. We present an example

in Fig. 2.6 with results for the initial state $|+0\rangle$ for just the principal quadrant of the Bloch sphere, where we first rotate the measurement on A and then on B (for better data reading, the heat map is usually more clear).

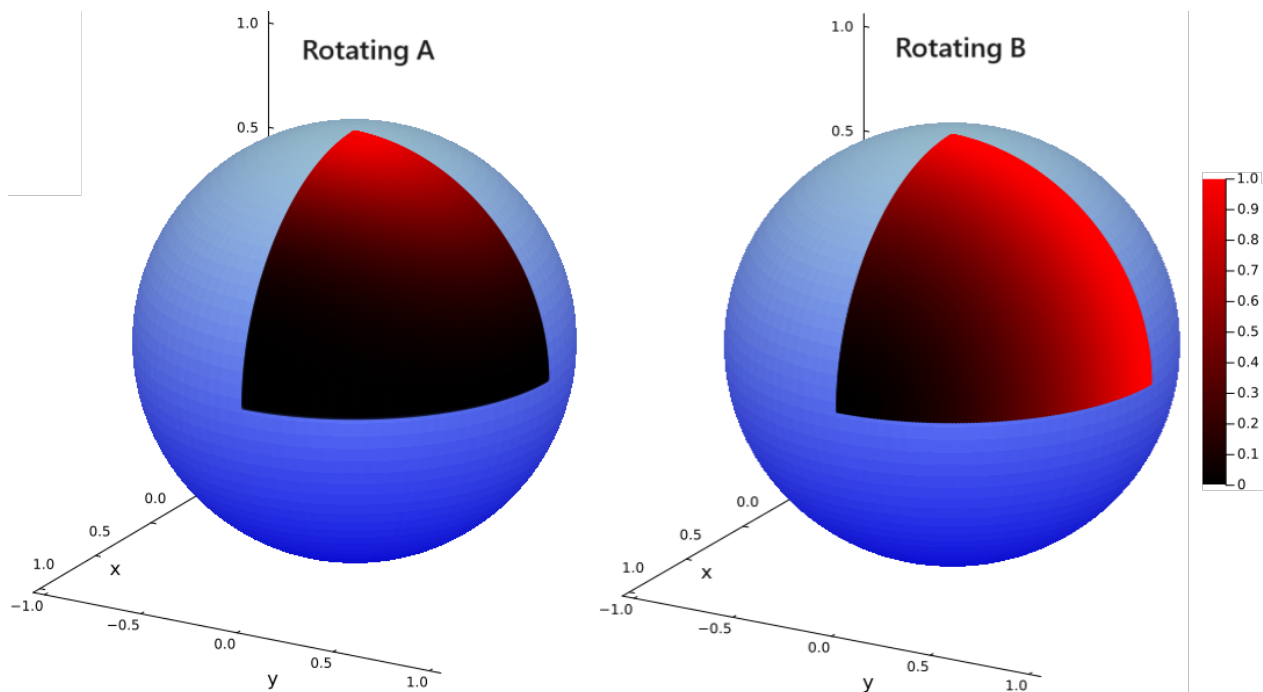


Figure 2.6: Value of the transfer entropy rotating respectively the measurements on qubit A and B for initial state $|+0\rangle$.

2.2.3 Properties

We will use a different notation, expressed in Tab. 2.3, to keep this section light and easy to read. Most of the parameters are fixed after choosing our measurement apparatus and we will not note them. We drop the indexing for first or second measure of B since it will be clear from the position of the projector in the trace.

A, B	Observable measured on qubit A and B
$ b^i\rangle$	The eigenvector, associated to observable B , with eigenvalue i
$A^i = a^i\rangle \langle a^i \otimes \mathbb{I}_2$	The projector onto $ a^i\rangle$ that operates in the Hilbert space of qubit A

Table 2.3: Notation for subsection 2.2.3

Considering the transfer entropy's analytic formulation, we can look for some property and case of interests.

Theorem 2.2.1. *When the channel acts only locally, $U = U_A \otimes U_B$ then the transfer entropy is zero.*

Proof. This is a direct consequence of the Non-communication theorem expressed in subsection 2.1.3. Since the operations performed on the initial state are local, regardless of the initial state, $p(b_2|a_1, b_1) = p(b_2|b_1)$, due to the statistical independence of system B respect to A . \square

Reviewing what has been done for the classical part, we want to see when the transfer entropy it is equal to the bound found in the classical section. what theorem 1.2.1 means for our process. Starting with the former we reformulate it for the analytical version of the transfer entropy:

Theorem 2.2.2. *The transfer entropy is equal to its bound, when:*

$$\text{tr}(B^k U A^i B^j \rho B^j A^i U^\dagger) = \text{tr}(A^i B^j \rho) \quad \forall i, j, k \quad \text{s.t.} \quad \text{tr}(B^k U A^i B^j \rho B^j A^i U^\dagger) \neq 0.$$

Remembering that in our case:

$$\begin{aligned} p(b_2, b_1, a_1) &= \text{tr}(B^{b_2} U A^{a_1} B^{b_1} \rho B^{b_1} A^{a_1} U^\dagger), \\ p(b_1, a_1) &= \text{tr}(A^{a_1} B^{b_1} \rho), \end{aligned}$$

We will consider the proof in the classical case valid, and consider what implication it has for the processes we observe. To start we want to understand when $p(b_2, b_1, a_1) = 0$. We analyze

$$\text{tr}(B^k U A^i B^j \rho B^j A^i U^\dagger) = 0,$$

where i, j, k are representing the values that our variables can assume, namely $+$ and $-$. We will consider that the same observable is measured on qubit B before and after the evolution through the channel. We represent operators in the basis $\{|a^+\rangle, |a^-\rangle\}$ and $\{|b^+\rangle, |b^-\rangle\}$, eigenvectors for observable measured on qubit A and B respectively:

$$\begin{aligned} \rho &= \sum_{m,f,r,t} c_{mf} c_{rt}^* |a^m b^f\rangle \langle a^r b^t|, \\ U &= \sum_{l,s,h,c} u_{lshc} |a^l b^s\rangle \langle a^h b^c|. \end{aligned}$$

We first consider the central term in the trace and expand it:

$$A^i B^j \rho B^j A^i = \tag{2.12}$$

$$= |a^i\rangle \langle a^i| \otimes |b^j\rangle \langle b^j| \sum_{m,f,r,t} c_{mf} c_{rt}^* |a^m b^f\rangle \langle a^r b^t| |a^i\rangle \langle a^i| \otimes |b^j\rangle \langle b^j| = \tag{2.13}$$

$$= \sum_{m,f,r,t} c_{mf} c_{rt}^* |a^i b^j\rangle \langle a^i b^j| |a^m b^f\rangle \langle a^r b^t| |a^i b^j\rangle \langle a^i b^j| = \tag{2.14}$$

$$\stackrel{a)}{=} \sum_{m,f,r,t} c_{mf} c_{rt}^* \delta_{im} \delta_{jf} \delta_{ir} \delta_{jt} |a^i b^j\rangle \langle a^i b^j| = \tag{2.15}$$

$$= |c_{ij}|^2 |a^i b^j\rangle \langle a^i b^j| = |c_{ij}|^2 A^i B^j, \tag{2.16}$$

where in a) the orthogonality of the basis was used.

The equation $p(b_2, b_1, a_1) = 0$ reduces to:

$$|c_{ij}|^2 \text{tr}(B^k U A^i B^j U^\dagger) = 0.$$

We further expand the term inside the trace, and just with same calculation as before we obtain:

$$UA^i B^j U^\dagger = \sum_{s,l,m,f} u_{s,l,j,i} u_{j,i,m,f}^* |a^s b^l\rangle \langle a^m b^f|.$$

We multiply both sides by B^k (it does not affect the trace because a projector is equal to itself squared):

$$\begin{aligned} B^k UA^i B^j U^\dagger B^k &= \\ &= |b^k\rangle \langle b^k| \otimes \mathbb{I} \sum_{s,l,m,f} u_{s,l,i,j} u_{i,j,m,f}^* |a^s b^l\rangle \langle a^m b^f| |b^k\rangle \langle b^k| \otimes \mathbb{I} = \\ &\stackrel{a)}{=} \sum_{s,l,m,f} u_{s,l,i,j} u_{i,j,m,f}^* \langle b^k|b^l\rangle \langle b^f|b^k\rangle |a^s b^k\rangle \langle a^m b^k| = \\ &= \sum_{s,l,m,f} u_{s,l,i,j} u_{i,j,m,f}^* \delta_{kl} \delta_{fk} |a^s b^k\rangle \langle a^m b^k| = \\ &= \sum_{s,m} u_{s,k,i,j} u_{i,j,m,k}^* |a^s b^k\rangle \langle a^m b^k|, \end{aligned}$$

where in $a)$ we use the fact that we choose to measure the same observable for qubit B before and after the channel, otherwise so the scalar product would not have resulted into a Kronecker delta and the computation would be way more complicated, but it would include as a parameter the angle between eigenvectors of the two observables measured for B . The trace is easy to compute because it only takes diagonal elements, the elements in the sum associated to identical kets and bras. In our case it reduces to a δ_{ms} in the calculation, obtaining in the end:

$$\sum_s u_{s,k,i,j} u_{i,j,s,k}^* = 0. \quad (2.17)$$

Before analyzing with more attention the term, we consider the condition in theorem 2.2.2, having already calculated the first term, Eq. 2.17, and the second, Eq. 2.12, we can rewrite the condition as:

$$\begin{aligned} \text{tr}(B^k UA^i B^j \rho B^j A^i U^\dagger) &= \text{tr}(A^i B^j \rho), \\ |c_{ij}|^2 \sum_s u_{s,k,i,j} u_{i,j,s,k}^* &= |c_{ij}|^2 \\ \sum_s u_{s,k,i,j} u_{i,j,s,k}^* &= 1. \end{aligned}$$

We can notice that the condition imposes a restriction on the form of the channel written in the basis of eigenvectors of our observables. This does not mean that, given a channel, transfer entropy remains constant varying the initial state, as we have seen previously. It only means that the transfer entropy for channel respecting the previous constraint is equal to its bound, which will probably be dependent on the state, measurement apparatus and channel all together. For the purpose of understanding, we try to rewrite the two conditions

in a more practical form, even though more intricate:

$$\begin{aligned}
 & \sum_s u_{s,k,i,j} u_{i,j,s,k}^* = \\
 & = \sum_s \langle a^s b^k | U | a^i b^j \rangle \left(\langle a^i b^j | U | a^s b^k \rangle \right)^* = \\
 & = \sum_s \langle a^s b^k | U | a^i b^j \rangle \langle a^s b^k | U^\dagger | a^i b^j \rangle = \\
 & = \langle a^+ b^k | U | a^i b^j \rangle \langle a^+ b^k | U^\dagger | a^i b^j \rangle + \langle a^- b^k | U | a^i b^j \rangle \langle a^- b^k | U^\dagger | a^i b^j \rangle.
 \end{aligned}$$

Calling $U_{i,j}^s = U | a^i b^j \rangle \langle a^s b^k | U^\dagger$:

$$\langle a^+ b^k | U_{ij}^+ | a^i b^j \rangle + \langle a^- b^k | U_{ij}^- | a^i b^j \rangle.$$

Introducing the permutation operator $P_{l,m}$ that exchanges row l with row m of the matrix, we can rewrite everything compactly:

$$\langle a^+ b^k | U_{ij}^+ + P_{1,3} P_{2,4} U_{ij}^- | a^i b^j \rangle \quad (2.18)$$

Combining our condition together we can see that the transfer entropy is equal to its bound when:

$$\langle a^+ b^k | U_{ij}^+ + P_{1,3} P_{2,4} U_{ij}^- | a^i b^j \rangle \in [0, 1] \quad \forall i, j, k \in \{+1, -1\} \quad (2.19)$$

and if all elements are zero, then transfer entropy and its bound are zero. Practically, once $U_{ij}^+ + P_{1,3} P_{2,4} U_{ij}^-$ is computed, we have to check the values of the first two rows to see if the transfer entropy is equal to its bound.

Chapter 3

Conclusions and outlooks

This thesis aimed to introduce and investigate the behavior of the transfer entropy in both classical and quantum systems.

After providing a concise introduction to information theory and transfer entropy, we investigated some of its properties. We successfully demonstrated the monotonicity of the transfer entropy respect to the number of considered variables as a source of information flow, establishing a recursive representation through a sum of relative entropies. Additionally, we explored transfer entropy's boundary properties, revealing its monotonically decreasing nature respect to the number of known variables in the target's past. We found a representation of the boundary as a sum of transfer entropies that represented the information flow from the target's past to the target itself.

We identified the conditions necessary for a dynamic between two systems to saturate transfer entropy's bound and achieve its maximum value. It became evident that the dynamics of the source directly influenced the value of the transfer entropy. Specifically, to attain a maximal value for the transfer entropy, the source's dynamics should be non-interacting with its past.

We reported the behavior of the transfer entropy for some interesting cases and provided a critique of its interpretation as a measure of causality. It appears that the transfer entropy may not accurately capture the true source of causality when more than two systems are interacting, leading to overestimation or underestimation of the causal effect from a source.

Furthermore, we developed a code to simulate the dynamics of two interacting systems. With these simulations, we calculated the transfer entropy, with a focus on scenarios where the target variables were randomly extracted with a probability of p , while copying the source's value with a probability of $1 - p$. The evolution of the source system variables was governed by two distinct dynamics: one involved random extraction from a set of values, while the other multiplied the values of the previous variable by minus one. Remarkably, the behavior predicted theoretically was observed. Particularly, in cases where the source evolved deterministically, it became apparent that the transfer entropy decreased with an increase in conditioned variables, reflecting the effect of what can be described as "stored information."

Subsequently, we applied the transfer entropy on measurements of spin on quantum circuits. Initially, we designed a code to simulate two-qubit quantum circuits, offering flexibility in selecting measurement types, channels, and circuit lengths. Our implementation utilized a tree structure to depict the circuit, where nodes represented density matrices and edge weights encoded transition probabilities between density matrices. Using this structure, we conducted numerous simulations and computed the transfer entropy effectively.

The utilized circuit involved two qubits initialized with an arbitrary state. It started by measuring the spin of the qubits, followed by their evolution through the C-NOT gate, and concluded with another round of measurements. To compute the transfer entropy, we used the measurement of qubit A as the source and those of qubit B as the target.

Initially, we simulated the transfer entropy for spin measurements in the z direction. Notably, we observed that the choice of the initial state strongly influenced the behavior of the transfer entropy. Different initial states yield distinct probability distributions of measurement outcomes. When the source qubit's initial state was an eigenstate of its measurement, the transfer entropy was equal to zero, as it could not detect the effect of the source on the target without observing varied source values. It was also observed the linearity of the transfer entropy with the squared coefficients of the initial state.

In order to have a better description of the behaviour of the transfer entropy on quantum circuit we tried to provide an analytical formulation of how the code computed the transfer entropy, and use the Born's rule to derive the probability distributions. It is shown that the analytic formulation reproduces the same results of the simulations. It can be used to infer constraints on our circuit that present certain properties of the transfer entropy.

Subsequently It would be interesting to measure how the transfer entropy behaves respect to entanglement creating channels, to understand if the transfer entropy is an appropriate measure of causality in quantum processes and to find a formulation that uses density matrices instead of probability distributions.

References

- [1] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, “Quantifying causal influences”, *The Annals of Statistics*, vol. 41, no. 5, Oct. 2013. DOI: [10.1214/13-aos1145](https://doi.org/10.1214/13-aos1145).
- [2] X. S. Liang, “Information flow and causality as rigorous notions ab initio”, *Phys. Rev. E*, vol. 94, p. 052201, 5 Nov. 2016. DOI: [10.1103/PhysRevE.94.052201](https://doi.org/10.1103/PhysRevE.94.052201).
- [3] J. Pearl *et al.*, “Models, reasoning and inference”, *Cambridge, UK: Cambridge University Press*, vol. 19, no. 2, p. 3, 2000.
- [4] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods”, 1969. DOI: <https://doi.org/10.2307/1912791>.
- [5] T. Schreiber, “Measuring information transfer”, *Phys. Rev. Lett.*, vol. 85, pp. 461–464, 2 Jul. 2000. DOI: [10.1103/PhysRevLett.85.461](https://doi.org/10.1103/PhysRevLett.85.461).
- [6] R. V. Michael Wibral and M. Lindner, *Transfer Entropy in Neuroscience*. Springer, Berlin, Heidelberg, 2014.
- [7] M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser, “Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks”, *Progress in biophysics and molecular biology*, vol. 105, no. 1-2, pp. 80–97, 2011.
- [8] T. Dimpfl and F. J. Peter, “Using transfer entropy to measure information flows between financial markets”, *Studies in Nonlinear Dynamics and Econometrics*, vol. 17, no. 1, pp. 85–102, 2013.
- [9] L. Sandoval Jr, “Structure of a global network of financial companies based on transfer entropy”, *Entropy*, vol. 16, no. 8, pp. 4443–4482, 2014.
- [10] S. Stramaglia, G.-R. Wu, M. Pellicoro, and D. Marinazzo, “Expanding the transfer entropy to identify information circuits in complex systems”, *Physical Review E*, vol. 86, no. 6, p. 066211, 2012.
- [11] J. Nichols, M. Seaver, S. Trickey, M. Todd, C. Olson, and L. Overbey, “Detecting nonlinearity in structural systems using the transfer entropy”, *Physical Review E*, vol. 72, no. 4, p. 046217, 2005.
- [12] N. Ciolli, *Transfer Entropy*, Jan. 2024. [Online]. Available: <https://github.com/niccogc/TransferEntropy/>.
- [13] V. Vedral, “The role of relative entropy in quantum information theory”, *Reviews of Modern Physics*, vol. 74, no. 1, pp. 197–234, Mar. 2002. DOI: [10.1103/revmodphys.74.197](https://doi.org/10.1103/revmodphys.74.197).
- [14] C. Shannon, “A mathematical theory of communication”, 1948.

- [15] V. Vedral, “The role of relative entropy in quantum information theory”, *Reviews of Modern Physics*, vol. 74, no. 1, pp. 197–234, Mar. 2002. DOI: [10.1103/revmodphys.74.197](https://doi.org/10.1103/revmodphys.74.197).
- [16] L. Barnett, A. B. Barrett, and A. K. Seth, “Granger causality and transfer entropy are equivalent for gaussian variables”, *Physical review letters*, vol. 103, no. 23, 2009.
- [17] T. Bossomaier, L. Barnett, M. Harré, and J. Lizier, *An Introduction to Transfer Entropy: Information Flow in Complex Systems*. Springer International Publishing, 2016.
- [18] N. B. Ryan G. James and J. P. Crutchfield, “Information flows? a critique of transfer entropies”, 2016. DOI: [10.1103/PhysRevLett.116.238701](https://doi.org/10.1103/PhysRevLett.116.238701).
- [19] E. M. B. Jie Sun, “Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings”, 2015. DOI: <https://doi.org/10.1016/j.physd.2013.07.001>.
- [20] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [21] B. Kraus and J. I. Cirac, “Optimal creation of entanglement using a two-qubit gate”, 2000. DOI: [10.1103/PhysRevA.63.062309](https://doi.org/10.1103/PhysRevA.63.062309).